SHIFTS AND ERGODIC THEORY

ANTONIO DONALD

Abstract. This paper introduces the basics of ergodic theory and applies it to probabilistic shift spaces. We define Bernoulli schemes by formally constructing the product space given a base space X. We define measure-preserving and ergodic maps, and show that Bernoulli schemes are both measure-preserving and ergodic. We then define subshifts of finite type and Markov measures, and show the conditions under which the Markov shift is ergodic.

Contents

1. Introduction

Measure-preserving maps are maps where the measure of the pre-image of a set is the same as the measure of the original set. In this paper, we will study and define a sub-class of measure-preserving maps called ergodic maps. Ergodic maps are measure-preserving maps which cannot be further decomposed into smaller measure-preserving maps (we formalize this definition in Section 4). We will use ergodic theory to study sequence spaces (topological space of infinite sequences) from a probabilistic perspective. Ergodic theory originated in statistical mechanics, to study the states of a closed system through time. In this paper, we will use Birkhoff's Ergodic Theorem to study sequence spaces and shift maps. We assume that the reader has working knowledge of measure theory. We begin by introducing the definition of measure-preserving maps and some examples. We will formally define sequence spaces and Bernoulli schemes and then present Birkhoff's ergodic

Date: September 2021.

theory. The overall goal of this paper is to use Birkhoff's Ergodic Theorem to analyze when the shift map on different sequence spaces is ergodic.

2. Measure-Preserving Maps

This paper uses ergodic theory to analyze shift spaces from a probabilistic viewpoint. To that end, we define probability spaces:

Definition 2.1. A *probability space* is a triple $(X, \mathcal{S}, \mu)$, where X is a topological space, $\mathcal{S}$ is a $\sigma-$algebra on X, and μ is a measure on $(X, \mathcal{S})$ such that $\mu(X) = 1$.

Throughout this paper, we will always let $(X, \mathcal{S}, \mu)$ denote a probability space. Given $(X, \mathcal{S}, \mu)$, the probability of an event in some set B occurring is $\mu(B)$. Before defining a measure-preserving transformation, we first recall the definition of a measurable transformation. Let $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{M}, \nu)$ be probability spaces.

Definition 2.2. We say that $T : X \to Y$ is *measurable* if for all $A \in \mathcal{M}$, $T^{-1}(A) \in \mathcal{S}$. In plain English, the pre-image of every measurable set in Y is measurable in X.

Definition 2.3. A map $T : X \to Y$ is *measure-preserving* if for all $A \in \mathcal{M}$, $\mu(T^{-1}(A)) = \nu(A)$. In other words, the pre-image of a measurable set has the same measure.

For an intuitive example of a measure-preserving map, let us consider the circle S^1.

Example 2.4. Consider S^1 as $\mathbb{R}/\mathbb{Z}$. We can put the normal Lebesgue measure from the real line onto S^1. Taking $\alpha \in \mathbb{R}$, rotation of the circle by α is the map

$$\begin{aligned} \mathcal{R}_\alpha : S^1 &\to S^1 \\ x &\mapsto x + \alpha. \end{aligned}$$

It is easy to see that $\mathcal{R}_\alpha$ is measurable. To see that $\mathcal{R}_\alpha$ is measure-preserving, notice that rotation does not change arc length. Thus, the pre-image of any measurable set is just a collection of arcs of the same length. Hence, $\mathcal{R}_\alpha$ is measure-preserving. To visualize this argument, consider the rotation on S^1 by $\frac{\pi}{2}$ depicted in Figure 1. Let A be the arc from π to $\frac{3\pi}{2}$. Then, the pre-image of A under T is the arc from 0 to $\frac{\pi}{2}$.

Notice that when we defined measure-preserving transformations we required that the pre-image of a measurable set have the same measure as the set itself. We will now work through an example that shows that this definition is not equivalent to the forward image of a measurable set having equal measure.

Example 2.5. Let's consider the doubling map (on the circle)

$$\begin{aligned} T : [0, 1) &\to [0, 1) \\ T(x) &= 2x \mod 1 \end{aligned}$$

where we endow $[0, 1)$ with the usual Borel $\sigma-$algebra and Lebesgue measure. T is indeed measure-preserving. Figure 2 illustrates $T^{-1}([0, \frac{1}{2})$. Notice that the pre-image of $[0, \frac{1}{2})$ is

$$\left[0, \frac{1}{4}\right) \bigcup \left(\frac{3}{4}, 1\right].$$

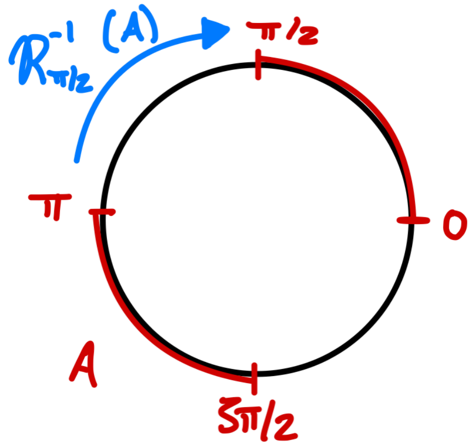

FIGURE 1. Rotation of the circle is measure-preserving.

Clearly both $[0, \frac{1}{2})$ and its pre-image have measure $\frac{1}{2}$. For any set $A \in \mathcal{S}$, $T^{-1}(A)$ will be two disjoint sets with measure $\frac{\mu(A)}{2}$. Hence, T is measure-preserving. Notice that $T([0, \frac{1}{4})) = [0, \frac{1}{2}]$. So, for $A \in \mathcal{S}$, $\mu(A)$ does not necessarily equal $\mu(T(A))$. Therefore, the pre-image of a measurable set preserving measure is not equivalent to the forward image preserving measure.

Why do we care if maps are measure-preserving? We can answer this question from a variety of perspectives. The origins of ergodic theory lie in statistical mechanics. The state of a closed system is a measure-preserving transformation, and ergodic theory was designed to analyze these systems over time. For further reading about statistical mechanics and ergodic theory, [1] contains a great essay on why ergodic theory and shifts are important. For our purposes we care about measure-preserving maps because shifts on probability spaces are measure-preserving! Hence, we will use the ergodic theory we introduce in Section 4 to analyze different shift spaces.

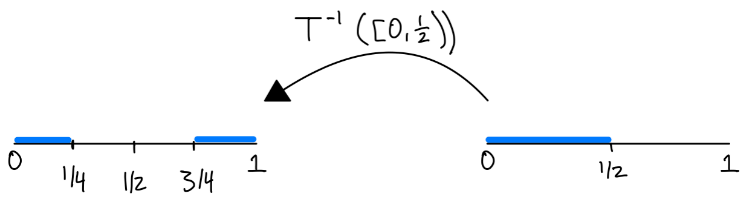

FIGURE 2. The map $T = 2x \mod 1$ is measure-preserving.

3. Bernoulli Shifts

In this section, we will formally define sequence spaces given an abstract topological space X. For probability spaces, we will construct a measure on the sequence space and define the shift map on these spaces. The combination of a sequence space and a probability measure on that space is is called a Bernoulli shift space, which we formally define at the end of this section.

3.1. **Sequence Spaces.**

Let X be a topological space. For any such X, we can define a sequence space. Following [2] we define the set of all (integer) sequences on X as $B(X)$ where $\theta \in B(X)$ is a map $\theta : \mathbb{Z} \to X$. Written in another way,

$$\theta = (\ldots, \theta_{-1}, \theta_0, \theta_1, \ldots) \text{ where } \theta_i \in X \text{ for } i \in \mathbb{Z}.$$

So, elements $\theta \in B(X)$ are just infinite sequences of elements of X. We can consider $B(X)$ to be the infinite cartesian product

$$\prod_{i=-\infty}^{\infty} X_i.$$

Because we are working with an infinite product, we will follow standard practice and apply the product topology to $B(X)$. The product topology is the coarsest topology (smallest number of open sets) such that the projections onto each X_i are continuous [3].

Definition 3.1. The basis of the product topology are sets $U \in B(X)$ of the form

$$U = \prod_{i=-\infty}^{\infty} U_i$$

where each U_i is open in X and $U_i \neq X_i$ for only finitely many i. Every open set $O \in B(X)$ can be expressed as a union of basis sets U. [1]

We will consider sequence spaces from a probabilistic perspective. Thus, our spaces will always be of the form

$$X = \{1, \dots, n\},$$

where we can interpret each element of X as a possible outcome of an experiment. Endow X with the discrete topology (every point is open). Let each outcome $1, \dots, n$ have probability

$$p_1, p_2, \dots, p_n$$

respectively and

$$\sum_{i=1}^{n} p_i = 1.$$

We will use these probabilities to construct a probability measure on X. Let $\mathcal{S}$, our σ-algebra, be the power set of X (the set of all subsets of X) and for $A \in \mathcal{S}$ we define

$$\mu(A) = \sum_{x \in A} p_x,$$

where p_x is the probability of event x. For example, let us describe rolling a die as a probability space X. We have

$$X = \{1, 2, 3, 4, 5, 6\},$$

where each number represents a possible roll of the die. For a fair die,

$$p_1 = p_2 = \dots = p_6 = \frac{1}{6}.$$

Then, the measure of any set $A \in \mathcal{S}$ is just the number of faces in that set. For example, one set in $\mathcal{S}$ is $A = \{1, 3, 5\}$. Then, $\mu(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$, which aligns with the probability of rolling an odd number.
Now let $B(X)$ have the product topology of X as described above. We will now construct a probability measure on $B(X)$ by extending our measure on X.

Following [2], we construct the desired measure space on $B(X)$. Consider finitely many Borel sets $A_i \in X$ (when $X = \{1, \dots, n\}$ the Borel sets are just any combination of events), $0 \leq i \leq m$, where $m \in \mathbb{N}$. Take some integer $z \in \mathbb{Z}$. Then, we can define

Definition 3.2. A *cylinder set* $C(j, A_0, \dots, A_m)$ is the set in $B(X)$ whose jth through $j+m$th coordinates are in A_i. That is,

$$C(j, A_0, \dots, A_m) = \{\theta \in B(X) : \theta(j+i) \in A_i, 0 \leq i \leq m\}$$

[1]This is not the most obvious way to define a topology on an infinite product space, however issues arise when we consider infinite products of open sets. The topology where infinite products of open sets are open is called the box topology. For more reading on why it is standard to use the product topology instead of the box topology, I refer the reader to Chapter 19 of [3].

We can check that disjoint unions of such cylinder sets form an algebra which generates the Borel $\sigma-$algebra on our product space $B(X)$. Given a measure μ_0 on our original space X, we can define the measure μ of some cylinder set as

$$\mu(C(j, A_0, \ldots, A_m)) = \prod_{i=0}^{m} \mu_0(A_i).$$

This measure on a cylinder set follows nicely from our original probability measure and does not run into issues because we do not have to take infinite products! For a formal construction of μ from μ_0 using the Cartheodory Extension Theorem see [2]. We have now extended our probability spaces $(X, \mathcal{S}_0, \mu_0)$ to a measure space (a probability space, so $\mu(B(X)) = 1$) on the infinite product: $(B(X), \mathcal{S}, \mu)$.

Consider our die example from earlier where $X = \{1, \ldots, 6\}$ and $p_1 = \ldots = p_6 = 1/6$. We will give an example of a cylinder set on $B(X)$. Consider the probability that our first and second rolls of the die are both 1. From elementary probability we know that the probability of rolling two 1s in a row is $\frac{1}{36}$. Let $A = \{1\}$. We can write this cylinder set as

$$C(1, A, A).$$

We know that the measure of this cylinder should be $\frac{1}{36}$. From our construction of the measure of a cylinder in 3.2, we calculate that

$$\mu(C(1, A, A)) = \mu(A) \cdot \mu(A) = \frac{1}{36}.$$

It can also be verifies that for any finite chain of events our cylinder sets are the probability of these events occurring and our definition aligns with elementary probabilistic notions.

3.2. **Shift Maps.** The purpose of this paper is to explore ergodic theory through the lens of shift maps. Now, we can actually formally introduce shift maps! Let $X = \{1, \ldots, n\}$ be a probability space with associated probabilities $p_1, \ldots, p_n$. In Subsection 3.1 we constructed the sequence probability space $B(X)$, which is the space of all bi-infinite sequences in X. Then, we extended our probability measure on X to a probability measure on the cylinder sets of $B(X)$. We will now consider a class of maps from $B(X) \to B(X)$ called shift maps. The idea of a shift map is we will shift the position of our sequence to the left or the right by some integer amount. The shift we will analyze most is the left shift:

Definition 3.3. The *left shift* map is the transformation $\sigma : B(X) \to B(X)$ where $\sigma(\theta)(n) = \theta(n + 1)$. That is, we shift all events over to the left by one and return a new sequence.

The left shift takes a sequence and moves every element to the left one position. Let's work through an example to visualize exactly what is going on:

Example 3.4. Let $X = \{1, 2\}$. Then, $B(X)$ is just all possible sequences of 1s and 2s. Take $\theta = \{\ldots, 1, 1, 1, 2, 1, 1, 1, \ldots\}$ with 2 in the 1st position and 1s everywhere else. Then, $\sigma(\theta)$ has 2 at position 0 and 1 at every other position.

We can generalize the idea of a shift map. There is no particular reason why we chose the left shift to analyze. It would make just as much sense to shift a sequence to the right. The *right shift* is the map

$$\sigma_r : B(X) \to B(X) \qquad \sigma_r(\theta)(n) = \theta(n - 1).$$

Applying σ_r to θ from Example 3.4, $\sigma_r(\theta)$ would have 2 at the second position and 1s everywhere else. Moreover, we can also have shift maps which shift an arbitrary integer amount to the left or right instead of shifting by one position. For example, we could consider the left shift by 3 transformation. Formally, this would be the map

$$\sigma_{3l} : B(X) \to B(X) \qquad \sigma(\theta)(n) = \theta(n+3).$$

Applying σ_{3l} to θ from Example 3.4 we would have 2 at the negative second position and $1s$ everywhere else.

We stated in Section 2 that we care about measure-preserving transformations because they help us analyze shift maps and sequence spaces. To that end, we will now show that the left shift map is measure-preserving (the same argument holds true for any general shift map). Fix $X = \{1, \dots, n\}$, with associated probabilities $p_1, \dots, p_n$. Fix $j \in \mathbb{Z}$ and $m \in \mathbb{N}$. Let $A_0 \dots, A_m$ be measurable sets in $\mathcal{S}$ (the power set of X). Let $C(j, A_0, \dots, A_m)$ be a cylinder set. Then, applying the left shift to our cylinder we notice that

$$\sigma^{-1}(C(j, A_0, \dots, A_m)) = C(j+1, A_0, \dots, A_m).$$

Hence, the pre-image of our cylinder is a cylinder with the same exact sets $A_0, \dots, A_m$ just shifted over one position to the right. Therefore, we can see that

$$\mu(C(j+1, A_0, \dots, A_m)) = \prod_{i=0}^{m} \mu_0(A_i) = \mu(C(j, A_0, \dots, A_m)).$$

Hence, the shift map σ is measure-preserving! To see that a general shift map is measure-preserving, notice that the measure of a cylinder is the product of the measure of the sets $A_0, \dots, A_m$ which are not the whole space X. Since there are only finitely many such sets in a cylinder, their position in the sequence does not alter the measure of the cylinder. Therefore, any integer shift to the left or the right will not change the measure of a cylinder, and all general shift maps are measure-preserving. Moreover, we will show in Section 4 that σ is ergodic. We now end this section by combining probability space $B(X)$ and a general shift map σ to define what is known as a *Bernoulli scheme*. Given some set $X = \{1, \dots, n\}$, a Bernoulli scheme is the sequence space $B(X)$ along with a shift map σ. We can interpret this probabilistically by saying that Bernoulli schemes are infinite sequences of finite possible outcomes of an experiment, where the shift map changes our starting point in time. Formally, we define Bernoulli schemes as:

Definition 3.5. Let $X = \{1, \dots, n\}$. Endow X with the discrete topology. Let $p = (p_1, \dots, p_n)$ be a probability vector where p_i is the probability of event i. Let $B(X)$ be the sequence space of X as defined above. Let $\mathcal{S}$ be the Borel σ-algebra of cylinders on $B(X)$, and let μ be our product measure on $\mathcal{S}$. Let σ be the left shift map on $B(X)$. Then, the quadruple

$$(B(X), \mathcal{S}, \mu, \sigma)$$

is the *Bernoulli scheme defined by* p. We will denote this Bernoulli scheme by $BS(p)$. For an alternate construction of $BS(p)$, please look at [1].

Now let's translate flipping a fair coin into a Bernoulli scheme. We define $X = \{1, 2\}$ where 1 is heads and 2 is tails [2] . For a fair coin we have

$$p = (\frac{1}{2}, \frac{1}{2}).$$

We construct $B(X)$ to be the sequence space of all sequences of heads and tails and apply the probability measure on cylinder sets that we defined in Section 3.1. Let σ be the left shift map. Then,

$$BS(\frac{1}{2}, \frac{1}{2}) = (\{1, 2\}, \mathcal{S}, \mu, \sigma)$$

is the Bernoulli scheme of all infinite sequences of coin flips along with the left shift. Our measure μ is the probability of getting heads or tails in an integer position of flips and our left shift changes our starting time (consider 0 as the start of time).

4. Ergodicity

4.1. **Defining Ergodicity.**

We have defined measure-preserving maps and Bernoulli shifts so far. In this section we will introduce ergodicity (which is a stronger condition for maps than being measure-preserving) and show that Bernoulli shifts are ergodic. There are many equivalent definitions for ergodicity. We begin with a very intuitive definition. Let $(X, \mathcal{S}, \mu)$ be a probability space.

Definition 4.1. A map $T : X \to X$ is *ergodic* if T is measure-preserving and for $A \in \mathcal{S}$ if $T^{-1}(A) = A$, then $\mu(A) = 0$ or 1.

This may seem like an arbitrary definition, but following [5] we will arrive at a very reasonable explanation. Let's consider what happens if T is measure-preserving, but not ergodic. Then, there is at least one set $A \in \mathcal{S}$ such that $\mu(A) \in (0, 1)$ and $T^{-1}(A) = A$. Moreover, $\mu(X\backslash A) \in (0, 1)$ and $T^{-1}(X\backslash A) = X\backslash A$. Thus, we can break such a map T into two simpler maps $T|_A$ and $T|_{X\backslash A}$. In a sense, this means ergodic maps are measure-preserving maps which we cannot decompose further. So, we can think of ergodic maps as the base case of measure-preserving maps, and thus the interesting ones to study. Let's revisit rotations of the circle. Following [4] we can use our definition of ergodicity to prove that irrational rotations of the circle are ergodic.

Theorem 4.2. *Consider the circle S^1 as $\mathbb{R}/\mathbb{Z}$ with Lebesgue measure and the standard Borel sigma algebra. Let α be an irrational number. Then, rotation by α,*

$$\mathscr{R}_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z} \qquad \mathscr{R}_\alpha(x) = x + \alpha \mod \mathbb{Z}$$

is ergodic.

Proof. It is an elementary result in dynamics that the orbit under $\mathscr{R}_\alpha$ of every point in $\mathbb{R}/\mathbb{Z}$ is dense in $\mathbb{R}/\mathbb{Z}$. Thus, $n\alpha \mod \mathbb{Z}$ is dense in $\mathbb{R}/\mathbb{Z}$ for $n \in \mathbb{Z}$. Let $B \subset \mathbb{R}/\mathbb{Z}$ such that $\mathscr{R}_\alpha(B) = B$. Fix $\varepsilon > 0$. Then, there exists a continuous function $f \in L^1(\mathbb{R}/\mathbb{Z})$ (because continuous functions are dense in L^1) such that

$$\|f - \chi_B\|_1 < \varepsilon.$$

[2] A binary Bernoulli scheme is also known as a *Bernoulli process*. Bernoulli schemes are a generalization of Bernoulli processes.

Because B is invariant under $\mathscr{R}_\alpha$, we can apply the triangle inequality to see that

$$\|f \circ \mathscr{R}_\alpha^n - f\|_1 \leq \|f \circ \mathscr{R}_\alpha^n - \chi_B\|_1 + \|\chi_B - f\|_1 \leq 2\varepsilon.$$

Notice that $R_\alpha^n = x + n\alpha \mod \mathbb{Z}$, so we have that

$$\|f(x+n\alpha) - f(x)\|_1 \leq 2\varepsilon$$

for all $n \in \mathbb{N}$ and $x \in \mathbb{R}/\mathbb{Z}$. Because the orbit $n\alpha$ is dense in $\mathbb{R}/\mathbb{Z}$ and f is continuous, we have that

$$\|f(x+t) - f(x)\|_1 \leq 2\varepsilon \tag{4.3}$$

for all $t \in \mathbb{R}/\mathbb{Z}$. We can apply Fubini's theorem to see that

$$\begin{aligned} \left\| f - \int f(t)dt \right\|_1 &= \int \left| \int f(x) - f(x+t)dt \right| dx \\ &\leq \int \int |f(x) - f(x+t)dxdt \\ &\leq 2\varepsilon \end{aligned}$$

where the last inequality holds by (4.3). Notice that our choice of f (because f is arbitrarily close to χ_B) gives us that

$$\|\chi_B - f\|_1 < \varepsilon \text{ and } \left\| \int f(t)dt - \mu(B) \right\|_1 < \varepsilon.$$

Therefore, applying the triangle inequality we have

$$\|\chi_B - \mu(B)\|_1 \leq \|\chi_B - f\|_1 + \left\| f - \int f(t)dt \right\|_1 + \left\| \int f(t)dt - \mu(B) \right\|_1 < 4\varepsilon.$$

Because our choice of ε was arbitrary, we must have that χ_B is constant almost everywhere. Hence, $\mu(B)$ is either 0 or 1 and $\mathscr{R}_\alpha$ is ergodic. □

It would be a good exercise to verify that rational rotations are not ergodic. However, we will not prove that in this paper. Our goal is to show that Bernoulli schemes are ergodic. It is not always easy to determine whether a map is ergodic using only Definition 4.1. So, we will now present some equivalent definitions of ergodicity which are proved in Theorem 1.5 of [5].

Theorem 4.4. *Let $(X, \mathcal{S}, \mu)$ be a probability space, and let $T : X \to X$ be a measure-preserving transformation. Then, the following statements are equivalent*

1. *T is an ergodic map.*
2. *For $A \in \mathcal{S}$, if $\mu(T^{-1}(A) \triangle A) = 0$ then $\mu(A) = 0$ or $\mu(A) = 1$.*
3. *For $A \in \mathcal{S}$ such that $\mu(A) > 0$, then $\mu(\bigcup_{n=1}^\infty T^{-n}A) = 1$.*
4. *If $A, B \in \mathcal{S}$ and $\mu(A), \mu(B) > 0$, then there is some $n \in \mathbb{N}$ so that*

$$\mu(T^{-n}A \cap B) > 0.$$

We will now use these definitions of ergodicity to prove that Bernoulli schemes are ergodic.

Theorem 4.5. *Let $B(X), \mathcal{S}, \mu, \sigma$ be a Bernoulli scheme. Then, the left shift map, σ, is ergodic.*

Proof. Suppose $A \in \mathcal{S}$ is a set such that $T^{-1}(A) = A$. We will show that $\mu(A) = 0$ or 1 by showing that $\mu(A)^2 = \mu(A)$. We defined $\mathcal{S}$ to be the σ-algebra generated by unions of cylinder sets in Section 3. Fix $\varepsilon > 0$. We can choose a finite union of cylinder sets B so that

$$\mu(A \triangle B) < \varepsilon.$$

Then, we can calculate that

$$\begin{aligned} |\mu(A) - \mu(B)| &= |\mu(A \backslash B) - \mu(B \backslash A) + \mu(A \cap B) - \mu(A \cap B) \\ &\leq \mu(A \backslash B) + \mu(B \backslash A) \\ &\leq \varepsilon, \end{aligned} \tag{4.6}$$

where the first equality is just expanding the difference of A and B and the last inequality holds by definition of symmetric difference.

Notice that B is just a union of cylinders. Hence, the measure of B is determined only by the finitely many cylinder sets which are not the whole space X. We will say that Bs constituent coordinates are the positions in the sequence which contribute to the measure of B. Because B only has finitely many constituent coordinates, there is some $n \in \mathbb{N}$ so that $C = \sigma^{-n}(B)$ has constituent coordinates which are all to the right of Bs constituent coordinates. For example, if $X = \{1, 2\}$, and B is the cylinder where the $0th$ coordinate has to be 2, then n would be 1 and C would have to be 2 in the 1st coordinate. Because B and C are disjoint in the portions that contribute to their measures, we can see that

$$\mu(B \cap C) = \mu(B)\mu(C) = \mu(B)^2, \tag{4.7}$$

where the last equality follows because B and C have the same measure. Applying (4.7) and the fact that $\sigma^{-1}(A) = A$, we can see that

$$\begin{aligned} \mu(A \triangle C) &= \mu(A \triangle \sigma^{-n}(B)) \\ &= \mu(\sigma^{-n}(A) \triangle \sigma^{-n}(B)) \\ &= \mu(A \triangle B) \\ &< \varepsilon \end{aligned} \tag{4.8}$$

Notice that by expanding the symmetric difference, we have that $A \triangle (B \cap C) \subset (A \triangle B) \cup (A \triangle C)$. So, applying (4.6) and (4.8) we get

$$\mu(A \triangle (B \cap C)) < 2\varepsilon. \tag{4.9}$$

Putting everything together, we get

$$\begin{aligned} |\mu(A) - \mu(A)^2| &\leq |\mu(A) - \mu(B \cap C)| + |\mu(B \cap C) - \mu(A)^2| \\ &\leq 2\varepsilon + |\mu(B)^2 - \mu(A)^2| \\ &\leq 2\varepsilon + \mu(B)|\mu(B) - \mu(A)| + \mu(A)|\mu(B) - \mu(A)| \\ &\leq 4\varepsilon, \end{aligned}$$

where the first inequality is the triangle inequality, the second uses (4.9) and (4.7), the third uses the triangle inequality, and the fourth uses the fact that $B(X)$ is a probability space. Because our choice of ε was arbitrary, $\mu(A) = \mu(A)^2$, and σ is ergodic. $\square$

4.2. **The Ergodic Theorem.**
Consider a probability space $(X, \mathcal{S}, \mu)$ and a measure-preserving transformation T. Let us suppose that $X = \{1, \ldots, n\}$ with associated probabilities $p_1, \ldots, p_n$. It is natural to ask how often the orbit of a point or a set under T visits other points and sets. For example, if we consider a set $A = \{1, 2\} \subset X$, for given $x \in X$, what is the average as $n \to \infty$ of

$$\chi_A(T^n(x)).$$

Does this average align with the probability of event A happening? More generally, consider some $f : X \to \mathbb{R}$. The ergodic theorem answers the question; when does

$$\lim_{n\to\infty} \frac{f(x) + f(T(x)) + f(T^2(x)) + \ldots + f(T^{n-1}(x))}{n}$$

converge? Does the convergence depend on choice of T? Does it depend on choice of $x \in X$? These are the motivating questions behind Birkhoff's Ergodic Theorem. Birkhoff proved that for any $f \in L^1(X)$ and any choice of measure-preserving T, the limit converges almost everywhere. Moreover, if T is ergodic, then the limit is constant almost everywhere and does not depend on the choice of ergodic transformation.

We will not prove this beautiful theorem. Many proofs are widely available. Also, every proof that I have seen is very technical and lengthy. If interested in the full details of the proof, I would strongly recommend either the proof in [2] or [5]. That being said, it will be important for our purposes to accurately state the theorem:

Theorem 4.10 (Birkhoff's Ergodic Theorem)**.** *Let $(X, \mathcal{S}, \mu)$ be a probability space and $T : X \to X$ a measure-preserving map. Then, for $f \in L^1(X)$:*

(1) The limit

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$$

converges almost everywhere in X. We can define a function $\hat{f}(x) : X \to \mathbb{R}$ as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)).$$

(2) For $f \in L^p(X), 1 \leq p < \infty$, the limit converges to $\hat{f}$ in L^p:

$$\lim_{n\to\infty} \left\| \hat{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \right\|_p = 0$$

(3) Defined as above, we have that $\hat{f}(T(x)) = \hat{f}$ almost everywhere.

(4) Finally, for $f \in L^p(X)$:

$$\int_X \hat{f} d\mu = \int_X f d\mu.$$

There is a lot to unpack in Birkhoff's Ergodic Theorem! We can think of the map $\hat{f}$ as an orbital average of f, because $\hat{f}$ averages the values achieved by f on a T orbit of x. If we consider when $f = \chi_A$ for some $A \in \mathcal{S}$. Then, $\hat{f}$ is the average amount of time spent by x in A (interpreting T as some sort of time map). More officially, following [2], we have,

Definition 4.11. Take $A \in \mathcal{S}$, define

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

Notice that Birkhoff's Ergodic Theorem tells us that

$$\int_X \tau_A(x) d\mu = \mu(A).$$

We can now introduce another equivalent definition of Ergodicity based on properties of $\hat{f}$:

Theorem 4.12. *Let $(X, \mathcal{S}, \mu)$ be a probability space. Let $T : X \to X$ be a measure-preserving map. Then, the following conditions are equivalent:*

(1) T is ergodic;
(2) If $f \in L^1(X)$ and $f \circ T(x) = f(x)$ almost everywhere, then f is constant almost everywhere;
(3) If $f \in L^p(X)$ and $f \circ T(x) = f(x)$ almost everywhere, then f is constant almost everywhere;
(4) For all $A, B \in \mathcal{S}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-m}(A) \cap B) = \mu(A)\mu(B);$$

(5) For every $f \in L^1(X)$,

$$\hat{f} = \int_X f d\mu$$

almost everwhere.

Proof. (1) $\Rightarrow$ (2). Suppose that T is ergodic. Let $f \in L^1(X)$ be such that $f \circ T = f$ almost everywhere. Then, define a set

$$A_m = \{x \in X : f(x) \leq m\} \text{ for } m \in \mathbb{R}.$$

We know that $f \circ T = f$ almost everywhere, so $f(T(A_m)) = f(A_m)$ almost everywhere. Therefore (taking A_m without the above measure zero set), $T(A_m) = A_m$ and thus $\mu(A_m) = 0$ or 1. This implies that there exists some $m \in \mathbb{R}$ such that

$$\mu(A_m) = 1 \text{ and } \mu(A_{m-\varepsilon}) = 0 \text{ for all } \varepsilon > 0.$$

Thus, f is constant almost everywhere. Notice that if $f \in L^p(X)$, $1 < p < \infty$, then $f \in L^1(X)$ so (1) $\Rightarrow$ (3) as well.
(3) $\Rightarrow$ (1) Suppose that $f \circ T = f$ almost everywhere, for $f \in L^p(X)$, implies that f is constant almost everywhere. For $A \in \mathcal{S}$, if $T^{-1}(A) = A$, then $\chi_(A) = \chi_(T(A))$ almost everywhere and $\chi_A \in L^p(X)$. Thus, by assumption χ_A is constant almost everywhere. That is only true for a characteristic function if A has zero measure or full measure. This exactly implies that if $T^{-1}(A) = A$, then $\mu(A) = 0$ or 1. Thus, T is ergodic and (3) $\Rightarrow$ (1).
(5) $\Rightarrow$ (4). Suppose that for every $f \in L^1(X), \hat{f} = \int_X f d\mu$ almost everywhere. Then, for any $A \in \mathcal{S}$, we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) = \hat{\chi}_A = \int_X \chi_A = \mu(A).$$

Therefore, substituting $\mu(A)$ for the above, we have

$$\mu(A)\mu(B) = \int_X (\hat{\chi}_A(x))\chi_B d\mu. \tag{4.13}$$

Notice taking

$$\chi_{A_n} = \frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i(x))$$

that we can apply the Dominated Convergence Theorem to (4.13) getting

$$\begin{aligned} \mu(A)\mu(B) &= \lim_{n\to\infty} \frac{1}{n}\int_X \sum_{i=0}^{n-1} \chi_A(T^i(x))\chi_B d\mu \\ &= \lim_{n\to\infty} \sum_{i=0}^{n-1} \mu(T^{-j} \cap B) \end{aligned}$$

where in the last step we notice $\chi_A(T^j(x))\chi_B(x)$ is only 1 if $x \in T^{-j}A \cap B$. Thus, we have shown that (5) $\Rightarrow$ (4).
(2) $\Rightarrow$ (5). Birkhoff's Ergodic Theorem (Theorem 4.10) says that $\hat{f} \circ T = \hat{f}$ almost everywhere. Then, assuming (2) holds, $\hat{f}$ must be constant almost everywhere. Birkhoff's Ergodic Theorem tells us that

$$\int_X \hat{f} d\mu = \int_X f d\mu$$

So, because X is a probability space (and $\mu(X) = 1$), $\hat{f} = \int_X f d\mu$ almost everywhere, as needed.

Remark 4.14. If $0 < \mu(X) < \infty$, then

$$\hat{f} = \frac{1}{\mu(X)}\int_X f d\mu$$

almost everywhere.

(4) $\Rightarrow$ (1). Suppose that for $A \in \mathcal{S}$, $T^{-1}(A) = A$. Applying (4) to A and A^c we get

$$\mu(A)\mu(A^c) = \lim_{n\to\infty} \mu(T^{-j}(A) \cap A^c) = 0$$

because $T^{-j}(A) = A$ and $\mu(A \cap A^c) = 0$. Hence, either $\mu(A) or \mu(A^c)$ has measure zero, which implies $\mu(A) = 0$ or 1, as needed. ☐

5. Subshifts of Finite Type and Markov Shifts

In Section 4, we introduced multiple definitions of ergodicity and proved that Bernoulli shifts are ergodic. In this section, we will introduce subshifts of finite type and Markov shifts. Subshifts of finite type are a class of sequence spaces where only a certain set of transitions between elements of X are allowed in sequences. We will then introduce Markov shifts, where each transition between elements of X in a sequence has a certain probability conditional only on the previous element. Finally, we will prove the conditions that need to be satisfied for a Markov shift to be ergodic.

5.1. **Subshifts of finite type.**

We will introduce two equivalent definitions for subshifts of finite type. We begin this section by defining subshifts of finite type as a set of sequences determined by a directed graph, following [1]. We will then present an alternative definition of subshifts of finite type, using matrices, which is a little less intuitive but easier to perform manipulations with. We now begin without first definition. First, we have to understand what a directed graph is. Informally, a directed graph is a set of vertices with directional edges connecting some or all of these vertices. Formally, we have

Definition 5.1. A *directed graph* G is a pair (V, E), where V is a set of elements called vertices and E is a set of ordered pairs of vertices in V. If (a, b) is an ordered pair in E, then there is an edge from a to b.

To visualize this definition, let us consider an example. Let G be the graph

$$V = \{1, 2, 3\} \text{ and } E = \{(1, 2), (1, 3), (2, 3), (3, 1), (3, 3)\}.$$

Looking at Figure 3, we can see that G has three nodes and edges from 1 to 2, 2 to 3, 3 to 1, and 3 to itself.

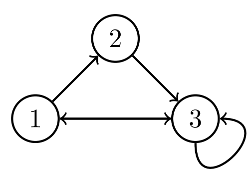

FIGURE 3. Graph $G = (V, E)$.

Let $X = \{1, \ldots, n\}$. Given a graph with n nodes, a subshift of finite type allows only transitions in sequences between nodes of the graph connected by edges. That is, some transitions between elements cannot occur. Thus, our subshifts of finite type are a subset of the sequence space we defined in Section 3, $B(X)$. From a probabilistic perspective, given a sequence of random variables, some events cannot occur given a certain previous event. Thus, subshifts of finite type allow us to study sequences where there are restrictions on transitions! Our first formal definition of subshifts of finite type is:

Definition 5.2. Suppose we have a directed graph G. Let V be the set of vertices and E be the set of edges for G. We will assume that E is finite ($|E| < \infty$). We define a space X to be the set of all infinite sequences θ such that each index of theta corresponds to a valid edge in E. That is, $\theta \in X$ is a infinite path of the graph G. Let σ be the left shift defined in 3. Then, the dynamical system (X, σ) is called a *subshift of finite type*.

Let us work through an example of a subshift of finite type to visualize what is going on:

Example 5.3. Take a graph G with two vertices $V = \{0, 1\}$ and $E = \{(0, 0), (0, 1), (1, 0)\}$. Our graph G, depicted in Figure 4, has three edges: one edge from 0 to itself, one edge from 0 to 1, and one edge from 1 to 0. Our graph G is only missing an edge from 1 to itself. An example of an infinite sequence of 0s and 1s which would NOT be in our space X is

$$\theta = \{1, 1, 1, \ldots\}$$

because our graph does not have an edge from 1 to 1, so $\theta \in X$ cannot have two or more 1s adjacent to each other. On the other hand,

$$\theta = \{0, 0, 0, 1, 0, 1, 0, 0, 1, 0, \ldots\}$$

is a valid member of X because all transitions between elements are edges in our graph G.

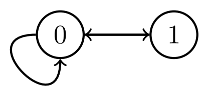

FIGURE 4. Graph with edges from vertex 0 to 1, 1 to 0, and 0 to itself.

Notice that the Bernoulli scheme we defined in Section 3 is a special case of a subshift of finite type, where G is the complete graph with edges between all vertices (each vertex is connected to every vertex) in G. We will call Bernoulli schemes the *full n-shift* where $n = |X|$. We will now introduce a second definition of subshifts of finite type which can be easier to manipulate.

Definition 5.4. Choose some $n \in \mathbb{N}$ and an $n \times n$ matrix A whose entries are all either 0 or 1. Given n and A, we construct:

(1) Our alphabet $\mathcal{A}$ of the subshift. We define

$$\mathcal{A} = \{1, \ldots, n\}.$$

(2) A subspace $X \subset B(n)$, where $B(n)$ is the Bernoulli scheme of some space $\{1, \ldots, n\}$ and

$$X := \{\theta = (\ldots, x_{-1}, x_0, x_1, \ldots) \in B(n) : A(x_t, x_{t+1}) = 1 \text{ for all } t \in \mathbb{Z}\}.$$

(3) A map (the left shift) $\sigma : X \to X$ such that $\sigma(\theta) = \theta'$, and $\theta'_t = \theta_{t+1}$.

Our alphabet $\mathcal{A}$, space X, and shift σ are called a *subshift of finite type*.

How does this more formal definition relate to Example 5.3? Instead of a graph with 2 vertices, our subshift of finite type will have a base space $n = \{1, 2\}$, and an $n \times n$ matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Where A is the matrix which allows transitions from 1 to 1, 1 to 2, and 2 to 1, but not 2 to 2. In fact if we choose our alphabet $\mathcal{A} = \{0, 1\}$, then we can see that our space X and our elements $\theta \in X$ are exactly the same at those in Example 5.3.

5.2. **Markov Shifts.** In the previous subsection, we defined subshifts of finite type. In a subshift of finite type, there is a binary choice whether one element can appear next to another element in a sequence. What if we want to consider the probability that a transition occurs between one element and any next element? Markov shifts are the sequence spaces which formalize this idea. In this section, we will define Markov measures and Markov shifts and will conclude by proving the conditions necessary for a Markov shift to be ergodic. We will be following [2] and [5]. Let us again consider the full n-shift space $B(n)$.

Definition 5.5. A measure μ on $(B(n), \mathcal{S}, \sigma)$, where $\mathcal{S}$ is the σ-algebra generated by finite unions of cylinder sets and σ is the left shift, is a *Markov measure* if μ is a σ-invariant measure such that there exist $p_{i,j} > 0, 1 \leq i, j \leq n$ so that

$$\mu(C(z, a_1, \ldots, a_l) = p_{a_1}, p_{a_1,a_2} \ldots p_{a_{l-1},a_l}$$

for each cylinder set in $B(n)$. We can interpret the probabilities $p_{i,j}$ as the conditional probability of a_j given a_{j-1}. Notice that for μ to be a Markov measure, the measure does not depend on indexing position $m \in \mathbb{Z}$, only on the sequence of events a_i.

To define a Markov shift, we first have to define what a stochastic matrix and its associated probability vectors are:

Definition 5.6. A pair called a probability vector and stochastic matrix make up what we call a *Markov shift*. An $n \times n$ *stochastic matrix* P has coordinates p_{ij} such that

(1) Each row satisfies $\sum_{j=1}^{n} p_{ij} = 1$;
(2) There exists a *probability vector* $\mathbf{p}$=$(p_1, \ldots, p_n)$, where $p_i > 0$ for $1 \leq i \leq n$, such that
$$\sum_{i=i}^{n} p_i = 1$$
and
$$\sum_{i=1}^{n} p_i p_{ij} = p_j \text{ for all } 1 \leq j \leq n.$$

Let us work through an example to visualize what a Markov shift is:

Example 5.7. We can verify that $\mathbf{p} = (\frac{1}{2}, \frac{1}{2})$ is the probability vector for stochastic matrix
$$P = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}.$$
The matrix P is a stochastic matrix with probability vector $\mathbf{p}$.

Given a stochastic matrix and its probability, it is possible to construct a unique Markov measure associated with the pair. We will not present the proof here, as it is fairly technical, but the construction can be found in Theorem 10.1 of [2]. Given a stochastic matrix, its probability vector and their associated Markov measure, we can define the Markov shift:

Definition 5.8. Given a stochastic matrix P and its associated probability vector $\mathbf{p}$, we can construct a unique Markov measure μ on $B(n)$. The left shift σ on this sequence space is called a *Markov shift*.

Ergodicity is a very useful property when analyzing sequence spaces. Unlike Bernoulli shifts, Markov shifts are not always ergodic. We will now start showing that Markov shifts are ergodic if and only if the stochastic matrix P is irreducible (or a few other equivalent conditions).

Definition 5.9. A matrix A is *irreducible* if there exists $t \in \mathbb{N}$ so that A^t has all strictly positive entries.

To prove the theorem, we will use a lemma from [5] showing that the $Q = \lim_{k=0}^{n} \frac{1}{n}\sum_{k=0}^{n-1} P^n$ exists whenever P is a stochastic matrix and $\mathbf{p}$ has strictly positive entries. We can interpret Q as being the average under iteration of P, which is a somewhat similar idea to the orbital average and Birkhoff's Ergodic Theorem.

Lemma 5.10. *Let* ($\mathbf{p}$,P) *be a probability vector and its Stochastic matrix, i.e.* $\mathbf{p}P = \mathbf{p}$. *Then,*

$$Q = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} P^n$$

exists. Moreover, Q is stochastic, $QP = PQ = Q$, $Q^2 = Q$ and if v is an eigenvector of P with eigenvalue 1 *then v is also an eigenvector of Q.*

Proof. The majority of our proof will be showing that Q exists. Once we've done this, the rest of the properties will follow immediately. Let μ be the Markov measure of $(\mathbf{p}, P)$. Let σ be the left shift on $B(n)$. We showed earlier that σ is measure-preserving, so we will be able to apply Birkhoff's Ergodic Theorem. Let's consider the cylinder $C(0, i)$ for $1 \le i \le n$. That is, the cylinder with $\theta \in B(n)$ such that $\theta_0 = 1$. Let $\chi_{C(0,i)}$ be the characteristic function of this cylinder. Fix both $1 \le i, j \le n$. Birkhoff's Ergodic Theorem tells us that for $\theta \in B(n)$

$$\frac{1}{n}\sum_{k=0}^{n-1} \chi_{C(0,j)}(\sigma^k(\theta) \to \hat{\chi}_{C(0,j)})(\theta) \in L^1(\mu).$$

Multiplying both sides by $\chi_{C(0,i)}$ and applying the Dominated Convergence Theorem we get

$$\lim_{n\to\infty} \int_X \frac{1}{n}\sum_{k=0}^{n-1} \chi_{C(0,i)}\chi_{C(0,j)}(\sigma^k(\theta))d\mu \to \int_X \hat{\chi}_{C(0,j)}(\theta)\chi_{C(0,i)}d\mu \tag{5.11}$$

By additivity of the integral and the fact that X is a probability space, the left integral in (5.11) is

$$\int_X \frac{1}{n}\sum_{k=0}^{n-1} \chi_{C(0,i)}\chi_{C(0,j)}(\sigma^k(\theta))d\mu = \frac{1}{n}\sum_{k=0}^{n-1} \chi_{C(0,i)}\chi_{C(0,j)}(\sigma^k(\theta))d\mu. \tag{5.12}$$

And (5.12) is just the probability given i in the 0th spot of θ that j is in the kth position of θ. We can write this as

$$\int_X \frac{1}{n}\sum_{k=0}^{n-1} \chi_{C(0,i)}\chi_{C(0,j)}(\sigma^k(\theta))d\mu = \frac{1}{n}\sum_{k=0}^{n} p_i p_{ij}^{(k)}. \tag{5.13}$$

Here $p_{ij}^{(k)}$ is the p_ij in P^k. Plugging (5.13) into (5.11) we get that

$$\frac{1}{n}\sum_{k=0}^{n} p_i p_{ij}^{(k)} \to \int_X \hat{\chi}_{C(0,j)}(\theta)\chi_{C(0,i)}d\mu. \tag{5.14}$$

Notice that $q_{ij} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} p_{ij}^{(k)}$. Hence, (5.14) tells us that

$$q_{ij} = \frac{1}{p_i}\int_X \hat{\chi}_{C(0,j)}(\theta)\chi_{C(0,i)}d\mu.$$

Therefore, Q exists. Because the limit exists, it's clear that $QP = Q = P^{-1}PQ$. Moreover, if $P(v) = v$, then $Q(v) = \frac{1}{n}nv = v$. Clearly, $Q^2 = Q$ because of uniqueness of limits. Hence, we are done. ☐

With the previous lemma, we will be able to show when a Markov shift is ergodic:

Theorem 5.15. *Let σ denote the $(\boldsymbol{p}, P)$ Markov shift. Let Q be the matrix in lemma 5.10. Then, the following conditions are equivalent:*

(1) The shift map σ is ergodic;
(2) All rows of Q are the same;
(3) Every entry q_{ij} of Q is strictly positive;
(4) P is irreducible;
(5) 1 is a simple Eigenvalue of P.

Proof. (1) $\Rightarrow$ (2). Suppose that σ is ergodic. We showed in Lemma 5.10 that

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(C(0,i)\cap C(i,j)) &= \lim_{n\to\infty}\sum_{k=0}^{n-1} p_i p_{ij}^(k) \\ &= \int_X \hat{\chi}_{C(0,j)}\chi_{C(0,i)}d\mu.\end{aligned}$$

Since σ is ergodic, $\hat{\chi}_{C(0,j)} = p_j$ almost everywhere, so our argument in Lemma 5.10 tells us that

$$p_i q_{ij} = p_i p_j.$$

Therefore, $q_{ij} = p_j$, which does not depend on i so all rows of Q are the same.

(2) $\Rightarrow$ (3). Suppose that the rows of Q are identical. We showed in Lemma 5.10 that if v is an eigenvector with eigenvalue 1 for P, then v is also an eigenvector with eigenvalue 1 for Q. Therefore,

$$\mathbf{p}Q = \mathbf{p}. \tag{5.16}$$

Since each row of Q is the same, we must have that $q_ij = p_j$ for (5.16) to hold. Since $\mathbf{p}$ has all positive entries, then each q_{ij} must be positive.

(3) $\Rightarrow$ (4). Suppose that each entry of Q is positive. Fix $1 \leq i, j \leq n$. We know that

$$\frac{1}{n}\sum_{k=0}^{n-1} p_{ij}^{(k)} \to q_{ij} > 0.$$

Take $\varepsilon = \frac{q_{ij}}{2}$. Then, there exists $t_{ij} \in \mathbb{N}$ such that

$$p_{ij}^{t_{ij}} > q_{ij} - \varepsilon > 0.$$

Take $t = \max_{1\le i,j\le n}\{t_{ij}\}$. Then, $p^t > 0$, as needed.
(3) ⇒ (2). Suppose that every entry of Q is positive. We will show that all rows of Q are the same. Fix $1 \le j \le n$. Let

$$q_j = \max_{1\le i\le n}\{q_{ij}\}.$$

We showed in Lemma 5.10 that $Q^2 = Q$. So if $q_i j < q_j$, for some $1 \le i \le n$, then

$$q_{lj} = \sum_i q_{li}q_{ij} < \sum_i q_{li}q_j = q_j.$$

for all $1 \le l \le n$. The equality comes from matrix multiplication, and the last inequality holds because Q is a stochastic matrix. But, then there should be some l such that $q_{lj} = q_l$. Hence, we have a contradiction and must have $q_{ij} = q_j$ for all i. Hence, all rows of Q are identical.
(2) ⇒ (1). Suppose that all rows of Q are identical. We showed in Theorem 4.12 that to show σ is ergodic, it suffices to chow that for all

$$A = C(a, i_0, \ldots, i_r), B = C(b, j_0, \ldots, j_s) \in S$$

that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(\sigma^{-k}A \cap B) = \mu(A)\mu(B).$$

For large enough k, i.e. $k > b + s - a$, we can calculate based on the formula for measure of a cylinder that

$$\mu(\sigma^{-k}(A) \cap B) = p_{j_0}p_{j_0j_1}\ldots p_{j_sj_{s-1}}p^{(k+a-b-s)}_{j_si_o}p_{i_0i_1}\ldots p_{i_ri_{r-1}}.$$

Because the rows of Q are identical, $q_{ij} = p_j$. So, $p^{(k+a-b-s)}_{j_si_0} \to p_{i_0}$. Therefore,

$$\begin{aligned}\mu(\sigma^{-k}(A) \cap B) &= p_{j_0}p_{j_0j_1}\ldots p_{j_sj_{s-1}}p_{i_0}p_{i_0i_1}\ldots p_{i_ri_{r-1}}\\ &= \mu(A)\mu(B).\end{aligned}$$

Hence, σ is ergodic.
(2) ⇒ (5). We know that Q having identical rows implies $q_{ij} = p_j$. Then, if v is an eigenvector of Q with eigenvalue 1, then v is a multiple of $\mathbf{p}$. Hence, Lemma 5.10 tells us that only multiples of $\mathbf{p}$ are eigenvalues of P with eigenvalue 1. Therefore, 1 is a simple eignenvalue of P.
(5) ⇒ (2). Suppose that 1 is a simple eigenvalue of P. We want to show Q has identical rows. Because $QP = Q$ by Lemma 5.10, each row of Q is an eigenvector of P. Therefore, each row of Q must be the same.
(4) ⇒ (3). Fix $1 \le i \le n$. Let

$$S_i = \{1 \le j \le n : q_{ij} > 0\}.$$

We showed in Lemma 5.10 that $Q = QP$. Hence, $q_{ij} = \sum_{l=1}^n q_{il}p_{lj}$ and $q_{ij} \ge q_{il}p_{lj}$ for each l. Thus, if $l \in S_i$ and $p_l j$ is positive, then q_{ij} must also be positive, which implies $j \in S_i$. Hence, if $l \in S_i$, then

$$\sum_{j\in S_i} p_{lj} = 1.$$

Because P is irreducible, $|S_i| = n$ for each i, so $q_{ij} > 0$ for each ij, as needed. □

Thus, we have proved that Markov shifts are ergodic only when the stochastic matrix P is irreducible. This is the last major result we will prove. The theory of shift maps and sequences is diverse and beautiful. We have only scratched the surface of the probabilistic and topological applications of ergodic theory on sequence spaces. For further reading, [5],[2],[1] explore beautiful areas of ergodic theory.

Acknowledgments

Thank you so much Meg! I could not have done this without your help. It has been a pleasure talking about math and life with you throughout the REU! Thank you for guiding me through reading, suggesting great material, and always helping me when I was stuck. I would also like to thank Professor Peter May for organizing a fantastic REU. Finally, I would like to thank all of the speakers for taking time out of their lives to give some great talks!

6. Bibliography

References

[1] T. Bedford, M. Keane, C. Series Ergodic Theory and Subshifts of Finite Type Oxford University Press, 1991
[2] Ricardo Mañé Ergodic Theory and Differentiable Dyamics Springer, 1987
[3] James Munkres Topology Prentice Hall, 2000
[4] Siming Tu (https://math.stackexchange.com/users/20611/siming-tu). Ergodicity of irrational rotation. URL (version: 2014-12-31): https://math.stackexchange.com/q/1086185
[5] Peter Walters. An Introduction to Ergodic Theory Springer, 2000.