MARTINGALE APPROXIMATIONS FOR ERGODIC SYSTEMS

ELI BUSSEL

Abstract. Chaotic dynamical systems are difficult to predict exactly, but methods from probability theory can be used in approximation to describe these systems. This paper presents a method due to Gordin for approximating a system by a martingale, a construct from probability theory which models a "fair game" or "random walk." We use martingale approximation to prove a central limit theorem for dynamical systems. Though this result is known, the current literature on the topic is cursory and informal. We present a complete discussion of the martingale approximation method and a self-contained proof of the central limit theorem, following notes by Vaughn Climenhaga.

Contents

Introduction

Many dynamical systems exhibit *chaotic* behavior, an informal term often used to describe unpredictability or a high degree of complexity. If a chaotic system is treated as a "random" process, however, methods from probability theory can describe its large-scale behavior. This paper focuses on approximating systems as martingales, a method originally formulated by Mikhail Gordin in 1969 [1]. Vaughn Climenhaga outlined a proof of the central limit theorem for dynamical systems using martingale approximation, based on a 2013 lecture by Matt Nicol at the University of Houston [2]. Following notes from Climenhaga's blog, we present a complete proof of the central limit theorem for dynamical systems using a martingale approximation.

Sections 1-3 define preliminary terminology from measure theory and probability theory, culminating in the definition of a martingale. Section 4 introduces concepts from ergodic theory and begins discussion of the main result. Section 5 is devoted to the proof of the central limit theorem.

Date: August 28^{th}, 2018.

1. Basic probability theory

Probability theory describes the outcomes of an experiment, and how to best predict which outcomes will occur. The term "experiment" here refers to any activity with uncertain outcomes, such as rolling a die or playing a game of cards. We are primarily interested in modeling chaotic dynamical systems as probabilistic experiments.

We refer to the set of possible outcomes of a given experiment as Ω. Events such as "the die came up even" or "Bob won the game" are formally described as subsets of Ω. We say that an event A has happened if one of the outcomes in A has happened.

It is not reasonable to assume that every possible event is observable. An event may not be detectable with the amount of information available to the observer. We must therefore define which events are observable and which are not.

Definition 1.1. Let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a *σ-algebra on* Ω if $\mathcal{F}$ satisfies the following conditions:

i. Ω is in $\mathcal{F}$.
ii. If A is in $\mathcal{F}$, then its complement $A^c = \Omega \setminus A$ is in $\mathcal{F}$.
iii. If A and B are in $\mathcal{F}$, then $A \cup B$ is in $\mathcal{F}$.
iv. if $\{A_n\}$ is a countable subset of $\mathcal{F}$, then $\bigcup_{n\in\mathbb{N}} A_n$ is in $\mathcal{F}$.

We call $(\Omega, \mathcal{F})$ a *measurable space.* A set $A \in \mathcal{F}$ is called a *measurable set.*

A σ-algebra is a set of events which can be observed or verified, given some information or measuring tools. The set Ω can be thought of as the event "something happened." We always assume that some outcome occurs, so Ω is always measurable. The union of A and B is the event "A or B happened," and the complement of A is the event "A did not happen." Our definition of a σ-algebra is based on our intuitive understanding that these events are observable, given that A and B are observable.

Although the above definition only specifies closure under unions and complements, we can readily construct other set operations, such as intersections and set differences, from unions and complements. It is therefore important to note the following:

Remark 1.2. *A σ-algebra is closed under any finite or countable number of binary set operations.*

Now that we know which events are measurable, we can begin to quantify how likely they are to occur. We accomplish this by defining a function that maps measurable sets to real numbers.

Definition 1.3. Let $\mathcal{F}$ be a σ-algebra on Ω. A function $\mu\colon \mathcal{F} \to \mathbb{R}^+$ is a *measure* on Ω if μ satisfies the following conditions:

i. $\mu(\emptyset) = 0$.
ii. (Additivity) If A and B are disjoint sets in $\mathcal{F}$, then

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

iii. (Countable additivity) If $\{A_n\}$ is a countable collection of disjoint sets in $\mathcal{F}$, then

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

If $\mu(\Omega) < \infty$, we say that μ is *finite*. We call the triple $(\Omega, \mathcal{F}, \mu)$ a *measure space*.

In this paper, we will only consider finite measures.

The conditions listed above are based on a few axiomatic observations. Since we assume that some outcome has happened, the event "nothing happened," represented by the empty set, must have no chance of occurring. This requirement is formalized our first condition. Additivity and countable additivity are based on the observation that, for mutually exclusive events A and B, the probability that either A or B occurs is the sum of their individual probabilities.

For the moment, we are primarily interested in the measure that describes the probability of events, which we denote by $\mathbb{P}$. By convention, a probability is a real number between 0 and 1.

Definition 1.4. If $\mathbb{P}$ is a measure on Ω, and $\mathbb{P}(\Omega) = 1$, then we call $\mathbb{P}$ a *probability measure*. We then call the triple $(\Omega, \mathcal{F}, \mathbb{P})$ a *probability space*.

Non-empty events may occur with zero probability. Similarly, an event with probability 1 is not necessarily equal to Ω; it may have negligible exceptions.

Definition 1.5. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We say that $A \in \mathcal{F}$ is *μ-null* if $\mu(A) = 0$. If $\mu(A) = \mu(\Omega)$, then we say *A has full measure in μ.*

Remark 1.6. *A set A has full measure if and only if A^c is null.*

Remark 1.7. *Any finite or countable union of null sets is null.*

More thorough discussions of measure spaces and their properties can found in [3, ch. 1-2] [4, §1.1].

Only events have probabilities, and thus individual outcomes cannot be assigned probabilities. However, it is often useful to assign values to outcomes that coincide with some property of the outcomes themselves. We accomplish this with separate functions mapping Ω to $\mathbb{R}$ directly. In some circumstances, we interpret such a function as a *random variable*, a quantity whose value depends on the outcome of a process.

Eventually, we wish to discuss events that correspond to random variables. Perhaps a random variable X represents the amount of money a gambler wins or loses on a bet. We may wish to measure the probability that X lies within a given range. To discuss such probabilities, however, we need to create a σ-algebra on $\mathbb{R}$.

Definition 1.8. Let $\mathcal{T}$ be the topology on the reals; i.e. the collection of all open sets in $\mathbb{R}$. The *Borel σ-algebra* $\mathcal{B}$ is the smallest σ-algebra on $\mathbb{R}$ such that $\mathcal{T} \subset \mathcal{B}$. An element of $\mathcal{B}$ is called a *Borel set*. We call $(\mathbb{R}, \mathcal{B})$ the *Borel space*.

Constructing $\mathcal{B}$ with open sets conveniently guarantees that any arbitrary interval in $\mathbb{R}$ is measurable. Thus, we can easily predict if a gambler's earnings are, say, between \$10 and \$20.

When measuring the values of random variables as outcomes, we must be careful that these events always correspond to events in our probability space. Otherwise, there is no link between probabilities assigned in the Borel space and the actual probabilities we seek to measure. Thus, we will insist that every Borel set corresponds to a measurable set.

Definition 1.9. Let $f\colon \Omega \to \mathbb{R}$ be a function, and let $\mathcal{F}$ be a σ-algebra on Ω. We say that f is *$\mathcal{F}$-measurable* if, for all Borel sets B, the preimage $f^{-1}(B)$ is in $\mathcal{F}$. We equivalently say that f is *measurable in $\mathcal{F}$*, or simply *measurable* when there is no ambiguity.

We sometimes call a measurable function in a probability space a *random variable*, which we commonly denote by capital letters X, Y, Z, etc.

We briefly discuss a few basic properties of measurable functions. From now on, we work within the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, unless otherwise specified.

Lemma 1.10. *Let $f\colon \Omega \to \mathbb{R}$. Then the image and preimage of f preserve set operations. That is, for all Borel sets A, B,*

(a) $f(A \cup B) = f(A) \cup f(B)$.
(b) $f(\Omega \setminus A) = \mathbb{R} \setminus f(A)$.
(c) $f(A \cap B) = f(A) \cap f(B)$.

Analogous statements are also true for the preimage.

Proof. These are entirely properties of set operations. For instance, both $f(A\cup B)$ and $f(A) \cup f(B)$ are equal to $\{f(x) \mid x \in A \text{ or } x \in B\}$, and the similar expressions with the preimages are equal to $\{x \mid f(x) \in A \text{ or } f(x) \in B\}$. Analogous arguments hold for all other binary set operations. □

Remark 1.11. *By a similar argument, the image and preimage of f also preserve countably many set operations. It follows that, if f is measurable, then $f^{-1}(\mathcal{B}) := \{f^{-1}(B) \mid B \in \mathcal{B}\}$ is a σ-algebra on the domain of f.*

The following properties have more complicated proofs, whose specifics are beyond the scope of this paper. We do not prove them here, but complete proofs, along with more information on measurable functions, can be found in [3, ch. 3].

Theorem 1.12. *Let f and g be measurable.*

(a) The function $f + g$ is measurable.
(b) The function $f \cdot g$ is measurable.
(c) For all $\lambda \in \mathbb{R}$, λf is measurable.
(d) Any finite combination of the above operations on measurable functions is itself a measurable function.

We now define a probability measure on $\mathcal{B}$ that links $\mathbb{P}$ to a random variable X. The measure of a Borel set B will be the probability that the value of X is contained in B.

Definition 1.13. Let X be a random variable. The *distribution of X* is the function $\mu_X\colon \mathcal{B} \to [0, 1]$, where

$$\mu_X(B) = \mathbb{P}(X^{-1}(B))$$

for all Borel sets B.

We say that two random variables X, Y are *identically distributed* if $\mu_X = \mu_Y$. We sometimes write $X =_d Y$ for brevity.

Theorem 1.14. *Let X be a random variable with distribution μ. Then μ is a probability measure on the Borel space.*

Proof. The preimage of the empty set is empty, so

$$\mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0.$$

Let A, B be disjoint Borel sets. Theorem 1.10 gives us that

$$X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B),$$

and we have that

$$\mathbb{P}[X^{-1}(A) \cup X^{-1}(B))] = \mathbb{P}[X^{-1}(A)] + \mathbb{P}[X^{-1}(B)].$$

Thus,

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

A similar argument gives us countable additivity.

Since $X^{-1}(\mathbb{R}) = \Omega$, $\mu(\mathbb{R}) = \mathbb{P}(\Omega) = 1$. □

This result justifies the notation $\mathbb{P}(X \in B) := \mu_X(B)$.

It is often difficult to show that two random variables are equal. However, if the set of values where $X \neq Y$ is null (i.e. negligible), then X and Y are close enough for most purposes.

Definition 1.15. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Two measurable functions f and g are *equal almost everywhere* if the set $\{\omega \in \Omega \mid f(\omega) \neq g(\omega)\}$ is μ-null. If $(\Omega, \mathcal{F}, \mu)$ is a probability space, then we say that f and g are equal *almost surely.*

We sometimes use the abbreviation *a.e.* for "almost everywhere" and *a.s.* for "almost surely."

Of principal importance to us is that almost sure equality preserves distribution.

Theorem 1.16. *Let X, Y be random variables. If $X = Y$ almost surely, then X and Y are identically distributed.*

Proof. Let $A = \{\omega \mid X(\omega) \neq Y(\omega)\}$. Since $X = Y$ almost surely, A is null. Let B be a Borel set. By additivity,

$$\mathbb{P}[X^{-1}(B)] = \mathbb{P}(X^{-1}(B) \setminus A) + \mathbb{P}(A),$$

and the same applies for Y. Since A is null,

$$\mathbb{P}[X^{-1}(B)] = \mathbb{P}[X^{-1}(B) \setminus A],$$

and similarly for Y. By hypothesis, $X^{-1}(B) \setminus A = Y^{-1}(B) \setminus A$, so we conclude that X and Y are identically distributed. □

It is sometimes more practical to discuss the distribution as a function of reals rather than a function on a measure space. The following definition gives us the machinery to do so:

Definition 1.17. Let X be a random variable with distribution μ. The *distribution function* of X is the function $F \colon \mathbb{R} \to [0, 1]$, where $F(x) = \mu(-\infty, x]$.

Note that, for all $x \in \mathbb{R}$, $(-\infty, x] = (x, \infty)^c$ is a Borel set, so F is well-defined.

The similarity between between the terms "distribution" and "distribution function" is somewhat confusing. However, the distribution and distribution function of a random variable are essentially the same, only viewed from different perspectives. Thus, what applies to one almost always applies to the other.

Sequences and series of random variables are treated much like their real counterparts. Convergence, however, must be discussed carefully, as random variables by their nature do not often have well-defined limiting behavior.

Definition 1.18. Let $\{X_n\}$ be a sequence of random variables, where each X_n has distribution function F_n. We say that X_n *converges in distribution* to a function $F \colon \mathbb{R} \to [0,1]$ if, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Definition 1.18 is stronger than the usual definition of convergence in distribution, which restricts the limit to continuity points. However, this definition directly implies the usual one, and is more applicable to contexts relevant to this paper.

We are principally interested in sequences which converge to a specific distribution function called the *normal distribution,* often known informally as the "bell curve."

Definition 1.19. The *standard normal distribution function* N is the function

$$N(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{2\pi} dy.$$

Since the function N has no closed form, the normal distribution is often described by its integrand, which is called its *density function.* The graph of the normal density function is the familiar bell curve.

The most famous appearance of the standard normal distribution is in the Central Limit Theorem, which states that the sum of independent, identically distributed random variables with finite variance converges to standard normal when properly normalized. In reference to this result, many other constructs which converge to standard normal are said to have "central limit theorems." Of particular interest to us is the central limit theorem for martingales, which we discuss in Section 3.

More information on the properties of distributions and distribution functions can be found in [4, §1.2-3].

2. Lebesgue integration and expectation

Given a random variable X, we wish to calculate the most accurate prediction of its value in the long-term. We can think of this number as the "average" of X, or the value that is closest to the value of X most often. We will call this number the *expectation* of X.

If X has a finite set of values, calculating the expectation of X is simple. We take a weighted average, multiplying each value by its probability and summing the products. In many systems of interest, however, X can take on infinitely many values. In calculus, the integral serves as a method of "averaging" infinitely many numbers, but the Riemann integral normally encountered in calculus is not well-defined for many functions on measure spaces. This motivates the construction of the *Lebesgue integral*, an extension of the integral which we can apply to an arbitrary measure space.

In defining the Lebesgue integral, we begin with the simplest case, where a measurable function f takes on finitely many values. We can then express f in terms of functions which "detect" if an argument belongs to a given value. We first define how this "detection" works.

Definition 2.1. Let A be a subset of Ω. The *indicator function of A* is the function $\mathbf{1}_A : \Omega \to \mathbb{R}$, where

$$\mathbf{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}.$$

We briefly check that indicator functions are measurable.

Theorem 2.2. *The indicator function of A is measurable in every σ-algebra containing A.*

Proof. Let B be a Borel set. If B contains 1 but not 0, its preimage is A. If B contains 0 but not 1, its preimage is $\Omega \setminus A$. If B contains both 1 and 0, its preimage is Ω, and if B contains neither, its preimage is $\emptyset$. Any σ-algebra containing A must contain all four of these sets, and thus $\mathbf{1}_A$ is measurable on all such σ-algebras. $\square$

We now define a way to express functions with finite values as sums of indicator functions. We imagine dividing Ω into disjoint events, where outcomes in the same event have the same value under f. As noted above, the Lebesgue integral of f will simply be the weighted average of each value, where the weights are the probabilities of the corresponding events.

Definition 2.3. A measurable function $f : \Omega \to \mathbb{R}^+$ is *simple* if there exists a finite collection of disjoint, measurable sets $\{A_1, A_2, \ldots, A_k\}$ and a corresponding set of positive reals $\{a_1, a_2, \ldots, a_k\}$, such that

$$f = \sum_{i=1}^{k} a_i \cdot \mathbf{1}_{A_i}.$$

The Lebesgue integral of a simple function f is defined to be

$$\int f d\mu = \sum_{i=1}^{k} a_i \mu(A_i).$$

The integrals of simple functions are somewhat similar to the rectangle method of integration encountered in calculus, with each disjoint set defining a rectangle. Just as the Riemann integral is defined as the limit of progressively smaller rectangle approximations, we define the Lebesgue integral of an arbitrary function as the limit of simple approximations.

Theorem 2.4. *Let f be a non-negative, measurable function. Then there exists an increasing sequence of simple functions f_n such that $\lim_{n\to\infty} f_n = f$.*

We do not prove this theorem here, but [5, Prop. 3.10] provides a proof.

Definition 2.5. Let $f : \Omega \to \mathbb{R}$ be a non-negative, measurable function. Let $\{s \leq f\}$ be the set of simple functions such that $s(x) \leq f(x)$ for all $x \in S$. The Lebesgue integral of f is the least upper bound of the integrals of these simple functions.

$$\int f d\mu = \sup \left\{ \int s d\mu \mid s \leq f \right\}.$$

For functions with positive and negative values, we rewrite them in terms of non-negative functions.

Lemma 2.6. *Let $f\colon \Omega \to \mathbb{R}$ be a measurable function. Define the functions*

$$f^+(x) = \max(f(x), 0)$$
$$f^-(x) = \max(-f(x), 0).$$

Then f^+ and f^- are measurable, non-negative functions, and $f = f^+ - f^-$.

Proof. This is trivial to check. □

Definition 2.7. Let $f\colon \Omega \to \mathbb{R}$ be a measurable function. The Lebesgue integral of f is

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Certain vocabulary, such as "expectation," is specific to the random variable interpretation of measurable functions.

Definition 2.8. The *expectation* of a random variable X is its Lebesgue integral.

$$\mathbb{E}(X) = \int X d\mathbb{P}.$$

Definition 2.9. The *variance* of a random variable is

$$\sigma^2 = \mathbb{E}([X - \mathbb{E}(X)]^2).$$

The *standard deviation* of a random variable is the square root of its variance.

$$\sigma = \sqrt{\sigma^2}.$$

Variance and standard deviation are useful indicators of how "spread out" the values of X are. They also feature prominently in central limit theorems as normalization factors.

Not all measurable functions have a finite Lebesgue integral. We wish to treat integrals as well-defined real numbers, so it is convenient to define a set of functions with finite integrals.

Definition 2.10. Let $f\colon \Omega \to \mathbb{R}$ be a measurable function. We say that f is *Lebesgue integrable*, or simply *integrable*, if $\int |f| \, d\mu < \infty$.

The *Lebesgue space* $L^1(\Omega, \mathcal{F}, \mu)$ is the set of all integrable functions $f\colon \Omega \to \mathbb{R}$.

We often drop the parenthetical arguments when discussing L^1 spaces, but sometimes write them to avoid ambiguity.

We extend the concept of an L^1 space to functions whose squares, cubes, etc. are integrable.

Definition 2.11. For $p \geq 1$, we define the L^p *space* as the set of all functions $f\colon \Omega \to \mathbb{R}$ such that $\int |f|^p \, d\mu < \infty$.

Since each L^p space is a subset of the spaces before it, we commonly refer to a function as L^p for the largest p applicable. In this paper, we will only consider L^1 and L^2 functions.

We briefly establish some properties concerning integrable functions. Like the properties of measurable functions, they are presented here without proof.

Lemma 2.12. *Let f and g be measurable and integrable.*

i. $f + g$ is integrable and $\int f + g d\mu = \int f d\mu + \int g d\mu$.

ii. If f and g are L^2, then $f \cdot g$ is L^1.

Theorem 2.13 (Monotone Convergence Theorem)**.** *Suppose $\{f_n\}$ is a sequence of integrable functions such that the following conditions are satisfied:*

(1) For all $x \in \Omega$, for all $n \in \mathbb{N}$, $f_n(x) \leq f_{n+1}(x)$.

(2) There exists an integrable function f such that $\lim_{n\to\infty} f_n(x) = f(x)$.

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

Just as we define the Riemann integral restricted to an interval, we can also define the Lebesgue integral restricted to a measurable set.

Definition 2.14. For $A \subseteq S$, we define the integral of f on A as

$$\int_A f d\mu = \int \mathbf{1}_A f d\mu.$$

We note that $|\mathbf{1}_A f|^p \leq |f|$ for all p, so $\mathbf{1}_A f$ is integrable if f is integrable. Thus, $\int_A f d\mu$ is well defined for all integrable functions f and all $A \subseteq \Omega$.

The concept of integration on a specific set gives a method of showing that functions are equal almost everywhere. When used with Theorem 1.16, this method can also show that random variables are identically distributed.

Theorem 2.15. *Let f, g be $\mathcal{F}$-measurable and integrable. Suppose that, for all $A \in \mathcal{F}$, $\int_A f d\mu = \int_A g d\mu$. Then $f = g$ almost everywhere.*

Proof. Let $A_{\neq} = \{x \mid f(x) \neq g(x)\}$. We note that $A_{\neq}$ is equal to the complement of $(f-g)^{-1}(\{0\})$. Since $f - g$ is measurable, $\{f = g\} \in \mathcal{F}$, and so $A_{\neq} \in \mathcal{F}$. We further note that

$$A_{\neq} = \bigcup_{n\in\mathbb{N}} \{x \mid |f(x) - g(x)| \geq 1/n\}$$

Let A_n denote a member of the above union. By Remark 1.7, it is sufficient for $A_{\neq}$ to be null that each A_n is null.

By hypothesis, $\int_{A_n} f d\mu = \int_{A_n} g d\mu$. It follows that

$$\int_{A_n} f - g d\mu = \int_{A_n} f d\mu - \int_{A_n} g d\mu = 0.$$

Since $\frac{1}{n} \leq |f-g|$, it follows that

$$\int \frac{\mathbf{1}_{A_n}}{n} d\mu \leq \int_{A_n} |f-g| \, d\mu = 0.$$

The function $\frac{\mathbf{1}_{A_n}}{n}$ is simple, so we evaluate the integral to yield $\mu(A_n)/n \leq 0$. Since n is a natural and μ is non-negative by definition, we conclude that $\mu(A_n) = 0$. $\square$

[3, ch. 5] discusses the construction of the Lebesgue integral further, and [4, §1.4-5] contains detailed descriptions of the properties discussed here. The proof of Theorem 2.15 is found in [4, §5.1], in reference to conditional expectation.

3. Conditional expectation and martingales

Expectation is a useful tool for predicting the value of a random variable. We may, however, have additional information about which values X holds, and this information may influence our prediction. Conditional expectation is the formal construct that specifies how information affects such predictions.

Suppose we are given that an event A has occurred, and we wish to calculate the expected value of some random variable X. Only outcomes in A are relevant to our prediction, and so our most accurate prediction of X is its integral on A. More generally, we can construct a function $f\colon A \mapsto \int_A X d\mathbb{P}$. The values of f will then represent the conditional expectation of X given the event A.

We can be given more information than just a single event, however. Perhaps we are told that two events A and B have occurred. Let $\mathcal{G}$ be the collection of events either known to have occurred, or known to have not. We immediately note the following:

- We trivially know that Ω has occurred.
- If A has occurred, we know that A^c has not occurred.
- If one of A and B has occurred, we know that $A \cup B$ has occurred.
- If one of a countable number of events has occurred, we know the union of these events has occurred.

It follows that $\mathcal{G}$ is a σ-algebra on Ω. Thus, when we are given information about which events have occurred, we are really given a σ-algebra representing the events we are allowed to use in our prediction. We will therefore define conditional probability in terms of σ-algebras rather than individual events.

Definition 3.1. Let X be a random variable, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra on Ω. The *conditional expectation of X given $\mathcal{G}$* is a $\mathcal{G}$-measurable random variable $\mathbb{E}(X \mid \mathcal{G})$ such that, for all events $E \in \mathcal{G}$,

$$\int_E \mathbb{E}(X \mid \mathcal{G}) d\mathbb{P} = \int_E X d\mathbb{P}.$$

Remark 3.2. *By Theorem 2.15, if two random variables Y and Z satisfy Definition 3.1, then $Y = Z$ almost surely. We then refer to Y and Z as* versions *of* $\mathbb{E}[X \mid \mathcal{G}]$.

In light of Remark 3.2, we use $\mathbb{E}[X \mid \mathcal{G}]$ to refer to any version of the conditional expectation.

The existence of $\mathbb{E}[X \mid \mathcal{G}]$ is not discussed here, but is a well-known result following from the Radon-Nikodym theorem. [4, §5.1] contains an example of such a proof.

Expectation is a real number, but *conditional* expectation is defined as a random variable. We can interpret $\mathbb{E}[X \mid \mathcal{G}]$ as our best "model" of X given the available data. The numerical conditional expectation of X given an event A is the integral of $\mathbb{E}[X \mid \mathcal{G}]$ on A.

We are principally concerned with the conditional expectation of a variable in a dynamical system. In this interpretation, the system is in a certain state at time t, and we wish to predict its state at time $t + 1$. In many chaotic systems, however, no possible amount of knowledge about the current state can help us make an accurate prediction. Such systems are closely related to the probabilistic concept of a *martingale.*

Definition 3.3. Let $\{X_n\}$ be a sequence of random variables with finite expectation. Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \ldots$ be a sequence of σ-algebras of Ω. We say that $\{X_n\}$ is a *martingale with respect to* $\{\mathcal{F}_n\}$ if, for all $n \in \mathbb{N}$, X_n is $\mathcal{F}_n$-measurable, and $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ almost surely.

The σ-algebras represent the amount of information available at each time step. In our interpretation, this information tells us about the current and past states of the system. Since we gain additional information after each time step, while retaining information about the past, the sequence of σ-algebras increases over time. If our system is a martingale, then changes are essentially random. Since the system is a "fair game," equally likely to move in any direction, our best guess at each time step is the current state.

Often, martingales take the form of cumulative sums of some property over time. Consider a gambler betting on the outcome of coin flips, and suppose this gambler bets \$1 on heads every round. In the first round, the gambler's earnings are represented by a random variable X, which equals 1 if the outcome is heads and -1 if the outcome is tails. After n coin flips, the gambler's total earnings are given by the sum of n such random variables. The sequence of these sums are a martingale; since each X_n has expectation 0, we expect the gambler to neither win nor lose money. We formalize these statements in the following theorem.

Theorem 3.4. *Let $\{X_n\}$ be a sequence of random variables and $\{\mathcal{F}_n\}$ an increasing sequence of random variables, such that each X_n is $\mathcal{F}_n$-measurable. Suppose that $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = 0$ for all $n \in \mathbb{N}$. Then the sequence $S_n := \sum_{i=1}^{n} X_i$ is a martingale.*

Proof. By Remark 3.2, it is sufficient to show that $\mathbb{E}[S_{n+1} \mid \mathcal{F}_n]$ and S_n have the same integral on all measurable sets $A \in \mathcal{F}_n$.

Since $\{\mathcal{F}_n\}$ is increasing, for all $m \geq n$, X_n is $\mathcal{F}_m$-measurable. Thus, each S_n is $\mathcal{F}_n$ measurable.

By the definition of conditional expectation, for all $A \in \mathcal{F}_n$,

$$\int_A \mathbb{E}[S_{n+1}(X) \mid \mathcal{F}_n] d\mathbb{P} = \int_A \sum_{i=1}^{n+1} X_i d\mathbb{P}.$$

By the properties of integrals,

$$\int_A \mathbb{E}[S_{n+1}(X) \mid \mathcal{F}_n] d\mathbb{P} = \sum_{i=1}^{n+1} \int_A X_i d\mathbb{P}.$$

By hypothesis, $\int_A X_{n+1} d\mathbb{P} = 0$. Thus,

$$\int_A \mathbb{E}[S_{n+1}(X) \mid \mathcal{F}_n] d\mathbb{P} = \sum_{i=1}^{n} \int_A X_i d\mathbb{P}.$$

By the properties of integrals,

$$\int_A \mathbb{E}[S_{n+1}(X) \mid \mathcal{F}_n] d\mathbb{P} = \int_A \sum_{i=1}^{n} X_i d\mathbb{P} = \int_A S_n d\mathbb{P}.$$

□

If $\{X_n\}$ satisfies Theorem 3.4, we call $\{X_n\}$ a *sequence of martingale differences.* This is based on the observation that $X_n = S_{n+1} - S_n$.

Further information on conditional expectation and martingales can be found in [3, ch. 9-10] and [6, §6].

4. Measure-preserving systems and ergodicity

We now define the systems we are studying. We think of a dynamical system as a set Ω equipped with a function $T \colon \Omega \to \Omega$. For a point $x \in \Omega, T(x)$ is the "location" of that point at the next time step.

To apply probabilistic arguments to our dynamical system, we must equip Ω with a measure space. Furthermore, we need T to respect certain properties of the measure space.

Throughout this section, let $(\Omega, \mathcal{F}, \mu)$ be the measure space of interest.

Definition 4.1. Let $T \colon \Omega \to \Omega$ be $\mathcal{F}$-measurable (i.e. for all $A \in \mathcal{F}, T^{-1}(A) \in \mathcal{F}$). We say that T is *measure-preserving* if, for all $A \in \mathcal{F}$, $\mu[T^{-1}(A)] = \mu(A)$.

We call $(\Omega, \mathcal{F}, \mu, T)$ a *measure-preserving system.*

Consider a random variable X on a measure-preserving system. The value of X after n time steps is then given by $X \circ T^n$. The specific value of X changes over time, but $X \circ T^n$ should still have similar general properties as X. That the distribution and integral of X do not change over time is most important to us.

From now on, we work within the measure-preserving system $(\Omega, \mathcal{F}, \mu, T)$.

Theorem 4.2. *Let f be a measurable function. Then for all $n \in \mathbb{N}$, $f \circ T^n$ is a measurable function identically distributed with f.*

Proof. We consider the case where $n = 1$ first. A property of preimages gives us the following relation:

$$(f \circ T)^{-1} = T^{-1} \circ f^{-1}.$$

Let B be a Borel set. By hypothesis, both f and T are $\mathcal{F}$-measurable, and so $T^{-1}[f^{-1}(B)] \in \mathcal{F}$ by definition. Thus, $f \circ T$ is measurable.

Let μ_f be the distribution of f. By definition, $\mu_f = \mu \circ f^{-1}$. Similarly, the distribution of $f \circ T$, which we will denote by $\mu_{f \circ T}$, is

$$\mu_{f \circ T} = \mu \circ (f \circ T)^{-1} = \mu \circ T^{-1} \circ f^{-1}.$$

Since T is measure-preserving, $\mu = \mu \circ T^{-1}$, and so

$$\mu_{f \circ T} = \mu \circ f^{-1} = \mu_f.$$

An inductive argument easily extends the result to $n > 1$. $\square$

Theorem 4.3. *Let f be integrable. Then $\int f d\mu = \int f \circ T d\mu$.*

Proof. Suppose first that f is simple, with corresponding disjoint sets $(A_1, \dots A_k)$ and coefficients $(a_i, \dots a_k)$. It follows that

$$f \circ T = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i} \circ T = \sum_{i=1}^{k} a_i \mathbf{1}_{T^{-1}(A_i)}.$$

Since each A_i is disjoint, the preimages of each A_i must also be disjoint. Thus, $f \circ T$ is a simple function, and so

$$\int f \circ T d\mu = \sum_{i=1}^{k} a_i \mu[T^{-1}(A_i)].$$

Since T is measure-preserving,

$$\int f \circ T d\mu = \sum_{i=1}^{k} a_i \mu(A_i) = \int f d\mu.$$

Now suppose that f is non-negative, but not simple. Using Theorem 2.4, let $\{f_n\}$ be an increasing sequence of simple function where $\lim_{n\to\infty} f_n = f$. It follows that

$f_n \le f$ for all n, and thus $f_n \circ T \le f \circ T$. By the above argument, $\int f_n d\mu = \int f_n \circ T d\mu$ for all n. Thus

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

By the Monotone Convergence Theorem, $\int f d\mu = \int f \circ T d\mu$.

For functions f with negative values, we use Lemma 2.6 to split f into non-negative functions and apply the above argument to each individually. □

We are primarily interested in chaotic systems, which we like think of as randomly "scrambling" points with little discernable pattern. We will discuss this "scrambling" in terms of a property known as *ergodicity.*

Definition 4.4. Let T be measure-preserving. A set $A \in \mathcal{F}$ is *T-invariant* if $A = T^{-1}(A)$.

Definition 4.5. A measure-preserving transformation T is *ergodic* if, for all T-invariant sets $A \in \mathcal{F}$, A is either μ-null or has full measure. We call $(\Omega, \mathcal{F}, \mu, T)$ an *ergodic system.*

Let $(\Omega, \mathcal{F}, \mu, T)$ be an ergodic system with a probability measure. A sequence of random variables $\{X_n\}$ is *ergodic* if there is some measurable function f such that each $X_n = f \circ T^n$.

An ergodic system is not necessarily chaotic; such a system may be highly predictable. Ergodicity is a relatively weak description of a function that "mixes" a space. However, ergodicity provides enough complexity to approximate the system with probabilistic methods.

For more about ergodic theory, see [7] and [8].

5. The martingale approximation

Our primary goal is to prove a central limit theorem for ergodic systems. More specifically, we wish to show that, for a measurable function f in an ergodic system $(\Omega, \mathcal{F}, \mu, T)$, the sequence $S_n(f \circ T) := \sum_{i=1}^{n} f \circ T^i$ converges to a normal distribution when properly normalized. We accomplish this using a similar result for martingales, which is presented here without proof.

Theorem 5.1. *(Martingale Central Limit Theorem) Let $(\Omega, \mathcal{F}, \mu, T)$ be an ergodic system with a probability measure. Suppose $\{Y_n\}$ is an ergodic sequence of martingale differences, and that each Y_n is L^2 with finite variance σ^2. Then $\frac{S_n(Y)}{\sigma\sqrt{n}}$ converges to the standard normal distribution.*

The above formulation is adapted from [2, Thm. 2], but [4, §8.8] contains a more thorough discussion of the central limit theorem for martingales.

If we can show that $S_n(f \circ T)$ converges to the partial sums of a martingale difference sequence, then we can apply this result to obtain the central limit theorem. Before we can begin, however, we must discuss a tool necessary for our argument.

Definition 5.2. Let $(\Omega, \mathcal{F}, \mu, T)$ be a measure-preserving system. The *transfer operator* $\mathcal{P}$ is an L^2 operator (i.e. a function of L^2 functions) such that, for all $\phi, \psi \in L^2$,

$$\int \mathcal{P}(\phi)\psi d\mu = \int \phi \cdot (\psi \circ T) d\mu.$$

The transfer operator is, by definition, the adjoint of the L^2 operator $\phi \mapsto \phi \circ T$; thus, we are guaranteed its existence. More information on the construction of the transfer operator can be found in [9, §1.2.]. Our main purpose in introducing it here is to provide a link to the conditional expectation, which we will use to construct a sequence of martingale differences.

Lemma 5.3. *For all $\phi, \psi \in L^2, \mathcal{P}(\phi + \psi) = \mathcal{P}(\phi) + \mathcal{P}(\psi)$ almost everywhere.*

Proof. Let ϕ, ψ, L^2. Let $A \in \mathcal{F}$. Since the indicator function $\mathbf{1}_A$ is L^2, by the definition of the transfer operator,

$$\int \mathcal{P}(\phi + \psi)\mathbf{1}_A d\mu = \int (\phi + \psi) \cdot (\mathbf{1}_A \circ T) d\mu.$$

By the properties of integrals,

$$\int \mathcal{P}(\phi + \psi) \cdot \mathbf{1}_A d\mu = \int \phi \cdot (\mathbf{1}_A \circ T) d\mu + \int \psi \cdot (\mathbf{1}_A \circ T) d\mu.$$

By the definition of the transfer operator,

$$\int \mathcal{P}(\phi + \psi)\mathbf{1}_A d\mu = \int \mathcal{P}(\phi)\mathbf{1}_A d\mu + \int \mathcal{P}(\psi)\mathbf{1}_A d\mu.$$

Since our choice of $A \in \mathcal{F}$ is arbitrary, we conclude by Theorem 2.15 that $\mathcal{P}(\phi+\psi)$ and $\mathcal{P}(\phi) + \mathcal{P}(\psi)$ are equal almost everywhere. □

Lemma 5.4. *For all $\phi \in L^2, \mathcal{P}(\phi \circ T) = \phi$.*

Proof. Let $\phi \in L^2$. Let $A \in \mathcal{F}$. By definition,

$$\int \mathcal{P}(\phi \circ T)\mathbf{1}_A d\mu = \int (\phi \circ T) \cdot (\mathbf{1}_A \circ T) d\mu.$$

We note that $(\phi \circ T)(\mathbf{1}_A \circ T) = (\phi \mathbf{1}_A) \circ T$. Thus, by Theorem 4.3,

$$\int \mathcal{P}(\phi \circ T)\mathbf{1}_A d\mu = \int \phi \mathbf{1}_A d\mu.$$

By Theorem 2.15, $\mathcal{P}(\phi \circ T) = \phi$ almost everywhere. □

Lemma 5.5. *For all $\phi \in L^2(\mathcal{F}), \mathcal{P}(\phi) \circ T = \mathbb{E}[\phi \mid T^{-1}(\mathcal{F})]$.*

Proof. We first show that $\mathcal{P}(\phi) \circ T$ is $T^{-1}(\mathcal{F})$-measurable. Let $A \in \mathcal{F}$. We note that

$$(\mathcal{P}(\phi) \circ T)^{-1}(A) = T^{-1}[\mathcal{P}(\phi)^{-1}(A)].$$

By definition, $\mathcal{P}$ maps to $L^2(\mathcal{F})$, so $\mathcal{P}(\phi)$ must be $\mathcal{F}$-measurable. Thus, $\mathcal{P}(\phi)^{-1}(A) \in \mathcal{F}$, and so $T^{-1}[\mathcal{P}(\phi)^{-1}(A)] \in T^{-1}(\mathcal{F})$.

Keeping $A \in \mathcal{F}$, we have that

$$\int_{T^{-1}(A)} (\mathcal{P}(\phi) \circ T) d\mu = \int (\mathcal{P}(\phi) \circ T)\mathbf{1}_{T^{-1}(A)} d\mu.$$

We note that $\mathbf{1}_{T^{-1}(A)} = \mathbf{1}_A \circ T$. Thus,

$$\int_{T^{-1}(A)} (\mathcal{P}(\phi) \circ T) d\mu = \int (\mathcal{P}(\phi) \circ T) \cdot (\mathbf{1}_A \circ T) d\mu.$$

As in the proof of Lemma 5.4, it follows that

$$\int_{T^{-1}(A)} (\mathcal{P}(\phi) \circ T) \cdot (\mathbf{1}_A \circ T) d\mu = \int \mathcal{P}(\phi)\mathbf{1}_A d\mu.$$

By the definition of the transfer operator,

$$\int_{T^{-1}(A)} (\mathcal{P}(\phi) \circ T) d\mu = \int \phi \cdot (\mathbf{1}_A \circ T) d\mu,$$

and so, as noted above,

$$\int_{T^{-1}(A)} (\mathcal{P}(\phi) \circ T) d\mu = \int_{T^{-1}(A)} \phi d\mu.$$

We conclude that, by the definition of conditional expectation,

$$(\mathcal{P}(\phi) \circ T) = \mathbb{E}(\phi \mid T^{-1}(\mathcal{F})).$$

□

More information on the transfer operator, its construction, and it properties can be found in [9].

We now arrive at the central limit theorem for ergodic systems. A statement of the theorem adapted from [2, §4] is presented below.

Theorem 5.6. *Let $(\Omega, \mathcal{F}, \mu, T)$ be an ergodic system with a probability measure. Suppose that $X \colon \Omega \to \mathbb{R}$ is an L^2 random variable such that $\mathbb{E}(X) = 0$ and $\sum_{n=1}^{\infty} \mathcal{P}^n(X)$ is an L^2 random variable. Then $\frac{S_n(X \circ T)}{\sigma\sqrt{n}}$ converges to the standard normal distribution for some $\sigma \in \mathbb{R}$.*

To prove this theorem, we will construct a sequence of martingale differences whose partial sums converge in distribution to $S_n(X \circ T)$. Following [2], we first construct a sequence that meets most of our requirements, but a small technical detail prevents us from calling it a martingale difference sequence.

Construction 5.7. We assume the conditions for Theorem 5.6. Let $g = \sum_{n=1}^{\infty} \mathcal{P}^n(X)$. By hypothesis, g is L^2. Define $f := X + g - g \circ T$. We note that f is $\mathcal{F}$-measurable and L^2, as it is the sum of measurable, L^2 functions.

We consider the sequence $\{f \circ T^n\}$, starting with the $n = 0$ case. By Lemma 5.3,

$$\mathcal{P}(f) = \mathcal{P}(X) + \mathcal{P}(g) - \mathcal{P}(g \circ T), \text{a.s.}$$

Further use of Lemmas 5.3 and 5.4 yields

$$\mathcal{P}(f) = \mathcal{P}(X) + \sum_{n=1}^{\infty} \mathcal{P}^{n+1}(X) - g, \text{a.s.}$$

We now note that

$$\mathcal{P}(X) + \sum_{n=1}^{\infty} \mathcal{P}^{n+1}(X) = g.$$

Thus, $\mathcal{P}(f) = 0$ almost surely. Therefore, by Lemma 5.5, $\mathbb{E}[f \mid T^{-1}(\mathcal{F})] = 0$.

For $n \in \mathbb{N}$, $f \circ T^{n-1}$ is $T^{1-n}(\mathcal{F})$-measurable, and so a simple inductive argument gives us that $\mathbb{E}[f \circ T^{n-1} \mid T^{-n}(\mathcal{F})] = 0$.

We now show that $\{S_n(f \circ T^n)\}$ and $\{S_n(X \circ T^n)\}$ converge in distribution. By our definition,

$$X = f + g \circ T - g.$$

We note that

$$X \circ T^n = f \circ T^n + g \circ T^{n+1} - g \circ T^n.$$

Summation yields

$$S_n(X \circ T) = S_n(f \circ T) + \sum_{i=0}^{n} \left(g \circ T^{i+1} - g \circ T^i \right).$$

Evaluating the telescoping sum yields

$$S_n(X \circ T) = S_n(f \circ T) + g \circ T^{n+1} - g.$$

To normalize, we divide by $\sigma\sqrt{n}$, where σ is the standard deviation of f. (Since T does not affect the expectation of f, σ is also the standard deviation of all $f \circ T^n$.) We then have

$$\frac{S_n(X \circ T)}{\sigma\sqrt{n}} = \frac{S_n(f \circ T)}{\sigma\sqrt{n}} + \frac{g \circ T^{n+1}}{\sigma\sqrt{n}} - \frac{g}{\sigma\sqrt{n}}.$$

Since g is L^2, it is bounded. Thus, as $n \to \infty$, the last two terms approach 0. It follows that

$$\lim_{n\to\infty} \frac{S_n(X \circ T)}{\sigma\sqrt{n}} = \frac{S_n(f \circ T)}{\sigma\sqrt{n}},$$

assuming this limit exists.

For $\{f \circ T^n\}$ to be a martingale difference sequence, the sequence of σ-algebras $\{T^{-n}(\mathcal{F})\}$ must be increasing. However, Construction 5.7 does not guarantee this condition. Since T is measurable, we know that, for all $A \in \mathcal{F}, T^{-1}(A) \in \mathcal{F}$. Thus, an inductive argument readily shows that $T^{-n}(\mathcal{F}) \supseteq T^{-(n+1)}(\mathcal{F})$, so $\{T^{-n}(\mathcal{F})\}$ is *decreasing*.

As mentioned in [2], the solution to this problem is to construct a martingale difference sequence using $S_n(f \circ T^{-1})$. This sequence is only well-defined if T is invertible; that is, T has a well-defined inverse function T^{-1}. Our basic plan is as follows:

(1) Construct a measure-preserving transformation $\tilde{T}$ that is invertible. By the above argument, $S_n(f \circ \tilde{T}^{-1})$ is a martingale.
(2) Show that $S_n(f \circ \tilde{T}^{-1})$ and $S_n(f \circ \tilde{T})$ converge in distribution.
(3) Show that $S_n(f \circ \tilde{T})$ and $S_n(f \circ T^{-1})$ converge in distribution.

We first discuss (2), which can be quickly derived from the properties of measure-preserving systems.

Lemma 5.8. *Let f be a measurable function in a measure-preserving system $(\Omega, \mathcal{F}, \mu, T)$. Then for all $k, n \in \mathbb{N}$,*

$$\sum_{i=0}^{n} f \circ T^i \overset{d}{=} \sum_{i=k}^{n+k} f \circ T^i.$$

Proof. This follows directly from Theorem 4.2 when we note that

$$\sum_{i=k}^{n+k} f \circ T^i = \left(\sum_{i=0}^{n} f \circ T^i \right) \circ T^k.$$

□

Lemma 5.9. *Suppose T is invertible. Then for all $n \in \mathbb{Z}$, f and $f \circ T^n$ are identically distributed.*

Proof. Theorem 4.2 gives the result for non-negative integers. For negative integers, we note that the inverse of T is itself measure-preserving, with nth iterate T^{-n}. Thus, f and $f \circ T^{-n}$ are identically distributed by Theorem 4.2. $\square$

Theorem 5.10. *Suppose T is invertible. Then $S_n(f \circ T)$ and $S_n(f \circ T^{-1})$ converge in distribution, provided that a limit exists.*

Proof. We note that

$$S_n(f \circ T^{-n}) = \sum_{i=-n}^{0} f \circ T^i.$$

By the Corollary 5.9,

$$\sum_{i=0}^{n} f \circ T^i \stackrel{d}{=} \sum_{i=-n}^{0} f \circ T^i.$$

Thus, for all n, $S_n(f \circ T)$ and $S_n(f \circ T^{-1})$ are identically distributed. Therefore, the two sequences must converge to the same limit by definition. $\square$

We now seek to construct an invertible transformation $\tilde{T}$. To ensure that distribution is preserved, we want $\tilde{T}$ to closely mirror T. We accomplish this by creating a "copy" of the entire measure-preserving system where T is essentially the same, only invertible.

Definition 5.11. Let $(\Omega, \mathcal{F}, \mu, T)$ be a measure-preserving system. A measure-preserving system $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu}, \tilde{T})$ is a *natural extension* of $(\Omega, \mathcal{F}, \mu, T)$ if there exists some $X \subseteq \Omega$ and a map $\pi \colon \tilde{\Omega} \to X$ such that the following conditions are satisfied:

i. $\tilde{T}$ is invertible on $\tilde{A}$.
ii. π is $\tilde{\mathcal{F}}$-measurable and surjective.
iii. $\pi \circ \tilde{T} = T \circ \pi$.
iv. $\mu(X)$ and $\tilde{\mu}(\tilde{\Omega})$ both have full measure.
v. For all $A \in \mathcal{F}$, $\tilde{\mu}[\pi^{-1}(A)] = \mu(A)$.

We call π a *factor map.*

The existence of the natural extension is far from obvious, so we provide a construction based on [7, Def. 7] and [8, Def. 1.15].

Theorem 5.12. *Every measure-preserving system has a natural extension.*

Proof. We work in the measure-preserving system $(\Omega, \mathcal{F}, \mu, T)$. For each $x \in \Omega$, we consider all the possible "paths" that lead to x. Each "path" is essentially a possible reverse orbit of T, on which an inverse can be defined. More formally, an element of $\tilde{\Omega}$ is a sequence $\tilde{x} = (x_n) \subset \Omega$ such that for all $n \in \mathbb{N}, x_n = T(x_{n+1})$. Note that each x_n is contained in the preimage $T^{-n}(\{x_1\})$.

We define a factor map $\pi \colon (x_n) \mapsto x_1$. Note that π surjects onto $T^{-1}(\Omega)$, which has full measure since T is measure-preserving.

Define a collection of sets $\tilde{\mathcal{F}} :=: \pi^{-1}(\mathcal{F}) := \{\pi^{-1}(A) \mid A \subseteq T^{-1}(\Omega).\}$ Since π^{-1} preserves set operations, $\tilde{\mathcal{F}}$ is a σ-algebra. By definition, π is $\tilde{\mathcal{F}}$-measurable.

For $\tilde{A} \in \tilde{\mathcal{F}}$, let $A_1 = \pi(\tilde{A})$. By our definition of $\mathcal{F}$, $A_1 = \pi[\pi^{-1}(A)]$ for some $A \subseteq T^{-1}(\Omega)$. Since π is surjective, $A_1 = A$, and since T is measurable, $A_1 \in \mathcal{F}$. For $n \in \mathbb{N}$, let $A_n = T^{1-n}(A_1)$. Since T is measure-preserving, $A_n \in \mathcal{F}$ and $\mu(A_n) = \mu(A_1)$, for all $n \in \mathbb{N}$. Furthermore, for any collection $\{A_n \in \mathcal{F}\}$ where $A_n = T^{1-n}(A_1)$, $\{(x_n) \mid x_i \in A_i \text{ for all } i \in \mathbb{N}\}$ is an element of $\tilde{\mathcal{F}}$.

Define $\tilde{\mu}\colon \tilde{A} \mapsto \mu[\pi(A)]$. Suppose that $\tilde{A}, \tilde{B} \in \tilde{\mathcal{F}}$ are disjoint, and define corresponding collections $\{A_n = T^{1-n}[\pi(\tilde{A})]\}$ and $\{B_n = T^{1-n}[\pi(\tilde{B})]\}$. Since images and preimages preserve set operations, it follows that $\{A_n \cup B_n\}$ corresponds in the same way to $\tilde{A} \cup \tilde{B}$. Thus, by our definition of $\tilde{\mu}$,

$$\tilde{\mu}(\tilde{A} \cup \tilde{B}) = \mu(A_1 \cup B_1).$$

There must be some k for which A_k and B_k are disjoint; otherwise, $\tilde{A}$ and $\tilde{B}$ would not be disjoint. As noted above, $\mu(A_k \cup B_k) = \mu(A_1 \cup B_1)$. Thus,

$$\tilde{\mu}(\tilde{A} \cup \tilde{B}) = \mu(A_k \cup B_k).$$

Since A_k and B_k are disjoint, by the additivity of μ,

$$\tilde{\mu}(\tilde{A} \cup \tilde{B}) = \mu(A_k) + \mu(B_k).$$

Since $\mu(A_1) = \mu(A_k)$ and $\mu(B_1) = \mu(B_k)$, it follows that

$$\tilde{\mu}(\tilde{A} \cup \tilde{B}) = \mu(A_1) + \mu(B_1) = \tilde{\mu}(\tilde{A}) + \tilde{\mu}(\tilde{B}).$$

A similar argument gives countable additivity, so we conclude that $\tilde{\mu}$ is a measure on $\tilde{\mathcal{F}}$.

For $\tilde{x} \in \tilde{\Omega}$, we define $\tilde{T}$ as the image of T, so that $\tilde{T}(\tilde{x}) = (T(x_n))$. In other words, $\tilde{T}$ "shifts" each x_n over by applying T. We define the inverse of $\tilde{T}$ by "shifting" in the opposite direction; that is, $\tilde{T}^{-1}(\tilde{x}) = (x_{n+1})$. We note that $\tilde{T}^{-1} \circ \tilde{T}(\tilde{x}) = \tilde{T} \circ \tilde{T}^{-1}(\tilde{x}) = \tilde{x}$, so our inverse is well-defined.

We show that $\tilde{T}$ is $\tilde{\mathcal{F}}$-measurable and measure-preserving. Let $\tilde{A} \in \tilde{\mathcal{F}}$. As discussed before,

$$\tilde{T}^{-1}(\tilde{A}) = \{(x_n) \mid T(x_n) \in T^{1-n}[\pi(\tilde{A})], \text{ for all } n \in \mathbb{N}\}.$$

It follows that

$$\tilde{T}^{-1}(\tilde{A}) = \{(x_n) \mid x_n \in T^{-n}[\pi(\tilde{A})], \text{ for all } n \in \mathbb{N}\}.$$

As noted before, since each $\{T^{-n}[\pi(\tilde{A})]\}$, is in $\mathcal{F}$, $\tilde{T}^{-1}(\tilde{A}) \in \tilde{\mathcal{F}}$. We further note that, since T is measure-preserving,

$$\mu[T^{-1}(A_1)] = \mu(A_1),$$

and so $\tilde{\mu} \circ \tilde{T}^{-1} = \tilde{\mu}$. Therefore, $\tilde{T}$ is measure-preserving.

The remaining properties are quick to check. We note that, for $\tilde{x} = (x_n) \in \tilde{\Omega}$,

$$\pi[\tilde{T}(\tilde{x})] = \pi[(T(x_n))] = T(x_1).$$

Thus, $\pi \circ \tilde{T} = T \circ \pi$. Now let $A \in \mathcal{F}$. By definition,

$$\pi^{-1}(A) = \{(x_n) \in \tilde{\Omega} \mid x_1 \in A\}.$$

Since T is measurable, $\{T^{1-n}(A)\}$ is a sequence of measurable sets, and for each $(x_n) \in \pi^{-1}(A), x_n \in T^{1-n}(A)$. Thus, $\pi^{-1}(A) \in \tilde{\mathcal{F}}$, and so π is $\tilde{\mathcal{F}}$-measurable. By our definition, $\tilde{\mu}[\pi^{-1}(A)] = \mu[\pi[\pi^{-1}(A)]]$, and since π is surjective, $\tilde{\mu}[\pi^{-1}(A)] = \mu(A)$. □

To apply the martingale central limit theorem, we need to know that our natural extension is ergodic.

Theorem 5.13. *The natural extension of an ergodic system is ergodic.*

Proof. We work in the ergodic system $(\Omega, \mathcal{F}, \mu, T)$, with natural extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu}, \tilde{T})$ and factor map π. Let $\tilde{A}$ be $\tilde{T}$-invariant. By definition, $\tilde{T}^{-1}(\tilde{A}) = \tilde{A}$. It follows that $\pi(A) = \pi[\tilde{T}^{-1}(A)] = T^{-1}[\pi(\tilde{A})]$. Thus, $A = \pi(\tilde{A})$ is T-invariant, and since T is ergodic, A is either null or has full measure. If A is null, then by the definition of the natural extension, $\tilde{\mu}[\pi^{-1}(A)] = \mu(A)$, so $\pi^{-1}(A)$ is $\tilde{\mu}$-null. Since $\tilde{A} \subset \pi^{-1}(A)$, $\tilde{A}$ is also $\tilde{\mu}$-null. If A has full measure, then $\Omega \setminus A$ is null, and by the above argument and the measurability of π, $\tilde{\Omega} \setminus \tilde{A}$ is also null. Thus, $\tilde{A}$ has full measure. We conclude that $\tilde{T}$ is ergodic. □

We need to know that distributions, expectations, and conditional expectations in the natural extension have equivalent counterparts in the base system.

Theorem 5.14. *Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu}, \tilde{T})$ be a natural extension of $(\Omega, \mathcal{F}, \mu, T)$ with factor map π. Then for all $\mathcal{F}-$measurable functions f, $f \circ \pi$ is $\tilde{\mathcal{F}}-$measurable and identically distributed with f.*

Proof. Let $A \in \mathcal{F}$. We note that $\pi^{-1}[f^{-1}(A)] \in \tilde{\mathcal{F}}$ since f is $\mathcal{F}$-measurable and π is $\tilde{\mathcal{F}}$-measurable. Thus, $f \circ \pi$ is $\tilde{\mathcal{F}}$-measurable.

By the definition of the natural extension, $\tilde{\mu}[\pi^{-1}(f^{-1}(A))] = \mu[f^{-1}(A)]$. Thus, f and $f \circ \pi$ are identically distributed by definition. □

Theorem 5.15. *Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu}, \tilde{T})$ be a natural extension of $(\Omega, \mathcal{F}, \mu, T)$ with factor map π. Then for all $\mathcal{F}-$measurable functions f, $\int f d\mu = \int f \circ \pi d\tilde{\mu}$.*

Proof. Suppose first that f is simple. By definition, for some disjoint $A_1, \dots A_k \in \mathcal{F}$ and coefficients $a_1, \dots a_k$, $f = \sum_{i=1}^k a_i \mathbf{1}_{A_i}$. It follows that $f \circ \pi = \sum_{i=1}^k a_i(\mathbf{1}_{A_i} \circ \pi)$. We note that each $\mathbf{1}_{A_i} \circ \pi = \mathbf{1}_{\pi^{-1}(A_i)}$, and that $\pi^{-1}(A_1), \dots \pi^{-1}(A_k)$ are disjoint sets in $\tilde{\mathcal{F}}$. Thus, $f \circ \pi$ is a simple function with integral

$$\int f \circ \pi d\tilde{\mu} = \sum_{i=1}^k a_i \tilde{\mu}[\pi^{-1}(A_i)].$$

By the properties of the natural extension, $\tilde{\mu} \circ \pi^{-1} = \mu$. Thus,

$$\int f \circ \pi d\tilde{\mu} = \sum_{i=1}^k a_i \mu(A_i) = \int f d\mu.$$

If f is not simple, then we construct an increasing sequence of simple functions f_n such that $\lim_{n\to\infty} f_n = f$. It follows that $\lim_{n\to\infty} f_n \circ \pi = f \circ \pi$. By the above argument, $\int f_n d\mu = \int f_n \circ \pi d\tilde{\mu}$ for all n. Thus, the Monotone Convergence Theorem implies that $\int f d\mu = \int f \circ \pi d\tilde{\mu}$. □

Remark 5.16. *An immediate corollary of Theorem 5.15 is that for all $A \in \mathcal{F}$,*

$$\int_A f d\mu = \int_{\pi^{-1}(A)} f \circ \pi d\tilde{\mu}.$$

It follows that for all σ-algebras $\mathcal{G} \subseteq \mathcal{F}, \mathbb{E}[f \mid \mathcal{G}] = \mathbb{E}[f \circ \pi \mid \pi^{-1}(\mathcal{G})]$.

We can now present a complete proof of Theorem 5.6. We restate the theorem for convenience:

Let $(\Omega, \mathcal{F}, \mu, T)$ be an ergodic system with a probability measure. Suppose that $X\colon \Omega \to \mathbb{R}$ is an L^2 random variable such that $\mathbb{E}(X) = 0$ and $\sum_{n=1}^\infty \mathcal{P}^n(X)$ is an L^2 random variable. Then $\frac{S_n(X \circ T)}{\sigma\sqrt{n}}$ converges to the standard normal distribution.

Proof. We recall that Construction 5.7 gives us an L^2 function f such that, for all $n \in \mathbb{N}$, $f \circ T^n$ is $T^{-n}(\mathcal{F})$-measurable, and $\mathbb{E}[f \circ T^{n-1} \mid T^{-n}(\mathcal{F})] = 0$. Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu}, \tilde{T})$ be a natural extension of $(\Omega, \mathcal{F}, \mu, T)$ with factor map π. By Remark 5.16,

$$\mathbb{E}[f \circ T^{n-1} \circ \pi \mid \pi^{-1}(T^{-n}(\mathcal{F}))] = \mathbb{E}[f \circ T^{n-1} \mid T^{-n}(\mathcal{F})] = 0.$$

By the properties of preimages, $\pi^{-1} \circ T^{-1} = (T \circ \pi)^{-1}$. Since $T \circ \pi = \pi \circ \tilde{T}$, we have that

$$\mathbb{E}[f \circ T^{n-1} \circ \pi \mid \pi^{-1}(T^{-n}(\mathcal{F}))] = \mathbb{E}[f \circ \pi \circ \tilde{T}^{n-1} \mid \tilde{T}^{-n}(\pi^{-1}(\mathcal{F}))] = 0.$$

Since π is $\tilde{\mathcal{F}}$-measurable, $\pi^{-1}(\mathcal{F}) \subseteq \tilde{\mathcal{F}}$. We recall from the construction in Theorem 5.12 $= \tilde{\mathcal{F}} = \pi^{-1}(\mathcal{F})$, and so

$$\mathbb{E}[f \circ \pi \circ \tilde{T}^{n-1} \mid \tilde{T}^{-n}(\mathcal{F})] = \mathbb{E}[f \circ \pi \circ \tilde{T}^{n-1} \mid \tilde{T}^{-n}(\pi^{-1}(\mathcal{F}))] = 0.$$

Since $\tilde{T}$ is invertible, the inductive argument in Construction 5.7 that gave the conditional expectation now applies for all $n \in \mathbb{Z}$. We conclude that, for all $n \in \mathbb{N}, f \circ \pi \circ \tilde{T}^{-n}$ is $T^n(\mathcal{F})$-measurable, and that

$$\mathbb{E}[f \circ \pi \circ \tilde{T}^{-(n-1)} \mid \tilde{T}^n(\mathcal{F})] = 0.$$

Since $\{T^n(\mathcal{F})\}$ is an increasing sequence of σ-algebras, $f \circ \pi \circ \tilde{T}^{-n}$ is a martingale difference sequence with respect to $\{T^n(\mathcal{F})\}$. By Theorem 5.10, $\{S_n(f \circ \pi \circ \tilde{T}^{-n})\}$ converges in distribution to $\{S_n(f \circ \pi \circ \tilde{T}^n)\}$. By the definition of the natural extension, $\{S_n(f \circ \pi \circ \tilde{T}^n)\} = \{S_n(f \circ T^n \circ \pi)\}$, and by Theorem 5.14, $\{S_n(f \circ T^n \circ \pi)\}$ converges in distribution to $\{S_n(f \circ T^n)\}$. □

Acknowledgements

I would like to thank my mentors, Jonathan DeWitt and Isabella Scott, for their invaluable advice and guidance. I would also like to thank Peter May and the faculty of the REU program for making this paper possible.

References

[1] Mikhail Gordin. "The central limit theorem for stationary processes." Soviet Mathematics Doklady, vol. 10 (1969), no. 5.

[2] Vaughn Climenhaga. "Central Limit Theorem for dynamical systems using martingales." University of Houston, 2013.
`https://www.math.uh.edu/~climenha/blog-posts/martingale-clt.pdf`

[3] David Williams. *Probability with Martingales.* Cambridge University Press, 1991.

[4] Rick Durrett. *Probability: Theory and Examples.* Cambridge University Press, 2013.

[5] Gordan Zitkovic. "Lecture 3: the Lebesgue Integral." University of Texas at Austin, 2013.
`https://www.ma.utexas.edu/users/gordanz/notes/lebesgue_integration.pdf`

[6] Greg Lawler. "Notes on Probability." University of Chicago, 2016.
`http://www.math.uchicago.edu/~lawler/probnotes.pdf`

[7] Henk Bruin. "Notes on Ergodic Theory. " University of Vienna, 2014.
`http://www.mat.univie.ac.at/~bruin/ET2.pdf`

[8] Omri Sarig. "Lecture Notes on Ergodic Theory." Pennsylvania State University, 2008.
`http://www.weizmann.ac.il/math/sarigo/sites/math.sarigo/files/uploads/ergodicnotes.pdf`

[9] Omri Sarig. "Introduction to the transfer operator method." Weizmann Institute of Science, 2012.
`http://www.weizmann.ac.il/math/sarigo/sites/math.sarigo/files/uploads/transferoperatorcourse.pdf`