Ergod Th & Dynam Sys (1987), 7, 481-488
Printed in Great Britain

An ergodic theorem for iterated maps

JOHN H ELTON
School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, *USA*

(*Received* 10 *April* 1986 *and revised* 11 *August* 1986)

Abstract Consider a Markov process on a locally compact metric space arising from iteratively applying maps chosen randomly from a finite set of Lipschitz maps which, on the average, contract between any two points (no map need be a global contraction) The distribution of the maps is allowed to depend on current position, with mild restrictions Such processes have unique stationary initial distribution [**BE**], [**BDEG**]

We show that, starting at *any* point, time averages along trajectories of the process converge almost surely to a constant independent of the starting point This has applications to computer graphics

1 *Introduction*

Let (X, d) be a metric space in which sets of finite diameter are relatively compact. Let w_i $X \to X$ be Lipschitz maps, with $d(w_i x, w_i y) \le s_i\, d(x, y)$ for x, y in X, $i = 1, \quad , N$ A good example is affine maps on $\mathbb{R}^n$ Let p_i $X \to [0, 1]$ such that $p_i(x) \ge 0$ and $\sum_{i=1}^{N} p_i(x) = 1$, and assume that the p_i's are continuous Define a Markov transition probability by

$$p(x, B) = \sum_{i=1}^{N} p_i(x) 1_B(w_i x)$$

This is the probability of transfer from $x \in X$ into the Borel set B Intuitively, pick a number i between 1 and N according to the distribution $p_i(x)$ and go to $w_i x$

Such processes have been discussed in many places under the assumption that the maps are contractions and usually that the p_i's are constants [**BD**], [**DF**], [**DS**], [**H**], [**K**] (Karlin [**K**] discussed variable p_i's) It was shown recently [**BE**], [**BDEG**] that none of the w_i's need be contractions, but that if there is contraction 'on the average' between any two points, i e

$$\prod_{i=1}^{N} d(w_i x, w_i y)^{p_i(x)} \le r\, d(x, y) \qquad \forall x, y, \text{ where } r < 1,$$

and if the p_i's are bounded away from 0 and have moduli of continuity ϕ_i satisfying Dini's condition (i e $\phi_i(t)/t$ is integrable over $(0, \alpha)$ for some $\alpha > 0$), then there is a unique, attractive stationary initial probability distribution μ for the process This means

$$V\mu(B) = \int p(x, B)\, d\mu(x) = \mu(B)$$

for all Borel sets B and, for all initial probability distributions ν, $V^n\nu$ converges in distribution (that is, weakly) to μ, i e $\int f\,dV^n\nu \to \int f\,d\mu$ for all bounded continuous functions f on X

Our object is to show that starting at *any* $x \in X$, the trajectories (orbits) of the process converge in distribution to μ almost surely By this we mean that for almost all trajectories $x, x_1, x_2,$ of the process starting at x, the time averages

$$\frac{1}{n+1}\sum_{k=0}^{n} f(x_k)$$

converge to $\int f\,d\mu$ for all bounded continuous f, or in yet other terms, the empirical distribution

$$\nu_n = \frac{1}{n+1}\sum_{k=0}^{n} \delta_{x_k}$$

of the first $n+1$ points along the trajectory converges weakly to μ as $n\to\infty$

Let us explain why this is important It follows (see Lemma 1) from the classical pointwise ergodic theorem that for μ-almost all $x \in X$, almost all trajectories starting at x converge in distribution to μ (in the sense just explained) But in applications to computer graphics, for example (see [**BD**]) we may have no way of choosing the starting x according to the measure μ, in fact, the idea is to start at some x and let a computer-generated realization of the process 'draw a picture' of μ

A special case of this result, when the maps w_i are contractions with a special disjointness condition, and the p_i's are constants, was stated already by Diaconis and Shashahani [**DS**] Most of the difficulty of our proof arises from having non-contractions and variable p_i's

First we prove a general lemma about Markov processes, and then we state and prove the main theorem, using a martingale argument

2 *Markov processes with unique stationary distribution*

Let $(X, \mathscr{F})$ be an arbitrary measurable space, and let $p(\ ,\ )\ X\times\mathscr{F}\to[0,1]$ be a *transition probability*, i e $p(x,\ )$ is a probability measure for each x, and $p(\ ,A)$ is a measurable function for all $A\in\mathscr{F}$ A (discrete-time) stochastic process $\{Z_n, n = 0, 1, \}$ with values in X is called a *Markov process with transition probability* p if

$$P\{Z_{n+1}\in A \mid Z_0 = z_0, \quad , Z_n = z_n\} = p(z_n, A) \quad \text{a s}$$

X is called the *state space* Define the operator V on finite measures by

$$V\nu(A) = \int p(x, A)\, d\nu(x), \qquad A\in\mathscr{F}$$

A probability measure μ is called a *stationary initial distribution* if $V\mu = \mu$ If μ is a stationary initial distribution and if Z_0 has distribution μ, then $\{Z_n\}$ will be a *stationary stochastic process* Assume for the rest of this paragraph that Z_0 has a stationary initial distribution so that $\{Z_n\}$ is a stationary process A is called an *invariant event* if there exists $C\in\mathscr{F}_\infty$ such that $A = \{(Z_k, Z_{k+1}, \quad)\in C\}$ for all $k\ge 0$, where $\mathscr{F}_\infty$ is the σ-algebra in X^∞ generated by measurable cylinders A is called *almost invariant* if there is an invariant event B so that $P(A\,\triangle\,B) = 0$ Let $\mathscr{I}$ denote the σ-algebra of almost invariant events The process $\{Z_n\}$ is called *ergodic* if for every $A \in \mathscr{I}$, $P(A)=0$ or 1 A reference for the above definitions is [**D**]

The next lemma is surely known, but we were unable to find a statement of it for general Markov processes We did find it stated in [**FK**] for a special case In any case, it follows very easily from well-known results

LEMMA 1 *If μ is the unique stationary initial distribution (or just an extreme point of the set of stationary initial distributions), then the process $\{Z_n\}$ with Z_0 having distribution μ is ergodic*

Proof If not, there is $A \in \mathscr{I}$ with $0 < P(A) < 1$ Then there exists $C \in \mathscr{F}$ such that $A = \{Z_n \in C\}$ a e for all $n \geq 0$, since $\{Z_n\}$ is a stationary Markov process [see **S**] Define

$$\nu(B) = \mu(B \cap C)/\mu(C) \quad \text{and} \quad \lambda(B) = \mu(B \cap {\sim}C)/\mu({\sim}C)$$

(note that $\mu(C) = P(A)$ and $\mu({\sim}C) = P({\sim}A)$) Then $\mu = \mu(C)\nu + \mu({\sim}C)\lambda$, so the proof will be completed by showing that ν (and hence λ) is a stationary initial distribution, since clearly $\nu \neq \lambda$

Now

$$\begin{aligned}\nu(B) &= P(Z_1 \in B \cap C)/\mu(C) \\ &= P((Z_1 \in B) \cap (Z_1 \in C))/\mu(C) \\ &= P((Z_1 \in B) \cap (Z_0 \in C))/\mu(C) \\ &= \frac{1}{\mu(C)} \int_C P(Z_1 \in B \mid Z_0 = z)\, d\mu(z) \\ &= \frac{1}{\mu(C)} \int_C p(z, B)\, d\mu(z) \\ &= \int p(z, B)\, d\nu(z),\end{aligned}$$

since clearly $d\nu/d\mu = (1/\mu(C))1_C$ But this says that ν is a stationary initial distribution □

Remark The processes discussed in the introduction and the next section are *not* what is called *indecomposable* in [**B**] and *Markov ergodic* in [**S**], as the following simple example shows, so we could not quote the theorems in those references for our application

Example Let $X = [0, 1]$, $w_1 x = \frac{1}{2}x$, $w_2 x = \frac{1}{2} + \frac{1}{2}x$, $p_i = 1/2$, $i = 1, 2$ Then all trajectories starting at a rational number in $[0, 1]$ stay in the rationals, and all trajectories starting at an irrational number in $[0, 1]$ stay in the irrationals Thus the process is *not* indecomposable/Markov ergodic as defined in [**B**], [**S**] (some people use the word 'indecomposable' differently) However, the process *is* ergodic, since there is a unique stationary initial distribution

3 *Main results*

Let $\Omega = N^\infty = \{(i_1, i_2, \quad) \; 1 \leq i_j \leq N$ and i_j is an integer for each $j\}$ Let $\mathscr{A}$ be the σ-algebra generated by the cylinders in Ω

Return now to the setup of the introduction For each $x \in X$, let P_x be the probability measure on $\mathscr{A}$ defined on cylinders by

$$P_x((i_1, i_2, \quad, i_n)) = p_{i_1}(x) p_{i_2}(w_{i_1} x) p_{i_3}(w_{i_2} w_{i_1} x) \qquad p_{i_n}(w_{i_{n-1}} \qquad w_{i_1} x)$$

(we abuse notation by writing $P_x((i_1, i_2, \quad, i_n))$ when we mean $P_x(\{i_1, i_2, \quad, i_n\} \times N \times N \times N \times \quad))$ It is clear this is precisely the probability measure for realizations of the Markov process starting at x That is, if we consider a Markov process $\{Z_n, n = 0, 1, \quad\}$ with state space X and transition probability p as given in the introduction, then

$$P((Z_0, Z_1, \quad) \in B \mid Z_0 = x)$$
$$= P_x\{(i_1, i_2, \quad) \ (x, w_{i_1} x, w_{i_2} w_{i_1} x, \quad) \in B\}$$

for every $B \in \mathscr{F}_\infty$

THEOREM *Suppose there exists $r < 1$ such that*

$$\prod_{i=1}^{N} d(w_i x, w_i y)^{p_i(x)} \leq r d(x, y) \qquad \forall x, y \text{ in } X$$

Assume there is $\delta > 0$ such that $p_i(x) \geq \delta$ for all x and i, and that the moduli of continuity of the p_i's satisfy Dini's condition Let μ be the unique stationary initial distribution for the Markov process described above (see [**BDEG**]*) Then for every x in X, there exists $G_x \subset \Omega$ such that $P_x(G_x) = 1$ and for $(i_1, i_2, \quad) \in G_x$, we have*

$$\frac{1}{n+1} \sum_{k=0}^{n} f(w_{i_k} \qquad w_{i_1} x) \to \int f \, d\mu$$

for all $f \in C(X)$, that is, almost all trajectories $x, w_{i_1} x, w_{i_2} w_{i_1} x, \quad$ starting at x converge in distribution to μ (in the sense explained in the introduction)

COROLLARY 1 *Let ν be any probability measure, and let $\{Z_n\}$ be the Markov process with initial distribution ν and transition probability as above Assume the hypotheses of the Theorem Then for all $f \in C(X)$,*

$$\frac{1}{n+1} \sum_{k=0}^{n} f(Z_k) \to \int f \, d\mu \quad a \ s$$

Remark It is shown in [**FK**] that Corollary 1 holds, in case X is a *compact* metric space, for a general transition probability for which it is only required that $x \mapsto p(x, \ )$ is continuous with the measures being given the w^*-topology, and that there is a unique stationary initial distribution

COROLLARY 2 *If $B \subset X$ is such that $\mu(\partial B) = 0$, then for any $x \in X$, if $(i_1, i_2, \quad) \in G_x$, the average amount of time the trajectory spends in B converges to $\mu(B)$, that is,*

$$\lim_{k \to \infty} \frac{\#\{j \ \ 0 \leq j \leq k, w_{i_j} \qquad w_{i_1} x \in B\}}{k+1} = \mu(B)$$

This follows from a well-known consequence of weak convergence, and generalizes a statement of [**DS**]

We prove two lemmas and then the Theorem and Corollary 1 The first lemma uses a martingale argument

https://doi.org/10.1017/S0143385700004168 Published online by Cambridge University Press

LEMMA 2 *Let $x, y \in X$, $x \neq y$ Assume the hypotheses of the theorem Let $r < r_1 < 1$*

(i) *For all $\varepsilon > 0$, there exist n_ε and $S \subset \Omega$ with $P_x(S) < \varepsilon$ such that*

$$n \geq n_\varepsilon \Rightarrow d(w_{i_n} \cdots w_{i_1}x, w_{i_n} \cdots w_{i_1}y) \leq r_1^n \, d(x, y)$$

except for $(i_1, i_2, \ldots)$ in S,

(ii) $\lim_{n\to\infty} d(w_{i_n} \cdots w_{i_1}x, w_{i_n} \cdots w_{i_1}y) = 0$ *a s -P_x*

Proof Let $s = \max\{s_i, i = 1, \ldots, N\}$ *W l o g* assume $s \geq 1$ Define random variables X_n on Ω by

$$X_n(i_1, i_2, \ldots) = \begin{cases} \left[\log \dfrac{d(w_{i_n} \cdots w_{i_1}x, w_{i_n} \cdots w_{i_1}y)}{d(w_{i_{n-1}} \cdots w_{i_1}x, w_{i_{n-1}} \cdots w_{i_1}y)}\right] \vee \dfrac{1}{\delta}\log(r/s) \\ \quad \text{if } d(w_{i_{n-1}} \cdots w_{i_1}x, w_{i_{n-1}} \cdots w_{i_1}y) \neq 0, \\ \log r \qquad \text{otherwise} \end{cases}$$

The purpose of the $(1/\delta)\log(r/s)$ term is to keep X_n bounded below, it is already bounded above by $\log s$

Claim $E(X_n | i_1, \ldots, i_{n-1}) \leq \log r$ for all $n \geq 1$ The expectation means with respect to the probability measure P_x on Ω

Proof Assume $d(w_{i_{n-1}} \cdots w_{i_1}x, w_{i_{n-1}} \cdots w_{i_1}y) \neq 0$ Then

$E(X_n | i_1, \ldots, i_{n-1})$

$$= \sum_{i_n=1}^{N} p_{i_n}(w_{i_{n-1}} \cdots w_{i_1}x)\left[\log \frac{d(w_{i_n} \cdots w_{i_1}x, w_{i_n} \cdots w_{i_1}y)}{d(w_{i_{n-1}} \cdots w_{i_1}x, w_{i_{n-1}} \cdots w_{i_1}y)}\right] \vee \frac{1}{\delta}\log(r/s)$$

Assume the expression in brackets is $\geq (1/\delta)\log(r/s)$ for each i_n Then the hypothesis of the Theorem (take logarithms) implies that the above is $\leq \log r$ If for some i_n the expression in brackets is $\leq (1/\delta)\log(r/s)$ (which is negative), the fact that $p_{i_n} \geq \delta$ is easily seen to imply that the above is *still* $\leq \log r$ The claim is proved

Now let $D_n = X_n - E(X_n | i_1, \ldots, i_{n-1})$, so D_n is a *martingale difference sequence,* and $|D_n| \leq 2|X_n| \leq B$, say

Let $Y_n = \sum_{k=1}^{n} (1/k) D_k$, so Y_n is a *martingale* Now $E(Y_n^2) \leq B^2 \sum_{k=1}^{\infty} 1/k^2$ since $D_k \perp D_l$ for $k \neq l$ (because they are martingale differences) Thus Y_n is an L^2-bounded martingale, and so $Y_n \to$ a s Then by Kronecker's lemma,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} D_k = 0 \quad \text{a s}$$

Thus,

$$\overline{\lim} \frac{1}{n} \sum_{k=1}^{n} \log \frac{d(w_{i_k} \cdots w_{i_1}x, w_{i_k} \cdots w_{i_1}y)}{d(w_{i_{k-1}} \cdots w_{i_1}x, w_{i_{k-1}} \cdots w_{i_1}y)} - \log r \leq 0 \quad \text{a s}$$

This telescopes to

$$\overline{\lim} \log\left(\frac{d(w_{i_n} \cdots w_{i_1}x, w_{i_n} \cdots w_{i_1}y)}{d(x, y)}\right)^{1/n} \leq \log r \quad \text{a s},$$

that is,

$$\overline{\lim}\left(\frac{d(w_{i_n}\quad w_{i_1}x,\, w_{i_n}\quad w_{i_1}y)}{d(x,y)}\right)^{1/n}\le r<r_1\quad \text{a s}$$

It is now easy to get from this that (i) of the conclusion holds, (ii) follows immediately from (i) □

LEMMA 3 *Assume the hypotheses of the theorem Then for all* x, y *in* X, P_y *is absolutely continuous with respect to* P_x

Proof Let $P_x(E)=0$, and let $\varepsilon>0$ We shall show $P_y(E)<\varepsilon$ Let r_1 be as in Lemma 2

Let ϕ_i be the modulus of continuity of p_i, and let $\phi=\phi_1\vee\phi_2\vee\quad\phi_N$ Note ϕ is increasing

Claim $\sum_{k=1}^{\infty}\phi(r_1^k d(x,y))<\infty$

Proof

$$\infty>\int_0^{d(x,y)}\frac{\phi(t)}{t}\,dt=\sum_{k=1}^{\infty}\int_{r_1^k d(x,y)}^{r_1^{k-1}d(x,y)}\frac{\phi(t)}{t}\,dt$$

$$\ge\sum_{k=1}^{\infty}(r_1^{k-1}-r_1^k)\,d(x,y)\,\frac{\phi(r_1^k d(x,y))}{r_1^{k-1}d(x,y)}$$

$$=(1-r_1)\,d(x,y)\sum_{k=1}^{\infty}\phi(r_1^k d(x,y))$$

which proves the claim

Now choose m so large that $m>n_{\varepsilon/2}$ from (i) of Lemma 2 and also $\sum_{k=m+1}^{\infty}\phi(r_1^k d(x,y))<\delta/2$

Let $\mathscr{A}_n$ be the cylinder sets in Ω depending only on the first n coordinates By a standard approximation result, there exist sets $A_n\in\mathscr{A}_n$ such that $E\subset\bigcup A_n$, the sets A_n are *disjoint* and $P_x(\bigcup A_n)<(\varepsilon/4)(\delta/(1-\delta))^{-m}$

Let $Q_n=\{(i_1,i_2,\quad)\ d(w_{i_k}\quad w_{i_1}x,\, w_{i_k}\quad w_{i_1}y)\le r_1^k d(x,y)$ for $m\le k\le n\}$, $n\ge m$ Let $Q_n=\Omega$ for $n<m$ Thus $Q_n\in\mathscr{A}_n$ Let $Q=\bigcap_{n\ge1}Q_n$ By Lemma 2(i), $P_y(\sim Q)<\varepsilon/2$ Let $n\ge m$ Now if $(i_1,i_2,\quad)\in Q_n$,

$$p_{i_1}(y)\quad p_{i_n}(w_{i_{n-1}}\quad w_{i_1}y)$$

$$\le p_{i_1}(x)\quad p_{i_n}(w_{i_{n-1}}\quad w_{i_1}x)\left(\frac{1-\delta}{\delta}\right)^m$$

$$\times\prod_{k=m+1}^{n}\left[1+\frac{p_{i_k}(w_{i_{k-1}}\quad w_{i_1}y)-p_{i_k}(w_{i_{k-1}}\quad w_{i_1}x)}{p_{i_k}(w_{i_{k-1}}\quad w_{i_1}x)}\right]$$

$$\le p_{i_1}(x)\quad p_{i_n}(w_{i_{n-1}}\quad w_{i_1}x)\left(\frac{1-\delta}{\delta}\right)^m\prod_{k=m+1}^{n}\left[1+\frac{\phi(r_1^k d(x,y))}{\delta}\right]$$

But

$$\prod_{k=m+1}^{n}\left[1+\frac{\phi(r_1^k d(x,y))}{\delta}\right]\le 1+2\sum_{k=m+1}^{\infty}\frac{\phi(r_1^k d(x,y))}{\delta}\le 2,$$

https://doi.org/10.1017/S0143385700004168 Published online by Cambridge University Press

so

$$p_{i_1}(y) \quad p_{i_n}(w_{i_{n-1}} \quad w_{i_1}y) \le 2\left(\frac{1-\delta}{\delta}\right)^m p_{i_1}(x) \quad p_{i_n}(w_{i_{n-1}} \quad w_{i_1}x)$$

When $n < m$, this holds trivially for any $(i_1, i_2, \quad)$

Thus,

$$\begin{aligned} P_y(Q \cap A_n) &\le P_y(Q_n \cap A_n) \\ &= \sum_{(i_1 \quad i_n)} \sum_{(i_1\ i_2 \quad) \in Q_n \cap A_n} p_{i_1}(y) \quad p_{i_n}(w_{i_{n-1}} \quad w_{i_1}y) \\ &\le \sum_{(i_1, \quad i_n)} \sum_{(i_1\ i_2 \quad) \in Q_n \cap A_n} 2\left(\frac{1-\delta}{\delta}\right)^m p_{i_1}(x) \quad p_{i_n}(w_{i_{n-1}} \quad w_{i_1}x) \\ &\le 2\left(\frac{1-\delta}{\delta}\right)^m P_x(A_n) \end{aligned}$$

So $P_y(\bigcup (Q \cap A_n)) \le 2((1-\delta)/\delta)^m P_x(\bigcup A_n)$ since the A_n's are disjoint Now the right side is $<\varepsilon/2$ by construction

Also $P_y(\bigcup (\sim Q \cap A_n)) \le P_y(\sim Q) < \varepsilon/2$, so we have then $P_y(E) \le P_y(\bigcup A_n) < \varepsilon$

□

Proof of the Theorem Let $\{Z_n\}$ be the Markov process with transition probability p as given in the introduction and such that Z_0 has distribution μ Then the process is stationary since μ is a stationary initial distribution, and is ergodic by Lemma 1 since μ is unique Let $f \in C_c(X)$, the continuous functions with compact support Then $\{f(Z_n), n = 0, 1, \quad \}$ is also stationary and ergodic [**B**, p 119] Let

$$B = \left\{(x_0, x_1, \quad) \in X^\infty \quad \frac{1}{n+1}\sum_{k=0}^{n} f(x_k) \to \int f\,d\mu\right\}$$

By the classical pointwise ergodic theorem, $P((Z_0, Z_1, \quad) \in B) = 1$

But

$$\begin{aligned} P((Z_0, Z_1, \quad) \in B) &= \int P((Z_0, Z_1, \quad) \in B \mid Z_0 = x)\, d\mu(x) \\ &= \int P_x((i_1, i_2, \quad) \quad (x, w_{i_1}x, w_{i_2}w_{i_1}x, \quad) \in B)\, d\mu(x) \end{aligned}$$

Thus, for some $x_0 \in X$,

$$P_{x_0}((i_1, i_2, \quad) \quad (x_0, w_{i_1}x_0, w_{i_2}w_{i_1}x_0 \quad) \in B) = 1$$

Let $G = \{(i_1, i_2, \quad) \quad (x_0, w_{i_1}x_0, w_{i_2}w_{i_1}x_0 \quad) \in B\}$ Thus $P_{x_0}(G) = 1$ and for $(i_1, i_2, \quad) \in G$,

$$\frac{1}{n+1}\sum_{k=0}^{n} f(w_{i_k} \quad w_{i_1}x_0) \to \int f\,d\mu$$

By Lemma 3, $P_y(G) = 1$ for every $y \in X$ By Lemma 2(ii), for every $y \in X$, there exists H_y with $P_y(H_y) = 1$ and for $(i_1, i_2, \quad) \in H_y$,

$$\frac{1}{n+1}\sum_{k=0}^{n} f(w_{i_k} \quad w_{i_1}y) - f(w_{i_k} \quad w_{i_1}x_0) \to 0$$

(note f is uniformly continuous) Thus for $(i_1, i_2, \quad) \in G \cap H_y$,

$$\frac{1}{n+1}\sum_{k=0}^{n} f(w_{i_k} \quad w_{i_1} y) \to \int f\,d\mu,$$

and $P_y(G \cap H_y) = 1$

In the above, G and H_y depended on f But since $C_c(X)$ is separable (since X is σ-compact), we obtain that for each $y \in X$, there exists G_y with $P_y(G_y) = 1$ such that

$$\frac{1}{n+1}\sum_{k=0}^{n} f(w_{i_k} \quad w_{i_1} y) \to \int f\,d\mu$$

for each f in a countable dense subset of $C_c(X)$, and then a 3ε argument gives this for each $f \in C_c(X)$ Finally, since μ is a probability measure, it is easy to see (by Urysohn's lemma) that this holds for all $f \in C(X)$ □

Proof of Corollary 1 As in the proof of the theorem, let

$$B = \left\{ (x_0, x_1, \quad) \in X^\infty \ \frac{1}{n+1}\sum_{k=0}^{n} f(x_k) \to \int f\,d\mu \right\}$$

Then

$$\begin{aligned} P((Z_0, Z_1, \quad) \in B) &= \int P((Z_0, Z_1, \quad) \in B \mid Z_0 = x)\, d\nu(x) \\ &= \int P_x((i_1, i_2, \quad)\ (x, w_{i_1}x, w_{i_2}w_{i_1}x, \quad) \in B)\, d\nu(x) \end{aligned}$$

But the integrand is 1 for each x by the theorem □

Acknowledgments We would like to thank Michael Barnsley for suggesting the problem, Marc Berger for pointing out the work of Furstenberg and Kifer, and Marc Berger and Jamie Goode for helpful discussions

REFERENCES

[B] L Breiman *Probability* Addison-Wesley (Reading, Massachusetts, 1968)

[BD] M F Barnsley & S Demko Iterated function systems and the global construction of fractals *Prod R Soc Lond* A **399** (1985), 243–275

[BDEG] M F Barnsley, S Demko, J Elton & J Geronimo Markov processes arising from function iteration with place-dependent probabilities Preprint, Dec 1985

[BE] M F Barnsley & J Elton Stationary attractive measures for a class of Markov chains arising from function iteration Preprint, Aug 1985

[D] J L Doob *Stochastic Processes* Wiley (New York, 1953)

[DF] L Dubins & D Freedman Invariant probabilities for certain Markov processes *Ann Math Stat* **32** (1966), 837–848

[DS] P Diaconis & M Shashahani Products of random matrices and computer image generation Stanford U preprint (1984)

[FK] H Furstenberg & Y Kifer Random matrix products and measures on projective spaces *Israel J Math* **46** (1983), 12–32

[H] J Hutchinson Fractals and self-similarity *Indiana U Journal of Math* **30** (1981), 713–747

[K] S Karlin Some random walks arising in learning models *Pacific J of Math* **3** (1953), 725–756

[S] W F Stout *Almost Sure Convergence* Academic Press (New York, 1974)