EQUIVALENT NOTIONS OF ENTROPY UNDER ERGODICITY

Abstract. In this paper, we first discuss ergodicity and give a proof of the Birkhoff Ergodic Theorem. We then show how the Birkhoff Ergodic Theorem affects results in measure-theoretical entropy, giving a proof of the Brin-Katok formula for local entropy. Finally, we show that on expanding C^2 maps on the unit circle, the Lyapunov exponent is equal to the entropy. We assume the reader has an understanding of Lebesgue integration and measure theory, but an appendix is included with background material.

Contents

1. Introduction

We begin with a motivating exercise: the flipping of a fair coin. Although coin flipping is the archetypal probability exercise, how it can be rigorously modelled may be unclear. As we will demonstrate, measure theory presents various structures to represent the events of coin flipping, and ergodic theory gives further results about the probabilities of those events.

Question 1.1. *How can we model the flipping of a fair coin?*

Suppose we were to flip a fair coin every second for eternity[1]. Representing heads as 0 and tails as 1, the series of outcomes can be written as a sequence of 0s and 1s i.e. a binary sequence. The probability of heads is equivalent to how frequently we encounter 0 in any binary sequence.

[1] Sisyphus: The Probabilist Version

Let $x_1x_2x_3...$ be a binary sequence. The average frequency of 0s in this sequence can be expressed as

$$\lim_{n\to\infty} \frac{\#\text{ of times 0 appears in first } n \text{ digits of } x_1x_2x_3...}{n}. \tag{1.2}$$

To represent the binary sequence as a single real number, let

$$x = \sum_{i=1}^{\infty} \frac{x_i}{2^i} = \frac{x_1}{2} + \frac{x_2}{2^2} + \frac{x_3}{2^3} + ..., \tag{1.3}$$

Call x the *unit representation* of $x_1x_2x_3....$ We now introduce the doubling map, the model of coin flipping. Let $T : [0, 1) \to [0, 1)$ be the *doubling map*, defined as

$$T(x) = \begin{cases} 2x & 0 \leq x < \frac{1}{2} \\ 2x - 1 & \frac{1}{2} \leq x < 1. \end{cases}$$

By applying T to x, we can "shift" the binary sequence to the left by one space i.e. $T(x)$ is the unit representation of the binary sequence $x_2x_3x_4...$. To check the number of times 0 appears in $x_1x_2x_3...$, we can iterate through the sequence with the doubling map and check if the first digit of each iterate is 0. If we let $\phi_x = 1$ if x_1 is 0, and 0 otherwise, we can simplify (1.2) to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi_{T^i(x)}. \tag{1.4}$$

One of the main results of this paper, the Birkhoff Ergodic Theorem, shows that (1.4) exists and equals $\frac{1}{2}$. Coincidentally, notice that the unit representation of every binary sequence that starts with 0 lies in the half interval $[0, \frac{1}{2})$. As we will see, the half interval and its length is intimately related to (1.4).

In the material to follow, we first introduce ergodicity and prove the Birkhoff Ergodic Theorem. Then, we present measure-theoretical entropy and use the Birkhoff Ergodic Theorem to prove the Brin-Katok local entropy formula. Finally, we show that a different characterization of entropy, the Lyapunov exponent, is equivalent to the Brin-Katok formulation of entropy in the context of expanding C^2 maps on the unit circle.

2. Ergodicity

Consider a set that is "self-contained" i.e. all of its points travel within itself, and no points go in or out of it. Ergodicity is the property that if a "self-contained" set exists, then it is equal in measure to the entire set or a null set. Imagine a barista making a latte. He/she pours milk into a dark espresso, and swirls until the color becomes homogenous. The swirling is an ergodic transformation, because the milk does not stay in one place - instead, it permeates throughout the entire liquid.

Below, we define the notion of a "self-contained" (or almost "self-contained") set and formally define ergodicity. For the two definitions, let $(X, S(X), \mu)$ be a probability space.

Definition 2.1. Let $T : X \to X$ be a transformation. Then $A \subset X$ is called *strictly T-invariant*[2] if $A = T^{-1}(A)$. If $A = T^{-1}(A) \mod \mu$, then A is called *strictly T-invariant* $\mod \mu$.[3]

Definition 2.2. Let $T : X \to X$ be a measure-preserving transformation. Then T is said to be ergodic if for any measurable $A \subset X$ that is strictly T-invariant, $\mu(A) = 0$ or $\mu(A^c) = 0$. (equivalently, $\mu(A) = 1$ or $\mu(A^C) = 1$.)

The doubling map presented in the introduction is an example of an ergodic transformation.

Theorem 2.3. *Let $T : [0, 1) \to [0, 1)$ be defined as $T(x) = 2x \mod 1$. Then T is an ergodic measure-preserving transformation on $[0, 1)$.*

Proof. We first show that T is a measure-preserving transformation.

To show that T is measurable, let A be a Lebesgue measurable set. Define $A_0 = T^{-1}(A) \cap [0, \frac{1}{2})$, and $A_1 = T^{-1}(A) \cap [\frac{1}{2}, 1)$. For real numbers s, t, let $sA + t$ denote the set $\{sa + t \mid a \in A\}$. Since $A_0 = \frac{1}{2}A$, and $A_1 = \frac{1}{2}A + \frac{1}{2}$, both A_0 and A_1 are measurable. Thus $T^{-1}(A) = A_0 \cup A_1$ is measurable as well. To show that T preserves measure, observe that A_0 and A_1 are disjoint. This implies that $\lambda(T^{-1}(A)) = \lambda(A_0)+\lambda(A_1)$ by countable additivity, so since both $\lambda(A_0) = \lambda(A_1) = \frac{1}{2}\lambda(A)$, we have

$$\lambda(T^{-1}(A)) = \frac{1}{2}\lambda(A) + \frac{1}{2}\lambda(A) = \lambda(A).$$

We now show that T is ergodic. Define the dyadic interval $D_{n,k} = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$, where n, k are integers such that $n \geq 0$, $0 \leq k \leq 2^n$. Let A be a strictly invariant set. By similar proof to the above, it follows from induction that for any $n \geq 0$,

$$\lambda(T^{-n}(A) \cap D_{n,k}) = \frac{1}{2^n}\lambda(A) = \lambda(D_{n,k})\lambda(A). \tag{2.4}$$

As A is strictly invariant, $T^{-n}(A) = A$, for any $n \geq 0$. Therefore, (2.4) leads to

$$\lambda(A \cap D_{n,k}) = \lambda(A)\lambda(D_{n,k}). \tag{2.5}$$

We have two cases on the measure of A. If A has zero measure, then we are done. If A has positive measure, then observe that the dyadic intervals form a sufficient semi-ring on $[0, 1)$.[4] For any $\delta > 0$, there exists a dyadic interval $D_{n,k}$ such that $\lambda(A \cap D_{n,k}) > (1 - \delta)\lambda(D_{n,k})$. Because $\delta > 0$ is arbitrary, we have $\lambda(D_{n,k}) \leq \lambda(A \cap D_{n,k})$. But since $A \cap D_{n,k}$ is a subset of $D_{n,k}$, $\lambda(A \cap D_{n,k}) \leq \lambda(D_{n,k})$, which means that $\lambda(A \cap D_{n,k}) = \lambda(D_{n,k})$. (2.5) finally gives that $\lambda(A) = 1 = \lambda([0, 1))$, so T is ergodic. □

The lemma below gives a useful fact for ergodic transformations and invariant functions.

Lemma 2.6. *Let $(X, S(X), \mu)$ be a probability space. If $T : X \to X$ is an ergodic measure-preserving transformation, then for any measurable function $f : X \to \mathbb{R}$ that is invariant i.e. $f(x) = f(T(x))$ for a.e. x, f is constant a.e.*

[2] Also called *strictly invariant* or *invariant*. The same applies for strict T-invariance mod μ.

[3] In short, the points that go in or out of A are negligible in measure.

[4] See Appendix A.2.

Proof. Let $f : X \to \mathbb{R}$ be a measurable, invariant function. For any $t \in \mathbb{R}$, define the set

$$A_t = \{x \mid f(x) > t\}.$$

The invariance of f implies that A_t is also strictly invariant. By ergodicity, $\mu(A_t) = 1$ or $\mu(A_t) = 0$. If f is not constant, then there exists a $t_0 \in \mathbb{R}$ such that $0 < \mu(A_{t_0}) < 1$, and this contradicts the ergodicity assumption. □

3. Birkhoff Ergodic Theorem

The Birkhoff Ergodic theorem asserts that given an ergodic transformation, the average number of times that a particle passes through a set equals the size of the set itself. Therefore, averaging over a particle's behavior can tell much about an ergodic dynamical system. From an applied standpoint, this means that we can approximate the expected value of a function by computing its time average for large values.

We introduce some notation first. From now on, given a function f and a transformation T, let f_n, f_*, and f^* be defined as:

(1) $f_n(x) = \sum_{i=0}^{n-1} f(T^i(x))$, for $n \geq 1$.
(2) $f_*(x) = \liminf\limits_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$.
(3) $f^*(x) = \limsup\limits_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$.

Furthermore, if f is integrable, then define $||f||_1 := \int |f| d\mu$.

We give a combinatorially-flavored proof of the Maximal Ergodic Theorem from [1], which uses the following definition.

Definition 3.1. Let $a_1, a_2, ..., a_n$ be a finite sequence of real numbers. Let $m \leq n$ be an integer. Then the term a_k is called a *m-leader* if there exists an integer p with $1 \leq p \leq m$ such that $a_k + a_{k+1} + ... + a_{k+p-1} \geq 0$.

Lemma 3.2. *Let $a_1, ..., a_n$ be a finite sequence of real numbers. Then the sum of all m-leaders is nonnegative.*

Proof. If no m-leaders exist, then the sum of all m-leaders is 0. If m-leaders do exist, let a_k be the first m-leader. Let p_1 with $1 \leq p_1 \leq m$ be the least integer such that $a_k + ... + a_{k+p_1-1} \geq 0$.

We claim that for every h such that $k < h \leq k+p_1-1$, a_h is a m-leader. Assume that there exists h such that a_h is not an m-leader i.e. $a_h + ... + a_{k+p_1-1} < 0$. But $a_k + ... + a_{k+p_1-1} \geq 0$, so $a_k + ... + a_{h-1} > 0$, which contradicts that p_1 is the least integer such that the m-leader definition is satisfied. Continue inductively through the remaining sequence $a_{k+p}, ..., a_n$ to collect all the m-leaders. The sum of the m-leaders at each step satisfies the nonnegative condition, so the assertion holds. □

Lemma 3.3 (Maximal Ergodic Theorem)**.** *Let $(X, S(X), \mu)$ be a probability space, and let $T : X \to X$ be a measure-preserving transformation. Let $f : X \to \mathbb{R}$ be an integrable function, and define*

$$G(f) = \{x \mid f_n(x) \geq 0 \text{ for some } n > 0\}. \tag{3.4}$$

Then $\int_{G(f)} f d\mu \geq 0$.

Proof. Let $n \geq 1$ be an integer, and $m \leq n$. Define

$$G_m = \{x \mid f_p(x) \geq 0 \text{ for some } p \text{ with } 1 \leq p \leq m\}.$$

Define $s_m(x)$ to be the sum of all the m-leaders of the finite sequence $f(x)$, $f(T(x))$, ..., $f(T^{n+m-1}(x))$. Next, define

$$S_k = \{x \mid f(T^k(x)) \text{ is a } m\text{-leader of } f(x), ..., f(T^{n+m_1}(x))\}.$$

Using S_k and the characteristic function, $s_m(x) = \sum_{k=0}^{n+m-1} f(T^k(x)) \cdot \mathbb{1}_{S_k} \geq 0$. As both the measurability of S_k and the integrability of s_m follow from the fact that f is integrable, we get from Lemma 3.2 that

$$0 \leq \sum_{k=0}^{n+m-1} \int_{S_k} f \circ T^k d\mu. \tag{3.5}$$

Observe that for $k = 0, ..., n-1$, $f(T^k(x))$ is an m-leader if and only if $f(T^{k-1}(T(x)))$ is also an m-leader. This means that for any $0 \leq k \leq n-1$, $S_k = T^{-1}(S_{k-1})$. By iterating through the $k's$, we get that $S_k = T^{-k}(S_0)$. Hence, for $k = 0, ..., n-1$, we have

$$\int_{S_k} f \circ T^k d\mu = \int (f \circ T^k) \cdot \mathbb{1}_{T^{-k}(S_0)} d\mu = \int_{S_0} f d\mu. \tag{3.6}$$

Since $f(x)$ is a m-leader of $f(x), ..., f(T^{n+m-1}(x))$ if and only if $f_p(x) > 0$ for some $1 \leq p \leq m$, it follows that $S_0 = G_m$. Because $f \leq |f|$, from (3.5) and (3.6) follows that

$$\begin{aligned} 0 &\leq \sum_{k=0}^{n-1} \int_{G_m} f d\mu + \sum_{j=n}^{n+m-1} \int_{S_j} f \circ T^j d\mu \\ &\leq n \int_{G_m} f d\mu + m \int_{S_j} |f| d\mu. \end{aligned}$$

Dividing by n and taking the limit as $n \to \infty$ on both sides of the inequality gives $\int_{G_m} f d\mu \geq 0$. Now, observe that $G(f) = \bigcup_{m \geq 1} G_m$. This means that $\lim\limits_{m\to\infty} f \cdot \mathbb{1}_{G_m} = f \cdot \mathbb{1}_{G(f)}$. Furthermore, for any $m \geq 1$, $|f \cdot \mathbb{1}_{G_m}| \leq |f \cdot \mathbb{1}_{G(f)}|$. By the Dominated Convergence Theorem, we have

$$0 \leq \lim_{m\to\infty} \int f \cdot \mathbb{1}_{G_m} d\mu = \int \lim_{m\to\infty} f \cdot \mathbb{1}_{G_m} d\mu = \int_{G(f)} f d\mu.$$

□

Finally, armed with the Maximal Ergodic Theorem, we prove the Birkhoff Ergodic Theorem. The following proof comes from [5].

Theorem 3.7 (Birkhoff Ergodic Theorem)**.** *Let $(X, S(X), \mu)$ be a probability space, and let $T : X \to X$ be a measure-preserving transformation. If $f : X \to \mathbb{R}$ is integrable, then the following are true:*

(1) $\tilde{f}(x) = \lim\limits_{n\to\infty} \sum_{i=0}^{n-1} f(T^i(x))$ exists a.e..

(2) $\tilde{f}(T(x)) = \tilde{f}(x)$ a.e.

(3) For any measurable, strictly invariant set A, $\int_A f d\mu = \int_A \tilde{f} d\mu$.

Furthermore, if T is ergodic, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) = \int f d\mu \ a.e. \tag{3.8}$$

Proof of (1). For any $\alpha, \beta \in \mathbb{R}$, define

$$M_{\alpha,\beta} = \{x \mid f_*(x) < \alpha < \beta < f^*(x)\}.$$

The set $M_{\alpha,\beta}$ marks the points of X where f_* and f^* differ. By definition, $f_* \leq f^*$, but if $M_{\alpha,\beta}$ is a null set, then $f_* = f^*$ almost everywhere. Hence, it suffices to show that $M_{\alpha,\beta}$ has measure 0 to show that the limit exists almost everywhere.

Let $\alpha, \beta \in \mathbb{R}$. Assume for the sake of contradiction that $\mu(M_{\alpha,\beta}) > 0$. In the style of (3.4), consider the set

$$G(f-\beta) = \{x \mid (f-\beta)_n \geq 0 \text{ for some } n > 0\}.$$

We claim that $M_{\alpha,\beta} \subset G(f-\beta)$. If $x \in M_{\alpha,\beta}$, then there exists an $N \in \mathbb{N}$ such that $\frac{1}{N}\sum_{i=0}^{N-1} f(T^i(x)) > \beta$. Hence it follows that $\sum_{i=0}^{N-1} f(T^i(x)) - N\beta = (f-\beta)_N > 0$, which means that $M_{\alpha,\beta} \subset G(f-\beta)$. Also, from part (2), $M_{\alpha,\beta}$ is T-invariant, so we can restrict T to $M_{\alpha,\beta}$ and apply the Maximal Ergodic Theorem to get

$$\int_{M_{\alpha,\beta}} (f-\beta) d\mu \geq 0, \text{ which implies that } \int_{M_{\alpha,\beta}} f d\mu \geq \beta\mu(M_{\alpha,\beta}).$$

We apply similar logic to $G(\alpha - f)$ to get $\int_{M_{\alpha,\beta}} f d\mu \leq \alpha\mu(M_{\alpha,\beta})$, which means that $\beta\mu(M_{\alpha,\beta}) \leq \alpha\mu(M_{\alpha,\beta})$. However, if $\mu(M_{\alpha,\beta}) > 0$, $\alpha < \beta$ cannot be true, so we have a contradiction. Since $M_{\alpha,\beta}$ is a null set for any two rationals α, β, the limit exists a.e.

□

Proof of (2). We show that both f^* and f_* are invariant. Observe that

$$\begin{aligned}\frac{1}{n} f_n(T(x)) &= \frac{1}{n}\Big(\sum_{i=0}^{n} f(T^i(x)) - f(x)\Big) \\ &= \frac{n+1}{n} \cdot \frac{1}{n+1} f_{n+1}(x) - \frac{1}{n} f(x).\end{aligned}$$

By taking the lim inf as n goes to ∞ on both sides, it follows that

$$f_*(T(x)) = \liminf_{n\to\infty} \frac{1}{n} f_n(T(x)) = \liminf_{n\to\infty} \frac{1}{n+1} f_{n+1}(x) = f_*(x).$$

Similar logic shows that f^* is invariant, and the result follows. □

Proof of (3). Let $f^+(x) = f(x)$ when $f(x) \geq 0$ and 0 otherwise. Similarly, let $f^-(x) = -f(x)$ when $f(x) \leq 0$ and 0 otherwise. Because $f = f^+ - f^-$, it suffices to show part (3) for nonnegative integrable functions.

The following cases on the bounded-ness of f can be made. First, consider if f is a bounded, nonnegative function almost everywhere. By the Dominated

Convergence Theorem,

$$\int_A \tilde{f} d\mu = \int_A \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \lim_{n\to\infty} \frac{1}{n} \int_A \sum_{i=0}^{n-1} f(T^i(x)). \tag{3.9}$$

Because A is strictly invariant, we have from (3.9) that

$$\int_A \tilde{f} d\mu = \lim_{n\to\infty} \frac{1}{n} \int_A \sum_{i=0}^{n-1} f(T^i(x)) = \int_A f d\mu.$$

Hence, part (3) holds for nonnegative integrable functions that are bounded almost everywhere. Next, consider a nonnegative integrable function f with no condition for boundedness. By approximation of simple functions[5], for any $\epsilon > 0$, there exists a bounded function g such that $||f - g||_1 < \epsilon$. Consider $|\int_A f d\mu - \int_A \tilde{f} d\mu|$. From the triangle inequality, we have

$$\left|\int_A f d\mu - \int_A \tilde{f} d\mu\right| \leq \left|\int_A f d\mu - \int_A g d\mu\right| + \left|\int_A g d\mu - \int_A \tilde{f} d\mu\right|. \tag{3.10}$$

As f and g are both integrable, $|f - g|$ is also integrable. By using the triangle inequality again on the second absolute value expression, we have that the latter expression of (3.10) is less than or equal to

$$\int_A |f - g| d\mu + \left|\int_A g d\mu - \int_A \tilde{g} d\mu\right| + \left|\int_A \tilde{g} d\mu - \int_A \tilde{f} d\mu\right|. \tag{3.11}$$

Since g is bounded, the second expression in (3.11) evaluates to 0 by the first case. To set an upper bound on the third expression of (3.11), observe that by Fatou's Lemma,

$$||\tilde{f}||_1 \leq \int \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |f|(T^i(x)) \leq \liminf_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} |f|(T^i(x)). \tag{3.12}$$

Since T is measure-preserving, the last expression of (3.12) equals $\int |f| d\mu = ||f||_1$. This implies that $||\tilde{f}||_1 \leq ||f||_1$. Returning to (3.11), $\int_A |f - g| d\mu = ||f - g||_1$ and $\left|\int_A \tilde{g} d\mu - \int_A \tilde{f} d\mu\right| \leq ||f - g||_1$, so

$$\left|\int_A f d\mu - \int_A \tilde{f} d\mu\right| \leq ||f - g||_1 + ||f - g||_1 < 2\epsilon.$$

Because ϵ can be made arbitrarily small, $\int_A \tilde{f} d\mu = \int_A f d\mu$, and this completes the proof of part (3).

Finally, the addition of ergodicity allows for the application of Lemma 2.6. Part (2) shows that $\tilde{f}$ is T-invariant and T is ergodic by assumption, so

$$\int f d\mu = \int \tilde{f} d\mu = \tilde{f}\mu(X) = \tilde{f} \text{ a.e.}$$

□

[5]To be precise, we use that $\int f$ is defined as the supremum of the integrals of simple functions that are less than f.

Welcome back to coin flipping. Because the doubling map is ergodic, and the function $\mathbb{1}_{[0,\frac{1}{2})}$ is integrable, the Birkhoff Ergodic Theorem implies that the average frequency of heads in a binary sequence converges to the length of $[0, \frac{1}{2})$. In other words, the probability of heads being $\frac{1}{2}$ can be approximated by calculating the average frequency of heads in a binary sequence of large length N.

4. Entropy

In this section, we define measure-theoretical entropy, and give a proof of the Brin-Katok Formula for local entropy.

Entropy measures the complexity of a system. To understand complexity for our purposes, imagine a baker thoroughly placing brown sugar balls into dough and kneading the dough. The complexity of a kneading technique can be measured by how quickly the sugar balls break down. For example, a technique that keeps the original balls intact does not mix the dough as well as one that crushes the balls very quickly. Furthermore, notice that the rate of break-down is equivalent to the rate at which the number of distinguishable sugar fragments grows over time. We will show that in an ergodic process such as kneading, the entropy equals the rate of growth of the number of distinguishable orbits as time increases.

4.1. **Entropy Definitions.** Given a probability space $(X, S(X), \mu)$, define a *partition* of X as a countable collection of pairwise disjoint measurable sets such that their union has full measure.[6]

Definition 4.1. Let $\alpha = \{A_1, ..., A_n\}$ be a finite partition of X. Then the *entropy of the partition*[7] α is defined as

$$H(\alpha) = -\sum_{A\in\alpha} \mu(A)\log\mu(A).$$

Given two partitions α, β, denote $\alpha \vee \beta$ as the partition consisting of sets of the form $A \cap B$, where $A \in \alpha$, and $B \in \beta$. Also, if $\alpha = \{A_1, ..., A_n\}$, then given a transformation $T : X \to X$, define $T^{-k}(\alpha)$ as the partition $\{T^{-k}(A_1), ..., T^{-k}(A_n)\}$.

Definition 4.2. Let $(X, S(X), \mu, T)$ be a measure-preserving dynamical system. Let $\alpha = \{A_1, ..., A_n\}$ be a finite partition of X. Then the *entropy of the measure-preserving dynamical system with respect to* α is defined as

$$\begin{aligned} h_\mu(T, \alpha) &= \lim_{n\to\infty} \frac{1}{n} H(\alpha \vee T^{-1}(\alpha) \vee ... \vee T^{-(n-1)}(\alpha)) \\ &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)). \end{aligned}$$

Note that the limit exists from Fekete's Subadditivity Lemma (see [6]).

Definition 4.3. Let $(X, S(X), \mu, T)$ be a measure-preserving dynamical system. Then the *entropy of the measure-preserving dynamical system* is defined as

$$h_\mu(T) = \sup\{h_\mu(T, \alpha) \mid \alpha \text{ is a finite partition}\}.$$

[6]The measure of their union equals the measure of X.

[7]Also called Shannon entropy.

Calculating the entropy of a dynamical system is difficult; the supremum must be taken over *all possible partitions* of the measure space. Thankfully, the Brin-Katok Formula for local entropy greatly simplifies entropy calculation in the contexts where X is also a metric space.

4.2. **The Brin-Katok Formula for Local Entropy.** For this section, assume all spaces to be a topological measure-preserving dynamical system with metric d.

Define $d^n(x,y) := \max_{0\le i\le n-1} d(T^i(x),T^i(y))$. With this, define $B^n(x,\epsilon) := \{y \in X \mid d^n(x,y) < \epsilon\}$. Essentially, $B^n(x,\epsilon)$ is the set of points that are indistinguishable from x up to error ϵ in n iterates. The Brin-Katok formula tells us that the size of this set decays at an exponential rate, and this rate equals the entropy of the system.

Theorem 4.4 (Brin-Katok Formula for Local Entropy)**.** *Let (X,T) be a topological dynamical system with a metric d, and a measure μ that is ergodic and T-invariant with entropy h. Then for a.e. $x \in X$,*

$$h_\mu = \lim_{\epsilon\to 0}\left(\limsup_{n\to\infty}\frac{-\log\mu(B^n(x,\epsilon))}{n}\right) = \lim_{\epsilon\to 0}\left(\liminf_{n\to\infty}\frac{-\log\mu(B^n(x,\epsilon))}{n}\right) \tag{4.5}$$

Notice that calculating the measure of a specific set $B^n(x,\epsilon)$ is much easier than computing the supremum over the set of all possible partitions.

Below are several combinatorial lemmas from [2] and [3] that are used in the proof of the Brin-Katok Formula.

Lemma 4.6. *If $A_1,...,A_N$ are sets in a probability space, and $\mu(A_i) > c$, and each $x \in X$ belongs to at most k of $\{A_i\}_{i=1}^N$, then $N \le \frac{k}{c}$.*

Proof. Since each $x \in X$ belongs to at most k of the sets $A_1,...,A_N$, we have $\sum_{i=1}^N \mathbb{1}_{A_i}(x) \le k$. Integrating on both sides, it follows that

$$\begin{aligned}\int k\,d\mu &\ge \int \sum_{i=1}^N \mathbb{1}_{A_i}\,d\mu \\ &= \sum_{i=1}^N \mu(\mathbb{1}_{A_i}) > Nc.\end{aligned}$$

Since we are in a probability space, $\int k\,d\mu = k$, and the result follows. □

Lemma 4.7. *Define $\binom{n}{i}$ as the number of subsets of $\{1,2,...,n\}$ that are of size less than i. Also, define $H(\alpha) = -\alpha\log_2(\alpha) - (1-\alpha)\log_2(1-\alpha)$. Then for every $0 \le \alpha < \frac{1}{2}$,*

$$\binom{n}{\alpha n} \le 2^{n(H(\alpha))}.$$

Proof. By the binomial theorem, we have

$$1 = (\alpha + (1-\alpha))^n \ge \sum_{i\le\alpha n}\binom{n}{i}\alpha^i(1-\alpha)^{n-i}. \tag{4.8}$$

Because $\alpha \le \frac{1}{2}$, $\alpha^k(1-\alpha)^k$ decreases as k increases, which means that

$$\min_{i\le\alpha n}\{\alpha^i(1-\alpha)^{n-i}\} = \alpha^{\alpha n}(1-\alpha)^{n-\alpha n}.$$

Therefore, from (4.8) follows that

$$1 \geq \sum_{i \leq \alpha n} \binom{n}{i} \alpha^{\alpha n}(1-\alpha)^{n-\alpha n} = \sum_{i \leq \alpha n} \binom{n}{i} 2^{-n(H(\alpha))}.$$

Dividing by $2^{-n(H(\alpha))}$ gives the desired result. $\square$

Finally, we arrive at the proof for the Brin-Katok Formula. We follow the proof in [2].

Proof of Brin-Katok - Part I. The goal is to show that

$$h_\mu \geq \lim_{\epsilon \to 0} \left(\limsup_{n \to \infty} \frac{-\log \mu(B^n(x, \epsilon))}{n} \right).$$

Let $\epsilon > 0$, and let α be a partition with atoms of diameter less than ϵ. Then for a.e. $x \in X$, $\alpha^n(x) \subseteq B^n(x, \epsilon)$.[8] Since $\mu(\alpha^n(x)) \leq \mu(B^n(x, \epsilon))$ and log is an increasing function, it follows that

$$\liminf_{n \to \infty} \frac{-\log \mu(B^n(x, \epsilon))}{n} \leq \liminf_{n \to \infty} \frac{-\log \mu(\alpha^n(x))}{n} = h_\mu(T, \alpha) \leq h_\mu.$$

Because ϵ is arbitrary, the result follows. $\square$

Proof of Brin-Katok - Part II. We will show that for a.e. $x \in X$,

$$h_\mu \leq \lim_{\epsilon \to 0} \left(\liminf_{n \to \infty} \frac{-\log \mu(B^n(x, \epsilon))}{n} \right). \tag{4.9}$$

To prove (4.9) for a.e. $x \in X$, it suffices to show that for any $\rho > 0$,

$$\mu\left(\left\{ x \;\middle|\; \lim_{\epsilon \to 0} \left(\liminf_{n \to \infty} \frac{-\log \mu(B^n(x, \epsilon))}{n} \right) < h_\mu - \rho \right\} \right) = 0.$$

To this end, let $\rho > 0$, and fix $\epsilon > 0$. Let $\alpha = \{A_1, ..., A_k\}$ be a partition of X such that $h_\mu(\alpha, T) > h_\mu - \frac{\rho}{4}$, and $\mu(\partial A_i) = 0$ for all $1 \leq i \leq k$. Define $E_\epsilon = \bigcup_{A \in \alpha} (\partial A)^{(\epsilon)}$, which is the union of the boundaries of $A \in \alpha$ up to error ϵ. Also, let $I_n(x) = \{0 \leq i \leq n-1 \mid T^i(x) \notin E_\epsilon\}$, and define

$$\gamma_n(x) = \bigcap_{i \in I_n} (T^{-i}\alpha)(x).$$

We claim that $B^n(x, \epsilon) \subseteq \gamma_n(x)$. Let $y \in B^n(x, \epsilon)$. Without loss of generality, assume that for $1 \leq i \leq n-1$, $T^i(x) \notin E_\epsilon$. We have that $d(T^i x, T^i y) < \epsilon$, and $d(x, \partial((T^{-i}\alpha)(x))) \geq \epsilon$, so $T^i(y) \in (T^{-i}\alpha)(x)$. Therefore, $y \in \gamma_n(x)$. From this, $\mu(B^n(x, \epsilon)) \leq \mu(\gamma_n(x))$, which means that

$$\lim_{\epsilon \to 0} \left(\liminf_{n \to \infty} \frac{-\log \mu(B^n(x, \epsilon))}{n} \right) \geq \lim_{\epsilon \to 0} \left(\liminf_{n \to \infty} \frac{-\log \mu(\gamma_n(x))}{n} \right).$$

Thus it suffices to show that for a.e. x,

$$\lim_{\epsilon \to 0} \left(\liminf_{n \to \infty} \frac{-\log \mu(\gamma_n(x))}{n} \right) \geq h_\mu - \rho.$$

Now, define another partition

$$\beta = \{A_1 \cap E_\epsilon, ..., A_k \cap E_\epsilon, X \setminus E_\epsilon\}.$$

[8] Given $\alpha = \{A_1, ..., A_n\}$ as a partition of X, define $\alpha(x) = A_i$ if $x \in A_i$ for some $1 \leq i \leq n$.

Consider the following property of Shannon entropy: if ξ_1, ξ_2 are partitions, then $H(\xi_1 \vee \xi_2) \le H(\xi_1) + H(\xi_2)$. This and the fact that T is measure-preserving gives that $h_\mu(T, \beta) \le -\sum_{B \in \beta} \mu(B) \log \mu(B)$. As ϵ approaches 0, since $\mu(E_\epsilon)$ approaches 0, $\mu(A_i \cap E_\epsilon)$ and $\log \mu(X \setminus E_\epsilon)$ also approach 0. Hence, for sufficiently small ϵ, $h_\mu(T, \beta) < \frac{\rho}{4}$.

Consider the following set of sets

$$\begin{aligned} U_n &= \{\alpha^n(x) \mid \mu(\alpha^n(x)) < 2^{-n(h_\mu(T,\alpha) - \frac{\rho}{4})}\}, \\ V_n &= \{\beta^n(x) \mid \mu(\beta^n(x)) > 2^{-n(h_\mu(T,\beta) + \frac{\rho}{4})}\}, \\ W_n &= \{\gamma_n(x) \mid \mu(\gamma_n(x)) > 2^{-n(h_\mu - \rho)}\}, \text{ and} \\ Z_n &= \{\gamma_n(x) \cap \beta^n(x) \mid \alpha^n(x) \in U_n, \beta^n(x) \in V_n, \gamma_n(x) \in W_n\}. \end{aligned}$$

With these sets, the goal is to show that by the Borel-Cantelli Lemma, for a.e. $x \in X$, there are only finitely many n such that $\gamma_n(x) \in W_n$. Showing this means that for a.e. x and for large N's,

$$\mu(\gamma_N(x)) \le 2^{-N(h_\mu - \rho)}, \text{ which means that } \frac{-\log(\mu(\gamma_N(x)))}{N} \ge h_\mu - \rho.$$

First, we show that the conditions of the Borel-Cantelli Lemma are met. Every element of W_n has measure of at least $2^{-n(h_\mu - \rho)}$, and we are in a probability space. Assuming that $\{\gamma_n(x)\}_{x \in X}$ is a partition (which it isn't, but we will establish a workaround later), it follows that

$$1 \ge \sum_{\gamma_n(x) \in W_n} \mu(\gamma_n(x)) > |W_n| \cdot 2^{-n(h_\mu - \rho)}.$$

Therefore, $|W_n| \le 2^{n(h_\mu - \rho)}$. By similar logic, $|V_n| \le 2^{n(h_\mu(T,\beta) + \frac{\rho}{4})}$.

Now, observe that if $D \in Z_n$, then since $D = \gamma_n(x) \cap \beta^n(x) \subseteq \alpha^n(x)$, it must be that $\mu(D) \le \mu(\alpha^n(x) < 2^{-n(h_\mu(T,\alpha) - \frac{\rho}{4})}$. Because every element of Z_n is the intersection of an element of W_n and an element of V_n,

$$\begin{aligned} \mu(\cup Z_n) &\le \sum_{D \in Z_n} \mu(D) \\ &< |W_n| \cdot |V_n| \cdot 2^{-n(h_\mu(T,\alpha) - \frac{\rho}{4})} \\ &\le 2^{n(h_\mu - \rho)} \cdot 2^{n(h_\mu(T,\beta) + \frac{\rho}{4})} \cdot 2^{-n(h_\mu(T,\alpha) - \frac{\rho}{4})}. \end{aligned}$$

Recall that $h_\mu(T, \alpha) > h_\mu - \frac{\rho}{4}$. Then $\mu(\cup Z_n) < 2^{n(h_\mu(T,\beta) - \frac{\rho}{4})}$. Also recall that for sufficiently small ϵ, $h_\mu(T, \beta) < \frac{\rho}{4}$. This means that $2^{n(h_\mu(T,\beta) - \frac{\rho}{4})}$ decreases exponentially with growing n. Therefore, $\sum_n \mu(\cup Z_n) < \infty$, and the conditions for Borel-Cantelli are satisfied.

By Borel-Cantelli, a.e. $x \in X$ is in finitely many Z_n. This means that for a.e. x, there exists an integer N_x such that if $n > N_x$, then $x \notin Z_n$. However, for sufficiently large n, it is true that for a.e. x, $\alpha^n(x) \in U_n$ and $\beta^n(x) \in V_n$. By definition of Z_n, $\gamma_n(x)$ must be in finitely many W_n as well. Hence, we have shown what we set out to prove.

However, $\{\gamma_n(x)\}_{x \in X}$ is not a partition, which means that the cardinality argument for W_n does not hold. We will show that that the argument above still holds regardless. Define

$$\Gamma_n = \{\gamma_n(x) \mid x \in X \text{ and } |I_n(x)| > n(1 - 2\mu(E_\epsilon))\}.$$

We want to show that for a.e. x, $\gamma_n(x) \in \Gamma_n$ for large n. By the Birkhoff Ergodic theorem, for a.e. x,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{E_\epsilon}(T^i(x)) = \mu(E_\epsilon).$$

This is to say that the average number of times that the iterates of x visit E_ϵ approaches $\mu(E_\epsilon)$. Recall that $|I_n|$ is the number of iterates that are *not* in E_ϵ. Therefore, there exists an integer N such that if $n > N$, then

$$\left| \frac{n - |I_n(x)|}{n} - \mu(E_\epsilon) \right| < \mu(E_\epsilon).$$

From this follows that for sufficiently large n and a.e x, $|I_n(x)| > n(1 - 2\mu(E_\epsilon)$, so $\gamma_n(x) \in \Gamma_n$. By definition of $\gamma_n(x)$, each element of Γ_n is the intersection of at least $n(1 - 2\mu(E_\epsilon))$ of the sets $\alpha(x), (T^{-1}\alpha)(x), ...(T^{-n+1}\alpha)(x)$. Furthermore, $x \in \gamma_n(x)$, so each x belongs to at most $\binom{n}{n(1-2\mu(E_\epsilon))} = \binom{n}{n(2\mu(E_\epsilon))}$ elements of Γ_n. By Lemma 4.7, since $2\mu(E_\epsilon) \leq \frac{1}{2}$ for sufficiently small ϵ, x belongs to at most

$$2^{-n(H(2\mu(E_\epsilon)))} \text{ elements of } \Gamma_n.$$

Since $\mu(E_\epsilon)$ goes to 0 as ϵ goes to 0, the exponent approaches 0 as well. Finally, re-define $W_n = \{C \in \Gamma_n \mid \mu(C) > 2^{-n(h_\mu - \rho)}\}$. By Lemma 4.6 and the logic showed above for finding cardinalities,

$$|W_n| \leq 2^{n(h_\mu-\rho)} \cdot 2^{-n(H(2\mu(E_\epsilon)))}.$$

As ϵ goes to 0, $|W_n| \leq 2^{n(h_\mu-\rho)}$, and the argument holds. □

Example 4.10. As an example of applying the Brin-Katok Formula, we compute the entropy of the doubling map. The goal is to calculate

$$\lim_{\epsilon\to 0}\left(\lim_{n\to\infty} \frac{-\log \mu(B^n(x,\epsilon))}{n}\right).$$

As long as two points are $\frac{\epsilon}{2^n}$ apart from each other from the outset, all of their iterates up to the nth iteration will be at most ϵ apart from each other. Therefore, $\mu(B^n(x,\epsilon)) = \frac{\epsilon}{2^n}$. Hence, it follows that

$$\begin{aligned}\lim_{\epsilon\to 0}\left(\lim_{n\to\infty} \frac{-\log \mu(B^n(x,\epsilon))}{n}\right) &= \lim_{\epsilon\to 0}\left(\lim_{n\to\infty} \frac{n\log 2 - \log \epsilon}{n}\right) \\ &= \log 2.\end{aligned}$$

5. Entropy Calculation for Expanding C^2 Maps on S^1

In this section, we show that the entropy of an expanding C^2 map on S^1 is equal to its Lyapunov exponent, which we define below.

5.1. **Lyapunov Exponent.**

Along with entropy, the Lyapunov exponent gives another characterization of chaos. In particular, it measures the average rate at which the iterates of two infinitesimally close points diverge (or converge).

Consider the following heuristic derivation of the Lyapunov exponent. Let x and $x + dx$ be points that are arbitrarily close to each other. We will observe how the iterates of x and $x + dx$ travel *relative to each other*. Using the derivative of f, we approximate the distance between each iterate of x and $x + dx$ inductively:

$$\begin{aligned}
&|f(x) - f(x+dx)| \approx |f'(x)||dx| \\
&|f^2(x) - f^2(x+dx)| \approx |f'(f(x))||f(x) - f(x+dx)| = |f'(f(x))||f'(x)||dx| \\
&\ldots \\
&|f^n(x) - f^n(x+dx)| \approx |dx| \prod_{i=0}^{n-1} |f'(f^i(x))|.
\end{aligned}$$

The *Lyapunov exponent* λ at x, the average rate at which the iterates of x and $x + dx$ diverge, is therefore defined as

$$\lambda(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log(|f'(f^i(x))|). \tag{5.1}$$

The right-hand side of (5.1) looks suspiciously like a Birkhoff sum. Indeed, if f is ergodic, then by the Birkhoff Ergodic Theorem, (5.1) is guaranteed to converge almost everywhere to $\int \log|f'|$.

5.2. **Connecting Entropy and Lyapunov Exponents.** From now on, assume all maps are expanding C^2 maps on S^1. We say f is C^2 expanding on S^1 if there exists $\tilde{f} \in C^2(\mathbb{R}, \mathbb{R})$, with $\tilde{f}' > 1$ and $\tilde{f}(x+1) = \tilde{f}(x)$, such that $f(x) = \tilde{f}(x \mod 1) \mod 1$.

These two lemmas from [4] are used in the main proof.

Lemma 5.2 (Existence of Unique Continuous Invariant Measure)**.** *There exists a unique, continuous function $\phi : S^1 \to [0, +\infty)$ such that $d\mu = \phi \cdot d\lambda$ is ergodic and invariant under f, where λ is the Lebesgue measure.*

Proof. See [4, p. 191]. □

Lemma 5.3 (Distortion Lemma)**.** *There exists $c \in \mathbb{R}$ such that for all $x \in S^1$ and $y \in B^n(x, \epsilon)$,*

$$\frac{1}{c} < \frac{(f^n)'(x)}{(f^n)'(y)} < c. \tag{5.4}$$

Proof. By the chain rule, we have

$$\frac{(f^n)'(x)}{(f^n)'(y)} = \prod_{i=0}^{n-1} \frac{f'(f^i(x))}{f'(f^i(y))} = \prod_{i=0}^{n-1} \left(1 + \frac{f'(f^i(x)) - f'(f^i(y))}{f'(f^i(y))}\right).$$

Because f is a C^2 map, for $0 \le i \le n-1$, there exists some point z_i in between $f^i(x)$ and $f^i(y)$ such that $f'(f^i(x)) - f'(f^i(y)) = f''(z_i)(f^i(x) - f^i(y))$. Furthermore, since f'' is continuous and S^1 is compact, let $M = \sup_{x\in S^1}(f''(x))$. It follows that

$$\frac{(f^n)'(x)}{(f^n)'(y)} \le \prod_{i=0}^{n-1} \left(1 + \frac{M|f^i(x) - f^i(y)|}{f'(f^i(y))}\right).$$

Since f' is also continuous, let $\lambda^{-1} > 1$ be the lower bound of f'. Again, by the Mean Value Theorem, it follows that for $0 \le i \le n-1$, $\frac{|f^i(x) - f^i(y)|}{f'(f^i(y))} \le \lambda^{n-i}|f^n(x) -$

$f^n(y)|$, which means that

$$\prod_{i=0}^{n-1}\left(1+\frac{M|f^i(x)-f^i(y)|}{f'(f^i(y))}\right) \le \prod_{i=0}^{n-1}\left(1+M\lambda^{n-i}|f^n(x)-f^n(y)|\right). \tag{5.5}$$

We want to establish an upper bound on the right hand side of (5.5). Observe that given a sequence $\{a_n\}$, $\log\left(\prod_{n=1}^{\infty}(1+a_n)\right) = \sum_{n=1}^{\infty}(1+a_n)$. Furthermore, $\log(1+a_n) \le a_n$ if a_n is positive, so it follows that $\sum_{n=1}^{\infty}(1+a_n) \le \sum_{n=1}^{\infty} a_n$.

Now, observe that in (5.5), M is a constant value, and $|f^n(x)-f^n(y)|$ is bounded above by the fact that $y \in B^n(x,\epsilon)$. Since $\lambda^{-1} > 1$, the argument above leads to an infinite geometric series with a finite sum, so we have a global upper bound c_0. By similar logic, there exists c_1 such that $\frac{1}{c_1} < \frac{(f^n)'(x)}{(f^n)'(y)}$. Let $c = \max(c_0, c_1)$ and the result follows. □

Theorem 5.6. *Let $f : S^1 \to S^1$ be an expanding C^2 map, and μ be the unique, ergodic, and invariant measure given by Lemma 5.2. Then*

$$h_\mu = \int \log|f'|d\mu.$$

Proof. The goal is to use the Brin-Katok Formula to connect entropy with the Lyapunov exponent. To this end, we will first calculate $\mu(B^n(x,\epsilon))$ for any $x \in S^1$.

Let $x \in S^1$. Because f' is continuous and S^1 is compact, let $M = \sup_{x\in S^1} f'(x)$, and set $\epsilon < \frac{1}{2M}$. For any $n \ge 1$, we first show that $f^n|_{B^n(x,\epsilon)} : B^n(x,\epsilon) \to B(f^n(x),\epsilon)$ is bijective and that its inverse exists.

For surjectivity, the definition of $B^n(x,\epsilon)$ gives $f^n(B^n(x,\epsilon)) \subset B(f^n(x),\epsilon)$. To show the other inclusion, it suffices to prove that $B(f(x),\epsilon) \subset f(B(x,\epsilon))$, because this implies that for all k such that $0 \le k \le n$, $B(f(x),\epsilon) \subset f^{n-k}(B(f^k(x),\epsilon)$ by an inductive argument.

Recall that there exists $\tilde{f} \in C^2(\mathbb{R},\mathbb{R})$, with $\tilde{f}' > 1$ and $\tilde{f}(x+1) = \tilde{f}(x)$, such that $f(x) = \tilde{f}(x \mod 1) \mod 1$. It thus suffices to show that $\tilde{f}$ maps a ball of diameter ϵ to a ball of diameter greater than ϵ. Observe that because $\tilde{f}$ is continuous and monotone, $\tilde{f}(B(x,\epsilon)) = (\tilde{f}(x-\epsilon), \tilde{f}(x+\epsilon))$. Showing that $\tilde{f}(x+\epsilon) > \tilde{f}(x)+\epsilon$ and $\tilde{f}(x-\epsilon) < \tilde{f}(x)-\epsilon$ proves that $B(f(x),\epsilon) \subset f(B(x,\epsilon))$. Using the Fundamental Theorem of Calculus and the fact that $\tilde{f}' > 1$, we have

$$\tilde{f}(x+\epsilon) - \tilde{f}(x) = \int_x^{x+\epsilon} \tilde{f}'(y)\ d\lambda(y) > \int_x^{x+\epsilon} d\lambda(y) = \epsilon.$$

The other inequality follows by similar logic, so the surjectivity of $f^n|_{B^n(x,\epsilon)}$ follows. We now show injectivity. It suffices to show that, for $0 \le k \le n-1$, $f|_{f^k(B^n(x,\epsilon))}$ is injective, because $f^n|_{B^n(x,\epsilon)}$ is the composition of these restricted f's. Because $f^k(B^n(x,\epsilon)) \subset B(f^k(x),\epsilon)$, $f|_{f^k(B^n(x,\epsilon))}$ is restricted to a set that is smaller than a ball of diameter ϵ in $\mathbb{R}$. Working with $\tilde{f}$ again, if a ball of diameter ϵ in $\mathbb{R}$ maps to a ball of diameter less than 1, then f must be injective on the unit circle. To show this, we have

$$\tilde{f}(x+\epsilon) - \tilde{f}(x-\epsilon) = \int_{x-\epsilon}^{x+\epsilon} f'(y)\ d\lambda(y) < (2M)\epsilon < 1.$$

Therefore, $f^n|_{B^n(x,\epsilon)}$ is bijective onto $B(f^n,\epsilon)$.

Now, let $g = f^n|_{B^n(x,\epsilon)}$. Lemma 5.3 gives a c_0 such that for all $z \in B^n(x,\epsilon)$, $\frac{1}{c_0} < \frac{g'(x)}{g'(z)} < c_0$. By bijectivity, g^{-1} exists, so $\frac{1}{c_0} < \frac{(g^{-1})'(g(z))}{(g^{-1})'(g(x))} < c_0$. Therefore, for any $y \in B(f^n(x),\epsilon)$, we have that

$$\frac{1}{c_0} < \frac{(g^{-1})'(y)}{(g^{-1})'(g(x))} < c_0.$$

By Lemma 5.2, let $\phi : S^1 \to [0, +\infty)$ be the unique, continuous function such that $d\mu = \phi \cdot d\lambda$ is invariant and ergodic. The continuity of ϕ implies that ϕ is bounded on $B(f^n(x),\epsilon)$, so there exists c_1 such that for all $y \in B(f^n(x),\epsilon)$, $\frac{1}{c_1} < \phi(y) < c_1$.

Set $c = \max(c_0, c_1)$. We return to the estimation of $\mu(B^n(x,\epsilon))$ with the change of variables formula, which gives

$$\begin{aligned}\mu(B^n(x,\epsilon)) &= \int_{B(f^n(x),\epsilon)} (g^{-1})'(y)\ d\mu(y)\\ &= \int_{B(f^n(x),\epsilon)} (g^{-1})'(y)\cdot\phi(y)\ d\lambda(y).\end{aligned}$$

By using the lower bound $\frac{1}{c}$ and the upper bound c, it follows that

$$\frac{2\epsilon}{c^2}\cdot (g^{-1})'(g(x)) \le \int_{B(f^n(x),\epsilon)} (g^{-1})'(y)\cdot\phi(y)\ d\lambda(y) \le 2\epsilon c^2 \cdot (g^{-1})'(g(x)),$$

Applying the log and dividing by n on all sides, we get

$$\frac{\log\left(\frac{2\epsilon}{c^2}\right)}{n} + \frac{\log\left(\frac{1}{(f^n)'(x)}\right)}{n} \le \frac{\log(\mu(B^n(x,\epsilon)))}{n} \le \frac{\log\left(2\epsilon c^2\right)}{n} + \frac{\log\left(\frac{1}{(f^n)'(x)}\right)}{n}.$$

Taking the limit as n goes to infinity, we get

$$\lim_{n\to\infty} \frac{-\log(\mu(B^n(x,\epsilon)))}{n} = \lim_{n\to\infty} \frac{\log((f^n)'(x))}{n}. \tag{5.7}$$

Keeping in mind that $f' > 1$, applying the chain rule equates the right hand side of (5.7) to

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \log(|f'(f^i(x))|),\ \text{which is a Birkhoff sum.}$$

The Birkhoff Ergodic theorem gives that (5.7) equals $\int \log|f'|d\mu$. By the Brin-Katok Formula, h_μ is the limit of (5.7) as ϵ approaches zero, so the desired result follows. □

Remark 5.8. Notice that this result gives an extremely easy proof that the entropy of the doubling map is $\log 2$.

Acknowledgments

I would like to thank my mentor, Daniel Mitsutani, for his constant guidance during this project; this would not have been possible without his help in the writing of this paper and in the learning of the material. Furthermore, I would like to thank Professor Peter May and all of those who are involved in the organization of The University of Chicago Mathematics REU for allowing me to participate in this wonderful learning opportunity.

References

[1] Paul R Halmos. Lectures on Ergodic Theory. Mathematical Society of Japan, 1956.

[2] Mike Hochman. Dimension of 1-dimensional expanding maps. 2015.

[3] Mike Hochman. Lectures on Dynamical Systems and Entropy 2014.

[4] Anatole Katok, Boris Hasselblatt. Introduction to the Modern Theory of Dynamical Systems. Cambridge University Press, 1997.

[5] Cesar Ernesto Silva. Invitation to Ergodic Theory. American Math. Soc., 2008.

[6] Proof of Fekete's Subadditive Lemma. Planet Math, 18 Mar.2014. planetmath.org/proofoffeketessubadditivelemma.

Appendix A. Measure Theory and Integration

A.1. **Introduction to Measure Theory.** As mentioned in the introduction, the length of the half interval is related to the probability of heads turning up in a series of coin flips. However, what is "length", and can we measure the length or size of other objects? Although "size" and "length" are easy concepts to grasp as far as intervals are concerned, defining them in generality proves to be more difficult.

Definition A.1. A *σ-algebra* on a nonempty set X is a collection $S(X)$ of subsets of X such that

- $S(X)$ is nonempty.
- $S(X)$ is closed under countable unions.
- $S(X)$ is closed under complements.

That $S(X)$ is closed under countable intersections follows from the second and third conditions.

With the σ-algebra forming the sets that can be measured, the natural next question is what unit of measure can be used for the σ-algebra.

Definition A.2. Let $S(X)$ be a σ-algebra. Then the function $\mu : S(X) \to [0, \infty]$ is called a *measure* on $S(X)$ if

(1) $\mu(\emptyset) = 0$.
(2) μ is countably additive i.e. for any collection of disjoint sets $\{X_n\}_{n\geq 1}$ in $S(X)$,

$$\mu\left(\bigcup_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mu(X_n).$$

Definition A.3. Let X be a nonempty set, $S(X)$ be a σ-algebra on X, and μ be a measure on $S(X)$. A *measure space* is a tuple $(X, S(X), \mu)$. The members of $S(X)$ are called *measurable sets*.

A measure space $(X, S(X), \mu)$ is called *σ-finite* if there is a countable collection $\{A_n\}_{n\geq 1}$ of measurable sets of finite measure such that $X = \bigcup_{n=1}^{\infty} A_n$. If $\mu(X) = 1$, then $(X, S(X), \mu)$ is called a *probability space*. Notice that in a probability space, μ can be seen as the probability, and X as the event space.

Definition A.4. Let $A, B \subset X$. Then $A = B \mod \mu$ if $\mu(A \triangle B) = 0$, where $\triangle$ denotes the symmetric difference[9].

Having established measurable sets and measure, we turn to functions that preserve measure-related properties while moving the points within a space.

Definition A.5. Let $(X, S(X), \mu)$ be a measure space. A transformation $T : X \to X$ is *measurable* if for any $A \in S(X)$, $T^{-1}(A) \in S(X)$, where $T^{-1}(A) = \{x \in X \mid T(x) \in A\}$. T is called *measure-preserving* if it is measurable and if for any $A \in S(X)$, $\mu(A) = \mu(T^{-1}(A))$.

As far as ergodicity and entropy are concerned, measure spaces can be viewed as dynamic objects. When observing the movement of a point x throughout space, $T^n(x)$ gives the position of x at second n.[10] Thus, the sequence $x, T(x), T^2(x), \ldots$ captures the movement of x at every second.

[9] $A \triangle B = (A \setminus B) \cup (B \setminus A)$.
[10] Composing T with itself n times is denoted as T^n.

Below is the Borel-Cantelli Lemma, which is central in the proof of the Brin-Katok formula. It asserts that if the sum of the measures of a countable collection's measurable sets is finite, then almost every point in the space lies in at most finitely many of those measurable sets.

Lemma A.6 (Borel-Cantelli Lemma). *Let $(X, S(X), \mu)$ be a probability space. Let $\{E_n\}_{n\geq 1}$ be a countable collection of measurable sets. If $\sum_{n=1}^{\infty} \mu(E_n) < \infty$, then*

$$\mu(\bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} E_j) = 0.$$

Proof. By measure properties, we have that for all $i \in \mathbb{N}$,

$$\mu(\bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} E_j) \leq \mu(\bigcup_{j=i}^{\infty} E_j) \leq \sum_{j=i}^{\infty} \mu(E_j). \tag{A.7}$$

By the assertion, $\lim_{i\to\infty} \sum_{j=i}^{\infty} \mu(E_j) = 0$. The LHS of (A.7) is a lower bound of the RHS expression, so it follows that $\mu(\bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} E_j) \leq 0$. But measure are nonnegative only, so the result follows. □

A.2. **Approximations with Sufficient Semi-Rings.** Sometimes, proving properties about measurable sets with a specific collection of sets is difficult. The concepts below allow us to manipulate properties of measurable sets by using approximations or generalizations of a specific collection of sets.

Definition A.8. A *semi-ring* on a nonempty set X is a collection $R(X)$ of subsets of X such that

- $R(X)$ is nonempty.
- $R(X)$ is closed under intersection.
- If $A, B \in R(X)$, then $A \setminus B = \bigcup_{j=1}^{n} E_j$, where $E_j \in R$ are disjoint.

Definition A.9. Let $(x, S(X), \mu)$ be a measure space. Then a semi-ring R of subsets of X with finite measure is called a *sufficient semi-ring* of $(X, S(X), \mu)$ if for every $A \in S(X)$,

$$\mu(A) = \inf\left\{\sum_{j=1}^{\infty} \mu(E_j) \mid A \subset \bigcup_{j=1}^{\infty} E_j, \text{ where } E_j \in R \text{ for all } j \geq 1\right\}.$$

Sufficient semi-rings allow us to approximate the main property with another, more suitable collections of sets. For instance, consider the unit interval $[0, 1)$. The collection of intervals on $[0, 1)$ is a σ-algebra, but to prove a result with it may be cumbersome. Instead, consider the dyadic intervals $[\frac{k}{2^n}, \frac{k+1}{2^n})$, which form a sufficient semi ring on $[0, 1)$. They are approximations of the possible intervals on $[0, 1)$ but more specific. Often times, proving a result on a sufficient semi-ring is sufficient to prove it for the σ-algebra.

A.3. **Lebesgue Integration Theorems.** We now present the most important results of Lebesgue integration. These results aid in showing the required limiting arguments in the Birkhoff Ergodic Theorem. The proofs of Fatou's Lemma and the Dominated Convergence Theorem can be found in [5].

Lemma A.10 (Fatou's Lemma)**.** *Let $\{f_n\}$ be a sequence of nonnegative measurable functions. Then*

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Theorem A.11 (Dominated Convergence Theorem)**.** *Let h be an integrable non-negative function. Let $\{f_n\}$ be a sequence of measurable functions such that $\lim_{n\to\infty} f_n$ exists a.e.*

Set $f(x) = \lim_{n\to\infty} f_n(x)$. If $|f_n| \leq h$ a.e. for any $n > 0$, then f is integrable, and

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$