

ASYMPTOTIC MEASURES FOR SKEW PRODUCTS OF BERNOULLI SHIFTS WITH GENERALIZED NORTH POLE – SOUTH POLE DIFFEOMORPHISMS

D. K. MOLINEK

ABSTRACT. We study asymptotic measures for a certain class of dynamical systems. In particular, for $T: \Sigma_2 \times M \rightarrow \Sigma_2 \times M$, a skew product of the Bernoulli shift with a generalized north pole–south pole diffeomorphism, we describe the limits of the following two sequences of measures:

(1) iterates under T of the product of Bernoulli measure with Lebesgue measure, $T_*^n(\mu \times m)$, and

(2) the averages of iterates of point mass measures, $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$.

We give conditions for the limit of each sequence to exist. We also determine the subsequential limits in case the sequence does not converge.

We exploit several properties of null recurrent Markov Chains and apply them to the symmetric random walk on the integers. We also make use of Strassen's Theorem as an aid in determining subsequential limits.

1. INTRODUCTION

A main objective in the study of dynamical systems is to determine the long-term behavior of the orbits of the system. One way this is accomplished is through the study of asymptotic measures for the system. These measures provide a good qualitative description of the long-term behavior of most points. In this paper we study the limits of two sequences of measures for a specific class of dynamical systems. The systems we study are skew products of two-sided Bernoulli shifts with generalized north pole–south pole diffeomorphisms; namely,

$$T: \Sigma_2 \times M \rightarrow \Sigma_2 \times M$$

defined by

$$T(w, x) = (\sigma w, f^{w_0} x).$$

In this paper, the Bernoulli shift is on the two symbols $\{1, -1\}$. Thus, if $w \in \Sigma_2$, $w = \cdots w_{-1} w_0 w_1 \cdots$, with each $w_i = \pm 1$. Thus, the fiber map is either f or f^{-1} . The sequences of measures we consider are:

(i) $T_*^n(\mu \times m)$, where $\mu = B(\alpha, 1 - \alpha)$ is the Bernoulli measure on Σ_2 and m is normalized Lebesgue measure on M ; and,

(ii) averages of iterates of pointmass measures, $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$, for most points (w, x) of the system.

Received by the editors September 20, 1993.

1991 *Mathematics Subject Classification*. Primary 28D99, 60J15; Secondary 40A05, 58F11.

© 1994 American Mathematical Society
0002-9947/94 \$1.00 + \$.25 per page

The use of the skew product of two “nice” systems to increase our understanding of dynamical systems is not new. For example, Marcus and Newhouse [11] and Newhouse and Young [14] showed that measures of maximal entropy exist for certain skew products. It is a fact that these measures exist for basic sets of Axiom A diffeomorphisms. One reason for using skew products is that although the resulting dynamical system is no longer uniformly hyperbolic, perhaps enough of the “good” hyperbolic structure from the original two systems persists to gain insight into the new system.

The convergence properties and limit points of the above two sequences of measures are well understood for systems with uniform hyperbolicity. Sinai [20] and Ruelle [17] have proved the main theorems describing this. We discuss this in the next section. It is our desire to increase the knowledge of the convergence properties and limit points of these sequences of measures for systems without uniform hyperbolicity, as in “the real world” most systems are not uniformly hyperbolic.

We now state the results proved in this paper. Let $\sigma: \Sigma_2 \rightarrow \Sigma_2$ be the Bernoulli shift on the two symbols $\{1, -1\}$ with Bernoulli measure $\mu = B(\frac{1}{2}, \frac{1}{2})$.

Let $f: M \rightarrow M$ be a generalized north pole–south pole diffeomorphism of the compact manifold M with normalized Lebesgue measure, m . A diffeomorphism of this type has exactly one attracting fixed point, p , and one repelling fixed point, q .

We give a more precise definition in §2. T is the skew product of $\Sigma_2 \times M$ given by

$$T(w, x) = (\sigma w, f^{w_0} x).$$

The following theorems are proved in §5.

Theorem 5. $\lim_{n \rightarrow \infty} T_*^n(\mu \times m) = \mu \times (\frac{1}{2}\delta_p + \frac{1}{2}\delta_q)$.

We will show that for $\mu \times m$ almost every (w, x) , $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$ does not exist by describing all possible limit points of this sequence.

Theorem 8. *There is a measurable set $\Omega \subset \Sigma_2$ with $\mu(\Omega) = 1$, such that for every $w \in \Omega$ and Lebesgue almost every $x \in M$, the set of accumulation points of the equation $\nu_n(w, x) = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$ is precisely the set of all measures ν of the form $\nu = \mu \times (\lambda\delta_p + (1 - \lambda)\delta_q)$, $0 \leq \lambda \leq 1$.*

We note here the contrast with Ruelle’s theorem, which states that the sequence in Theorem 8 converges.

The convergence (or not) of the sequences of measures above relies on several properties of the symmetric random walk, in particular, its null recurrence and the Arcsine Law. These properties are discussed in §3.

In addition to the above-mentioned topics from Probability Theory, we also utilize Strassen’s Theorem in the proof of Theorem 8.

If we change the measure in the base to the biased measure $\mu = B(\alpha, 1 - \alpha)$, $\alpha > 1/2$, with a generalized north pole–south pole diffeomorphism of M as the fiber map, then we prove the following:

Theorem 11. $\lim_{n \rightarrow \infty} T_*^n(\mu \times m) = \mu \times \delta_p$.

Theorem 12. $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)} = \mu \times \delta_p$ for $\mu \times m$ almost every (w, x) .

We actually prove a slightly stronger result in Theorem 5, namely, that the sequence $T_*^n(\mu \times \delta_x)$ converges for Lebesgue almost every $x \in M$. A discussion of various sequences of measures and relationships between them is given in §4. The last section is a summary and a discussion of open questions. We now precisely define the system being studied and give background for hyperbolic theory. In addition, we give notation and definitions.

2. BACKGROUND, DEFINITIONS, AND NOTATION

In this section, we state Ruelle's theorem which describes the asymptotic measures examined in this paper for systems exhibiting uniform hyperbolicity. We also give a brief background of this theorem. We then define the systems being studied in this paper. Notation and definitions for terms we use are also presented.

The main theorem which describes the asymptotic measures which we study for a large class of dynamical systems is the following.

Theorem 1 (Ruelle [17]). *Let M be a compact manifold. Let Λ be an attracting compact hyperbolic set for $f: U \rightarrow U$ where f is C^r , $r \geq 2$. U is an open subset of M such that $\bigcap_{n \geq 0} f^n U = \Lambda$ and $fU \subset U$. There is a finite family (μ_α) of f -invariant probability measures with disjoint supports $\Omega_\alpha \subset \Lambda$ such that the following properties hold.*

(a) Ω_α is a hyperbolic attractor, and the family (Ω_α) contains all hyperbolic attractors in U . One can write

$$\mu_\alpha = \frac{1}{N_\alpha} \sum_{\beta=1}^{N_\alpha} \mu_{\alpha\beta}$$

where the probability measures $\mu_{\alpha 1}, \dots, \mu_{\alpha N_\alpha}$ are cyclically permuted by f and their supports $\Omega_{\alpha 1}, \dots, \Omega_{\alpha N_\alpha}$ are the connected components of Ω_α .

(b) Let the measure ν have support in U and be absolutely continuous with respect to Lebesgue measure. If N is the least common multiple of the N_α , then

$$\lim_{k \rightarrow \infty} f_*^{kN} \nu$$

exists and is a linear combination of the $\mu_{\alpha\beta}$.

(c) For Lebesgue almost every point x in U ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \delta_{f^k x} = \mu_{\alpha(x)}.$$

The measure μ_α is called the Bowen-Ruelle-Sinai (BRS) measure. Sinai first showed the existence of a BRS measure for Anosov diffeomorphisms, and Ruelle extended the results of the Axiom A case. Bowen and Ruelle then extended this to Axiom A flows. In the above, the existence of Markov partitions, symbolic dynamics, and techniques from statistical mechanics are used in showing the existence of the BRS measure (see, for example, [3, 17, 20]). A proof of the above theorem and a construction of the BRS measure without using Markov partitions can be found in Newhouse [3].

The measure δ_p is the point mass measure at p . If $A \subset M$, then $\delta_p(A) = 1$ if $p \in A$ and 0 otherwise.

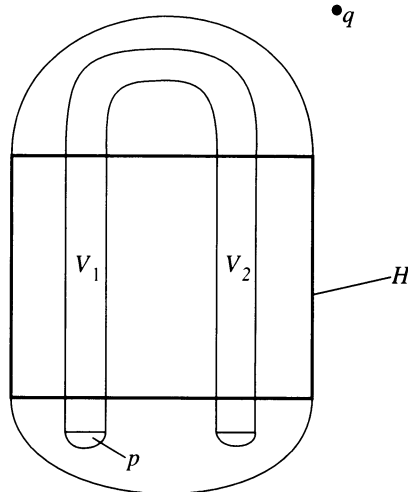


FIGURE 1

The measure f_*m is defined by $(f_*m)(A) = m(f^{-1}A)$.

A sequence of measures, $\{\nu_n\}$, converges to the measure ν if for every continuous function $\varphi: M \rightarrow \mathbf{R}$ we have that $\int \varphi d\nu_n$ converges to $\int \varphi d\nu$. All convergence of sequences of measures is with respect to this weak topology.

We now give the definition of the system studied in this paper.

The base space. Let $\sigma: \Sigma_2 \rightarrow \Sigma_2$ be the full shift on the two symbols $\{1, -1\}$. Elements in Σ_2 are denoted by $w = \cdots w_{-2}w_{-1}w_0w_1w_2\cdots$, each $w_i = 1$ or -1 , and $\sigma(w)_i = w_{i+1}$. The measures we use for this system are the Bernoulli measures, $\mu = B(\alpha, 1 - \alpha)$, with $0 < \alpha < 1$.

This system can be realized as the restriction to the nonwandering set (minus the two points p and q) of the Smale Horseshoe diffeomorphism of S^2 . (See Figure 1.)

$\Lambda = \bigcap_{n \in \mathbf{Z}} h^n(H)$ is a Cantor set consisting of those points whose forward and backward iterates remain in H . Figure 1 shows $h(S) \subset S$ where S is the “stadium” attached to the square H . Let $V_1 \cup V_2 = h(H) \cap H$ be as shown. Note that any point $z \in \Lambda$ has the property that $h^i z \in V_1$ or V_2 for every $i \in \mathbf{Z}$. Then $h|_\Lambda$ is topologically conjugate to $\sigma: \Sigma_2 \rightarrow \Sigma_2$. The conjugacy $c: \Lambda \rightarrow \Sigma_2$ is defined by

$$(c(z))_i = \begin{cases} 1 & \text{if } h^i z \in V_1, \\ -1 & \text{if } h^i z \in V_2. \end{cases}$$

The fiber space. Let $f: M \rightarrow M$ be a generalized north pole–south pole diffeomorphism on the compact manifold M . By this we mean that the nonwandering set of f , $\Omega(f)$, consists of one fixed source, q , one fixed sink, p , and saddle-type periodic orbits. Let $\mathcal{B}$ be the set of all the saddle-type periodic points of f . Then

$$\Omega(f) = \{p\} \cup \{q\} \cup \mathcal{B}.$$

Let m be the Lebesgue measure. Put $\mathcal{L} = W^s(\mathcal{B}) \cup W^u(\mathcal{B}) \cup \{p, q\}$. Note that $m(\mathcal{L}) = 0$.

The salient feature of the generalized north pole–south pole diffeomorphism is that for every $x \in M \setminus \mathcal{L}$, $f^n x \rightarrow p$ and $f^{-n} x \rightarrow q$ as $n \rightarrow \infty$. Figure 2

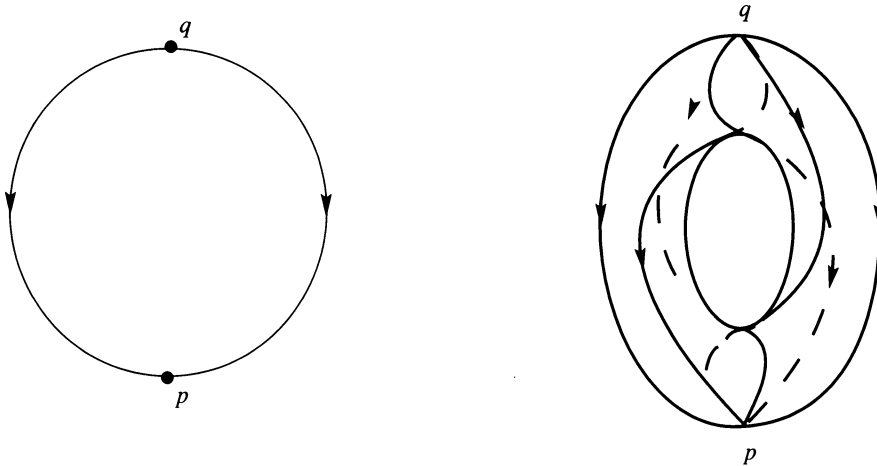


FIGURE 2

shows “typical” diagrams of such diffeomorphisms. The one on the left represents the usual north pole–south pole diffeomorphism of S^1 , and the one on the right is the negative of the gradient of the height function on the torus.

The *skew product* of the two systems above is defined as follows:

$$T: \Sigma_2 \times M \rightarrow \Sigma_2 \times M \quad \text{by} \quad T(w, x) = (\sigma w, f^{w_0} x).$$

Note that

$$T^n(w, x) = (\sigma^n w, f^{S_n(w)} x) \quad \text{where} \quad S_n(w) = \sum_{k=0}^{n-1} w_k.$$

Throughout this paper, we refer to the “unbiased” and “biased” case. The system is unbiased if the measure on the base space Σ_2 is the equally weighted Bernoulli measure, $\mu = B(\frac{1}{2}, \frac{1}{2})$. The system is biased if the measure on Σ_2 is the Bernoulli measure $\mu = B(\alpha, 1 - \alpha)$, $\alpha \neq 1/2$.

We now proceed to discuss the symmetric random walk on the integers.

3. THE SYMMETRIC RANDOM WALK

The results in this paper rely heavily on properties of the random walk. The random walk is realized as the quantity S_n which arises in the definition of the skew product. We first review some basic ideas about Markov chains and state results which will be used in later chapters. Specifically, we show that for almost every point in the shift space, the average occupation time of its cumulative sums $(S_n(w))$ in any finite state tends to zero as n increases. Also, we recall the Arcsine Law which describes the amount of time a random walk spends in positive states.

Let $\xi_1, \xi_2, \dots$ be independent identically distributed (i.i.d.) random variables on Σ_2 . For example, if we let $\xi_k = w_k$ ($w \in \Sigma_2$), then the sequence $\xi_1, \xi_2, \dots$ is a sequence of i.i.d. random variables from Σ_2 to the set $\{-1, 1\}$, with $P\{\xi_k = -1\} = P\{\xi_k = 1\} = \frac{1}{2}$.

Let $S_n = \xi_1 + \dots + \xi_n$. Then the sequence $\{S_n\}$ is a Markov chain with transition probabilities:

$$p_{ij} = P\{S_{n+1} = j | S_n = i\} = \begin{cases} \frac{1}{2} & \text{if } |j - i| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, the sequence $\{S_n\}$ is called the symmetric random walk on the integers. The symmetric random walk is an example of a null recurrent Markov chain. Before giving the definition, we furnish some notation.

The value of S_n is called the state of the chain at step n . Let $p_{ij}^{(n)}$ denote the probability that the chain goes from state i to state j in exactly n steps.

As we will see, the results we state are independent of the starting position of the chain. Unless stated otherwise, we will assume the initial value of the chain is zero. Let $T_j = \min\{n \geq 1 | S_n = j\}$ be the first time that the chain is in state j .

Let $f_{0j}^{(n)}$ be the probability that the first time the chain is in state j is at time n . We write $f_{0j}^{(n)} = P\{T_j = n | S_0 = 0\}$. The probability that the state is ever in state j is

$$f_{0j}^* = P\{S_k = j \text{ for some } k \geq 1 | S_0 = 0\} = \sum_{n=1}^{\infty} f_{0j}^{(n)}.$$

Let m_{jj} be the expected return time to state j . So

$$m_{jj} = E(T_j | S_0 = j) = \sum_{k=1}^{\infty} k f_{jj}^{(k)}.$$

Definition. A state j is called *recurrent* if $f_{0j}^* = 1$.

Definition. A state j is *null recurrent* if $m_{jj} = \infty$.

Theorem 2 (Feller [6]). *A state i is null recurrent if and only if*

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty \quad \text{and} \quad p_{ii}^{(n)} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

The symmetric random walk is an irreducible¹ Markov chain. This implies that any recurrence property which holds for one state holds for every state. We say that the symmetric random walk is recurrent if any state is.

Lemma 1. *The symmetric random walk is null recurrent.*

Proof. In view of the above theorem and remark, it suffices to calculate $p_{00}^{(n)}$. First, note that a return of the chain to zero must be accomplished in an even number of steps, so $p_{00}^{(2n+1)} = 0 \forall n \geq 0$. The number of ways for the chain to return back to zero after $2n$ steps is $\binom{2n}{n}$. Thus, we see that

$$p_{00}^{(2n)} = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n}.$$

Applying Stirling's formula, $(n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n})$ gives $p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}$, where the $\sim$ means that the ratio of the two sides tends to 1 as $n \rightarrow \infty$.

Since $\sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}}$ diverges and $\frac{1}{\sqrt{\pi n}} \rightarrow 0$, we have the desired result that the state zero and hence, the random walk are null recurrent. $\square$

¹ Every state is accessible to and from every other state in a finite number of steps.

We now calculate the average expected occupation time of the random walk in any state. Let

$$\mathbf{I}_{[S_k=i]} = \begin{cases} 1 & \text{if } S_k = i, \\ 0 & \text{if } S_k \neq i. \end{cases}$$

be the indicator function of the state i . Then $N_n(i) = \sum_{k=0}^n \mathbf{I}_{[S_k=i]}$ is the number of times the chain is in state i during the first n steps. Its expected value, given that the chain started in state j , is

$$E_j(N_n(i)) = E_j \left(\sum_{k=0}^n \mathbf{I}_{[S_k=i]} \right) = \sum_{k=0}^n E_j(\mathbf{I}_{[S_k=i]}) = \sum_{k=0}^n p_{ji}^{(k)}.$$

In light of the above, the quantity

$$\frac{1}{n+1} \sum_{k=0}^n p_{0i}^{(k)}$$

is the average expected occupation time of the chain in state i during the first n steps.

Theorem 3 (Chung [5]). *For all i*

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n p_{0i}^{(k)} = \frac{1}{m_{ii}}.$$

Note again the independence of the starting position of the chain. Because the symmetric random walk is null recurrent (i.e., $m_{jj} = \infty$), we conclude from the above discussion and theorem that the average expected occupation time of the chain in any finite state is zero. In addition, we see that the average expected occupation time in any finite interval is also zero. The next result, which states that the actual average occupation time of the random walk in any state i is zero for μ almost every point w , is important in what follows.

Proposition 1. *For any state i ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{I}_{[S_k=i]} = 0 \quad \mu\text{-almost surely (a.s.).}$$

Proof. The proof of this proposition uses standard techniques from Probability Theory. We show this proposition holds for the state zero. The result that the average occupation time in any finite state i is zero follows from the remark after Theorem 3. Let

$$Y_n = \frac{1}{n} \sum_{k=1}^n \mathbf{I}_{[S_k=0]}.$$

To prove the proposition, we first show that

$$\sum_{n=1}^{\infty} P(Y_n > \varepsilon) < \infty$$

and then use a Borel-Cantelli type theorem to conclude that $Y_n \rightarrow 0$ a.s.

Step 1. Define the inter-returning times: $T_0 = 0$ and

$$T_i = \inf\{n > T_{i-1} | S_n = 0\} - T_{i-1}, \quad \text{for } i \in \mathbb{N}.$$

Then $\{T_i | i \in \mathbf{N}\}$ are i.i.d. with

- $P(T_i < \infty) = 1 \forall i$ by the recurrence property of the symmetric random walk,
- $E(T_i) = \infty \forall i$ by the null recurrence property of the symmetric random walk.

For $c > 0$, define $T_i^c = T_i \cdot \mathbf{I}_{[T_i \leq c]}$, $i \in \mathbf{N}$. The sequence $\{T_i^c\}$ is also i.i.d. For each i , T_i^c is a monotone increasing sequence in c , $T_i^c \leq c$, and $\lim_{c \rightarrow \infty} T_i^c = T_i$.

By the Monotone Convergence Theorem and the above discussion, we have

$$\begin{aligned} \lim_{c \rightarrow \infty} E(T_i^c) &= \lim_{c \rightarrow \infty} \int T_i^c d\mu = \int \lim_{c \rightarrow \infty} T_i^c d\mu \\ &= \int T_i d\mu = E(T_i) = \infty. \end{aligned}$$

Step 2. Let $\varepsilon > 0$ be given. Choose c such that $E(T_i^c) > 2/\varepsilon$. This can be done since $E(T_i^c) \rightarrow \infty$. Choose $N = N(\varepsilon, c)$ to be such that

$$N > \frac{\varepsilon E(T_i^c)^2}{2\varepsilon E(T_i^c) - 4}.$$

Then

$$\begin{aligned} P(Y_n > \varepsilon) &= P\left(\sum_{k=1}^n \mathbf{I}_{[S_k=0]} > n\varepsilon\right) \quad (\text{definition of } Y_n) \\ &\leq P\left(\sum_{i=1}^{\lceil n\varepsilon \rceil} T_i < n\right) \leq P\left(\sum_{i=1}^{\lceil n\varepsilon \rceil} T_i^c < n\right) \\ &= P\left(\sum_{i=1}^{\lceil n\varepsilon \rceil} \frac{T_i^c}{\sqrt{n}} < \sqrt{n}\right) = P\left(\sqrt{n} - \sum_{i=1}^{\lceil n\varepsilon \rceil} \frac{T_i^c}{\sqrt{n}} > 0\right) \\ &= P\left(\exp\left(\sqrt{n} - \sum_{i=1}^{\lceil n\varepsilon \rceil} \frac{T_i^c}{\sqrt{n}}\right) > 1\right) \\ &\leq e^{\sqrt{n}} E\left(\exp\left(-\sum_{i=1}^{\lceil n\varepsilon \rceil} \frac{T_i^c}{\sqrt{n}}\right)\right) \quad (\text{Chebyshev's inequality}) \\ &= e^{\sqrt{n}} E\left(\exp\left(-\frac{T_i^c}{\sqrt{n}}\right)\right)^{\lceil n\varepsilon \rceil} \\ &\leq e^{\sqrt{n}} \left(1 - \frac{E(T_i^c)}{\sqrt{n}} + \frac{E(T_i^c)^2}{2n}\right)^{\lceil n\varepsilon \rceil} \\ &\leq e^{\sqrt{n}} \exp\left(-\lceil n\varepsilon \rceil \left(\frac{E(T_i^c)}{\sqrt{n}} - \frac{E(T_i^c)^2}{2n}\right)\right) \\ &= \exp\left(-\sqrt{n} \left(-1 + \lceil n\varepsilon \rceil \frac{E(T_i^c)}{n} - \lceil n\varepsilon \rceil \frac{E(T_i^c)^2}{2n^{3/2}}\right)\right) \\ &\leq e^{-\sqrt{n}} \quad \text{for } n > N(\varepsilon, c). \end{aligned}$$

Thus, $P(Y_n > \varepsilon) \leq e^{-\sqrt{n}}$ for $n > N$ and so

$$\sum_{n=N+1}^{\infty} P(Y_n > \varepsilon) \leq \sum_{n=N+1}^{\infty} e^{-\sqrt{n}}.$$

Since $\sum_{n=1}^{\infty} e^{-\sqrt{n}}$ converges, so does $\sum_{n=1}^{\infty} P(Y_n > \varepsilon)$.

Step 3. See [19, p. 254]. We show that

$$\sum_{n=1}^{\infty} P(Y_n > \varepsilon) < \infty \Rightarrow Y_n \rightarrow 0 \quad \text{a.s.}$$

Let $B_n^\varepsilon = \{w | Y_n > \varepsilon\}$, and let $B^\varepsilon = \limsup B_n^\varepsilon = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k^\varepsilon$. Then $\{w | Y_n \not\rightarrow 0\} = \bigcup_{\varepsilon>0} B^\varepsilon$.

We first note that

$$\sum_{n=1}^{\infty} P\{Y_n > \varepsilon\} < \infty \Rightarrow \lim_{n \rightarrow \infty} P\left\{\sup_{k \geq n} Y_k > \varepsilon\right\} = 0.$$

Since

$$\begin{aligned} P\left\{\sup_{k \geq n} Y_k > \varepsilon\right\} &= P\left\{\bigcup_{k=n}^{\infty} (Y_k > \varepsilon)\right\} \leq \sum_{k=n}^{\infty} P\{Y_k > \varepsilon\} \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ since } \sum_{k=1}^{\infty} P\{Y_k > \varepsilon\} \text{ converges.} \end{aligned}$$

We now show that

$$\lim_{n \rightarrow \infty} P\left\{\sup_{k \geq n} Y_k > \varepsilon\right\} = 0 \Rightarrow \lim_{n \rightarrow \infty} Y_n = 0 \quad \text{a.s.}$$

$$\begin{aligned} P\left\{\sup_{k \geq n} Y_k > \varepsilon\right\} \rightarrow 0 &\Rightarrow P\left\{\bigcup_{k \geq n} B_k^\varepsilon\right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty \\ &\Rightarrow P\{B^\varepsilon\} = 0 \quad \text{for each } \varepsilon > 0 \\ &\Rightarrow P\{B^{1/m}\} = 0 \quad \forall m \geq 1 \\ &\Rightarrow P\left\{\bigcup_{m=1}^{\infty} B^{1/m}\right\} > 0 \\ &\Rightarrow P\left\{\bigcup_{\varepsilon>0} B^\varepsilon\right\} = 0 \\ &\Rightarrow P\{w | Y_n \not\rightarrow 0\} = 0 \\ &\Rightarrow Y_n \rightarrow 0 \quad \text{a.s.} \end{aligned}$$

So we have the desired result, $Y_n \rightarrow 0$ μ -almost surely. $\square$

As a consequence of this proposition, we have that for μ almost every w in Σ_2 , for every $\varepsilon > 0$, and for every integer i , there is an $N \in \mathbb{N}$ so that if $n > N$, $\frac{1}{n} \sum_{k=0}^{n-1} I_{[S_k(w)=i]} < \varepsilon$.

Next, we examine the proportion of time the random walk spends on one side of the origin. Our intuition might indicate that this quantity should be $1/2$. We will see that this is not the case as shown in the next theorem.

Theorem 4 (The Arcsine Law). *Let $\gamma(n)$ = the number of times $S_k > 0$ for $0 \leq k \leq n$. Denote the probability that during the first $2n$ steps the chain is positive for $2k$ of those steps by $P_{2k, 2n}$. Suppose $0 < x < 1$; then*

$$\begin{aligned} P \left\{ \frac{\gamma(2n)}{2n} \leq x \right\} &= \sum_{\{k | \frac{k}{n} \leq x\}} P_{2k, 2n} \rightarrow \frac{1}{\pi} \int_0^x \frac{1}{\sqrt{x(1-x)}} dx \\ &= \frac{2}{\pi} \arcsin \sqrt{x}. \end{aligned}$$

Proof. See [6, p. 82] or [19, p. 100].

We note that the integrand above is a U-shaped curve with minimum value at $1/2$. This means that it is the least likely case that the random walk will spend half its time positive and the other half negative. As an illustration, note that in one out of ten experiments, the chain will spend 99.4% of its time on one side of zero; that is,

$$P \left\{ \frac{\gamma(2n)}{2n} \leq 0.006 \right\} + P \left\{ \frac{\gamma(2n)}{2n} \geq 0.994 \right\} \rightarrow 0.1.$$

We note that all statements concerning probability can be restated in terms of the measure $\mu = B(\alpha, 1 - \alpha)$. For example,

$$P \left\{ \frac{\gamma(2n)}{2n} \leq x \right\} \rightarrow \frac{2}{\pi} \arcsin \sqrt{x}$$

is equivalent to

$$\mu \left\{ w \in \Sigma_2 \mid \frac{1}{2n} \sum_{k=0}^{2n-1} \mathbf{I}_{[S_k(w) > 0]} \leq x \right\} \rightarrow \frac{2}{\pi} \arcsin \sqrt{x}.$$

4. SEQUENCES OF MEASURES

In this section we consider certain sequences of measures on the product space $X \times Y$ and the relationships among them. Let $g: X \rightarrow X$ be a homeomorphism of the compact metric space X , with Borel probability measure μ . Let $f_x: Y \rightarrow Y$ be a family of homeomorphisms depending continuously on $x \in X$, of the compact metric space Y , with Borel probability measure m . Define the skew product of the product space, $T: X \times Y \rightarrow X \times Y$ by $T(x, y) = (gx, f_x y)$.

We recall some basic notation and facts from Ergodic Theory (see Petersen [15] or Walters [23]):

- (1) $T_*^n(\mu \times m)$ is defined by $T_*^n(\mu \times m)(A) = (\mu \times m)(T^{-n}A)$.
 - (2) A sequence of measures $\{\nu_n\}$ on X converges to the measure ν if and only if for every continuous $\varphi: X \rightarrow \mathbf{R}$, the sequence $\{\int \varphi d\nu_n\}$ converges to $\int \varphi d\mu$.
 - (3) $\int \varphi(s, t) d(\delta_x \times \delta_y) = \varphi(x, y)$ for every continuous $\varphi: X \times Y \rightarrow \mathbf{R}$.
 - (4) $\int \varphi(s, t) d(\delta_{T^k(x, y)}) = \int \varphi \circ T^k(s, t) d(\delta_x \times \delta_y)$, which by (3), equals $\varphi \circ T^k(x, y)$.
 - (5) $\int \varphi(s, t) d(T_*^k(\mu \times m)) = \int \varphi \circ T^k(s, t) d(\mu \times m)$.
- Consider the following sequences of measures:
- (a) $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(x, y)}$,

- (b) $\frac{1}{n} \sum_{k=0}^{n-1} T_*^k(\mu \times m)$,
- (c) $T_*^k(\mu \times m)$,
- (d) $T_*^k(\mu \times \delta_y)$.

Note that if $T: X \times Y \rightarrow X \times Y$ has a hyperbolic attractor Λ on which T is topologically mixing and if $\mu \times m$ is supported on the basin of the attractor and is absolutely continuous with respect to Lebesgue measure on $X \times Y$, then the sequences in (a), (b), and (c) all converge, as we are in the setting of Ruelle's theorem. We now give the relationships among these sequences in the general setting we have described in the opening paragraph of this section.

Proposition 2. *If the sequence in (a) converges to the measure ν , for $\mu \times m$ almost every (x, y) , then the sequence in (b) also converges to ν .*

Proof. Suppose $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(x, y)} \rightarrow \nu$ for $\mu \times m$ almost every (x, y) . From (2), (3), and (4), this means that for every continuous $\psi: X \times Y \rightarrow \mathbf{R}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \psi \circ T^k(s, t) d(\delta_x \times \delta_y) \rightarrow \int \phi d\nu.$$

We need to show that for every continuous function $\phi: X \times Y \rightarrow \mathbf{R}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \int \phi(T^k(s, t)) d(\mu \times m) \rightarrow \int \phi d\nu.$$

Let $\phi: X \times Y \rightarrow \mathbf{R}$ be continuous. Define $g_n: X \times Y \rightarrow \mathbf{R}$ by

$$g_n(x, y) = \frac{1}{n} \sum_{k=0}^{n-1} \int \phi \circ T^k(s, t) d(\delta_x \times \delta_y).$$

Note that $g_n(x, y) = \frac{1}{n} \sum_{k=0}^{n-1} \phi \circ T^k(x, y)$, by (4). Each g_n is measurable and $|g_n|$ is bounded by $K = \max_{(x, y) \in X \times Y} |\phi(x, y)|$. Furthermore, by hypothesis $g_n(x, y) \rightarrow \int \phi d\nu$ for $\mu \times m$ a.e. (x, y) . By the Bounded Convergence Theorem, we have

$$\int g_n(x, y) d(\mu \times m) \rightarrow \int \left(\int \phi d\nu \right) d(\mu \times m) = \int \phi d\nu.$$

Then we have

$$\begin{aligned} & \frac{1}{n} \sum_{k=0}^{n-1} \int \phi(T^k(x, y)) d(\mu \times m) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \int \left(\int \phi(T^k(s, t)) d(\delta_x \times \delta_y) \right) d(\mu \times m) \\ &= \int \left(\frac{1}{n} \sum_{k=0}^{n-1} \int \phi(T^k(s, t)) d(\delta_x \times \delta_y) \right) d(\mu \times m) \\ &= \int g_n(x, y) d(\mu \times m) \\ &\rightarrow \int \phi d\nu \quad \text{as desired.} \quad \square \end{aligned}$$

Proposition 3. *If the sequence in (c) converges to ν , then so does the sequence in (b).*

Proof. This is a standard fact from analysis which we include for completeness.

We show $a_k \rightarrow a \Rightarrow \frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow a$.

Let $b_n = \frac{1}{n} \sum_{k=0}^{n-1} a_k$. Let $\varepsilon > 0$ be given, let N be such that $|a_k - a| < \frac{\varepsilon}{2}$ $\forall k \geq N$, and let $M \geq N$ be such that $\frac{1}{n} |a_1 + a_2 + \cdots + a_N - Na| < \frac{\varepsilon}{2}$ $\forall n \geq M$. Then

$$\begin{aligned} |b_n - a| &= \frac{1}{n} |a_1 + a_2 + \cdots + a_n - na| \\ &\leq \frac{1}{n} \{ |a_1 + a_2 + \cdots + a_N - Na| + |a_{N+1} - a| + \cdots + |a_n - a| \} \\ &< \frac{\varepsilon}{2} + \frac{1}{n} (n - N) \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{1}{n} \left(n \frac{\varepsilon}{2} \right) = \varepsilon. \end{aligned}$$

Thus, $b_n \rightarrow a$ as desired. $\square$

Proposition 4. *If the sequence in (d) converges to ν for m almost every y , then the sequence in (c) also converges to ν .*

Proof. This proof is similar to the one for Proposition 2. Let $\varphi: X \times Y \rightarrow \mathbf{R}$. Here, define $h_n: Y \rightarrow \mathbf{R}$ to be $h_n(y) = \int \varphi \circ T^n(x, z) d(\mu \times \delta_y)$. Then h_n is measurable, $|h_n|$ is bounded, and by hypothesis $h_n(y) \rightarrow \int \varphi d\nu$ for m almost every y . Again, by the Bounded Convergence Theorem, $\int h_n(y) dm \rightarrow \int (\int \varphi d\nu) dm = \int \varphi d\nu$.

We use Fubini's Theorem to interchange the order of integration and show that $\int \varphi \circ T^n(x, y) d(\mu \times m) \rightarrow \int \varphi d\nu$.

$$\begin{aligned} \int_{X \times Y} \varphi \circ T^n(x, y) d(\mu \times m) &= \int_Y \left(\int_X \varphi \circ T^n(x, y) d\mu \right) dm \\ &= \int_Y \left(\int_{X \times Y} \varphi \circ T^n(x, z) d(\mu \times \delta_y) \right) dm \\ &= \int h_n(y) dm \rightarrow \int \varphi d\nu. \end{aligned}$$

Thus, $T_*^n(\mu \times m) \rightarrow \nu$ as desired. $\square$

We now prove the theorems relating the above sequences of measures and the skew product described in the introduction.

5. THE UNBIASED NORTH POLE-SOUTH POLE DIFFEOMORPHISM

In this section we show that iterates of the measure $\mu \times \delta_x$ converge for Lebesgue almost every point x in M but that the averages of iterates of point mass measures do not converge. In §5.1, we give the set-up of the example being studied and state the main proposition. In the next section we recall some basic facts which allow us to simplify the proof. We prove the proposition in the third section. Section 5.4 is devoted to showing that the averages of iterates of point mass measures do not converge by exhibiting all possible weak limit points.

5.1. Introduction. Let $\sigma: \Sigma_2 \rightarrow \Sigma_2$ be the full shift on the two symbols $+1$, -1 with Bernoulli measure $\mu = B(\frac{1}{2}, \frac{1}{2})$ as described earlier.

Let $f: M \rightarrow M$ be a generalized north pole–south pole diffeomorphism on the compact manifold M .

We recall the definition of the skew product $T: \Sigma_2 \times M \rightarrow \Sigma_2 \times M$

$$T(w, x) = (\sigma w, f^{w_0} x).$$

Note that

$$T^n(w, x) = (\sigma^n w, f^{S_n(w)} x), \quad \text{where } S_n(w) = \sum_{k=0}^{n-1} w_k.$$

Recall the definition of $\mathcal{L} = \{p, q\} \cup W^s(\mathcal{B}) \cup W^u(\mathcal{B})$, where p and q are fixed attracting and repelling (respectively) points and $\mathcal{B}$ is the set of saddle type periodic points. We now state the main theorem:

Theorem 5. For every $x \in M \setminus \mathcal{L}$,

$$(1) \quad \lim_{n \rightarrow \infty} T_*^n(\mu \times \delta_x) = \mu \times \left(\frac{1}{2} \delta_p + \frac{1}{2} \delta_q \right).$$

Equivalently, for any continuous function $\varphi: \Sigma_2 \times M \rightarrow \mathbf{R}$,

$$(2) \quad \lim_{n \rightarrow \infty} \int \varphi \circ T^n(w, y) d(\mu \times \delta_x) = \int \varphi(w, y) d\left(\mu \times \left(\frac{1}{2} \delta_p + \frac{1}{2} \delta_q\right)\right).$$

Note that by the results in §4, Theorem 5 implies that $T_*^n(\mu \times m)$ and $\frac{1}{n} \sum_{k=0}^{n-1} T_*^k(\mu \times m)$ also converge to $\mu \times (\frac{1}{2} \delta_p + \frac{1}{2} \delta_q)$.

5.2. Simplification. The first step in the proof of Theorem 5 is to show that it suffices to replace φ in (2) with functions of the form $\chi_E \cdot \psi$, where $E \subset \Sigma_2$ is a cylinder set and $\psi: M \rightarrow \mathbf{R}$ is continuous.

We first recall

Theorem 6. Let M be a compact space and $\mathcal{A}$ an algebra of continuous real-valued functions on M which separates points of M and which contains the constant functions. Then, given any continuous real-valued function F on M and any $\varepsilon > 0$, there is a function G in $\mathcal{A}$ such that for all $x \in M$ we have $|G(x) - F(x)| < \varepsilon$, (that is, $\mathcal{A}$ is a dense subset of $C(M)$).

See, for example, Royden [16].

The set $\mathcal{a} = \{\sum_{i=1}^n \theta_i \cdot \psi_i | \theta_i: \Sigma_2 \rightarrow \mathbf{R}, \psi_i: M \rightarrow \mathbf{R} \text{ are continuous}\}$ is an algebra which satisfies the hypotheses of the Stone-Weierstrass Theorem; so to prove Theorem 5, it suffices to show that for all continuous functions, $\theta: \Sigma_2 \rightarrow \mathbf{R}$, $\psi: M \rightarrow \mathbf{R}$ we have

$$\lim_{n \rightarrow \infty} \int \theta(\sigma^n w) \cdot \psi(f^{S_n(w)} y) d(\mu \times \delta_y) = \int \theta(w) \cdot \psi(y) d\left(\mu \times \left(\frac{1}{2} \delta_p + \frac{1}{2} \delta_q\right)\right).$$

Any continuous function can be approximated by simple functions. i.e. functions of the form $\sum_{k=1}^n c_i \chi_{C_i}$ where the c_i are real numbers and the C_i are measurable sets (cf. [16]).

Theorem 7. *If $(M, \mathcal{B}, \mu)$ is a probability space, then a subalgebra $\mathcal{B}_0 \subset \mathcal{B}$ generates $\mathcal{B} \pmod{0}$ if and only if for every $B \in \mathcal{B}$ and $\varepsilon > 0$ there exists a $B_0 \in \mathcal{B}_0$ such that $\mu(B \Delta B_0) \leq \varepsilon$.*

This theorem can be found in various ergodic theory books; see, for example, Mañé [10].

The Borel σ -algebra for the shift space is generated by the algebra of centered cylinder sets. Thus, any measurable set A can be approximated by an element in this algebra, i.e., by a finite union of disjoint cylinder sets, $\bigcup_{i=1}^k \sigma^{j_i}(E_i)$, where E_i is a centered cylinder set of length $2l_i + 1$.

In light of these facts, to show that (2) is true, it suffices to show that for every cylinder set

$$E = \{w \in \Sigma_2 \mid w_{-l} = i_{-l}, \dots, w_0 = i_0, \dots, w_k = i_k \text{ each } i_j = \pm 1\}$$

in Σ_2 , for every continuous function $\psi: M \rightarrow \mathbf{R}$ and for every $x \in M \setminus \mathcal{L}$, we have

$$(3) \quad \begin{aligned} & \lim_{n \rightarrow \infty} \int \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)} y) d(\mu \times \delta_x) \\ &= \int \chi_E(w) \cdot \psi(y) d\left(\mu \times \left(\frac{1}{2}\delta_p + \frac{1}{2}\delta_q\right)\right). \end{aligned}$$

5.3. Iterates of $\mu \times \delta_x$. We now prove Theorem 5 by proving (3). Fix $x \in M \setminus \mathcal{L}$. Let $\varepsilon > 0$ be given, and let E and ψ be as above.

Since ψ is continuous, we can choose $\delta > 0$ so that $|\psi(z_1) - \psi(z_2)| < \varepsilon/6$ whenever $d(z_1, z_2) < \delta$.

For every $y \in M \setminus \mathcal{L}$, there is an $N_1 \in \mathbf{N}$ such that $f^{N_1} y \in B_\delta(p)$ and there is an $N_2 \in \mathbf{N}$ such that $f^{-N_2} y \in B_\delta(q)$.

Therefore, we can choose $N_0 = N_0(\varepsilon, x) \in \mathbf{N}$ so that $f^{N_0} x \in B_\delta(p)$ and $f^{-N_0} x \in B_\delta(q)$.

For each set E described above, we define a partition of Σ_2 into three sets:

$$\begin{aligned} R_n^+ &= R^+(n, l, N_0) = \begin{cases} \{w \in \Sigma_2 \mid S_{n-l}(w) \geq N_0 + l\} & \text{if } l \leq 0, \\ \{w \in \Sigma_2 \mid S_n(w) \geq N_0\} & \text{if } l > 0; \end{cases} \\ R_n^- &= R^-(n, l, N_0) = \begin{cases} \{w \in \Sigma_2 \mid S_{n-l}(w) \leq -N_0 - l\} & \text{if } l > 0, \\ \{w \in \Sigma_2 \mid S_n(w) \leq -N_0\} & \text{if } l \leq 0; \end{cases} \\ R_n^0 &= R^0(n, l, N_0) = \begin{cases} \{w \in \Sigma_2 \mid |S_{n-l}(w)| < N_0 + l\} & \text{if } l > 0, \\ \{w \in \Sigma_2 \mid |S_n(w)| < N_0\} & \text{if } l \leq 0. \end{cases} \end{aligned}$$

As we note later, the sets of the partition are chosen so that each is independent (in the measure-theoretic sense) of $\sigma^{-n}E$.

We remark that

$$(4) \quad \text{if } w \in R_n^+, \text{ then } S_n(w) \geq N_0 \text{ and so } f^{S_n(w)}(x) \in B_\delta(p)$$

and

$$(5) \quad \text{if } w \in R_n^-, \text{ then } S_n(w) \leq -N_0 \text{ and so } f^{S_n(w)}(x) \in B_\delta(q).$$

In the next lemma, we calculate the measures of the sets described above as $n \rightarrow \infty$.

Lemma 2. *For the partition defined above, we have*

$$(6) \quad \lim_{n \rightarrow \infty} \mu(R_n^0) = 0,$$

$$(7) \quad \lim_{n \rightarrow \infty} \mu(R_n^+) = \frac{1}{2},$$

$$(8) \quad \lim_{n \rightarrow \infty} \mu(R_n^-) = \frac{1}{2}.$$

Proof. First, we show that $\mu(R_n^0) \rightarrow 0$. The proof is for $l > 0$, but a similar argument works for $l \leq 0$.

We rewrite R_n^0 as a finite union of disjoint sets, $r_n(i)$, where $r_n(i) = \{w \in \Sigma_2 | S_{n-l}(w) = i\}$.

$$\begin{aligned} R_n^0 &= \{w \in \Sigma_2 | |S_{n-l}(2)| < N_0 + l\} \\ &= \bigcup_{i=-N_0-l+1}^{N_0+l-1} \{w \in \Sigma_2 | S_{n-l}(w) = i\} \\ &= \bigcup_{i=-N_0-l+1}^{N_0+l-1} r_n(i) \end{aligned}$$

We now calculate the measure of each $r_n(i)$:

$$\begin{aligned} \mu(r_n(i)) &= \binom{n-l}{\frac{n-l+i}{2}} (2^{-n(-l)}) \quad \text{where } \binom{n-l}{\frac{n-l+i}{2}} = 0 \text{ if } \frac{n-l+i}{2} \notin \mathbf{N} \\ &\sim \sqrt{\frac{2}{\pi}} \left(\frac{n-l}{n-l-i} \right)^{n-l} \left(\frac{n-l-i}{n-l+i} \right)^{(n-l+i)/2} \left(\frac{n-l}{(n-l)^2 - i^2} \right)^{1/2} \\ &\quad \text{(Stirling's formula).} \end{aligned}$$

Summing over $i \in (-N_0-l, N_0+l)$ and letting $n \rightarrow \infty$, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu(R_n^0) &= \lim_{n \rightarrow \infty} \sum_{i=-N_0-l+1}^{N_0+l-1} \mu(r_n(i)) \\ &\sim \lim_{n \rightarrow \infty} \sum_{i=-N_0-l+1}^{N_0+l-1} \sqrt{\frac{2}{\pi}} \left(\frac{n-l}{n-l-i} \right)^{n-l} \\ &\quad \times \left(\frac{n-l-i}{n-l+i} \right)^{(n-l+i)/2} \left(\frac{n-l}{(n-l)^2 - i^2} \right)^{1/2} \\ &= \sum_{i=-N_0-l+1}^{N_0+l-1} e^i e^{-i} \sqrt{\frac{2}{\pi}} \lim_{n \rightarrow \infty} \left(\frac{n-l}{(n-l)^2 - i^2} \right)^{1/2} \\ &= 0. \end{aligned}$$

Thus, (6) holds. To show (7) and (8), we define the map $\tau: \Sigma_2 \rightarrow \Sigma_2$ by $\tau(w) = -w$; that is, τ is the map which interchanges the 1's and -1's. It is clear that $\tau(R_n^+) = R_n^-$, that τ is a measure-preserving transformation with

respect to the measure μ , and hence that $\mu(R_n^+) = \mu(R_n^-)$. We have

$$\begin{aligned}
 2\mu(R_n^+) &= \mu(R_n^+) + \mu(R_n^+) \\
 &= \mu(R_n^+) + \mu(\tau R_n^+) \quad \text{since } \tau \text{ is measure preserving} \\
 &= \mu(R_n^+) + \mu(R_n^-) \quad \text{since } \tau(R_n^+) = R_n^- \\
 &= \mu(R_n^+ \cup R_n^-) \quad \text{since the two sets are disjoint} \\
 &\rightarrow 1 \text{ as } n \rightarrow \infty \text{ since the } R_n \text{'s partition } \Sigma_2 \text{ and by (6).}
 \end{aligned}$$

So $\mu(R_n^+) \rightarrow 1/2$ and since $\mu(R_n^+) = \mu(R_n^-)$, we have $\mu(R_n^-) \rightarrow 1/2$ also. Thus, Lemma 2 is proved. $\square$

From this lemma and the fact that the measure μ is Bernoulli, we have the following strong mixing type lemma.

Lemma 3.

$$\lim_{n \rightarrow \infty} \int_{R_n^+} \chi_E(\sigma^n w) d\mu(w) = \frac{1}{2} \int \chi_E(w) d\mu(w)$$

and

$$\lim_{n \rightarrow \infty} \int_{R_n^-} \chi_E(\sigma^n w) d\mu(w) = \frac{1}{2} \int \chi_E(w) d\mu(w).$$

Proof. We only give the proof for R_n^+ , as the proof for R_n^- is identical.

$$\begin{aligned}
 \int_{R_n^+} \chi_E(\sigma^n w) d\mu(w) &= \int \chi_{R_n^+}(w) \cdot \chi_{\sigma^{-n}(E)}(w) d\mu(w) \\
 &= \mu(R_n^+ \cap \sigma^{-n}E) \\
 &= \mu(R_n^+) \mu(\sigma^{-n}E) \text{ since } R_n^+ \text{ and } \sigma^{-n}E \text{ are independent} \\
 &= \mu(R_n^+) \mu(E) \text{ since } \sigma \text{ is measure preserving} \\
 &= \mu(R_n^+) \int \chi_E(w) d\mu(w) \\
 &\rightarrow \frac{1}{2} \int \chi_E(w) d\mu(w) \text{ by Lemma 2.}
 \end{aligned}$$

The third equality, that R_n^+ and $\sigma^{-n}E$ are independent, is justified by the following: if $l > 0$, then R_n^+ depends on the coordinates $[0, n-l-1]$ and $\sigma^{-n}E$ depends on the coordinates $[n-l, n+k]$. If $l \leq 0$, then R_n^+ depends on the coordinates $[0, n-1]$ and $\sigma^{-n}E$ depends on $[n-l, n+k]$ (note that $n-l \geq n$, since $l \leq 0$). In either case, the two sets R_n^+ and $\sigma^{-n}E$ are independent and the measure of their intersection is the product of their measures. Thus, Lemma 3 is proved. $\square$

We rewrite the left integral in (3) as three integrals by using the partition of Σ_2 :

$$\begin{aligned}
 & \int \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)} y) d(\mu \times \delta_x) \\
 &= \int_{R_n^+} \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)} x) d\mu(w) \\
 & \quad + \int_{R_n^-} \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)} x) d\mu(w) \\
 & \quad + \int_{R_n^0} \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)} x) d\mu(w) \\
 &= I_n^+ + I_n^- + I_n^0.
 \end{aligned}$$

Let $K = \max_{y \in M} |\psi(y)|$. Choose $J \in \mathbb{N}$ so that if $n \geq J$,

$$(i) \quad \left| \int_{R_n^+} \chi_E(\sigma^n w) d\mu(w) - \frac{1}{2} \int \chi_E(w) d\mu(w) \right| < \varepsilon/6K \text{ (Lemma 3)},$$

$$(ii) \quad \mu(R_n^0) < \varepsilon/3K \text{ (Lemma 2)}.$$

We continue with the proof of (3). By the triangle inequality,

$$\begin{aligned}
 (9) \quad & \left| \int \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)} y) d(\mu \times \delta_x) \right. \\
 & \quad \left. - \int \chi_E(w) \cdot \psi(y) d\left(\mu \times \left(\frac{1}{2}\delta_p + \frac{1}{2}\delta_q\right)\right) \right| \\
 & \leq |I_n^0| + \left| I_n^+ - \int \chi_E(w) \cdot \psi(y) d\left(\mu \times \frac{1}{2}\delta_p\right) \right| \\
 & \quad + \left| I_n^- - \int \chi_E(w) \cdot \psi(y) d\left(\mu \times \frac{1}{2}\delta_q\right) \right|.
 \end{aligned}$$

We need to show that each of the three terms in (9) are small, that is, less than $\varepsilon/3$ for large n .

First for $|I_n^0|$: Let $n \geq J$; then

$$\begin{aligned}
 |I_n^0| &\leq \int_{R_n^0} |\chi_E(\sigma^n w)| \cdot |\psi(f^{S_n(w)} x)| d\mu(w) \\
 &\leq K \cdot \mu(R_n^0) < \varepsilon/3 \text{ as desired.}
 \end{aligned}$$

Next, we show the middle term of (9) is less than $\varepsilon/3$ for $n \geq J$:

$$\begin{aligned}
& \left| I_n^+ - \int \chi_E(w) \cdot \psi(y) d\left(\mu \times \frac{1}{2}\delta_p\right) \right| \\
& \leq \left| \int_{R_n^+} \chi_E(\sigma^n w) \cdot \psi(f^{S_n(w)}x) d\mu(w) - \int_{R_n^+} \chi_E(\sigma^n w) \cdot \psi(p) d\mu(w) \right| \\
& \quad + \left| \int_{R_n^+} \chi_E(\sigma^n w) \cdot \psi(p) d\mu(w) - \int \chi_E(w) \cdot \psi(y) d\left(\mu \times \frac{1}{2}\delta_p\right) \right| \\
& \leq \int_{R_n^+} \chi_E(\sigma^n w) \cdot |\psi(f^{S_n(w)}x) - \psi(p)| d\mu(w) \\
& \quad + \left| \int_{R_n^+} \chi_E(\sigma^n w) \cdot \psi(p) d\mu(w) - \frac{1}{2} \int \chi_E(w) \cdot \psi(p) d\mu(w) \right| \\
& \leq \int_{R_n^+} \varepsilon/3 d\mu(w) \\
& \quad + |\psi(p)| \left| \int_{R_n^+} \chi_E(\sigma^n w) d\mu(w) - \frac{1}{2} \int \chi_E(w) d\mu(w) \right| \\
& < \varepsilon/6 + K \left(\frac{\varepsilon}{6K} \right) \\
& = \varepsilon/3 \quad \text{as desired.}
\end{aligned}$$

That the last term of (9) is also small is clear. So, finally, we have shown that for every point x in $M \setminus \mathcal{L}$, $T_*^n(\mu \times \delta_x)$ converges to $\mu \times (\frac{1}{2}\delta_p + \frac{1}{2}\delta_q)$. By Proposition 4 in §4, we have that $T_*^n(\mu \times m) \rightarrow \mu \times (\frac{1}{2}\delta_p + \frac{1}{2}\delta_q)$ also. We remark here that if $x = p$, then $T_*^n(\mu \times \delta_p)$ converges to $\mu \times \delta_p$, and if $x = q$, then $T_*^n(\mu \times \delta_q)$ converges to $\mu \times \delta_q$. This is easily seen as both $\mu \times \delta_p$ and $\mu \times \delta_q$ are invariant measures for T . $\square$

5.4. Averages of iterates of point mass measures. In this section, we show that the sequence of averages of iterates of point mass measures does not converge by computing all possible weak limit points. We show that the set of limit points of

$$\nu_n(w, x) = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$$

coincides with the set $\{\mu \times (\lambda\delta_p + (1-\lambda)\delta_q), 0 \leq \lambda \leq 1\}$.

Thus, we see that the symmetric skew product being discussed does not have a measure with the property that $\mu \times m$ almost every point is generic with respect to it. Recall the definition, $\mathcal{L} = \{p, q\} \cup W^s(\mathcal{B}) \cup W^u(\mathcal{B})$, where p and q are fixed attracting and repelling (respectively) points and $\mathcal{B}$ is the set of saddle-type periodic points.

Theorem 8. *There is a measurable set $\Omega \subset \Sigma_2$ with $\mu(\Omega) = 1$, such that for every $w \in \Omega$ and every $x \in M \setminus \mathcal{L}$, the set of accumulation points of the sequence $\nu_n(w, x)$ is precisely the set of all measures ν of the form $\nu = \mu \times (\lambda\delta_p + (1-\lambda)\delta_q)$, $0 \leq \lambda \leq 1$.*

Proof. We first show that any limit point of $\nu_n(w, x)$ is a product measure with μ the measure in the base and a linear combination of point masses of q and p as the fiber measure. We then show that any measure of the form $\mu \times (\lambda\delta_p + (1-\lambda)\delta_q)$ can be realized as a subsequential limit of $\nu_n(w, x)$.

For the first part of the proof, we first show the convergence of the sequence $\nu_n(w, x)$ is independent of $x \in M \setminus \mathcal{L}$ (Claim 1). We then compute $(\pi_2)_* \nu$ where ν is any limit point of ν_n . That $(\pi_1)_* \nu = \mu$ is clear. In the course of these proofs we will note that the projection of the limiting measure on the fiber is independent of the base point, hence any limit point of ν_n is product measure. First we show the independence of $x \in M \setminus \mathcal{L}$.

Claim 1. For each $w \in \Sigma_2$ and for every $x, y \in M \setminus \mathcal{L}$, $\nu_n(w, x)$ converges if and only if $\nu_n(w, y)$ converges; that is, the convergence does not depend on the choice of $x \in M \setminus \mathcal{L}$.

Proof of Claim 1. We show that the difference between the two sequences tends to zero as n tends to infinity. By the discussion in §5.2, it suffices to show that for any cylinder set $E \subset \Sigma_2$, any continuous function $\psi: M \rightarrow \mathbb{R}$, and any $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that for $n \geq N$,

$$\left| \int \chi_E(t) \cdot \psi(z) d\nu_n(w, x) - \int \chi_E(t) \cdot \psi(z) d\nu_n(w, y) \right| < \varepsilon.$$

Let $\varepsilon > 0$ be given. Let $K = \max_{z \in M} |\psi(z)|$.

Choose $\delta > 0$ so that if $d(x, y) < \delta$, then $|\psi(x) - \psi(y)| < \varepsilon/2$.

Let N_1 be such that if $n \geq N_1$, then $f^n x, f^n y \in B_{\delta/2}(p)$ and $f^{-n}x, f^{-n}y \in B_{\delta/2}(q)$.

Let N_2 be such that if $n \geq N_2$, then $\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} < \varepsilon/4K$ ($2N_1 = 1$) for every $i \in (-N_1, N_1)$. That such an N_2 can be chosen follows for Proposition 1 in §3.

Note that $\sum_{i \in \mathbb{Z}} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} = 1$, so for any $J \subset \mathbb{Z}$,

$$\sum_{i \in J} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \leq 1.$$

Let $N = \max(N_1, N_2)$, and suppose $n \geq N$; then

$$\begin{aligned} & \left| \int \chi_E(t) \cdot \psi(z) d\nu_n(w, x) - \int \chi_E(t) \cdot \psi(z) d\nu_n(w, y) \right| \\ &= \left| \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) \cdot \psi(f^{S_k(w)} x) - \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) \cdot \psi(f^{S_k(w)} y) \right| \\ &\leq \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) |\psi(f^{S_k(w)} x) - \psi(f^{S_k(w)} y)| \\ &\leq \frac{1}{n} \sum_{k=0}^{n-1} |\psi(f^{S_k(w)} x) - \psi(f^{S_k(w)} y)| \end{aligned}$$

$$\begin{aligned}
&= \sum_{i \in \mathbb{Z}} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} |\psi(f^i x) - \psi(f^i y)| \\
&\leq \sum_{i \geq N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} |\psi(f^i x) - \psi(f^i y)| \\
&\quad + \sum_{i \leq -N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} |\psi(f^i x) - \psi(f^i y)| \\
&\quad + \sum_{i=-N+1}^{N-1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} |\psi(f^i x) - \psi(f^i y)| \\
&< \sum_{i \geq N_1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \varepsilon/2 + \sum_{i \leq -N_1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \varepsilon/2 \\
&\quad + 2K \sum_{i=-N_1+1}^{N_1-1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \\
&< \varepsilon/2 + 2K(2N_1 - 1) \cdot \varepsilon/4K(2N_1 - 1) \\
&= \varepsilon.
\end{aligned}$$

Thus, the convergence of the sequence does not depend on the choice of $x \in M \setminus \mathcal{L}$. $\square$

We now compute the projections of the limit measure ν . The first step is to derive an explicit form for $(\pi_2)_* \nu$, where ν is any limit of $\nu_n(w, x)$. That is, $\nu = \lim_{k \rightarrow \infty} \nu_{n_k}(w, x)$. We do this by considering the support of the measures ν_n .

From the definition of ν_n , we see that

$$\text{supp } \nu_n \subset \{\sigma^k w\}_{k=0}^{n-1} \times \{f^{S_k(w)} x\}_{k=0}^{n-1}.$$

Taking limits, we have $\text{supp } \nu \subset \Sigma_2 \times \overline{\{f^{S_k(w)} x\}_{k \in \mathbb{N}}} = \Sigma_2 \times \overline{\{f^i x\}_{i \in \mathbb{Z}}}$, since for almost every $w \in \Sigma_2$, $S_n(w)$ takes on every integer value infinitely often. Note that the support of $(\pi_2)_* \nu$ is in the closure of the orbit of x , which is a countable set. Hence, $(\pi_2)_* \nu$ is a linear combination of point mass measures, i.e.,

$$(10) \quad (\pi_2)_* \nu = \sum_{i \in \mathbb{Z}} \lambda_i(w) \delta_{f^i x} + \lambda_p(w) \delta_p + \lambda_q(w) \delta_q.$$

We now find the $\lambda_i(w)$. Again, recall the definition of ν_n :

$$\nu_n = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)} = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{\sigma^k w} \cdot \delta_{f^{S_k(w)} x}.$$

Therefore,

$$\begin{aligned}
 (\pi_2)_* \nu_n &= \frac{1}{n} \sum_{k=0}^{n-1} \delta_{f^{S_k(w)}x} = \frac{1}{n} \sum_{i=-n}^n \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x} \\
 &= \sum_{i \geq N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x} + \sum_{i \leq -N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x} \\
 &\quad + \sum_{i=-N+1}^{N-1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x}
 \end{aligned}$$

where $\mathbf{I}_{[S_k(w)=i]} = 1$ if $S_k(w) = i$ and is zero otherwise.

Claim 2. For each $j \in \mathbf{Z}$,

$$(11) \quad \lambda_j(w) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=j]}.$$

Proof. By assumption, we have that

$$\begin{aligned}
 (\pi_2)_* \nu &= \lim_{n \rightarrow \infty} \left(\sum_{i \geq N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x} \right. \\
 &\quad \left. + \sum_{i \leq -N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x} + \sum_{i=-N+1}^{N-1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \delta_{f^i x} \right).
 \end{aligned}$$

Then, by combining this with (10), we have, for every continuous function $\varphi: M \rightarrow \mathbf{R}$,

$$\begin{aligned}
 (12) \quad \lim_{n \rightarrow \infty} &\left(\sum_{i \geq N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \varphi(f^i x) + \sum_{i \leq -N} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \varphi(f^i x) \right. \\
 &\quad \left. + \sum_{i=-N+1}^{N-1} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=i]} \varphi(f^i x) \right) \\
 &= \sum_{i \in \mathbf{Z}} \lambda_i(w) \varphi(f^i x) + \lambda_p(w) \varphi(p) + \lambda_q(w) \varphi(q).
 \end{aligned}$$

To show that Claim 2 is true, we define a continuous “bump” function at $f^j x$. Let $j \in \mathbf{Z}$, and let $\varepsilon > 0$ be such that $\overline{B_\varepsilon(f^j x)} \cap \{f^i x\}_{i \in \mathbf{Z}} = f^j x$; that is, $B_\varepsilon(f^j x)$ is a neighborhood of $f^j x$ which does not contain any other point in the closure of the orbit of x .

Let

$$\theta(t) = \begin{cases} 1 & \text{if } t \leq 0, \\ 1-t & \text{if } 0 \leq t \leq 1, \\ 0 & \text{if } t \geq 1. \end{cases}$$

Define $\varphi_j: M \rightarrow \mathbf{R}$ by $\varphi_j(y) = \theta(\frac{1}{\varepsilon} \cdot d(y, f^j x))$. So φ_j is a continuous function which is zero outside of $B_\varepsilon(f^j x)$ and is one at $f^j x$. Substituting φ_j in (12)

we have, for each $j \in \mathbb{Z}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w)=j]} = \lambda_j(w) \quad \text{as desired.} \quad \square$$

Recall Proposition 1 in §3, which states that the average occupation time of the symmetric random walk in any finite state j tends to zero as n tends to infinity. Hence, $\lambda_j(w) = 0$ for each finite j ; that is, the random walk spends on average, all its time either at $+\infty$ or $-\infty$. Therefore,

$$(\pi_2)_* \nu = \lambda(w) \delta_p + (1 - \lambda(w)) \delta_q,$$

where

$$(13) \quad \lambda(w) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{I}_{[S_k(w) \geq N]} \quad \text{for some large integer } N,$$

is the average occupation time of the symmetric random walk “at infinity”. Then $1 - \lambda(w)$ is the average occupation time at negative infinity. To see that any limit is indeed product measure, we use the notion of the disintegration of a measure.

Theorem 9 [7, p. 108]. *If $\varphi \in L^1(\Sigma_2 \times M, \nu)$, then*

$$\int \varphi d\nu = \int \left\{ \int \varphi d\nu_t \right\} d\mu(t).$$

From this and the preceding discussion of the projection measures, it is clear that ν_t is independent of t and depends only on the point $w \in \Sigma_2$ which is being iterated. Hence, ν is the product measure.

To compute the proof of Theorem 8, we show that for μ -almost every w , and every $\lambda \in [0, 1]$, we can construct a subsequence $\{n_i\}$ depending on w and λ , such that

$$\lim_{i \rightarrow \infty} \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) > 0]} = \lambda \quad \text{and} \quad \lim_{i \rightarrow \infty} \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) < 0]} = 1 - \lambda.$$

To accomplish this, we make use of Strassen's Theorem [22]. Before stating this theorem, we introduce some notation.

As before, $S_n(w) = \sum_{k=0}^{n-1} w_k$.

For each w in Σ_2 , define the linear interpolation of $S_n(w)$ at n to be

$$\eta^w(u) = ([u] + 1 - u)S_{[u]}(w) + (u - [u])S_{[u]+1}(w),$$

where $[u]$ is the greatest integer less than u .

The graph of η^w is the same as the piecewise linear graph obtained by connecting the points $\{(n, S_n(w))\}_{n \geq 0}$ in the plane. For each n , define $\eta_n^w: [0, 1] \rightarrow [-1, 1]$ by

$$\eta_n^w(t) = (2n \log \log n)^{-1/2} \eta^w(nt).$$

The graph of η_n^w is the same as that of η^w except that it is rescaled by $1/n$ in the horizontal direction and by $(2n \log \log n)^{-1/2}$ in the vertical. That the

range of η_n^w is $[-1, 1]$ for μ almost every w follows from the Law of the Iterated Logarithm (see, for example, [15, p. 98]). This theorem asserts that

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1$$

and

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = -1.$$

If we denote by C the Banach space of continuous maps from $[0, 1]$ to $\mathbb{R}$ with the sup norm $\|\cdot\|$, then η_n is a random variable on Σ_2 with values in C ; that is, for each $w \in \Sigma_2$, $\eta_n(w) = \eta_n^w$ is in C . Define the subset K of C as follows:

$$K = \left\{ g \in C \mid g \text{ is absolutely continuous, } g(0) = 0, \text{ and } \int_0^1 (g'(t))^2 dt \leq 1 \right\}.$$

In what follows, we suppress the w in the function η for typing ease. In this setting, we can now state

Theorem 10 (Strassen [22]). *With probability one, the sequence $\{\eta_n\}_{n \geq 3}$ is relatively norm compact and the set of its norm limit points coincides with K .*

In other words, there is a set $\Omega \subset \Sigma_2$, $\mu(\Omega) = 1$, such that for every $w \in \Omega$, for every $\varepsilon > 0$, and for every $g \in K$, there is a subsequence $\{n_i\}$ and an N such that if $n_i \geq N$, then $\|\eta_{n_i} - g\| < \varepsilon$.

Let $w \in \Omega$, and let $\lambda \in (0, 1)$. The cases $\lambda = 0, 1$ are treated separately. We now construct a subsequence $\{n_i\}$ so that

$$\frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) > 0]} \rightarrow \lambda \quad \text{and} \quad \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) < 0]} \rightarrow 1 - \lambda.$$

Let $\varepsilon > 0$ be given. Choose J larger than both $\frac{\ln(6/\varepsilon)}{\ln 9}$ and $\frac{\ln(6/\lambda)}{\ln 9}$. For each $i \geq J$, define g_i as follows:

$$g_i(t) = \begin{cases} \frac{a_i}{\delta_i} t & \text{if } 0 \leq t \leq \delta_i, \\ a_i & \text{if } \delta_i \leq t \leq \lambda - \delta_i, \\ \frac{-a_i}{\delta_i} t + \frac{a_i}{\delta_i} \lambda & \text{if } \lambda - \delta_i \leq t \leq \lambda + \delta_i, \\ -a_i & \text{if } \lambda + \delta_i \leq t \leq 1, \end{cases}$$

where $a_i = \frac{1}{3^i}$ and $\delta_i = \frac{3}{9^i}$.

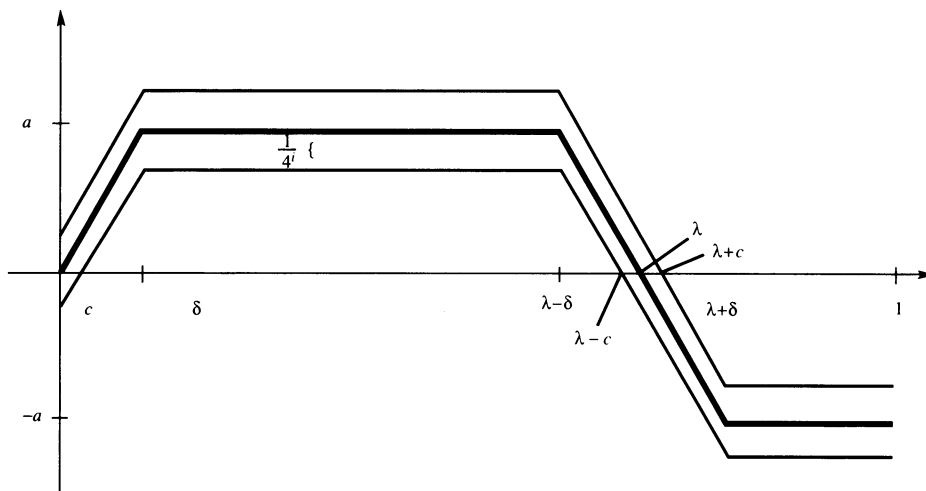
It is easy to verify that $g_i \in K$. We sketch the graph of a typical g_i in Figure 3.

Choose n_i so that $\|\eta_{n_i} - g_i\| < 1/4^i$. Then the largest set on which η_{n_i} can be positive is the interval $(0, \lambda + \delta_i/a_i 4^i)$ and the smallest is the interval $(\delta_i/a_i 4^i, \lambda - \delta_i/a_i 4^i)$.

The largest set on which η_{n_i} can be negative is the union of the two intervals $(0, \delta_i/a_i 4^i)$ and $(\lambda - \delta_i/a_i 4^i, 1)$, and the smallest is the interval $(\lambda + \delta_i/a_i 4^i, 1)$.

Thus,

$$\lambda - \frac{2\delta_i}{a_i 4^i} \leq m\{t \in [0, 1] \mid \eta_{n_i}(t) > 0\} \leq \lambda + \frac{\delta_i}{a_i 4^i}$$



$$a = a_i, \delta = \delta_i, \text{ and } c = \delta_i / a_i 4^i$$

FIGURE 3

and

$$1 - \lambda - \frac{\delta_i}{a_i 4^i} \leq m\{t \in [0, 1] | \eta_{n_i}(t) > 0\} \leq 1 - \lambda + \frac{2\delta_i}{a_i 4^i}.$$

It is clear that if t is in the interval $(\delta_i / a_i 4^i, \lambda - \delta_i / a_i 4^i)$, then $\eta_{n_i}(t) > 0$ and also $\eta(n_i t) > 0$. However, if t is very close to $\delta_i / a_i 4^i$, then it is possible that $\eta(\lceil n_i t \rceil) \leq 0$. To avoid this, we restrict t to the interval $(\delta_i, \lambda - \delta_i / a_i 4^i)$. Then, as long as $|\delta_i - \delta_i / a_i 4^i| > 1/n_i$, we have $\eta(\lceil n_i t \rceil) > 0$.

If $k \in (\lceil \eta_i \delta_i \rceil, \lceil (\lambda - \delta_i / a_i 4^i) \eta_i \rceil)$, then $S_k > 0$. This implies that

$$(14) \quad \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) > 0]} \geq \lambda - \frac{\delta_i}{a_i 4^i} - \delta_i.$$

On the other hand, $S_k > 0$ for $0 \leq k < n_i$ implies that $\eta(k) > 0$ and hence that $\eta_{n_i}(k/n_i) > 0$. The length of the largest interval for which $\eta(k) > 0$ is $\lambda + \delta_i / a_i 4^i$, so

$$(15) \quad \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) > 0]} \leq \lambda + \frac{\delta_i}{a_i 4^i}.$$

Note that $\delta_i / a_i 4^i = 3/12^i$. Combining this with (14) and (15) gives

$$\left| \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) > 0]} - \lambda \right| < \frac{6}{9^i}.$$

Using similar arguments, we can also show that

$$\left| \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{I}_{[S_k(w) < 0]} - (1 - \lambda) \right| < \frac{6}{9^i}.$$

As $6/9^i < \varepsilon$, we have the desired conclusion. The cases $\lambda = 0, 1$ are treated similarly; however, different functions from K must be used. For example, if

$\lambda = 1$, then $h_i \in K$ defined below works:

$$h_i(t) = \begin{cases} \frac{a_i}{\delta_i} t & \text{if } 0 \leq t \leq \delta_i, \\ a_i & \text{if } \delta_i \leq t \leq 1. \end{cases}$$

This completes the proof of Theorem 8. $\square$

6. THE BIASED NORTH POLE–SOUTH POLE DIFFEOMORPHISM

In this section we turn to the case where the measure in the base space μ is biased; that is, $\mu = B(\alpha, 1 - \alpha)$ with $\alpha \neq 1/2$. In §6.1 we show the convergence of the sequence $T_*^n(\mu \times m)$, and in §6.2 we show the convergence of the sequence $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$.

Theorem 11. *Let the hypotheses be the same as in Theorem 5 in §5, except now let $\mu = B(\alpha, 1 - \alpha)$. Then for $x \in M \setminus \mathcal{L}$,*

$$(16) \quad T_*^n(\mu \times \delta_x) \rightarrow \mu \times \delta_p \quad \text{if } 1/2 < \alpha \leq 1$$

and

$$(17) \quad T_*^n(\mu \times \delta_x) \rightarrow \mu \times \delta_q \quad \text{if } 0 \leq \alpha < 1/2.$$

We first prove the analog of Lemma 2 for this nonsymmetric case. The following lemma states that if the measure on the shift space is not symmetric, then the cumulative sums of the first n coordinates of most of the elements in Σ_2 tend to either ∞ (in case $\alpha > 1/2$) or $-\infty$ (in case $\alpha < 1/2$) as $n \rightarrow \infty$.

Lemma 4. *Let (Σ_2, σ) be the shift on the two symbols $+1, -1$ with Bernoulli measure $\mu = B(\alpha, 1 - \alpha)$. Let $N \in \mathbb{Z}$, and let $S_n(w) = \sum_{k=0}^{n-1} w_k$.*

For each $n \in \mathbb{N}$, define the following partition of Σ_2 : $\Sigma_2 = R_n^+ \cup R_n^-$, where $R_n^+ = R^+(n, N) = \{w \in \Sigma_2 | S_n(w) \geq N\}$ and $R_n^- = R^-(n, N) = \{w \in \Sigma_2 | S_n(w) < N\}$.

$$(18) \quad \text{If } 1/2 < \alpha \leq 1, \text{ then } \lim_{n \rightarrow \infty} \mu(R_n^+) = 1 \text{ and } \lim_{n \rightarrow \infty} \mu(R_n^-) = 0.$$

$$(19) \quad \text{If } 0 \leq \alpha < 1/2, \text{ then } \lim_{n \rightarrow \infty} \mu(R_n^+) = 0 \text{ and } \lim_{n \rightarrow \infty} \mu(R_n^-) = 1.$$

Proof. To prove this lemma we show that (18) is true and by symmetry we easily see that (19) is also true.

To prove (18), we show that $\mu(R_n^+) \rightarrow 1$; and since for each n , $\{R_n^+, R_n^-\}$ is a partition of Σ_2 , it follows that $\mu(R_n^-) \rightarrow 0$.

The expected value of S_n is positive since

$$E(S_n) = \sum_{k=0}^{n-1} \alpha - (1 - \alpha) = n(2\alpha - 1)$$

and $\alpha > 1/2$. Thus, in this biased case, the random walk is transient. We have the following lemma:

Lemma 5 (Schmidt [18]). *The following are equivalent:*

1. S_n is transient.
2. $\lim_{n \rightarrow \infty} S_n(w) = \infty$ for μ almost every $w \in \Sigma_2$.

In the above lemma, ∞ is either $\pm\infty$ depending on whether α is greater than $1/2$ or less than $1/2$. From this, it is clear that $\mu\{w \in \Sigma_2 | S_n(w) \geq N\} \rightarrow$

1 as n increases without bound and hence (18) holds. By symmetry, (19) is also proved. $\square$

Using this lemma, we can now prove the theorem. We prove (16) and observe that (17) follows by symmetry. Let $x \in M \setminus \mathcal{L}$. We show that for any continuous function $\varphi: \Sigma_2 \times M \rightarrow \mathbf{R}$,

$$\int \varphi(\sigma^n w, f^{S_n(w)} x) d\mu(w) \rightarrow \int \varphi(w, y) d(\mu(w) \times \delta_p(y)) \quad \text{as } n \rightarrow \infty.$$

Let $\varepsilon > 0$ be given. Let $\varphi: \Sigma_2 \times M \rightarrow \mathbf{R}$ be continuous.

Let $K = \max\{|\varphi(t, y)| : (t, y) \in \Sigma_2 \times M\}$.

Choose $\delta > 0$ such that $|\varphi(t, y) - \varphi(w, z)| < \varepsilon/2$ whenever

$$d((t, y), (w, z)) < \delta.$$

Choose $N_0 \in \mathbf{N}$ such that $f^{N_0} x \in B_\delta(p)$, and define $R^+(n, N_0) = R_n^+$ and $R^-(n, N_0) = R_n^-$ as in Lemma 4. Note that if $w \in R_n^+$, then $f^{S_n(w)} x \in B_\delta(p)$. Choose $L \in \mathbf{N}$ such that if $n \geq L$, $\mu(R_n^-) < \varepsilon/4K$. This can be done by Lemma 4. The following string of inequalities proves Theorem 11:

$$\begin{aligned} & \left| \int \varphi(\sigma^n w, f^{S_n(w)} x) d\mu(w) - \int \varphi(w, y) d(\mu(w) \times \delta_p(y)) \right| \\ & \leq \left| \int \varphi(\sigma^n w, f^{S_n(w)} x) - \varphi(\sigma^n w, p) d\mu(w) \right| \\ & \quad + \left| \int \varphi(\sigma^n w, p) d\mu(w) - \int \varphi(w, p) d\mu(w) \right| \\ & \leq \left| \int_{R_n^+} \varphi(\sigma^n w, f^{S_n(w)} x) - \varphi(\sigma^n w, p) d\mu(w) \right| \\ & \quad + \left| \int_{R_n^-} \varphi(\sigma^n w, f^{S_n(w)} x) - \varphi(\sigma^n w, p) d\mu(w) \right| + 0 \\ & \leq \int_{R_n^+} |\varphi(\sigma^n w, f^{S_n(w)} x) - \varphi(\sigma^n w, p)| d\mu(w) + 2K\mu(R_n^-) \\ & < \mu(R_n^+) \left(\frac{\varepsilon}{2} \right) + 2K\mu(R_n^-) \\ & < \frac{\varepsilon}{2} + 2K \left(\frac{\varepsilon}{4K} \right) \quad \text{if } n \geq L \\ & = \varepsilon. \quad \square \end{aligned}$$

6.1. Averages of iterates of point mass measures. We now show that the analog of the BRS measure exists for the biased case.

Theorem 12. For μ almost every w in Σ_2 and for every $x \in M \setminus \mathcal{L}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)} = \mu \times \delta_p \quad \text{if } \alpha > \frac{1}{2}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)} = \mu \times \delta_q \quad \text{if } \alpha < \frac{1}{2}.$$

Proof. We show the first part of the theorem ($\alpha > 1/2$).

Let $\varepsilon > 0$ be given. Let E be a cylinder set in Σ_2 and $\psi: M \rightarrow \mathbf{R}$ a continuous function. Let δ be such that if $d(x, y) < \delta$, then $|\psi(x) - \psi(y)| < \varepsilon/2$. Let L be such that for every $n \geq L$,

- (i) $f^n x \in B_\delta(p)$ and
- (ii)

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) - \int \chi_E(t) d\mu \right| < \frac{\varepsilon}{K}.$$

This can be done as the measure μ is ergodic.

Then, using Lemma 4, choose N depending on L to be such that $S_n(w) \geq L$ for every $n \geq N$.

Let $K = \max_{x \in M} |\psi(x)|$. Let J be such that if $n \geq J$, $1/n \leq \varepsilon/(4K \cdot N)$.

Now suppose n is larger than L , N , and J , then we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) \psi(f^{S_n(w)} x) - \int \chi_E(t) \psi(p) d\mu \right| \\ & \leq \left| \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) \psi(f^{S_n(w)} x) - \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) \psi(p) \right| \\ & \quad + \left| \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) \psi(p) - \int \chi_E(t) \psi(p) d\mu \right| \\ & = A + B. \end{aligned}$$

We show that A and B are both less than ε , and hence the theorem will be proved. First for A :

$$\begin{aligned} A & \leq \left| \frac{1}{n} \sum_{k=0}^{N-1} \chi_E(\sigma^k w) (\psi(f^{S_n(w)} x) - \psi(p)) \right| \\ & \quad + \left| \frac{1}{n} \sum_{k=N}^{n-1} \chi_E(\sigma^k w) (\psi(f^{S_n(w)} x) - \psi(p)) \right| \\ & \leq \frac{1}{n} \sum_{k=0}^{N-1} \chi_E(\sigma^k w) |\psi(f^{S_n(w)} x) - \psi(p)| \\ & \quad + \frac{1}{n} \sum_{k=N}^{n-1} \chi_E(\sigma^k w) |\psi(f^{S_n(w)} x) - \psi(p)| \\ & \leq \frac{1}{n} \sum_{k=0}^{N-1} 2K + \frac{1}{n} \sum_{k=N}^{n-1} \frac{1}{n} \varepsilon/2 \\ & = \frac{1}{n} (2K)N + \frac{\varepsilon}{2n} (n - N) \\ & < \varepsilon. \end{aligned}$$

Now for B :

$$\begin{aligned} B &\leq \left| \psi(p) \left(\frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) - \int \chi_E(t) d\mu(t) \right) \right| \\ &\leq K \left| \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(\sigma^k w) - \int \chi_E(t) d\mu(t) \right| \\ &< K \left(\frac{\varepsilon}{K} \right) = \varepsilon. \end{aligned}$$

The second part of Theorem 12 is proved similarly. $\square$

Thus, we have convergence of the averages of iterates of point mass measures for the biased case, which is the analog of the existence of the BRS measure for this example.

7. CONCLUSIONS

We have shown in the preceding pages that if the skew product of the shift space with the generalized north pole–south pole diffeomorphism is biased, then we have the analog of the BRS measure; that is, we have shown that not only does the sequence $T_*^n(\mu \times m)$ converge, but so does the sequence $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(w, x)}$ for $\mu \times m$ almost every $(w, x) \in \Sigma_2 \times M$. So, just as in the hyperbolic theory, we have that the averages of iterates of point mass measures converge. On the other hand, if the system is unbiased, we have shown that the analog of the BRS measure (the limit of averages of point mass measures) does not exist. We do, however, construct all possible weak limit points. Here, we still have the convergence of the sequence $T_*^n(\mu \times m)$.

In [12], some of these results have been generalized to the case where the fiber space is any Morse-Smale diffeomorphism. It is shown that in the unbiased case $T_*^n(\mu \times \nu)$ converges under certain conditions. Otherwise, we have shown that there are at most two weak limit points. Here ν is a measure absolutely continuous with respect to Lebesgue. Several “residue” results have been established; specifically, it is shown that the values of the symmetric random walk are equally distributed over the residue classes of a positive integer.

ACKNOWLEDGMENT

This paper is part of the author’s thesis completed at the University of North Carolina at Chapel Hill. I wish to acknowledge several people for their assistance. I thank my advisor, Sheldon Newhouse, for giving me access to this exciting area of mathematics and supplying me with a great research problem. He has been an invaluable resource to me. I am grateful to the members of my committee, S. Goodman, J. Hawkins, C. Ji, and K. Petersen, for always being helpful, available, and willing to listen. I also thank Máté Wierdl for numerous discussions which were immensely helpful. I owe a special debt to Dan Rudolph who took the time and effort to be interested in my work. In addition to many useful and interesting conversations, he provided several valuable ideas as well as inspiration.

REFERENCES

1. J. Aaronson, M. Denker, and A. Fisher, *Second order ergodic theorems for ergodic transformations of infinite measure spaces*, Proc. Amer. Math. Soc. (to appear).
2. R. Abraham and S. Smale, *Nongenericity of Ω -stability*, Proc. Sympos. Pure Math., vol. 14, Amer. Math. Soc., Providence, RI, 1970, pp. 5–8.
3. R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Math., vol. 470, Springer-Verlag, 1975.
4. ———, *On Axiom A diffeomorphisms*, CBMS Regional Conf. Ser. Math., no. 35, Amer. Math. Soc., Providence, RI, 1977.
5. K. L. Chung, *Elementary probability theory with stochastic processes*, 3rd ed., Springer-Verlag, 1979.
6. W. Feller, *An introduction to probability theory and its applications*, Vol. 1, 3rd ed., Wiley, 1968.
7. H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton Univ. Press, 1981.
8. M. Hirsch, C. Pugh, and M. Shub, *Invariant manifolds*, Lecture Notes in Math., vol. 583, Springer-Verlag, 1970.
9. Y. Kifer, *Random perturbations of dynamical systems*, Birkhäuser, 1988.
10. R. Mañé, *Ergodic theory and differentiable dynamics*, Springer-Verlag, 1987.
11. B. Marcus and S. Newhouse, *Measures of maximal entropy for a class of skew products*, Lecture Notes in Math., vol. 729, Springer-Verlag, 1979, pp. 105–124.
12. D. K. Molinek, *Asymptotic measures for skew products of Bernoulli shifts with Morse-Smale diffeomorphisms*, Ph.D. Thesis, Univ. of North Carolina at Chapel Hill, 1992.
13. S. Newhouse, *Lectures on dynamical systems*, Progr. in Math., vol. 8, Birkhäuser, 1978, pp. 1–114.
14. S. Newhouse and L.-S. Young, *Dynamics of certain skew products*, Lecture Notes in Math., vol. 1007, Springer-Verlag, 1981, pp. 611–629.
15. K. Petersen, *Ergodic theory*, Cambridge Studies in Adv. Math., vol. 2, Cambridge Univ. Press, 1983.
16. H. L. Royden, *Real analysis*, 2nd ed., Macmillan, 1968.
17. D. Ruelle, *A measure associated with Axiom-A attractors*, Amer. J. Math. **98** (1976), 619–654.
18. K. Schmidt, *Cocycles on ergodic transformation groups*, Macmillan, 1977.
19. A. N. Shiryaev, *Probability*, Graduate Texts in Math., vol. 95, Springer-Verlag, 1984.
20. Ya. G. Sinai, *Gibbs measures in ergodic theory*, Russian Math. Surveys **166** (1972), 21–69.
21. S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.
22. V. Strassen, *An invariance principle for the law of the iterated logarithm*, Z. Wahrsch. **3** (1964), 211–226.
23. P. Walters, *An introduction to ergodic theory*, Graduate Texts in Math., vol. 79, Springer-Verlag, 1982.
24. L.-S. Young, *Stochastic stability of hyperbolic attractors*, Ergodic Theory Dynamical Systems **6** (1986), 311–319.

DEPARTMENT OF MATHEMATICS, DAVIDSON COLLEGE, DAVIDSON, NORTH CAROLINA 28036
 E-mail address: domolinek@apollo.davidson.edu