University of Bristol

School of Mathematics

Ergodicity of Stochastic Processes and the Markov Chain Central Limit Theorem

A Project of 30 Credit Points at Level 7

For the Degree of MSci Mathematics

Author:
Benjamin A. Robinson

Supervisor:
Dr Martón Balázs

May 1, 2015

Acknowledgement of Sources

For all ideas taken from other sources (books, articles, internet), the source of the ideas is mentioned in the main text and fully referenced at the end of the report.
All material which is quoted essentially word-for-word from other sources is given in quotation marks and referenced.
Pictures and diagrams copied from the internet or other sources are labelled with a reference to the web page, book, article etc.

Signed:

Dated:

Abstract

In this report we investigate how the well-known central limit theorem for i.i.d. random variables can be extended to Markov chains. We present some of the theory on ergodic measures and ergodic stochastic processes, including the ergodic theorems, before applying this theory to prove a central limit theorem for square-integrable ergodic martingale differences and for certain ergodic Markov chains. We also give an alternative proof of a central limit theorem for stationary, irreducible, aperiodic Markov chains on a finite state space. Finally, we outline some of the diverse applications of the Markov chain central limit theorem and discuss extensions of this work.

Contents

Introduction

A central limit theorem gives a scaling limit for the sum of a sequence of random variables. This controls the fluctuations of the sequence in the long run. It is well known that there is a central limit theorem for sequences of i.i.d. random variables; the theorem is given, for example, in Chapter III, Section 3 of [11]. This is a very useful result to have, and it is a natural question to ask whether this can be generalised to sequences of random variables which are not i.i.d.. In this report we show that we have a central limit theorem for functions of discrete-time Markov chains under certain conditions. This central limit theorem has many applications, some of which we discuss in later chapters. We now outline the structure of this report.

In Chapter 1, we collect preliminary material on measure theory and probability from various sources, which we will refer to later in the report. This includes definitions of modes of convergence of random variables, which we will need to be familiar with in order to understand in what sense the limit theorems in later chapters hold. We then recall elementary properties of Markov chains, which will be useful for us to have in mind when we consider limit theorems for Markov chains. We give rigorous definitions of conditional expectation and martingales, as well as some results on these, which we will use in the proofs of several theorems in later chapters. Here, we also define martingale differences, which are processes related to martingales, as we will study the limiting behaviour of these processes in Chapter 3, in order to prove a central limit theorem for Markov chains.

In Chapter 2, we prove the ergodic theorems from [15], define what it means for a measure or a stochastic process to be ergodic, and prove several results on ergodicity. We will see in Chapter 3 that the condition that a Markov chain is ergodic allows us, under a few additional conditions, to prove a central limit theorem for functions of that Markov chain.

The main results of this report are contained in Chapter 3. Here, we prove a central limit theorem for certain ergodic Markov chains in two ways. First, we prove a central limit theorem for square-integrable ergodic martingale differences and then, following [15], we deduce from this that we have a central limit theorem for functions of ergodic Markov chains, under some conditions. We then restrict ourselves to Markov chains which take values in a finite state space. In this setting, we use a different method, as in [13], to prove a central limit theorem for functions of ergodic Markov chains, where we have to impose fewer conditions than in the case of a general state space. In both cases we derive formulae for the variance of the limiting distribution.

In Chapter 4, we discuss some simple applications of the Markov chain central limit theorem which is proved in Chapter 3. We consider a simple random walk on a torus, started in its stationary distribution, and we show that there is a central limit theorem for the amount of time spent in the initial state. We treat the symmetric and asymmetric cases separately. We also look at an example of a random walk on the non-negative integers from [15], which could model the length of a queue. Here we show, by a similar method, that there is a central limit theorem for the amount of time during which the queue is empty.

We conclude the report by discussing extensions of the theory which we have presented and further applications of this.

Chapter 1

Preliminary Measure Theory and Probability

Here we present some theory on convergence of random variables, on Markov processes and on martingales that we will need in order to approach the topics which we discuss in the rest of this report. We assume knowledge of some basic definitions and properties from measure theory, including product sigma-fields and product measures, as could be found for example in Bartle's text [1]. We also assume knowledge of introductory probability theory, such as the material from Ross's book [10].
In this chapter we state results without proof that we will later apply in proving our main results.

1.1 Convergence of Random Variables

The main results of this report concern the convergence of sequences of random variables. Therefore, we need to recall some definitions and results on convergence which not all readers will be familiar with. There are four types of convergence which we will work with in this report: namely, they are almost sure convergence, weak convergence (convergence in distribution), convergence in $\mathcal{L}^p$ and convergence in measure. We will only consider real-valued random variables in this report. We start by recalling the definitions of these modes of convergence, as well as some useful facts, taken from the books of Bartle [1], Grimmett and Stirzaker [6], Ross [10], Shiryaev [11], and Varadhan [15].
Throughout this section, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1.1.1 Almost sure convergence

Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F})$. We define what it means for X_n to converge to X almost surely, where X is another random variable on $(\Omega, \mathcal{F})$, as in Section 7.2 of [6].

Definition 1.1.1 (Almost sure convergence)**.** We say that X_n converges to X almost surely if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

We write this as $X_n \to X$ a.s..

1.1.2 $\mathcal{L}^p$ convergence

For our next mode of convergence, we consider function spaces called $\mathcal{L}^p$ spaces. We will give a definition of these spaces as normed spaces and define what it means for random variables to converge in $\mathcal{L}^p$. $\mathcal{L}^p$ spaces are discussed in Chapter 6 of [1] in the context of general measure spaces. We reformulate the material here in terms of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
Recall that a random variable X is defined to be an $\mathcal{F}$-measurable function and that the expectation of X is defined as the integral of X with respect to $\mathbb{P}$, when this integral exists; i.e.

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

We start by defining the spaces $\mathcal{L}^p$, for $p \in [1, \infty)$ and then for $p = \infty$.

Definition 1.1.2 ($\mathcal{L}^p$ spaces)**.** Let $p \in [1, \infty)$. Then $\mathcal{L}^p\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ is the space of equivalence classes

$$[X] := \{Y : X = Y \ \mathbb{P}\text{-almost everywhere}\}$$

of $\mathcal{F}$-measurable functions X such that

$$\mathbb{E}(|X|^p) < \infty.$$

Definition 1.1.3 (The space $\mathcal{L}^\infty$)**.** We define $\mathcal{L}^\infty\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ to be the space of equivalence classes

$$[X] := \{Y : X = Y \ \mathbb{P}\text{-almost everywhere}\}$$

of $\mathcal{F}$-measurable functions X which are bounded $\mathbb{P}$-almost everywhere;
i.e. $[X] \in \mathcal{L}^\infty$ if and only if $\exists M \geq 0$ such that $|X| \leq M$ $\mathbb{P}$-almost everywhere.

Remark 1.1.1.

1. When it is clear which space, σ-field or probability measure we are working with, we will drop the arguments from $\mathcal{L}^p$ so that we may write

$$\mathcal{L}^p \equiv \mathcal{L}^p\left(\Omega, \mathcal{F}, \mathbb{P}\right) \equiv \mathcal{L}^p(\Omega) \equiv \mathcal{L}^p(\mathbb{P}).$$

2. Although technically the space $\mathcal{L}^p$ is a space of equivalence classes of functions, we will in practice say that a function X is an element of $\mathcal{L}^p$ if X is $\mathcal{F}$-measurable and $\mathbb{E}(|X|^p) < \infty$. This is a common convention, as remarked after Theorem 6.7 of [1].

For $1 \leq p \leq \infty$, we can define a norm on the space $\mathcal{L}^p$ as follows.

Definition 1.1.4 ($\mathcal{L}^p$ norm)**.** Suppose that $p \in [1, \infty)$. Then we define the $\mathcal{L}^p$ norm by

$$\|X\|_p := [\mathbb{E}(|X|^p)]^{\frac{1}{p}}, \tag{1.1.1}$$

for any $X \in \mathcal{L}^p$.
We define the $\mathcal{L}^\infty$ norm by

$$\|X\|_\infty := \inf\{M \geq 0 : |X| \leq M \text{ } \mathbb{P}\text{-almost everywhere}\}, \tag{1.1.2}$$

for any $X \in \mathcal{L}^\infty$.

These quantities are well-defined as norms on the $\mathcal{L}^p$ spaces, as we assert in the following theorem.

Theorem 1.1.1. *For $p \in [1, \infty)$, the space $\mathcal{L}^p$ equipped with the norm $\|\cdot\|_p$, as defined in* (1.1.1)*, is a normed space.*
Also, the space $\mathcal{L}^\infty$ equipped with the norm $\|\cdot\|_\infty$, as defined in (1.1.2)*, is a normed space.*

The proof that $\mathcal{L}^p$ is a normed space for $p \in [1, \infty)$ relies on Minkowski's inequality, which we will now state.

Lemma 1.1.1 (Minkowski's inequality)**.** *Suppose that $p \in [1, \infty)$ and $X, Y \in \mathcal{L}^p$. Then $X + Y \in \mathcal{L}^p$ and we have the following inequality:*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p. \tag{1.1.3}$$

We are now ready to define $\mathcal{L}^p$ convergence, as in Chapter 7 of [1].

Definition 1.1.5 ($\mathcal{L}^p$ convergence)**.** Let $p \in [1, \infty]$ and let $\{X_n\}$ be a sequence of random variables such that $X_i \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ for all $i \in \mathbb{N}$. Also let $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. Then we say that X_n converges to X in $\mathcal{L}^p$ if

$$\|X_n - X\|_p \xrightarrow{n \to \infty} 0.$$

We write $X_n \xrightarrow{\mathcal{L}^p} X$.

1.1.3 Weak convergence

Let us now define weak convergence of the sequence $\{X_n\}$ as in Section 7.2 of [6]. Let $X, X_1, X_2, \dots$ be random variables on $(\Omega, \mathcal{F})$.

Definition 1.1.6 (Weak convergence)**.** Let F_n be the distribution function of X_n, for each $n \in \mathbb{N}$, and F be the distribution function of X.
We say that X_n converges **weakly** to X, or X_n converges to X **in distribution** if, for every x at which $F(x)$ is continuous,

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

We will denote this as $X_n \xrightarrow{\mathcal{D}} X$.

We will now see two theorems which give us ways to prove weak convergence. We first state a necessary and sufficient condition for weak convergence, by defining the characteristic function of a random variable, as in Section 5.7 of [6].

Definition 1.1.7. Let X be a random variable. Then the characteristic function of X is the function $\phi : \mathbb{R} \to \mathbb{C}$, defined by

$$\phi(t) := \mathbb{E}\left(e^{itX}\right).$$

This function is well-defined for any random variable X.

The statement of the continuity lemma is taken from Chapter III, Section 3 of [11].

Theorem 1.1.2 (Continuity lemma)**.** *Let $\{X_n\}$ be a sequence of random variables with characteristic functions ϕ_n. Then we have the following implications.*

1. *If there exists a random variable X with characteristic function ϕ such that $X_n \xrightarrow{\mathcal{D}} X$, then $\phi_n(t) \xrightarrow{n\to\infty} \phi(t)$ for all $t \in \mathbb{R}$.*
2. *If $\phi(t) := \lim_{n\to\infty} \phi_n(t)$ exists for all $t \in \mathbb{R}$ and $\phi(t)$ is continuous at $t = 0$, then $\exists$ a random variable X with characteristic function ϕ such that $X_n \xrightarrow{\mathcal{D}} X$.*

The continuity lemma is very important in probability theory. For example, it is used by Shiryaev in [11] to prove the central limit theorem for sequences of i.i.d. random variables. We will use it to prove a central limit theorem for martingale differences in Section 3.1.
We now state a theorem which gives another criterion for weak convergence for certain sequences of random variables. We will first recall the definition of the moment generating function, from Chapter 7, Section 7 of [10].

Definition 1.1.8. For a random variable X, the moment generating function of X is the function $M : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, defined by

$$M(t) = \mathbb{E}\left(e^{tX}\right).$$

Note that this function may be infinite.

We now state the theorem, which is found in Chapter 8, Section 3 of [10].

Theorem 1.1.3. *Let $\{X_n\}$ be a sequence of random variables with moment generating functions M_n and let X be a random variable with moment generating function M. Suppose that $M_n \xrightarrow{n\to\infty} M$ for all $t \in \mathbb{R}$.*
Then $X_n \xrightarrow{\mathcal{D}} X$.

This theorem is less commonly used than the continuity lemma for characteristic functions, but we will use this form in our second proof of a Markov chain central limit theorem in Section 3.3, following [13].

1.1.4 Convergence in measure

The final mode of convergence which we define is convergence in measure, as in Section 1.3 of [15].

Definition 1.1.9. Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F})$ and X another random variable on $(\Omega, \mathcal{F})$. We say that X_n converges to X **in measure** or **in probability** if, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\right) \xrightarrow{n \to \infty} 0.$$

We will see how this mode of convergence is of use to us in Section 1.1.5. It will also be useful to note the following implication.

Theorem 1.1.4. *Let $p \in [1, \infty]$ and let $\{X_n\}$ be a sequence of random variables with $X_n \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ for each $n \in \mathbb{N}$. Suppose that $\exists X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \xrightarrow{\mathcal{L}^p} X$ or $X_n \to X$ a.s..*
Then X_n converges to X in measure.

The above theorem comes from Chapter 7 of [1], where Bartle discusses all of the relations between the modes of convergence which we have defined here. We should note here that the above theorem is false if we are working in a general measure space. For almost sure convergence to imply convergence in measure, we need the condition that the total measure of the space is finite. Of course, this is not a problem for us here, as a probability space has total measure 1.

1.1.5 The bounded convergence theorem

Let $X, X_1, X_2, \ldots$ be random variables on $(\Omega, \mathcal{F})$ and suppose that $X_n \to X$ in some sense. We are interested in whether we can interchange the order in which we take the limit and the expectation of these random variables.
There are a few results from measure theory which give sufficient conditions for this to be allowed (see, for example, [1] and [11]), but we are only going to use two of these in our work. These theorems both concern random variables which are bounded.
The first theorem which we state is a special case of Lebesgue's dominated convergence theorem, which can be found in Chapter II, Section 6 of [11].

Theorem 1.1.5 (Bounded convergence theorem)**.** *Suppose that there exists a constant C such that $|X_n| \leq C$, for all $n \in \mathbb{N}$, and suppose that $X_n \to X$ almost surely.*
Then $X, X_1, X_2, \ldots \in \mathcal{L}^1$,

$$\mathbb{E}(X_n) \xrightarrow{n \to \infty} \mathbb{E}(X)$$

and

$$\mathbb{E}(|X_n - X|) \xrightarrow{n \to \infty} 0.$$

We can relax the condition that $\{X_n\}$ converges almost surely to the weaker condition of convergence in measure and find that Lebesgue's dominated convergence theorem still holds, as shown in Section 1.3 of [15]. Hence we have a bounded convergence theorem for a sequence which converges in measure, as follows.

Theorem 1.1.6. *Suppose that there exists a constant C such that $|X_n| \leq C$ for all $n \in \mathbb{N}$ and suppose that $X_n \to X$ in measure.*
Then $X, X_1, X_2, \ldots \in \mathcal{L}^1$,

$$\mathbb{E}(X_n) \xrightarrow{n\to\infty} \mathbb{E}(X)$$

and

$$\mathbb{E}(|X_n - X|) \xrightarrow{n\to\infty} 0.$$

1.1.6 Useful inequalities

Proving convergence will involve approximating certain quantities. The inequalities which we list here are very commonly used in probability and will all be useful for us at some stage in this report.
The first inequality which we state is due to Chebyshev and is found in many textbooks, including in Chapter II, Section 6 of [11].

Theorem 1.1.7 (Chebyshev's inequality)**.** *Let X be a random variable and let $k \in (0, \infty)$. Define $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$. Then*

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}. \tag{1.1.4}$$

We now state an equally common inequality, known as the Cauchy-Schwarz inequality, from [11].

Theorem 1.1.8 (Cauchy-Schwarz inequality)**.** *Let $X, Y \in \mathcal{L}^2$. Then $XY \in \mathcal{L}^1$ and*

$$[\mathbb{E}(|XY|)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2). \tag{1.1.5}$$

Finally, we state and prove an immediate corollary of the Cauchy-Schwarz inequality.

Corollary 1.1.1. *Let $X, Y \in \mathcal{L}^2$. Then*

$$\mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}. \tag{1.1.6}$$

Proof. Set $\mu := \mathbb{E}(X)$ and $\nu := \mathbb{E}(Y)$. Then

$$\begin{aligned}
\mathrm{Cov}(X, Y)^2 &= [\mathbb{E}\,[(X - \mu)(Y - \nu)]]^2 \\
&\leq \mathbb{E}([X - \mu]^2)\mathbb{E}([Y - \nu]^2) \\
&= \mathrm{Var}(X)\,\mathrm{Var}(Y).
\end{aligned}$$

□

1.2 Markov Processes

Since a large proportion of this report will be spent on proving central limit theorems for Markov chains, we now take some time to review some basic Markov chain theory. In this section we recall the definition of a Markov chain, as well as some related definitions and properties, adapted from the material in Chapter 6 of [6].

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We recall that a stochastic process is defined to be a sequence of random variables on $(\Omega, \mathcal{F})$. Let $(S, \mathcal{B})$ be a measurable space in which the random variables can take values and suppose that S is countable. We call S the **state space** of the process.

Definition 1.2.1. A stochastic process $\{X_n\}$ is called a **Markov process** or a **Markov chain** if it satisfies the Markov condition:

$$\mathbb{P}(X_n = s | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s | X_{n-1} = x_{n-1}),$$

for all $n \in \mathbb{N}$ and $s, x_1, x_2, \ldots, x_{n-1} \in S$.

We are only interested in Markov chains such that

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) =: \pi_{i,j}, \tag{1.2.1}$$

for all $n = 0, 1, 2, \ldots$ and $i, j \in S$.

Definition 1.2.2 (Transition probabilities)**.** The **transition matrix** of a Markov chain which satisfies (1.2.1) is defined to be the matrix Π with entries $\pi_{i,j}$.

For any $m, n \in \mathbb{N}$, we define the **n-step transition probabilities** to be $\pi_{i,j}(m, m+n) = \mathbb{P}(X_{m+n} = j | X_m = i)$.

These probabilities are independent of m, so we write $\pi_{i,j}(m, m+n) =: \pi_{i,j}(n)$. We define $\Pi(n)$ to be the matrix with entries $\pi_{i,j}(n)$.

Lemma 1.2.1. *For any $n \in \mathbb{N}$, we have $\Pi(n) = \Pi^n$.*

We next define a stationary distribution of a Markov chain. This will be a very important definition for us later in the report, particularly when we study ergodicity of Markov chains in Section 2.5 and when we apply our theory to examples in Chapter 4.

Definition 1.2.3. A probability measure μ on $(S, \mathcal{B})$ is said to be a stationary distribution for a Markov chain with transition matrix Π if $\mu = \mu\Pi$.

The following two properties of a Markov chain are common and the study of Markov chains which have these properties is simpler than that of general Markov chains.

Definition 1.2.4. We say that a Markov chain with transition matrix Π is **irreducible** if, for each $i, j \in S$, $\exists m \in \mathbb{N}$ such that $\pi_{i,j}(m) > 0$.

The **period** $d(i)$ of a state $i \in S$ is defined to be the highest common factor of the set $\{n : \pi_{i,i} > 0\}$.
In an irreducible chain, $d(i) = d(j)$ for all $i, j \in S$.

Definition 1.2.5. We say that an irreducible chain is **aperiodic** if $d(i) = 1$ for all $i \in S$.

Finally, we state a result about Markov chains on finite state spaces, which we will use in Section 2.5 to say something about ergodicity of these chains.

Theorem 1.2.1. *Let S be a finite set and suppose that $\{X_n\}$ is an irreducible aperiodic Markov chain with state space S. Then the Markov chain $\{X_n\}$ has a unique stationary distribution.*

1.3 Martingales

Martingales are another essential tool in our work. For example, we prove a central limit theorem for Markov chains in Section 3.2 via the central limit theorem for martingale differences (Theorem 3.1.1), which are processes related to martingales and are defined below. Also, we will repeatedly use properties of martingales in our second proof of a central limit theorem for Markov chains in Section 3.3. We are now going to give a formal definition of a martingale and some basic properties which we will need, without proof, from Williams' book [16].
To define martingales, we need a rigorous definition of conditional expectation, as given in Chapter 9 of [16].
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X \in \mathcal{L}^1$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field.

Definition 1.3.1. [Conditional expectation] A version of the conditional expectation of X given $\mathcal{G}$ is defined to be a random variable Y such that

1. Y is $\mathcal{G}$-measurable,
2. $Y \in \mathcal{L}^1$,
3. For every $G \in \mathcal{G}$,
$$\mathbb{E}(\mathbb{1}_G X) = \mathbb{E}(\mathbb{1}_G Y).$$

It was proved by Kolmogorov that such a random variable exists and is almost surely unique.

Theorem 1.3.1. *There exists a random variable Y which satisfies the three conditions in Definition 1.3.1.*
Moreover, if Y_1, Y_2 both satisfy these conditions, then $Y_1 = Y_2$ almost surely.

Because of the above theorem, we refer to a random variable Y as in Defintion 1.3.1 as **the conditional expectation** of X given $\mathcal{G}$.
We write $Y = \mathbb{E}(X|\mathcal{G})$.
The conditional expectation has several nice properties which will be useful for

us. We take these from Chapter 9 of [16].
Let $X \in \mathcal{L}^1$ and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. Then we have the following properties.

Theorem 1.3.2 (Properties of conditional expectation)**.**

1. *$\mathbb{E}\left(\mathbb{E}[X|\mathcal{G}]\right) = \mathbb{E}(X)$.*
2. *If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X \qquad a.s.$.*
3. *(Tower property) If $\mathcal{H} \subseteq \mathcal{G}$, then*
$$\mathbb{E}\left(\mathbb{E}[X|\mathcal{G}]|\mathcal{H}\right) = \mathbb{E}(X|\mathcal{H}) \qquad a.s.. \tag{1.3.1}$$
4. *If Z is a bounded $\mathcal{G}$-measurable random variable, then*
$$\mathbb{E}(ZX|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G}) \qquad a.s..$$
5. *If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then*
$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G}) \qquad a.s..$$

We will now give the definition of what it means for a stochastic process to be a martingale with respect to some filtration, as in Chapter 10 of [16].

Definition 1.3.2 (Filtration)**.** A filtration is a family of σ-algebras $\{\mathcal{F}_n : n = 0, 1, \dots\}$ such that
$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}.$$

Definition 1.3.3. We say that a stochastic process $\{X_n\}$ is adapted to the filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable for all $n = 0, 1, 2, \dots$.

Definition 1.3.4. [Martingale] A stochastic process $\{X_n\}$ is a **martingale** with respect to $(\mathcal{F}_n)$ if

1. $\{X_n\}$ is adapted to $(\mathcal{F}_n)$,
2. $X_n \in \mathcal{L}^1$ for each $n = 0, 1, 2, \dots$, and
3. $\mathbb{E}(X_n|\mathcal{F}_{n-1}) = X_{n-1}$ a.s., for all $n \geq 1$.

When it is clear which filtration we are working with, we will just say that $\{X_n\}$ is a martingale.
An easy consequence of the definition of a martingale is the following lemma.

Lemma 1.3.1. *Let $\{X_n\}$ be a **martingale** with respect to $(\mathcal{F}_n)$. Then*
$$\mathbb{E}(X_n) = \mathbb{E}(X_0)$$
for all $n \in \mathbb{N}$.

We now see that that the above result can, under some conditions, be extended to the case where n is a random time.

Definition 1.3.5 (Stopping time)**.** We say that a non-negative integer-valued random variable T is a stopping time with respect to a filtration $(\mathcal{F}_n)$ if $\{T \leq n\} \in \mathcal{F}_n$ for every $n = 0, 1, 2, \ldots$.

The following theorem is a special case of the optional stopping theorem which is given in Chapter 10 of [16].

Theorem 1.3.3 (Optional stopping theorem)**.** *Let $(\mathcal{F}_n)$ be a filtration, T a stopping time and $\{X_n\}$ a martingale. Then we have that $X_T \in \mathcal{L}^1$ and $\mathbb{E}(X_T) = \mathbb{E}(X_0)$, if one of the following conditions holds.*

1. *T is bounded;*
2. *$\{X_n\}$ is bounded and T is almost surely finite;*
3. *$\mathbb{E}(T) < \infty$ and $\exists K \geq 0$ such that*

$$|X_n(\omega) - X_{n-1}(\omega)| \leq K$$

for every $n \in \mathbb{N}$ and $\omega \in \Omega$.

We will apply the optional stopping theorem several times in our second proof of a central limit theorem for Markov chains in Section 3.3.
Before closing this preliminary chapter, we give one more definition which will be of use to us later in this report. This comes from Section 5.1 of [15].

Definition 1.3.6. [Martingale difference] Let $\{X_n\}$ be a martingale with respect to a filtration $(\mathcal{F}_n)$. Define $Y_{n+1} := X_{n+1} - X_n$ for all $n \in \mathbb{N}$. Then we say that the process $\{Y_n\}$ is a **martingale difference**.

Lemma 1.3.2. *Let $(\mathcal{F}_n)$ be a filtration. A stochastic process $\{Y_n\}$ is a martingale difference with respect to $(\mathcal{F}_n)$ if and only if $\{Y_n\}$ satisfies the first two conditions in Definition 1.3.4 and, for all $n \in \mathbb{N}$,*

$$\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = 0 \qquad a.s. \tag{1.3.2}$$

In Section 3.1, we prove a central limit theorem for martingale differences and then go on to deduce a central limit theorem for Markov chains in the following section.

Chapter 2

Ergodicity of Stochastic Processes

In this chapter we prove the ergodic theorems, which concern convergence of random variables. These theorems will play a key role in proving the central limit theorems in the next chapter. We will define what it means for a probability measure or a stochastic process to be ergodic, and we will see that ergodicity is a sufficient condition for the limit in the ergodic theorems to be a constant.

2.1 Measure-Preserving Transformations

In this section we relate stationary stochastic processes to measure-preserving transformations, following Varadhan in [15].
A stationary stochastic process is a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ such that the joint distribution of $(X_{n_1}, X_{n_2}, \ldots, X_{n_k})$ is the same as the joint distribution of $(X_{n_1+m}, X_{n_2+m}, \ldots, X_{n_k+m})$ for any $n_1, n_2, \ldots, n_k, m \in \mathbb{Z}$.
Let Ω be the space of sequences which take values in some measurable space $(X, \mathcal{B})$ and let $\mathcal{F}$ be the product σ-field. Under certain consistency conditions, we can construct a measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ which describes the evolution of the process $\{X_n\}_{n\in\mathbb{N}}$ over time.
We can define the shift T on Ω by $(T\omega)(n) = \xi_{n+1}$, where $\omega(n) = \xi_n$. Then stationarity of the process is equivalent to invariance of $\mathbb{P}$ with respect to T; i.e. $\mathbb{P}T^{-1} = \mathbb{P}$.
In this case, we call T a **measure-preserving transformation** for $\mathbb{P}$. We also say that $\mathbb{P}$ is an **invariant measure** for T.
We will study general measure-preserving transformations and will later apply our results to stationary stochastic processes.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and T a measure-preserving transformation for $\mathbb{P}$. We will prove some general facts about these transformations, which are stated in Section 6.1 of [15], by relating T to a linear transformation on the space of functions on Ω, as described by the following lemma.

Lemma 2.1.1. *The measure-preserving transformation $T : \Omega \to \Omega$ induces a linear transformation U on the space of functions on Ω by*

$$(Uf)(\omega) = f(T\omega).$$

Proof. Clearly

$$U(\alpha f + \beta g)(\omega) = (\alpha f + \beta g)(T\omega) = \alpha f(T\omega) + \beta g(T\omega) = \alpha(Uf)(\omega) + \beta(Ug)(\omega).$$

$\square$

We will now study the transformation U. First we show that U is an **isometry** on the $\mathcal{L}^p$ spaces for $1 \leq p \leq \infty$. By definition (see [9]), U is an isometry on a normed space if it preserves the norm.

Lemma 2.1.2. *U is an isometry on $\mathcal{L}^p$ for $1 \leq p < \infty$.*

Proof. We just use the definition of U and invariance of T with respect to $\mathbb{P}$. Let $p \in [1, \infty)$ and $f \in \mathcal{L}^p$. Then

$$\begin{aligned}
\int_\Omega |f(\omega)|^p \, d\mathbb{P}(\omega) &= \int_\Omega |f(T\omega)|^p \, d\mathbb{P}(T\omega) \\
&= \int_\Omega |f(T\omega)|^p \, d\mathbb{P}(\omega) = \int_\Omega |(Uf)(\omega)|^p \, d\mathbb{P}(\omega).
\end{aligned}$$

$\square$

Remark 2.1.1. U is also an isometry on $\mathcal{L}^\infty$.

Proof. For any $\omega_0 \in \Omega$,

$$|Uf(\omega_0)| = |f(T\omega_0)| \leq \sup_{\omega \in \Omega} |f(\omega)| = \|f\|_\infty .$$

So $Uf \in \mathcal{L}^\infty$ and $\|Uf\|_\infty \leq \|f\|_\infty$. But $\exists \omega_1 \in \Omega$ such that $T\omega_1 = \omega_1$ and so $|Uf(\omega_1)| = |f(\omega_1)|$.
Thus $\|Uf\|_\infty = \|f\|_\infty$. $\square$

Next we show that U is invertible and we find its inverse.

Lemma 2.1.3. *U is invertible and the inverse of U is the transformation induced by T^{-1}, the inverse of T.*

Proof. Define U^{-1} by

$$(U^{-1}f)(\omega) = f(T^{-1}\omega) \text{ for any function } f \text{ on } \Omega.$$

Then

$$(UU^{-1}f)(\omega) = (U^{-1}f)(T\omega) = f(T^{-1}T\omega) = f(\omega).$$

So U^{-1} is the inverse of U. $\square$

If we consider U as a transformation on $\mathcal{L}^2$ and define the usual inner product on $\mathcal{L}^2$ by $\langle f, g \rangle = \int_\Omega f(\omega)g(\omega)d\mathbb{P}(\omega)$, for any $f, g \in \mathcal{L}^2$, then we can show that U is **unitary**. To define what it means for a transformation to be unitary, we introduce the **Hilbert adjoint operator** for a transformation A, which we denote A^*. These definitions are taken from Sections 3.9 and 3.10 of [9].
A^* is defined to be the transformation on $\mathcal{L}^2$ such that for any $f, g \in \mathcal{L}^2$,

$$\langle Af, g \rangle = \langle f, A^* g \rangle .$$

Then A is unitary if $A^* = A^{-1}$.

Remark 2.1.2. An equivalent condition for A to be unitary is that, for any $f, g \in \mathcal{L}^2$,

$$\langle Af, Ag \rangle = \langle f, g \rangle .$$

For a simple proof of this, see Section 3.10 of [9].

We now use the above remark to prove that U is unitary. This property will be useful for us in the proof of the $\mathcal{L}^2$ ergodic theorem.

Lemma 2.1.4. *U is unitary in $\mathcal{L}^2$, with inner product defined by $\langle f, g \rangle = \int_\Omega f(\omega)g(\omega)d\mathbb{P}(\omega)$; i.e. U preserves this inner product.*

Proof. Using the definition of U and invariance of T with respect to $\mathbb{P}$, we get

$$\begin{aligned}
\langle Uf, Ug \rangle = \int f(T\omega)g(T\omega)d\mathbb{P}(\omega) &= \int f(T\omega)g(T\omega)d\mathbb{P}(T\omega) \\
&= \int f(\omega)g(\omega)d\mathbb{P}(\omega) \\
&= \langle f, g \rangle .
\end{aligned}$$

□

Before moving on to prove the ergodic theorems, we note two more properties of U which will be of use to us.

Remark 2.1.3. $U1 = 1$, where $1(\omega) \equiv 1,\ \forall \omega \in \Omega$, since

$$(U1)(\omega) = 1(T\omega) = 1.$$

Remark 2.1.4. $U(fg) = U(f)U(g)$, for any functions f, g on Ω, since

$$U(fg)(\omega) = (fg)(T\omega) = f(T\omega)g(T\omega) = (Uf)(\omega)(Ug)(\omega).$$

2.2 Ergodic Theorems

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and T a measure-preserving transformation for $\mathbb{P}$. Define the invariant σ-field by $\mathcal{I} = \{A \subseteq \Omega : TA = A\}$.
The following theorems from [15] are key results in the proof of the central limit theorems of the next chapter. The first of these is alternately called the **Individual Ergodic Theorem** or **Birkhoff's Theorem**.

Theorem 2.2.1 (The Individual Ergodic Theorem)**.** *For any $f \in \mathcal{L}^1(\mathbb{P})$, the limit*

$$\lim_{n\to\infty} \frac{f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)}{n} = g(\omega)$$

exists for $\mathbb{P}$-almost all ω.
Moreover, the limit $g(\omega)$ is given by the conditional expectation

$$g(\omega) = \mathbb{E}^{\mathbb{P}}(f|\mathcal{I}).$$

We will first prove the **Mean Ergodic Theorems**, which concern $\mathcal{L}^p$ convergence. We will then develop the tools we need to complete the proof of the Individual Ergodic Theorem, following Varadhan in [15].

Theorem 2.2.2 (Mean Ergodic Theorems)**.** *Let $p \in [1, \infty)$. Then for any $f \in \mathcal{L}^p(\mathbb{P})$, the limit*

$$\lim_{n\to\infty} \frac{f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)}{n} = g(\omega)$$

exists in $\mathcal{L}^p(\mathbb{P})$.
Moreover, the limit $g(\omega)$ is given by the conditional expectation

$$g(\omega) = \mathbb{E}^{\mathbb{P}}(f|\mathcal{I}).$$

Proof. We first consider $p = 2$.
Define $H = \mathcal{L}^2$ and

$$H_0 = \{f : f \in H, Uf = f\} = \{f : f \in H, f(T\omega) = f(\omega)\},$$

where U is the operator induced by T.
We claim that H_0 is a closed non-trivial subspace of H.
Since $c \in H_0$ for any constant c, $H_0 \neq \emptyset$.
Suppose that $f, g \in H_0$, $\alpha \in \mathbb{R}$. Then $f + g, \alpha f \in H$, by an elementary property of $\mathcal{L}^p$ spaces, and we have

$$\begin{aligned} U(f+g) &= Uf + Ug = f + g \\ \text{and } U(\alpha f) &= \alpha Uf = \alpha f, \end{aligned}$$

by linearity of U. So $f + g,\ \alpha f \in H_0$.
We have that H_0 is a subspace of H. Now suppose that (f_j) is a sequence in

H_0 with $\mathcal{L}^2$ limit f. Then

$$\begin{aligned} Uf(\omega) = f(T\omega) &= \lim_{j\to\infty} f_j(T\omega) \\ &= \lim_{j\to\infty} Uf_j(\omega) = \lim_{j\to\infty} f_j(\omega) \\ &= f(\omega). \end{aligned}$$

Thus H_0 is closed. This proves our claim.
Now for each $n \in \mathbb{N}$, define $A_n : H \to \mathbb{R}$ by

$$A_n f = \frac{f + Uf + \cdots + U^{n-1}f}{n}.$$

Then

$$\begin{aligned} \|A_n f\|_2 &= \frac{1}{n} \left\| f + Uf + \cdots + U^{n-1}f \right\|_2 \\ &\leq \frac{1}{n} \left(\|f\|_2 + \|Uf\|_2 + \cdots + \left\| U^{n-1}f \right\|_2 \right) \\ &\qquad \text{by Minkowski's inequality (1.1.3)} \\ &= \|f\|_2 \qquad \text{since } U \text{ is an } \mathcal{L}^2 \text{ isometry by Lemma 2.1.2.} \end{aligned} \tag{2.2.1}$$

Hence $\|A_n\| \leq 1$.
Suppose that $f \in H_0$, so that $Uf = f$, and let $n \in \mathbb{N}$. Then

$$A_n f = \frac{f + Uf + \cdots + U^{n-1}f}{n} = \frac{f + f + \cdots + f}{n} = f.$$

So clearly we have that $\forall f \in H_0$,

$$A_n f \to f \qquad \text{in } H \text{ and almost surely.}$$

Now suppose that $f \in H_0^{\perp}$.
We claim that $H_0^{\perp} = \overline{\text{Range}\,(I - U)}$.
Since U is unitary by Lemma 2.1.4, we have the equivalence

$$Uf = f \qquad \Leftrightarrow \qquad U^{-1}f = U^* f = f.$$

Thus

$$H_0 = \{f : f \in H, (I - U^*)f = f\}.$$

We need the following remark to prove our claim.

Remark 2.2.1. The statement of Exercise 6.1 of [15] tells us that for any bounded linear transformation A on H, we have

$$\overline{\text{Range}\, A} = \{f : f \in H, A^* f = 0\}^{\perp}.$$

We do not prove this here, since this would require more discussion of Hilbert spaces and Hilbert adjoint operators, as could be found for example in Chapter 3 of [9].

We can see that $I - U$ is linear and bounded. Thus, by the above remark,

$$H_0^{\perp} = \{f : f \in H, (I - U^*)f = f\}^{\perp} = \overline{\text{Range}\,(I - U)},$$

as we claimed.
Therefore, we can approximate any $f \in H_0^{\perp}$ by a sequence (f^j) in Range $(I - U)$ such that $\left\|f - f^j\right\|_2 \xrightarrow{j\to\infty} 0$.
For each $j \in \mathbb{N}$, $\exists g^j \in H$ such that $f^j = (I - U)\, g^j$ and thus

$$\begin{aligned} A_n f^j &= \frac{g^j - Ug^j + U\left(g^j - Ug^j\right) + \cdots + U^{n-1}\left(g^j - Ug^j\right)}{n} \\ &= \frac{g^j - Ug^j + Ug^j - U^2 g^j - \cdots + U^{n-1}g^j - U^n g^j}{n} \\ &= \frac{g^j - U^n g^j}{n}. \end{aligned}$$

Hence

$$\begin{aligned} \left\|A_n f^j\right\|_2 = \left\|\frac{g^j - U^n g^j}{n}\right\|_2 &\leq \frac{1}{n}\left\|g^j\right\|_2 + \left\|U^n g^j\right\|_2 \\ &= \frac{2\left\|g^j\right\|_2}{n} \leq \frac{2}{n}\sup_{j\in\mathbb{N}}\left\|g^j\right\|_2. \end{aligned} \tag{2.2.2}$$

By the triangle inequality,

$$\|A_n f\|_2 = \left\|A_n(f - f^j) + A_n f^j\right\|_2 \leq \left\|A_n(f - f^j)\right\|_2 + \left\|A_n f^j\right\|_2.$$

As we noted in (2.2.1), by Minkowski's inequality, $\left\|A_n(f - f^j)\right\|_2 \leq \left\|f - f^j\right\|_2$.
Therefore

$$\|A_n f\|_2 \leq \left\|f - f^j\right\|_2 + \frac{2}{n}\sup_{j\in\mathbb{N}}\left\|g^j\right\|_2.$$

Taking $j \to \infty$, we get that

$$0 \leq \|A_n f\|_2 \leq \frac{2}{n}\sup_{j\in\mathbb{N}}\left\|g^j\right\|_2 \xrightarrow{n\to\infty} 0.$$

Hence

$$A_n f \xrightarrow[n\to\infty]{\mathcal{L}^2} 0 \qquad \forall f \in H_0^{\perp}.$$

If we denote by Π the orthogonal projection from H into H_0, as defined in Section 3.3 of [9] - i.e. $\Pi : H \to H_0$ satisfies

$$\Pi f = g, \text{ where } g \in H_0 \text{ is such that } f = g + h, \text{ for some } h \in H_0^{\perp}$$

- then

$$\Pi f = \begin{cases} f & \text{if } f \in H_0 \\ 0 & \text{if } f \in H_0^{\perp}, \end{cases}$$

and so

$$A_n f \to \Pi f \text{ in } \mathcal{L}^2 \text{ for any } f \in H.$$

We have a limit in $\mathcal{L}^2$ and next we need to show that this limit is in fact the conditional expectation claimed.

Let $f \in H$. It can be shown that $\Pi f = \mathbb{E}^{\mathbb{P}}(f|\mathcal{I})$, using a theorem from Section 9.5 of [9], as follows.

$\mathbb{E}^{\mathbb{P}}(\cdot|\mathcal{I})$ is an orthogonal projection if and only if for $f, g \in H$,

$$\mathbb{E}^{\mathbb{P}}(\mathbb{E}^{\mathbb{P}}(f|\mathcal{I})|\mathcal{I}) = \mathbb{E}^{\mathbb{P}}(f|\mathcal{I}) \text{ and } \left\langle \mathbb{E}^{\mathbb{P}}(f|\mathcal{I}), g \right\rangle = \left\langle f, \mathbb{E}^{\mathbb{P}}(g|\mathcal{I}) \right\rangle.$$

Let $f, g \in H$. Then the first equality which we require is clear from the tower property of conditional expectation (1.3.1), and we arrive at the second equality as follows:

$$\begin{aligned}
\left\langle \mathbb{E}^{\mathbb{P}}(f|\mathcal{I}), g \right\rangle &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}(f|\mathcal{I})g\right] \\
&= \mathbb{E}\left(\mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}(f|\mathcal{I})g|\mathcal{I}\right]\right) \\
&= \mathbb{E}\left(\mathbb{E}^{\mathbb{P}}(f|\mathcal{I})\mathbb{E}^{\mathbb{P}}\left(g|\mathcal{I}\right)\right) \quad \text{as } \mathbb{E}^{\mathbb{P}}(f|\mathcal{I}) \text{ is } \mathcal{I}\text{-measurable} \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}(g|\mathcal{I})f\right] \quad \text{by symmetry} \\
&= \left\langle f, \mathbb{E}^{\mathbb{P}}(g|\mathcal{I}) \right\rangle.
\end{aligned}$$

So $\mathbb{E}^{\mathbb{P}}(\cdot|\mathcal{I})$ is an orthogonal projection.

Since $\mathbb{E}^{\mathbb{P}}(f|\mathcal{I})$ is $\mathcal{I}$-measurable, $\mathbb{E}^{\mathbb{P}}(f|\mathcal{I}) \in H_0$ for all $f \in H$. In particular, for $f \in H_0$, $\mathbb{E}^{\mathbb{P}}(f|\mathcal{I}) = f$. So $\mathbb{E}^{\mathbb{P}}(\cdot|\mathcal{I}) : H \to H_0$ is surjective. Therefore $\mathbb{E}^{\mathbb{P}}(\cdot|\mathcal{I})$ is the orthogonal projection from H into H_0. Thus, for $f \in H$, $\mathbb{E}^{\mathbb{P}}(f|\mathcal{I}) = \Pi f$.

We have the required result for $\mathcal{L}^2$ and we now prove the mean ergodic theorems for general $\mathcal{L}^p$.

Note that the conditional expectation operator is well-defined on $\mathcal{L}^p$, so Π is an operator of norm one on $\mathcal{L}^p$, for $1 \leq p \leq \infty$.

First suppose that $f \in \mathcal{L}^\infty$. Then $\|A_n f\|_\infty \leq \|f\|_\infty$, by the triangle inequality for the $\mathcal{L}^\infty$ norm.

Since we have shown that $A_n f \to \Pi f$ in $\mathcal{L}^2$, we also have that $A_n f \to \Pi f$ in measure, by Theorem 1.1.4. Thus we can use the bounded convergence theorem for convergence in measure (Theorem 1.1.6) to get that

$$\|A_n f - \Pi f\|_p \xrightarrow{n\to\infty} 0 \qquad \text{for any } p \in [1, \infty). \tag{2.2.3}$$

Now let $p \in [1, \infty)$ and $f \in \mathcal{L}^p$. We know that $\mathcal{L}^\infty$ is dense in $\mathcal{L}^p$.

By Minkowski's inequality (1.1.3), we see that $\|A_n g\|_p \leq \|g\|_p$ for any $g \in \mathcal{L}^p$.

Let (f_j) be a sequence of functions in $\mathcal{L}^\infty$ which approximate f in $\mathcal{L}^p$; i.e. $\|f - f_j\|_p \to 0$ as $j \to \infty$. Fix $j \in \mathbb{N}$. Then

$$\begin{aligned}
0 \leq \|A_n f - \Pi f\|_p &\leq \|A_n f - A_n f_j\|_p + \|A_n f_j - \Pi f_j\|_p + \|\Pi f_j - \Pi f\|_p \\
&\leq \|f - f_j\|_p + \|A_n f_j - \Pi f_j\|_p + \|f_j - f\|_p \qquad (2.2.4) \\
&\xrightarrow{n\to\infty} 2\|f_j - f\|_p \qquad \text{by (2.2.3).}
\end{aligned}$$

By definition, $\|f - f_j\|_p \to 0$ as $j \to \infty$. Therefore, by letting $j \to \infty$ in (2.2.4), we have

$$\|A_n f - \Pi f\|_p \xrightarrow{n\to\infty} 0,$$

as required. $\square$

The proof of the almost sure convergence in the individual ergodic theorem is based on an inequality which we will be able to prove by means of the following theorem from [15].

Theorem 2.2.3 (Maximal Ergodic Theorem)**.** *Let $f \in \mathcal{L}^1(\mathbb{P})$ and, for $n \geq 1$, define*

$$E_n^0 := \left\{ \omega : \sup_{1\leq j\leq n} \big(f(\omega) + f(T\omega) + \cdots + f(T^{j-1}\omega)\big) \geq 0 \right\}.$$

Then

$$\int_{E_n^0} f(\omega) d\mathbb{P} \geq 0.$$

Proof. Let h_n be the function defined by

$$\begin{aligned} h_n(\omega) &:= \sup_{1\leq j\leq n} \big(f(\omega) + f(T\omega) + \cdots + f(T^{j-1}\omega)\big) \\ &= f(\omega) + \max\{0, h_{n-1}(T\omega)\} \\ &= f(\omega) + h_{n-1}^+(T\omega), \end{aligned}$$

where we define $h_n^+(\omega) := \max\{0, h_n(\omega)\}$.
Note that for any $\omega \in \Omega$,

$$\begin{gathered} h_n^+(\omega) \geq 0, \\ \text{and } h_{n-1}^+(\omega) \leq h_n^+(\omega). \end{gathered}$$

On E_n^0, $h_n(\omega) = h_n^+(\omega) \geq 0,\ \forall\omega$, so

$$\begin{aligned} f(\omega) &= h_n(\omega) - h_{n-1}^+(T\omega) \\ &= h_n^+(\omega) - h_{n-1}^+(T\omega), \quad \forall\omega \in E_n^0. \end{aligned}$$

Hence

$$\begin{aligned} \int_{E_n^0} f(\omega) d\mathbb{P} &= \int_{E_n^0} h_n^+(\omega) d\mathbb{P} - \int_{E_n^0} h_{n-1}^+(T\omega) d\mathbb{P} \\ &\geq \int_{E_n^0} h_n^+(\omega) d\mathbb{P} - \int_{E_n^0} h_n^+(T\omega) d\mathbb{P} \\ &= \int_{E_n^0} h_n^+(\omega) d\mathbb{P} - \int_{T(E_n^0)} h_n^+(\omega) d\mathbb{P}, \qquad \text{by invariance of } T \\ &\geq 0, \end{aligned}$$

since for any integrable function $h(\omega)$, $\displaystyle\int_E h(\omega) d\mathbb{P}$ is largest when
$E = \{\omega : h(\omega) \geq 0\}$. $\square$

We can now apply this theorem to prove the following inequalities which are adapted from Lemma 6.3 of [15].

Lemma 2.2.1. *Let* $f \in \mathcal{L}^1(\mathbb{P})$*,* $l > 0$ *and* $\tilde{E}_n = \left\{\omega : \sup_{1\le j\le n} (A_j f)(\omega) \ge l\right\}$. *Then*

$$\mathbb{P}(\tilde{E}_n) \le \frac{1}{l}\int_{\tilde{E}_n} f(\omega)d\mathbb{P}.$$

Proof. We have that

$$\begin{aligned}\tilde{E}_n &= \left\{\omega : \sup_{1\le j\le n} \frac{f(\omega)+f(T\omega)+\cdots+f(T^{j-1}\omega)}{j} \ge l\right\}\\ &= \left\{\omega : \sup_{1\le j\le n} \frac{(f(\omega)-l)+(f(T\omega)-l)+\cdots+\left(f(T^{j-1}\omega)-l\right)}{j} \ge 0\right\}.\end{aligned}$$

Thus, by the maximal ergodic theorem (Theorem 2.2.3),

$$\int_{\tilde{E}_n} (f(\omega)-l)\, d\mathbb{P} \ge 0,$$

which is equivalent to

$$\int_{\tilde{E}_n} f(\omega)d\mathbb{P} - l\mathbb{P}(\tilde{E}_n) \ge 0.$$

Rearranging gives us the required inequality.

□

Corollary 2.2.1. *For any* $f \in \mathcal{L}^1(\mathbb{P})$ *and* $l > 0$*,*

$$\mathbb{P}\left(\omega : \sup_{j\ge 1} |(A_j f)(\omega)| \ge l\right) \le \frac{1}{l}\int |f(\omega)|\, d\mathbb{P}.$$

Proof. Let $n \in \mathbb{N}$ and define

$$E_n := \left\{\omega : \sup_{1\le j\le n} |(A_j f)(\omega)| \ge l\right\}.$$

If we also define

$$\hat{E}_n := \left\{\omega : \sup_{1\le j\le n} (A_j\, |f|)(\omega) \ge l\right\},$$

then we see that $|(A_j f)(\omega)| = A_j\, |f(\omega)|$ for any $\omega \in \Omega$, by the triangle inequality, and so

$$E_n \subseteq \hat{E}_n.$$

We can now apply the previous lemma to get

$$\begin{aligned}\mathbb{P}(E_n) \le \mathbb{P}(\hat{E}_n) &\le \frac{1}{l}\int_{\hat{E}_n} |f|\, d\mathbb{P}\\ &\le \frac{1}{l}\int_{\Omega} |f|\, d\mathbb{P}.\end{aligned} \tag{2.2.6}$$

Now note that for any $n \in \mathbb{N}$,

$$\sup_{1 \leq j \leq n} |(A_j f)(\omega)| \leq \sup_{1 \leq j \leq n+1} |(A_j f)(\omega)|,$$

which implies that

$$E_n \subseteq E_{n+1}.$$

By monotonicity of the sequence of events (E_n), we have that

$$\begin{aligned}
\mathbb{P}\left(\sup_{j \geq 1} |(A_n f)(\omega)| \geq l\right) &= \mathbb{P}\left(\lim_{n\to\infty} \sup_{1 \leq j \leq n} |(A_n f)(\omega)| \geq l\right) \\
&= \mathbb{P}\left(\cup_{n=1}^{\infty} E_n\right) \\
&= \lim_{n\to\infty} \mathbb{P}(E_n) \\
&\leq \frac{1}{l} \int |f(\omega)| \, d\mathbb{P} \qquad \text{by (2.2.6).}
\end{aligned}$$

□

Now we are ready to prove the almost sure convergence in the individual ergodic theorem, as in [15].

Proof of the Individual Ergodic Theorem. Fix $f \in \mathcal{L}^1(\mathbb{P})$.
Note that the set $D := \{f_1 + f_2 : f_1 \in H_0, f_2 = (I - U)g, g \in \mathcal{L}^\infty\}$ is dense in $\mathcal{L}^1$. We do not prove this fact here.
For each $j \in \mathbb{N}$, let $f^j \in D$ be such that $\mathbb{E}\left(\left|f - f^j\right|\right) \to 0$ as $j \to \infty$;
i.e. $\left(f^j\right)$ approximates f in $\mathcal{L}^1$. Fix $j \in \mathbb{N}$. We claim that $A_n f^j$ converges almost surely.
We observed in the proof of the mean ergodic theorems that we have almost sure convergence for any $f_1 \in H_0$.
Now suppose $g \in \mathcal{L}^\infty$ and let $f_2 = (I - U)g$. Then, using (2.2.2),

$$\|A_n f_2\|_\infty = \left\|\frac{g - U^n g}{n}\right\|_\infty \leq \frac{\|g\|_\infty + \|U^n g\|_\infty}{n} = \frac{2\|g\|_\infty}{n} \xrightarrow{n\to\infty} 0.$$

So we also have almost sure convergence for f_2.
Since $f^j \in D$, for any fixed j, we have almost sure convergence for $A_n f^j$ as we claimed.
Thus we have convergence for f as follows.
For fixed j, we see that

$$\begin{aligned}
\limsup_{n\to\infty} A_n f - \liminf_{n\to\infty} A_n f &= \limsup_{n\to\infty} \left(A_n f^j - \left[A_n f^j - A_n f\right]\right) \\
&\quad - \liminf_{n\to\infty} \left(A_n f^j - \left[A_n f^j - A_n f\right]\right) \\
&\leq \limsup_{n\to\infty} A_n f^j - \liminf_{n\to\infty} A_n f^j \\
&\quad + \limsup_{n\to\infty} \left[A_n f^j - A_n f\right] \\
&\quad - \liminf_{n\to\infty} \left[A_n f^j - A_n f\right].
\end{aligned}$$

But, since $A_n f^j$ converges almost surely,

$$\limsup_{n\to\infty} A_n f^j = \liminf_{n\to\infty} A_n f^j \text{ a. s.}. \tag{2.2.7}$$

Also note that

$$\limsup_{n\to\infty} \left[A_n f^j - A_n f\right] - \liminf_{n\to\infty} \left[A_n f^j - A_n f\right] \le 2 \sup_{n\in\mathbb{N}} \left|A_n \left(f^j - f\right)\right|. \tag{2.2.8}$$

So, putting everything together and applying the previous corollary, we arrive at

$$\begin{aligned}
0 &\le \mathbb{P}\left(\limsup_{n\to\infty} A_n f - \liminf_{n\to\infty} A_n f \ge \varepsilon\right) \\
&= \mathbb{P}\left(\limsup_{n\to\infty} \left[A_n f^j - A_n f\right] - \liminf_{n\to\infty} \left[A_n f^j - A_n f\right] \ge \varepsilon\right) \quad \text{by (2.2.7)} \\
&\le \mathbb{P}\left(\sup_{n\in\mathbb{N}} \left|A_n \left(f^j - f\right)\right| \ge \frac{\varepsilon}{2}\right) \qquad \text{by (2.2.8)} \\
&\le \frac{2}{\varepsilon}\mathbb{E}\left(\left|A_n \left(f^j - f\right)\right|\right) \qquad \text{by Corollary 2.2.1} \\
&\le \frac{2}{\varepsilon}\mathbb{E}\left(\left|f^j - f\right|\right) \xrightarrow{j\to\infty} 0,
\end{aligned}$$

where the final inequality follows from the fact that $\|A_n g\|_1 \le \|g\|_1$, for any $g \in \mathcal{L}^1$. This is due to Minkowski's inequality (1.1.3), as we noted in the proof of the mean ergodic theorems.

If we now let $\varepsilon \to 0$, we see that we have an almost sure limit for $A_n f$ as $n \to \infty$.
By the result of the mean ergodic theorems, it must be the case that $\lim_{n\to\infty} A_n f = \mathbb{E}^{\mathbb{P}}(f|\mathcal{I})$ a.s.. □

We have shown almost sure and $\mathcal{L}^p$ convergence to a random variable, but the theorems will be of use to us when the limit is a constant. We will see in the following section that this is the case under certain conditions on the measure we are working with.

2.3 Ergodicity of Measures

We now define what it means for an invariant measure to be ergodic and see that this provides a sufficient condition for the limit in the ergodic theorems to be a constant.

Definition 2.3.1. Let $\mathbb{P}$ be an invariant measure for a transformation T on Ω. $\mathbb{P}$ is **ergodic** for T if $\mathbb{P}(A) \in \{0, 1\}$ for any invariant set $A \in \mathcal{I}$.

Remark 2.3.1. Let T be a transformation on Ω and $\mathbb{P}$ an ergodic measure for T. Then any function f which is invariant under T is almost surely constant with respect to $\mathbb{P}$.
Moreover, $\mathbb{E}\left(f|\mathcal{I}\right) = \mathbb{E}^{\mathbb{P}}(f)$.

Proof. Let f be invariant under T. Then for any $c \in \mathbb{R}$, $\{f(x) \leq c\} \in \mathcal{I}$. So by ergodicity, each of these sets is trivial. Thus $\exists c_0 \in \mathbb{R}$ such that $f(x) = c_0$ for $\mathbb{P}$-almost every x. That is, $f = \mathbb{E}^{\mathbb{P}}(f)$ almost surely.
Thus $\mathbb{E}(f|\mathcal{I}) = \mathbb{E}\left(\mathbb{E}^{\mathbb{P}}(f)|\mathcal{I}\right) = \mathbb{E}^{\mathbb{P}}(f)$. □

This remark tells us that, when $\mathbb{P}$ is ergodic for T, the limit in the ergodic theorems is a constant. We will use this result to prove the central limit theorems in the next chapter.

Next, we look at a simple example of when we have an ergodic measure.

Theorem 2.3.1. *Any product measure is ergodic for the shift.*

To prove this, we need to appeal to a result known as Kolmogorov's zero-one law.
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a product space with $\Omega = \ldots \times \Omega_{-2} \times \Omega_{-1} \times \Omega_0 \times \Omega_1 \times \Omega_2 \times \cdots$ and define $\mathcal{F}_i = \sigma\{\Omega_j : j \leq i\}$, for all $i \in \mathbb{N}$.
Define the **tail σ-algebra** to be $\mathcal{T} := \cap_{n\in\mathbb{N}} \mathcal{T}_n$, where $\mathcal{T}_n := \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots)$.

Lemma 2.3.1 (Kolmogorov's zero-one law)**.** *For any $A \in \mathcal{T}$, $\mathbb{P}(A) = 0$ or 1.*

Proof. Let $A \in \mathcal{T} \subseteq \mathcal{F}$ and let $\varepsilon > 0$.
Then $\exists a, b \in \mathbb{N}$ with $a < b$ and $\exists A' \in \sigma(\mathcal{F}_a, \mathcal{F}_{a+1}, \ldots, \mathcal{F}_b)$ such that

$$\mathbb{P}(A \triangle A') \leq \varepsilon,$$

where we use the notation $A \triangle A' = A \setminus A' \cup A' \setminus A$.
Note that $\mathbb{P}(A) \leq \mathbb{P}(AA') + \mathbb{P}(A \setminus A') \leq \mathbb{P}(AA') + \mathbb{P}(A \triangle A')$. So

$$\mathbb{P}(A) - \mathbb{P}(AA') \leq \varepsilon. \tag{2.3.1}$$

Since $A \in \mathcal{T}$ and $A' \in \sigma(\mathcal{F}_a, \mathcal{F}_{a+1}, \ldots, \mathcal{F}_b)$, A and A' are independent. Therefore we have

$$\mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(A') \leq \varepsilon.$$

Also, we have that $|\mathbb{P}(A) - \mathbb{P}(A')| \leq \varepsilon$, since $\mathbb{P}(A) - \mathbb{P}(A') \leq \mathbb{P}(A) - \mathbb{P}(AA') \leq \varepsilon$ by (2.3.1), and similarly $\mathbb{P}(A') - \mathbb{P}(A) \leq \varepsilon$. Therefore

$$\begin{aligned}
|\mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(A)| &\leq |\mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(A')| + |\mathbb{P}(A)\mathbb{P}(A') - \mathbb{P}(A)\mathbb{P}(A)| \\
&= (\mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(A')) + \mathbb{P}(A)\,|\mathbb{P}(A) - \mathbb{P}(A')| \\
&\leq \varepsilon + \mathbb{P}(A)\varepsilon \leq 2\varepsilon.
\end{aligned}$$

So we have

$$\mathbb{P}(A)\,(1 - \mathbb{P}(A)) \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\mathbb{P}(A)\,(1 - \mathbb{P}(A)) = 0$. So either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. □

We can now prove the theorem.

Proof of theorem. Let T be the shift, $\mathbb{P}$ a product measure and $A \in \mathcal{I}$ be an invariant set. We need to show that $\mathbb{P}(A) \in \{0,1\}$.
By the definition of $\mathcal{F}$, we can approximate A by sets A_n in the σ-field corresponding to the coordinates from $-n$ to n, in the sense that $\mathbb{P}(A \triangle A_n) \to 0$ as $n \to \infty$.
Equivalently, we can approximate $T^{\pm 2n}A$ by $T^{\pm 2n}A_n$, so by invariance of A, we can approximate A by $T^{\pm 2n}A_n$.
But $T^{2n}A_n$ is in the σ-field corresponding to coordinates from n to $3n$ and $T^{-2n}A_n$ in the σ-field corresponding to coordinates from $-n$ to $-3n$. So $T^{\pm 2n}A_n$ is independent of the σ-field corresponding to coordinates from $-n$ to n.
Since $T^{\pm 2n}A_n$ approximates A, we see that A belongs to the tail σ-field. So, by Kolmogorov's zero-one law (Lemma 2.3.1), $\mathbb{P}(A) \in \{0,1\}$. $\square$

2.4 Structure of Invariant Measures

We are now going to apply the individual ergodic theorem (Theorem 2.2.1) to prove a criterion for a probability measure to be ergodic. We will then see that any invariant measure can be obtained by taking a weighted average of ergodic measures.
Let T be a transformation on Ω and define

$$\mathcal{M} := \{\mathbb{P} : T\text{-invariant probability measure on } (\Omega, \mathcal{F})\}.$$

Note that $\mathcal{M}$ is a convex set which may be empty.
For any convex set C, we say that $x \in C$ is an **extreme** point of C (or x is **extremal**) if it cannot be written as a non-trivial convex combination of two other points in C.

Theorem 2.4.1. *$\mathbb{P} \in \mathcal{M}$ is ergodic for T if and only if it is an extreme point of $\mathcal{M}$.*

Proof. Let $\mathbb{P} \in \mathcal{M}$.
We first show that if $\mathbb{P}$ is not extremal, then $\mathbb{P}$ is not ergodic. Suppose that $\mathbb{P} \in \mathcal{M}$ is not extremal.
Then $\exists \mathbb{P}_1, \mathbb{P}_2 \in \mathcal{M}, a \in (0,1)$ such that $\mathbb{P}_1 \neq \mathbb{P}_2$ and

$$\mathbb{P} = a\mathbb{P}_1 + (1-a)\mathbb{P}_2.$$

Suppose for a contradiction that $\mathbb{P}$ is ergodic. Then we have that, for any $A \in \mathcal{I}$, $\mathbb{P}(A) \in \{0,1\}$.
But

$$\begin{aligned}
&\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{P}_1 = \mathbb{P}_2 = 0, \text{ and similarly} \\
&\mathbb{P}(A) = 1 \Leftrightarrow \mathbb{P}_1 = \mathbb{P}_2 = 1.
\end{aligned}$$

So $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{I}$.
Next we show that this implies that $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{F}$.

Let f be a bounded $\mathcal{F}$-measurable function. We will show that $\mathbb{E}^{\mathbb{P}_1}(f(\omega) = \mathbb{E}^{\mathbb{P}_2}(f(\omega))$. Define E to be the set where the following limit exists:

$$h(\omega) := \lim_{n\to\infty} \frac{1}{n}\left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)\right).$$

By the individual ergodic theorem (Theorem 2.2.1), $\mathbb{P}_1(E) = \mathbb{P}_2(E) = 1$ and h is $\mathcal{I}$-measurable.
Since $\mathbb{P}_1, \mathbb{P}_2$ are invariant, we have, for $i = 1, 2$ and $n \in \mathbb{N}$,

$$\begin{aligned}
&\int_E \frac{1}{n}\left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)\right) d\mathbb{P}_i \\
&\qquad = \int_E \frac{1}{n}\left(f(\omega) + f(\omega) + \cdots f(\omega)\right) d\mathbb{P}_i \\
&\qquad = \int_E f(\omega) d\mathbb{P}_i \\
&\qquad = \mathbb{E}^{\mathbb{P}_i}(f(\omega)), \quad \text{as } \mathbb{P}_i(E) = 1.
\end{aligned}$$

By the bounded convergence theorem (Theorem 1.1.5),

$$\begin{aligned}
\int_E h(\omega) d\mathbb{P}_i &= \lim_{n\to\infty} \int_E \frac{1}{n}\left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)\right) d\mathbb{P}_i \\
&= \mathbb{E}^{\mathbb{P}_i}(f(\omega)), \quad \text{for } i = 1, 2.
\end{aligned}$$

We now use that h is $\mathcal{I}$-measurable, $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{I}$, and $\mathbb{P}_1(E) = \mathbb{P}_2(E)$ to see that

$$\mathbb{E}^{\mathbb{P}_1}(f(\omega)) = \int_E h(\omega) d\mathbb{P}_1 = \int_E h(\omega) d\mathbb{P}_2 = \mathbb{E}^{\mathbb{P}_2}(f(\omega)).$$

So we have that $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{F}$. This is a contradiction.
Next we show the converse part of the theorem. Suppose that $\mathbb{P}$ is not ergodic. Then $\exists A \in \mathcal{I}$ with $0 < \mathbb{P}(A) < 1$.
We can therefore define probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ by

$$\begin{aligned}
\mathbb{P}_1(E) &= \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)} \quad \text{and} \\
\mathbb{P}_2(E) &= \frac{\mathbb{P}(A^c \cap E)}{\mathbb{P}(A^c)}.
\end{aligned}$$

Then we have

$$\begin{aligned}
\mathbb{P}_1(TE) &= \frac{\mathbb{P}(A \cap TE)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(TA \cap TE)}{\mathbb{P}(TA)} \qquad \text{as } A \in \mathcal{I} \\
&= \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)} \qquad \text{by invariance of } \mathbb{P}.
\end{aligned}$$

So $\mathbb{P}_1 \in \mathcal{M}$. Similarly, we can show that $\mathbb{P}_2 \in \mathcal{M}$.
Furthermore,

$$\begin{aligned}\mathbb{P} &= \mathbb{P}(A)\mathbb{P}_1 + \mathbb{P}(A^c)\mathbb{P}_2 \\ &= \mathbb{P}(A)\mathbb{P}_1 + [1 - \mathbb{P}(A)]\,\mathbb{P}_2.\end{aligned}$$

Therefore $\mathbb{P}$ is not extremal. This completes the proof. $\square$

The next theorem shows that any probability measure in the convex set $\mathcal{M}$ - that is any invariant measure - can be obtained by taking a weighted average of the extremal points of $\mathcal{M}$ - i.e. the ergodic measures.
We use the notation $\mathcal{M}_e := \{\mathbb{P} : \mathbb{P} \in \mathcal{M}, \mathbb{P} \text{ extremal}\}$.

Theorem 2.4.2. *For any invariant measure $\mathbb{P}$, there exists a probability measure $\mu_{\mathbb{P}}$ on the set $\mathcal{M}_e$ of ergodic measures such that*

$$\mathbb{P} = \int_{\mathcal{M}_e} Q\mu_{\mathbb{P}} dQ.$$

To prove this we need the following lemma.

Lemma 2.4.1. *Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$ and, for each $\omega \in \Omega$, define $\mathbb{P}_\omega$ by*

$$\mathbb{P}_\omega(E) := \mathbb{P}\left(E|\mathcal{I}\right)(\omega) = \mathbb{E}\left(\mathbb{1}_E|\mathcal{I}\right)(\omega), \text{ for any } E \in \mathcal{F}.$$

Suppose that $\mathbb{P}$ is invariant.
Then for $\mathbb{P}$-almost every ω, $\mathbb{P}_\omega$ is invariant and ergodic.

Proof. We want to show that $\mathbb{P}_\omega(TA) = \mathbb{P}_\omega(A)$ $\forall A \in \mathcal{F}$, for almost all ω. Then we will have invariance of $\mathbb{P}_\omega$ for almost all ω.
Let $E \in \mathcal{I}$. It is enough to show that $\int_E \mathbb{P}_\omega(A) d\mathbb{P}(\omega) = \int_E \mathbb{P}_\omega(TA) d\mathbb{P}(\omega)$, because $\mathbb{P}_\omega$ is $\mathcal{I}$-measurable.
We have that

$$\begin{aligned}\int_E \mathbb{P}_\omega(A) d\mathbb{P}(\omega) &= \int_\Omega \mathbb{1}_E \mathbb{P}_\omega(A) d\mathbb{P}(\omega) \\ &= \mathbb{E}^{\mathbb{P}}\left(\mathbb{1}_E \mathbb{E}\left(\mathbb{1}_A|\mathcal{I}\right)\right) \\ &= \mathbb{E}^{\mathbb{P}}\left(\mathbb{E}\left(\mathbb{1}_E \mathbb{1}_A|\mathcal{I}\right)\right) \text{ as } E \in \mathcal{I} \\ &= \mathbb{E}^{\mathbb{P}}\left(\mathbb{1}_E \mathbb{1}_A\right) \\ &= \mathbb{P}\left(E \cap A\right).\end{aligned} \tag{2.4.3}$$

On the other hand, by similar reasoning,

$$\begin{aligned}\int_E \mathbb{P}_\omega(TA) d\mathbb{P}(\omega) &= \mathbb{P}(E \cap TA) \\ &= \mathbb{P}(TE \cap TA), \text{ by invariance of } E \in \mathcal{I} \\ &= \mathbb{P}(E \cap A), \text{ by invariance of } \mathbb{P}.\end{aligned}$$

So we have invariance and now need to establish ergodicity.
Again let $E \in \mathcal{I}$. Then, repeating the argument in (2.4.3), we get

$$\int_E \mathbb{P}_\omega(E) d\mathbb{P}(\omega) = \mathbb{P}(E \cap E) = \mathbb{P}(E). \tag{2.4.4}$$

But, since $\mathbb{P}_\omega(E) \leq 1$,

$$\int_E \mathbb{P}_\omega(E) d\mathbb{P}(\omega) \leq \int_E d\mathbb{P}(\omega) = \mathbb{P}(E),$$

with equality if and only if $\mathbb{P}_\omega(E) = 1$ for $\mathbb{P}$-almost every $\omega \in E$.
Therefore, for the equality (2.4.4) to hold, we must have $\mathbb{P}_\omega(E) = 1$ for $\mathbb{P}$-almost every $\omega \in E$.
Repeating the same argument from (2.4.3) once more, we get

$$\int_{E^c} \mathbb{P}_\omega(E) d\mathbb{P}(\omega) = \mathbb{P}\left(E \cap E^c\right) = 0.$$

So for $\mathbb{P}$-almost every $\omega \in E^c$, $\mathbb{P}_\omega = 0$.
Thus for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\mathbb{P}_\omega(E) = \mathbb{1}_{\{\omega \in E\}} \in \{0, 1\}, \quad \forall E \in \mathcal{I};$$

that is, $\mathbb{P}_\omega$ is ergodic for $\mathbb{P}$-almost all ω. □

Proof of Theorem. As a consequence of the lemma, we can view $\mathbb{P}_\omega$ as a map $\Omega \to \mathcal{M}_e$. Take $\mu_{\mathbb{P}}$ to be the image of $\mathbb{P}$ under this map.
From the definition of $\mathbb{P}_\omega$, we have that

$$\mathbb{P} = \int_\Omega \mathbb{P}_\omega d\mathbb{P}(\omega).$$

By a change of variables we get

$$\mathbb{P} = \int_{\mathcal{M}_e} Q \mu_{\mathbb{P}} dQ$$

as required. □

2.5 Stationary Markov Processes

In this section we are going to show how an ergodic Markov process can be defined and how the theory from the previous sections can be applied to Markov processes, following Varadhan in Section 6.3 of [15].
Let $(X, \mathcal{B})$ be a measurable space and let $(\Omega, \mathcal{F})$ be the space of sequences which take values in X with the product σ-field.
Let $\{X_n\}$ be a stochastic process which takes values in the state space X.
For any $m, n \in \mathbb{Z}$ with $m < n$, define $\mathcal{F}_n^m = \sigma\{X_j : m \leq j \leq n\}$. Also define

$\mathcal{F}_n = \sigma\{X_j : j \le n\}$ and $\mathcal{F}^m = \sigma\{X_j : j \ge m\}$.
As we noted in the opening discussion of Section 2.1, we can, under certain consistency conditions, construct a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, which describes the evolution of $\{X_n\}$ over time. We assume from now on that these conditions are satisfied. Further discussion of these conditions can be found in [15].
We can then transfer all of the definitions and results from the previous section on measure-preserving transformations by taking the measure to be $\mathbb{P}$ and the transformation T to be the shift.
Suppose that $\{X_n\}$ is stationary; i.e. $\mathbb{P}$ is an invariant measure for T.

Definition 2.5.1. We say that $\{X_n\}$ is an **ergodic stochastic process** if $\mathbb{P}$ is an ergodic measure, as defined in Definition 2.3.1, for the shift T.

In particular, we want to consider Markov processes.
The measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ defines a Markov process with transition probabilities given by Π if, for any $n \in \mathbb{N} \cup \{0\}$ and $A \in \mathcal{B}$,

$$\mathbb{P}\{X_{n+1} \in A | \mathcal{F}_n\} = \Pi(X_n, A) \quad \mathbb{P}\text{-almost surely,}$$

whenever this measure exists and is unique.

The following theorem tells us that whenever the transition probabilities are independent of time n, we have the required existence and uniqueness.

Theorem 2.5.1. *Let $\mathbb{P}$ be a stationary Markov process with given transition probability Π.*
Then the one-dimensional marginal distribution μ, given by $\mu(A) = \mathbb{P}(X_n \in A)$ (independent of time by stationarity), is Π-invariant; i.e.

$$\mu(A) = \int \Pi(x, A)\mu(x)dx \quad \text{for every } A \in \mathcal{B}.$$

Conversely, given any such μ, there exists a unique stationary Markov process $\mathbb{P}$ with marginals μ and transition probability Π.

Proof. Suppose that $\mathbb{P}$ is a stationary Markov process with transition probabilities given by Π. Then

$$\begin{aligned}\mu(A) = \mathbb{P}(X_n \in A) &= \int \Pi(x, A)\mathbb{P}(X_{n-1} = x)dx \\ &= \int \Pi(x, A)\mu(x)dx, \quad \text{by stationarity.}\end{aligned}$$

Now take a measure μ on $(X, \mathcal{B})$. Then a unique stationary Markov process with marginals μ exists. We will not prove this fact, but refer the reader to Section 4.4 of [15] for further discussion of this. □

Let Π be a transition probability and define the set of invariant probability measures for Π by

$$\widetilde{\mathcal{M}} := \left\{\mu : \mu(A) = \int_X \Pi(x, A)\mu(x)dx \text{ for all } A \in \mathcal{B}\right\}.$$

$\widetilde{\mathcal{M}}$ is clearly a convex set of probability measures. Denote the set of extremals of $\widetilde{\mathcal{M}}$ by $\widetilde{\mathcal{M}}_e$, as we defined at the beginning of Section 2.4, and note that this set may be empty.
For each $\mu \in \widetilde{\mathcal{M}}$, denote the Markov process with marginals μ by $\mathbb{P}_\mu$.
Since the map $\mu \to \mathbb{P}_\mu$ is linear, we have that

$$\mu \in \widetilde{\mathcal{M}}_e \quad \Rightarrow \quad \mathbb{P}_\mu \in \mathcal{M}_e.$$

In fact we also have the opposite implication. To see this, we need the following theorem.

Theorem 2.5.2. *Let μ be an invariant measure for Π and $\mathbb{P} = \mathbb{P}_\mu$ the corresponding stationary Markov process.*
Let $\mathcal{I}$ be the invariant σ-field on Ω; $\mathcal{I} = \{A : TA = A\}$, where T is the shift.
Then $\mathcal{I} \subseteq \sigma(X_0)$, to within sets of $\mathbb{P}$ measure 0.

Proof. Let $E \in \mathcal{I} \subseteq \mathcal{F}$. Then, as in the proof of Theorem 2.3.1, we see that there are sets $E_n \in \mathcal{F}_n^{-n}$ which approximate E. So for any $k \in \mathbb{Z}$, $T^k E_n$ approximates $T^k E$ and, by invariance, approximates E. But $T^k E_n \in \mathcal{F}_{k+n}^{k-n}$. So $E \in \mathcal{T}^+ := \cap_{m\in\mathbb{N}} \mathcal{F}^m$ and $E \in \mathcal{T}^- := \cap_{n\in\mathbb{N}} \mathcal{F}_n$. Thus

$$\begin{aligned}
\mathbb{P}\left(E|\sigma(X_0)\right) &= \mathbb{P}\left(E \cap E|\sigma(X_0)\right) \\
&= \mathbb{P}\left(\mathbb{E}\left[E|\mathcal{T}^-\right] \cap \mathbb{E}\left[E|\mathcal{T}^+\right]|\sigma(X_0)\right) \\
&= \mathbb{P}\left(E|\sigma(X_0)\right)\mathbb{P}\left(E|\sigma(X_0)\right) \quad \text{by conditional independence.}
\end{aligned}$$

Therefore $\mathbb{P}\left(E|\sigma(X_0)\right) = 0$ or 1. It follows that $E \in \sigma(X_0)$ to within sets of $\mathbb{P}$ measure 0. $\square$

We can now show that ergodicity of a Markov process is equivalent to ergodicity of its marginals.

Theorem 2.5.3. *Let μ be a measure on $(X, \mathcal{B})$ and $\mathbb{P}_\mu$ the corresponding Markov process. Then the following equivalence holds:*

$$\mu \in \widetilde{\mathcal{M}}_e \quad \Leftrightarrow \quad \mathbb{P}_\mu \in \mathcal{M}_e.$$

Proof. For the first implication, suppose that μ is not extremal. Then $\exists \mu_1, \mu_2 \in \widetilde{\mathcal{M}}$ and $\exists \alpha \in \mathbb{R}$ such that $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$.
Therefore, $\mathbb{P}_\mu = \alpha\mathbb{P}_{\mu_1} + (1-\alpha)\mathbb{P}_{\mu_2}$, by linearity. So $\mathbb{P}_\mu$ is not extremal.
Now suppose that $\mathbb{P}$ is not extremal. Then $\mathbb{P}$ is not ergodic, by Theorem 2.4.1.
Therefore $\exists E \in \mathcal{I}$ such that $0 < \mathbb{P}(E) < 1$.
By the previous theorem, we can choose E such that $E \in \sigma(X_0)$. This means that $\exists A \subseteq X$ such that $0 < \mu(A) < 1$ and $E = \{X_0 \in A\}$.
By invariance, for any $n \in \mathbb{N}$, $E = \{X_n \in A\}$. Thus

$$E = \{\omega : X_n(\omega) \in A, \ \forall n\}.$$

Suppose that there exists a subset of A with positive measure on which $\Pi(x, A) < 1$. Then $\mathbb{P}(E) = 0$, which is a contradiction. Thus for μ-almost every $x \in A$,

$$\Pi(x, A) = 1 \text{ and } \Pi(x, A^c) = 0.$$

Now we can write

$$\begin{aligned}
\mu(A) &= \int \Pi(x, A) d\mu(x) \\
&= \int_A \Pi(x, A) d\mu(x) + \int_{A^c} \Pi(x, A) d\mu(x) \\
&= \mu(A) + \int_{A^c} \Pi(x, A) d\mu(x).
\end{aligned}$$

So $\displaystyle\int_{A^c} \Pi(x, A) d\mu(x) = 0$ and thus, for μ-almost every $x \in A^c$,

$$\Pi(x, A) = 0 \text{ and } \Pi(x, A^c) = 1.$$

For any measurable set B we have that

$$\begin{aligned}
\mu(B) &= \mu(B|A)\mu(A) + \mu(B|A^c)\mu(A^c), \quad \text{by the law of total probability} \\
&= \mu(B|A)\mu(A) + \mu(B|A^c)[1 - \mu(A)].
\end{aligned}$$

We claim that both $\nu_1 := \mu(\cdot|A)$ and $\nu_2 := \mu(\cdot|A^c)$ are stationary distributions. Then it will follow that μ is not extremal.

We want to show that $\nu_i(B) = \displaystyle\int \Pi(x, B) d\nu_i(x)$ for all measurable sets B, for $i = 1, 2$.

We have that

$$\begin{aligned}
&\int \Pi(x, B) d\nu_1(x) \\
&\quad = \int_A \Pi(x, B) \frac{d\mu(x)}{\mu(A)} \qquad \text{by a change of variables} \\
&\quad = \frac{1}{\mu(A)} \int_A \Pi(x, A \cap B) d\mu(x) \qquad \text{since } \Pi(x, A) = 1 \text{ for } \mu\text{-a.e. } x \in A \\
&\quad = \frac{1}{\mu(A)} \int \Pi(x, A \cap B) d\mu(x) \qquad \text{since } \Pi(x, A) = 0 \text{ for } \mu\text{-a.e. } x \in A^c \\
&\quad = \frac{1}{\mu(A)} \mu(A \cap B) \qquad \text{by stationarity of } \mu \\
&\quad = \mu(B|A) = \nu_1(B).
\end{aligned}$$

By a very similar argument,

$$\int \Pi(x, B) d\nu_2(x) = \nu_2(B).$$

This completes the proof. $\square$

Remark 2.5.1. The measure μ is always an invariant measure for the transition matrix Π of $\mathbb{P}_\mu$, by Theorem 2.5.1; i.e. μ is a stationary distribution of the Markov chain $\mathbb{P}_\mu$.
Suppose that μ is the unique invariant measure for Π. Then $\widetilde{\mathcal{M}} = \{\mu\}$ and so clearly μ is an extremal point of the set $\widetilde{\mathcal{M}}$. Therefore, by Theorem 2.5.3, $\mathbb{P}_\mu$ is ergodic.

We can now prove a simple criterion for ergodicity of a stationary Markov chain, in the case where the state space of the chain is finite.

Theorem 2.5.4. *Let $\{X_n\}$ be a stationary Markov chain which takes values on a finite state space. Suppose that the chain is irreducible and aperiodic. Then $\{X_n\}$ is ergodic.*

Proof. By Theorem 1.2.1, $\{X_n\}$ has a unique stationary distribution. Therefore, by Remark 2.5.1, $\{X_n\}$ is ergodic. $\square$

Chapter 3

Central Limit Theorems

In this chapter, we consider central limit theorems for ergodic stochastic processes. We will see that a central limit theorem holds for ergodic Markov chains, under some conditions. We present one proof, following [15], which will be a consequence of a central limit theorem for square-integrable ergodic martingale differences. We then present an alternative proof, following [13], for ergodic Markov chains on a finite state space.
These theorems concern weak convergence, which we defined in Section 1.1 and denote $X_n \xrightarrow{\mathcal{D}} X$, for random variables $X, X_1, X_2, \ldots$.

3.1 The CLT for Martingale Differences

Let $(X, \mathcal{B})$ be a measurable space and let $(\Omega, \mathcal{F})$ be the space of sequences which take values in X with the product σ-field.
Let $\{\xi_j\}$ be a sequence of square-integrable martingale differences with respect to a filtration $(\mathcal{F}_n)$, as we defined in Definition 1.3.6, which take values in X.
Let $\mathbb{P}$ be the probability measure on $(\Omega, \mathcal{F})$ which describes the evolution of the stochastic process $\{\xi_j\}$.
Suppose that $\{\xi_j\}$ is stationary and ergodic; i.e. $\mathbb{P}$ is stationary and ergodic on $(\Omega, \mathcal{F})$.

Remark 3.1.1. As stated in Section 6.5 of [15], it follows immediately from the individual ergodic theorem (Theorem 2.2.1) and Remark 2.3.1 that we have a strong law of large numbers,

$$\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} \to 0 \qquad \text{a.s..}$$

We will now prove that we have the following central limit theorem, as shown in [15].

Theorem 3.1.1.

$$Z_n := \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma^2\right)$*, for some* $\sigma > 0$*.*

Proof. This proof follows Varadhan in [15], using the continuity lemma for characteristic functions (Lemma 1.1.2).
The characteristic functions of Z_n and Z are

$$\begin{aligned} \phi_n(t) &= \mathbb{E}\left[\exp\left\{it\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{\sqrt{n}}\right\}\right] \\ \text{and} \quad \phi(t) &= \exp\left\{-\frac{\sigma^2 t^2}{2}\right\}, \end{aligned}$$

respectively.
Let us define

$$\psi(n, j, t) := \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \mathbb{E}\left[\exp\left\{it\frac{\xi_1 + \xi_2 + \cdots + \xi_j}{\sqrt{n}}\right\}\right].$$

We want to show that

$$|\psi(n, n, t) - 1| \xrightarrow{n \to \infty} 0.$$

Then the result will follow by Lemma 1.1.2, the continuity lemma, since,

$$|\psi(n, j, t) - 1| \xrightarrow{n \to \infty} 0 \quad \Rightarrow \quad \phi_n(t) \xrightarrow{n \to \infty} \phi(t).$$

First note that we can write

$$\begin{aligned} \psi(n, n, t) - 1 &= \psi(n, n, t) - \psi(n, 0, t) \\ &= \sum_{j=1}^{n} \left[\psi(n, j, t) - \psi(n, j-1, t)\right], \end{aligned}$$

so that we want to estimate the quantity

$$\Delta(n, t) := \left|\sum_{j=1}^{n} \left[\psi(n, j, t) - \psi(n, j-1, t)\right]\right|.$$

We will estimate this quantity in three steps.
For any $j \in \{1, 2, \ldots, n\}$, let us set $S_j = \xi_1 + \xi_2 + \cdots + \xi_j$. Then the j^{th} term in the sum is

$$\begin{aligned} &\psi(n, j, t) - \psi(n, j-1, t) \\ &= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_j}{\sqrt{n}}\right\}\right] - \exp\left\{\frac{\sigma^2 t^2 (j-1)}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\right] \\ &= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \left(\mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\} \exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\}\right]\right. \\ &\qquad \left. - \mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\} \exp\left\{\frac{-\sigma^2 t^2}{2n}\right\}\right]\right) \\ &= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\} \left(\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\} - \exp\left\{\frac{-\sigma^2 t^2}{2n}\right\}\right)\right]. \end{aligned}$$

For our first estimate, we show, by means of a Taylor expansion, that we can replace this term in the sum by

$$\theta(n,j,t) := \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\}\mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\left(\frac{\left(\sigma^2-\xi_j^2\right)t^2}{2n}\right)\right].$$

Take t to be in an arbitrary finite interval, say $|t| \leq T$. In this interval we have

$$\begin{aligned}
&|\psi(n,j,t)-\psi(n,j-1,t)-\theta(n,j,t)| \\
&= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \\
&\cdot\left|\mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\left(\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\}-\exp\left\{\frac{-\sigma^2t^2}{2n}\right\}-\frac{\left(\sigma^2-\xi_j^2\right)t^2}{2n}\right)\right]\right| \\
&= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\}\left|\mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\right.\right. \\
&\qquad\qquad \left.\left.\cdot\left(\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\}-\exp\left\{\frac{-\sigma^2t^2}{2n}\right\}-\frac{\left(\sigma^2-\xi_j^2\right)t^2}{2n}-\frac{it\xi_j}{\sqrt{n}}\right)\right]\right|,
\end{aligned}$$

where the second equality is due to the martingale difference property (1.3.2), as follows:

$$\begin{aligned}
\mathbb{E}\left(\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\frac{it\xi_j}{\sqrt{n}}\right) &= \frac{it}{\sqrt{n}}\mathbb{E}\left(\mathbb{E}\left(\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\xi_j|\mathcal{F}_{j-1}\right)\right) \\
&= \frac{it}{\sqrt{n}}\mathbb{E}\left(\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\mathbb{E}\left(\xi_j|\mathcal{F}_{j-1}\right)\right) \\
&= 0, \qquad \text{since } \mathbb{E}\left(\xi_j|\mathcal{F}_{j-1}\right)=0 \text{ by (1.3.2)}.
\end{aligned}$$

We now apply the triangle inequality for expectation and the fact that $|\exp\{is\}| = 1$ for $s \in \mathbb{R}$ to get

$$\begin{aligned}
&|\psi(n,j,t)-\psi(n,j-1,t)-\theta(n,j,t)| \\
&\leq \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \\
&\quad\cdot\mathbb{E}\left[\left|\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\right|\left|\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\}-\exp\left\{\frac{-\sigma^2t^2}{2n}\right\}-\frac{\left(\sigma^2-\xi_j^2\right)t^2}{2n}-\frac{it\xi_j}{\sqrt{n}}\right|\right] \\
&= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\}\mathbb{E}\left[\left|\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\}-\exp\left\{\frac{-\sigma^2t^2}{2n}\right\}-\frac{\left(\sigma^2-\xi_j^2\right)t^2}{2n}-\frac{it\xi_j}{\sqrt{n}}\right|\right] \\
&= \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \\
&\quad\cdot\mathbb{E}\left[\left|\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\}-1-\frac{it\xi_j}{\sqrt{n}}+\frac{\xi_j^2t^2}{2n}-\left(\exp\left\{\frac{-\sigma^2t^2}{2n}\right\}-1+\frac{\sigma^2t^2}{2n}\right)\right|\right].
\end{aligned}$$

Applying the triangle inequality again, and the fact that the exponential function is monotone increasing, we see that

$$
\begin{aligned}
&|\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)| \\
&\leq \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \\
&\cdot \left(\mathbb{E}\left[\left|\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\} - 1 - \frac{it\xi_j}{\sqrt{n}} + \frac{\xi_j^2 t^2}{2n}\right|\right] + \left|\left(\exp\left\{\frac{-\sigma^2 t^2}{2n}\right\} - 1 + \frac{\sigma^2 t^2}{2n}\right)\right|\right) \\
&\leq C_T \mathbb{E}\left[\left|\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\} - 1 - \frac{it\xi_j}{\sqrt{n}} + \frac{\xi_j^2 t^2}{2n}\right|\right] \\
&\quad + C_T \left|\left(\exp\left\{\frac{-\sigma^2 t^2}{2n}\right\} - 1 + \frac{\sigma^2 t^2}{2n}\right)\right|,
\end{aligned}
$$

with $C_T = \exp\left\{\frac{\sigma^2 T^2}{2}\right\}$.

This estimate is independent of j by stationarity.

We now take the following Taylor expansions up to the term linear in $\frac{1}{n}$:

$$
\begin{aligned}
\exp\left\{\frac{it\xi_j}{\sqrt{n}}\right\} &= 1 + \frac{it\xi_j}{\sqrt{n}} - \frac{\xi_j^2 t^2}{2n} + o\left(\frac{1}{n}\right) \\
\exp\left\{\frac{-\sigma^2 t^2}{2n}\right\} &= 1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{1}{n}\right).
\end{aligned}
$$

Substituting these into our equation above gives that for any $j \in \{1, 2, \ldots, n\}$ and $|t| \leq T$,

$$|\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)| = o\left(\frac{1}{n}\right),$$

so that

$$\sup_{\substack{|t|\leq T \\ 1\leq j\leq n}} |\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)| = o\left(\frac{1}{n}\right)$$

and

$$\sup_{|t|\leq T} \sum_{j=1}^{n} |\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)| = n \cdot o\left(\frac{1}{n}\right).$$

As this is true for arbitrary T, we have that for any $t \in \mathbb{R}$,

$$\sum_{j=1}^{n} |\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)| = n \cdot o\left(\frac{1}{n}\right) \xrightarrow{n\to\infty} 0. \tag{3.1.3}$$

It will now be enough to estimate $\left|\sum_{j=1}^{n} \theta(n,j,t)\right|$, since

$$\begin{aligned} 0 \leq \Delta(n,t) &= \left|\sum_{j=1}^{n} \left[\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)\right] + \sum_{j=1}^{n} \theta(n,j,t)\right| \\ &\leq \sum_{j=1}^{n} \left|\psi(n,j,t) - \psi(n,j-1,t) - \theta(n,j,t)\right| + \left|\sum_{j=1}^{n} \theta(n,j,t)\right|. \end{aligned} \tag{3.1.4}$$

In order to do this, we fix some large $k \in \mathbb{Z}$ and divide the set $\{1, 2, \dots, n\}$ into blocks of size k, with possibly an incomplete block at the end.
Let $j \in \mathbb{Z}$ and define $r(j) \geq 0$ to be the integer such that $kr(j)+1 \leq j \leq k(r(j)+1)$. We see that r indexes the block of size k in which we find any given j.
Define

$$\theta_k(n,j,t) := \exp\left\{\frac{\sigma^2 t^2 kr(j)}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_{kr(j)}}{\sqrt{n}}\right\} \frac{\left(\sigma^2 - \xi_j^2\right) t^2}{2n}\right].$$

Our second estimate will be on the $\theta_k(n,j,t)$. We will then show that the $\theta_k(n,j,t)$ approximate $\theta(n,j,t)$ sufficiently well.
Fix $r_1 \leq \frac{n}{k}$. Then, for any j such that $kr_1+1 \leq j \leq k(r_1+1)$, we have $r(j) = r_1$. Therefore

$$\begin{aligned} \left|\sum_{j=kr_1+1}^{k(r_1+1)} \theta_k(n,j,t)\right| &= \left|\sum_{j=kr_1+1}^{k(r_1+1)} \exp\left\{\frac{\sigma^2 t^2 kr_1}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_{kr_1}}{\sqrt{n}}\right\} \frac{\left(\sigma^2 - \xi_j^2\right) t^2}{2n}\right]\right| \\ &\leq C(t)\frac{1}{n}\mathbb{E}\left[\left|\sum_{j=kr_1+1}^{k(r_1+1)} \exp\left\{\frac{itS_{kr_1}}{\sqrt{n}}\right\} \left(\sigma^2 - \xi_j^2\right)\right|\right] \\ &= C(t)\frac{1}{n}\mathbb{E}\left[\left|\exp\left\{\frac{itS_{kr_1}}{\sqrt{n}}\right\}\right| \left|\sum_{j=kr_1+1}^{k(r_1+1)} \left(\sigma^2 - \xi_j^2\right)\right|\right] \\ &= C(t)\frac{1}{n}\mathbb{E}\left[\left|\sum_{j=kr_1+1}^{k(r_1+1)} \left(\sigma^2 - \xi_j^2\right)\right|\right], \end{aligned}$$

where $C(t) = \dfrac{t^2}{2} \exp\left\{\dfrac{\sigma^2 t^2}{2}\right\}$.

Set $\delta(k) := \dfrac{\mathbb{E}\left[\left|\sum_{j=kr_1+1}^{k(r_1+1)} \left(\sigma^2 - \xi_j^2\right)\right|\right]}{k}$, which is independent of r_1 by stationarity, so that

$$\left|\sum_{j=kr_1+1}^{k(r_1+1)} \theta_k(n,j,t)\right| \leq C(t)\frac{k}{n}\delta(k).$$

We can use the $\mathcal{L}^1$ ergodic theorem from Theorem 2.2.2 to show that $\delta(k) \xrightarrow{k\to\infty} 0$ as follows. Since $\{\xi_j\}$ is an ergodic process and $\mathbb{E}\left(\xi_j^2\right) = \sigma^2$, we have, setting $r_1 = 0$,

$$\begin{aligned}
\delta(k) = \frac{\mathbb{E}\left[\left|\sum_{j=1}^{k}\left(\sigma^2 - \xi_j^2\right)\right|\right]}{k} &\xrightarrow{k\to\infty} \mathbb{E}\left(\sigma^2 - \xi_j^2 | \mathcal{I}\right) \\
&= \mathbb{E}\left(\sigma^2 - \xi_j^2\right), \qquad \text{by Remark 2.3.1} \\
&= 0.
\end{aligned}$$

Therefore

$$\begin{aligned}
\left|\sum_{j=1}^{n} \theta_k(n,j,t)\right| &\leq \sum_{r=0}^{\lfloor \frac{n}{k} \rfloor} \left|\sum_{j=kr+1}^{k(r+1)} \theta_k(n,j,t)\right| \\
&\leq \frac{n}{k} C(t) \frac{k}{n} \delta(k) \\
&= C(t)\delta(k) \xrightarrow{k\to\infty} 0.
\end{aligned} \tag{3.1.5}$$

Next, for our final estimate, we consider

$$\begin{aligned}
&\sum_{j=1}^{n} |\theta_k(n,j,t) - \theta(n,j,t)| \leq n \sup_{1\leq j\leq n} |\theta_k(n,j,t) - \theta(n,j,t)| \\
&= \frac{t^2}{2} \sup_{1\leq j\leq n} \left| \exp\left\{\frac{\sigma^2 t^2 kr(j)}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_{kr(j)}}{\sqrt{n}}\right\}\left(\sigma^2 - \xi_j^2\right)\right] \right. \\
&\qquad\qquad \left. - \exp\left\{\frac{\sigma^2 t^2 j}{2n}\right\} \mathbb{E}\left[\exp\left\{\frac{itS_{j-1}}{\sqrt{n}}\right\}\left(\sigma^2 - \xi_j^2\right)\right] \right| \\
&= \frac{t^2}{2} \sup_{1\leq j\leq n} \left| \exp\left\{\frac{\sigma^2 t^2 kr(j)}{2n}\right\} \right. \\
&\qquad \cdot \mathbb{E}\left[\left(\sigma^2 - \xi_j^2\right) \exp\left\{\frac{itS_{kr(j)}}{\sqrt{n}}\right\} \right. \\
&\qquad\quad \left.\left. \cdot \left(1 - \exp\left\{\frac{\sigma^2 t^2 \left(j - kr(j)\right)}{2n}\right\} \exp\left\{\frac{it\left(S_{j-1} - S_{kr(j)}\right)}{\sqrt{n}}\right\}\right)\right]\right|.
\end{aligned}$$

Then, by the triangle inequality, we can estimate

$$
\begin{aligned}
&\sum_{j=1}^{n} |\theta_k(n,j,t) - \theta(n,j,t)| \\
&\le \frac{t^2}{2} \sup_{1\le j\le n} \Bigg\{ \exp\left\{ \frac{\sigma^2 t^2 kr(j)}{2n} \right\} \\
&\qquad \cdot \mathbb{E}\left[\left|\sigma^2 - \xi_j^2\right| \left| 1 - \exp\left\{ \frac{\sigma^2 t^2 \left(j - kr(j)\right)}{2n} \right\} \exp\left\{ \frac{it\left(S_{j-1} - S_{kr(j)}\right)}{\sqrt{n}} \right\} \right| \right] \Bigg\} \\
&\le \frac{t^2}{2} \exp\left\{ \frac{\sigma^2 t^2}{2} \right\} \sup_{1\le j\le n} \mathbb{E}\Bigg[\left|\sigma^2 - \xi_j^2\right| \\
&\qquad\qquad \cdot \left| 1 - \exp\left\{ \frac{\sigma^2 t^2 \left(j - kr(j)\right)}{2n} \right\} \exp\left\{ \frac{it\left(S_{j-1} - S_{kr(j)}\right)}{\sqrt{n}} \right\} \right| \Bigg].
\end{aligned}
\tag{3.1.6}
$$

We now claim that the expectation

$$
\mathbb{E}\left[\left|\sigma^2 - \xi_j^2\right| \left| 1 - \exp\left\{ \frac{\sigma^2 t^2 \left(j - kr(j)\right)}{2n} \right\} \exp\left\{ \frac{it\left(S_{j-1} - S_{kr(j)}\right)}{\sqrt{n}} \right\} \right| \right]
$$

is periodic in j with period k.
Let $m \in \mathbb{Z}$. First note that $r(j+mk) = r(j) + m$, by our definition of r, so that

$$
\begin{aligned}
S_{j+mk-1} - S_{kr(j+mk)} &= S_{j+mk-1} - S_{kr(j)+mk} \\
&= \xi_{kr(j)+mk+1} + \xi_{kr(j)+mk+2} + \cdots + \xi_{j+mk-1}.
\end{aligned}
\tag{3.1.7}
$$

and

$$
j + mk - kr(j+mk) = j + mk - kr(j) - mk = j - kr(j). \tag{3.1.8}
$$

Using (3.1.7), we write

$$
\begin{aligned}
&\mathbb{E}\Bigg[\left|\sigma^2 - \xi_{j+km}^2\right| \Bigg| 1 - \exp\left\{ \frac{\sigma^2 t^2 \left([j+km] - kr(j+km)\right)}{2n} \right\} \\
&\qquad\qquad \cdot \exp\left\{ \frac{it\left(S_{j+km-1} - S_{kr(j+km)}\right)}{\sqrt{n}} \right\} \Bigg| \Bigg] \\
&= \mathbb{E}\Bigg[\left|\sigma^2 - \xi_{j+km}^2\right| \Bigg| 1 - \exp\left\{ \frac{\sigma^2 t^2 \left([j+km] - kr(j+km)\right)}{2n} \right\} \\
&\qquad\qquad \cdot \exp\left\{ \frac{it\left(\xi_{kr(j)+mk+1} + \cdots + \xi_{j+mk-1}\right)}{\sqrt{n}} \right\} \Bigg| \Bigg].
\end{aligned}
$$

Then, by (3.1.8), we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\sigma^2-\xi_{j+km}^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2\left([j+km]-kr(j+km)\right)}{2n}\right\}\right.\right.\\
&\qquad\qquad\left.\left.\cdot\exp\left\{\frac{it\left(S_{j+km-1}-S_{kr(j+km)}\right)}{\sqrt{n}}\right\}\right|\right]\\
&=\mathbb{E}\left[\left|\sigma^2-\xi_{j+km}^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2\left(j-kr(j)\right)}{2n}\right\}\right.\right.\\
&\qquad\qquad\left.\left.\cdot\exp\left\{\frac{it\left(\xi_{kr(j)+mk+1}+\cdots+\xi_{j+mk-1}\right)}{\sqrt{n}}\right\}\right|\right]\\
&=\mathbb{E}\left[\left|\sigma^2-\xi_{j}^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2\left(j-kr(j)\right)}{2n}\right\}\right.\right.\\
&\qquad\qquad\left.\left.\cdot\exp\left\{\frac{it\left(\xi_{kr(j)+1}+\xi_{kr(j)+2}+\cdots+\xi_{j-1}\right)}{\sqrt{n}}\right\}\right|\right],
\end{aligned}$$

where in the last line we use stationarity to shift all indices of ξ by $-mk$.
Thus we have periodicity, and so we need only consider $j=1,2,\ldots,k$.
But, for $1\le j\le k$, we have that $r(j)=0$, by definition of r.
Therefore

$$\begin{aligned}
&\sup_{1\le j\le n}\mathbb{E}\left[\left|\sigma^2-\xi_j^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2\left(j-kr(j)\right)}{2n}\right\}\exp\left\{\frac{it\left(S_{j-1}-S_{kr(j)}\right)}{\sqrt{n}}\right\}\right|\right]\\
&=\sup_{1\le l\le k}\mathbb{E}\left[\left|\sigma^2-\xi_l^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2\left(l-kr(l)\right)}{2n}\right\}\exp\left\{\frac{it\left(S_{l-1}-S_{kr(l)}\right)}{\sqrt{n}}\right\}\right|\right]\\
&=\sup_{1\le l\le k}\mathbb{E}\left[\left|\sigma^2-\xi_l^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2l}{2n}\right\}\exp\left\{\frac{it\left(S_{l-1}\right)}{\sqrt{n}}\right\}\right|\right].
\end{aligned}$$

Substituting this back into (3.1.6), we obtain

$$\begin{aligned}
&\sum_{j=1}^{n}\left|\theta_k(n,j,t)-\theta(n,j,t)\right|\le n\sup_{1\le j\le n}\left|\theta_k(n,j,t)-\theta(n,j,t)\right|\\
&\le\frac{t^2}{2}\exp\left\{\frac{\sigma^2t^2}{2}\right\}\sup_{1\le l\le k}\mathbb{E}\left[\left|\sigma^2-\xi_l^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2l}{2n}\right\}\exp\left\{\frac{itS_{l-1}}{\sqrt{n}}\right\}\right|\right]\\
&=C(t)\sup_{1\le l\le k}\mathbb{E}\left(\left|\sigma^2-\xi_l^2\right|\left|1-\exp\left\{\frac{\sigma^2t^2l}{2n}\right\}\exp\left\{\frac{itS_{l-1}}{\sqrt{n}}\right\}\right|\right)\\
&\xrightarrow{n\to\infty}0\qquad\text{by bounded convergence (Theorem 1.1.5).}
\end{aligned}$$

We now have

$$0 \leq \left| \sum_{j=1}^{n} \theta(n,j,t) \right| = \left| \sum_{j=1}^{n} [\theta(n,j,t) - \theta_k(n,j,t)] + \sum_{j=1}^{n} \theta_k(n,j,t) \right|$$
$$\leq \sum_{j=1}^{n} |\theta(n,j,t) - \theta_k(n,j,t)| + \left| \sum_{j=1}^{n} \theta_k(n,j,t) \right|.$$

Fix k and let $n \to \infty$ to get

$$0 \leq \limsup_{n\to\infty} \left| \sum_{j=1}^{n} \theta(n,j,t) \right| \leq \limsup_{n\to\infty} \left| \sum_{j=1}^{n} \theta_k(n,j,t) \right|.$$

Now, letting $k \to \infty$ and recalling (3.1.5), we see that the right-hand side of the above inequality tends to 0 and so

$$\lim_{n\to\infty} \left| \sum_{j=1}^{n} \theta(n,j,t) \right| = 0. \tag{3.1.9}$$

Hence, putting together (3.1.3), (3.1.4) and (3.1.9), we have that

$$\Delta(n,t) \xrightarrow{n\to\infty} 0,$$

as required. □

We have proved a central limit theorem for square-integrable ergodic martingale differences. Next we will show that this result can be extended to a larger class of processes, and in particular to certain stationary ergodic Markov chains.

3.2 The CLT for Markov Chains I

We continue to work in the same setting as in the previous section. Thus we let $(X, \mathcal{B})$ be a measurable space and let $(\Omega, \mathcal{F})$ be the space of sequences which take values in X with the product σ-field.
In this section our aim is to prove a central limit theorem for functions of ergodic Markov chains, under some conditions, following the method used by Varadhan in [15]. We start by consdering the following processes.
Let $\{X_n\}$ be a stationary zero-mean process, adapted to a filtration $(\mathcal{F}_n)$, whose path is described by the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. If we can write $X_n = \xi_{n+1} + \eta_{n+1}$, where $\{\xi_n\}$ is a square-integrable ergodic martingale difference and $\{\eta_n\}$ is negligible in some sense, then we can show that we have a central limit theorem for $\{X_n\}$. We formalise this in the following theorem, which is a key step in proving a central limit theorem for Markov chains. This result is stated but not proved in [15]. Here we provide a detailed proof.

Theorem 3.2.1. *Let $\{X_n\}$ be a stationary process such that, for any n, $\mathbb{E}(X_n) = 0$ and $X_n = \xi_{n+1} + \eta_{n+1}$, where $\{\xi_n\}$ is an ergodic stationary sequence of square-integrable martingale differences and $\{\eta_n\}$ satisfies*

$$\mathbb{E}\left[\left(\sum_{j=1}^{n} \eta_j\right)^2\right] = o(n). \tag{3.2.1}$$

Then

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}(0, \sigma^2)$, for some $\sigma > 0$.

Proof. Fix $a > 0$. We need to show that

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \leq a\right) \xrightarrow{n \to \infty} \Phi(a),$$

where Φ is the distribution function of a $\mathcal{N}(0, \sigma^2)$ random variable for some $\sigma > 0$.
Now fix $\varepsilon > 0$ and $n \in \mathbb{N}$. We claim that

$$\begin{aligned}\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \leq a\right) \leq \mathbb{P}&\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a + \varepsilon\right) \\ &- \mathbb{P}\left(\frac{\sum_{i+1}^{n} \xi_{i+1}}{\sqrt{n}} < a + \varepsilon; \frac{\left|\sum_{i=1}^{n} \eta_{i+1}\right|}{\sqrt{n}} \geq \varepsilon\right) \\ &+ \mathbb{P}\left(\frac{\left|\sum_{i=1}^{n} \eta_{i+1}\right|}{\sqrt{n}} \geq \varepsilon\right)\end{aligned} \tag{3.2.2}$$

and

$$\begin{aligned}\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \geq a\right) \geq \mathbb{P}&\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a - \varepsilon\right) \\ &- \mathbb{P}\left(\frac{\left|\sum_{i=1}^{n} \eta_{i+1}\right|}{\sqrt{n}} \geq \varepsilon\right) \\ &+ \mathbb{P}\left(\frac{\sum_{I=1}^{n} \xi_{i+1}}{\sqrt{n}} + \frac{\sum_{I=1}^{n} \eta_{i+1}}{\sqrt{n}}; \frac{\left|\sum_{i=1}^{n} \eta_{i+1}\right|}{\sqrt{n}} \geq \varepsilon\right).\end{aligned} \tag{3.2.3}$$

We prove the upper bound (3.2.2) using a lemma which we now formulate.

Lemma 3.2.1. *For any $a, \varepsilon > 0$ and any random variables A, B, we have*

$$\mathbb{P}(A + B < a) \leq \mathbb{P}(A < a + \varepsilon) - \mathbb{P}(A < a + \varepsilon; |B| \geq \varepsilon) + \mathbb{P}(|B| \geq \varepsilon).$$

Proof of lemma. By the law of total probability,

$$\mathbb{P}(A + B < a) = \mathbb{P}(A + B < a; |B| < \varepsilon) + \mathbb{P}(A + B < a; |B| \geq \varepsilon). \tag{3.2.4}$$

Note that $|B| < \varepsilon \Rightarrow -B < \varepsilon$. So for the first term in the above equation we have

$$\begin{aligned}\mathbb{P}\left(A+B<a;|B|<\varepsilon\right) &= \mathbb{P}\left(A<a-B;|B|<\varepsilon\right)\\ &\leq \mathbb{P}\left(A<a+\varepsilon:|B|<\varepsilon\right)\\ &= \mathbb{P}\left(A<a+\varepsilon\right)-\mathbb{P}\left(A<a+\varepsilon;|B|\geq\varepsilon\right).\end{aligned}$$

We can bound the second term in equation (3.2.4) by

$$\mathbb{P}\left(A+B<a;|B|\geq\varepsilon\right)\leq\mathbb{P}\left(|B|\geq\varepsilon\right),$$

and so we have proved the lemma. $\square$

Note that

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\leq a\right)=\mathbb{P}\left(\frac{\sum_{i=1}^n \xi_{i+1}}{\sqrt{n}}+\frac{\sum_{i=1}^n \eta_{i+1}}{\sqrt{n}}\leq a\right)$$

The upper bound claimed in (3.2.2) now follows from the above lemma by substituting

$$A=\frac{\sum_{i=1}^n \xi_{i+1}}{\sqrt{n}} \quad \text{and} \tag{3.2.5a}$$

$$B=\frac{\sum_{i=1}^n \eta_{i+1}}{\sqrt{n}}. \tag{3.2.5b}$$

Now let us define an error term

$$\alpha(n):=-\mathbb{P}\left(\frac{\sum_{i+1}^n \xi_{i+1}}{\sqrt{n}}<a+\varepsilon;\frac{\left|\sum_{i=1}^n \eta_{i+1}\right|}{\sqrt{n}}\geq\varepsilon\right)+\mathbb{P}\left(\frac{\left|\sum_{i=1}^n \eta_{i+1}\right|}{\sqrt{n}}\geq\varepsilon\right),$$

so that the upper bound (3.2.2) implies that

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\leq a\right)\leq\mathbb{P}\left(\frac{\sum_{i=1}^n \xi_{i+1}}{\sqrt{n}}<a+\varepsilon\right)+\alpha(n).$$

By Chebyshev's inequality (1.1.4) and our assumption (3.2.1), we can see that

$$0\leq\mathbb{P}\left(\frac{\left|\sum_{i=1}^n \eta_{i+1}\right|}{\sqrt{n}}\geq\varepsilon\right)\leq\frac{\mathbb{E}\left(\left[\sum_{i=1}^n \eta_{i+1}\right]^2\right)}{\varepsilon n}\xrightarrow{n\to\infty}0,$$

and similarly,

$$0\leq\mathbb{P}\left(\frac{\sum_{i+1}^n \xi_{i+1}}{\sqrt{n}}<a+\varepsilon;\frac{\left|\sum_{i=1}^n \eta_{i+1}\right|}{\sqrt{n}}\geq\varepsilon\right)\leq\mathbb{P}\left(\frac{\left|\sum_{i=1}^n \eta_{i+1}\right|}{\sqrt{n}}\geq\varepsilon\right)\xrightarrow{n\to\infty}0.$$

Hence $\alpha(n)\xrightarrow{n\to\infty}0$.
We now turn our attention to the second claim (3.2.3) and formulate another lemma.

Lemma 3.2.2. *For any $a, \varepsilon > 0$ and any random variables A, B, we have*

$$\mathbb{P}(A + B < a) \geq \mathbb{P}(A < a - \varepsilon) - \mathbb{P}(|B| \geq \varepsilon) + \mathbb{P}(A + B < a; |B| \geq \varepsilon).$$

Proof of lemma. We note that $|B| < \varepsilon \Rightarrow -B > -\varepsilon$. Therefore

$$\begin{aligned}\mathbb{P}(A + B < a; |B| < \varepsilon) &\geq \mathbb{P}(A < a - \varepsilon; |B| < \varepsilon)\\ &= \mathbb{P}(A < a - \varepsilon) - \mathbb{P}(A < a - \varepsilon; |B| \geq \varepsilon)\\ &\geq \mathbb{P}(A < a - \varepsilon) - \mathbb{P}(|B| \geq \varepsilon).\end{aligned}$$

So, by the law of total probability,

$$\begin{aligned}\mathbb{P}(A + B < a) &= \mathbb{P}(A + B < a; |B| < \varepsilon) + \mathbb{P}(A + B < a; |B| \geq \varepsilon)\\ &\geq \mathbb{P}(A < a - \varepsilon) - \mathbb{P}(|B| \geq \varepsilon) + \mathbb{P}(A + B < a; |B| \geq \varepsilon),\end{aligned}$$

and the lemma is proved. ☐

We now make the same substitution as in (3.2.5) and the claimed lower bound (3.2.3) follows from the lemma that we have just proved.
Next we define another error term

$$\begin{aligned}\beta(n) := &-\mathbb{P}\left(\left|\frac{\sum_{i=1}^{n} \eta_{i+1}}{\sqrt{n}}\right| \geq \varepsilon\right)\\ &+ \mathbb{P}\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} + \frac{\sum_{i=1}^{n} \eta_{i+1}}{\sqrt{n}} < a; \left|\frac{\sum_{i=1}^{n} \eta_{i+1}}{\sqrt{n}}\right| \geq \varepsilon\right),\end{aligned}$$

so that the lower bound (3.2.3) implies

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \geq a\right) \geq \mathbb{P}\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a - \varepsilon\right) + \beta(n).$$

In the same way as for $\alpha(n)$, we can see that $\beta(n) \xrightarrow{n\to\infty} 0$.
At this point we have shown that

$$\begin{aligned}\mathbb{P}\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a - \varepsilon\right) + \beta(n) &\leq \mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \leq a\right)\\ &\leq \mathbb{P}\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a + \varepsilon\right) + \alpha(n),\end{aligned}$$

with $\alpha(n) \xrightarrow{n\to\infty} 0$, $\beta(n) \xrightarrow{n\to\infty} 0$. Since $\{\xi_n\}$ is an ergodic stationary sequence of square-integrable martingale differences, we have a central limit theorem for $\{\xi_n\}$ by Theorem 3.1.1. Therefore $\exists \sigma > 0$ such that

$$\begin{aligned}&\mathbb{P}\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a - \varepsilon\right) \xrightarrow{n\to\infty} \Phi(a - \varepsilon) \text{ and}\\ &\mathbb{P}\left(\frac{\sum_{i=1}^{n} \xi_{i+1}}{\sqrt{n}} < a + \varepsilon\right) \xrightarrow{n\to\infty} \Phi(a + \varepsilon),\end{aligned}$$

where Φ is the distribution function of a random variable with distribution $\mathcal{N}\left(0, \sigma^2\right)$.
Thus, by the sandwich rule,

$$\Phi(a-\varepsilon) \leq \liminf_{n\to\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \leq a\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \leq a\right) \leq \Phi(a+\varepsilon).$$

Finally, we let $\varepsilon \to 0$ to see that the required limit exists and

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \leq a\right) = \Phi(a).$$

□

The following remark from [15] helps us to find processes which satisfy the conditions of the above theorem.

Remark 3.2.1. Suppose that $\{Z_n\}$ is a stationary square-integrable sequence. Then if we define $\eta_n := Z_n - Z_{n+1}$ for each n, we have that $\mathbb{E}[(\sum_{j=1}^n \eta_j)^2] = o(n)$.

We provide a short proof of this fact.

Proof. Since $\{Z_n\}$ is square-integrable, the expectation we are interested in is well defined. To prove the remark, we use Minkowski's inequality (1.1.3) and stationarity.

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{j=1}^n \eta_j\right)^2\right] &= \mathbb{E}\left[(Z_1 - Z_{n+1})^2\right] \\
&= \|Z_1 - Z_{n+1}\|_2^2 \\
&\leq \left(\|Z_1\|_2^2 + \|Z_{n+1}\|_2^2\right)^2 \text{ by Minkowski's inequality (1.1.3)} \\
&= \left(2\|Z_1\|_2^2\right)^2 \text{ by stationarity} \\
&= o(n).
\end{aligned}$$

□

Given a zero-mean stationary ergodic process $\{X_n\}$, we can, under some conditions, construct $\{Z_n\}$ as in the remark above so that $X_n + (Z_{n+1} - Z_n)$ is a square-integrable ergodic martingale difference.
We would then have $X_n = \xi_{n+1} + \eta_{n+1}$ satisfying the conditions in Theorem 3.2.1 for a central limit theorem to hold.
Define

$$Z_n := \sum_{j=0}^{\infty} \mathbb{E}\left(X_{n+j} | \mathcal{F}_n\right), \text{ whenever this sum converges.}$$

We will now show that if Z_n exists and is square-integrable for each n, then $X_n + (Z_{n+1} - Z_n)$ is a square-integrable ergodic martingale difference and, by Theorem 3.2.1 and the previous remark, we have a central limit theorem for $\{X_n\}$, as is claimed in [15].

Lemma 3.2.3. *Suppose that*

$$Z_n := \sum_{j=0}^{\infty} \mathbb{E}\left(X_{n+j}|\mathcal{F}_n\right)$$

exists and is square-integrable for each n.
Then we have a central limit theorem for $\{X_n\}$.

Proof. We have that

$$\begin{aligned}
\mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left(\sum_{j=0}^{\infty} \mathbb{E}\left[X_{n+1+j}|\mathcal{F}_{n+1}\right]|\mathcal{F}_n\right) \\
&= \sum_{j=0}^{\infty} \mathbb{E}\left[X_{n+1+j}|\mathcal{F}_n\right] \quad \text{by the tower rule (1.3.1)} \\
&= \sum_{j=0}^{\infty} \mathbb{E}\left[X_{n+j}|\mathcal{F}_n\right] - \mathbb{E}\left[X_n|\mathcal{F}_n\right] \\
&= Z_n - X_n.
\end{aligned}$$

Therefore

$$\begin{aligned}
X_n &= Z_n - \mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] \\
&= \left(Z_n - Z_{n+1}\right) + \left(Z_{n+1} - \mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right]\right).
\end{aligned}$$

If we define

$$\begin{aligned}
\eta_{n+1} &:= Z_n - Z_{n+1} \text{ and} \\
\xi_{n+1} &:= Z_{n+1} - \mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right],
\end{aligned}$$

then we have

$$X_n = \eta_{n+1} + \xi_{n+1}.$$

By Remark 3.2.1, $\mathbb{E}[(\sum_{j=1}^{n} \eta_j)^2] = o(n)$. It is easy to see that $\{\xi_n\}$ is ergodic and square-integrable, so we only show that $\{\xi_n\}$ is a martingale difference. By Lemma 1.3.2, we just need to show that $\mathbb{E}\left[\xi_{n+1}|\mathcal{F}_n\right] = 0$. This follows immediately from our definition of ξ_n:

$$\begin{aligned}
\mathbb{E}\left[\xi_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[Z_{n+1} - \mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right]|\mathcal{F}_n\right] \\
&= \mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] - \mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] = 0.
\end{aligned}$$

Therefore, by Theorem 3.2.1, we have a central limit theorem for $\{X_n\}$. $\square$

We now see under what conditions we can apply this lemma to the specific case of a function of an ergodic Markov chain.
Let $\{X_n\}$ be a stationary ergodic Markov chain, adapted to a filtration $(\mathcal{F}_n)$, with state space X, transition probability Π and invariant measure μ. Let $\mathbb{P}$ be

the stationary ergodic probability measure on $(\Omega, \mathcal{F})$ which describes the path of this Markov process.
Let f be a square-integrable function with mean zero with respect to the invariant measure. We will show that, under further conditions, we have a central limit theorem for $f(X_n)$. We use the method outlined above, following [15].

Theorem 3.2.2. *Let $f \in \mathcal{L}^2(X, \mathcal{B}, \mu)$ be such that $\sum_{x \in X} f(x)\mu(x) = 0$. Suppose that $\exists U \in \mathcal{L}^2$ such that $[I - \Pi]\, U = f$. Then*

$$\frac{\sum_{j=1}^{n} f(X_j)}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma^2\right)$, with variance

$$\sigma^2 = \mathbb{E}^{\mathbb{P}_\mu}\left[\left(U(X_1) - U(X_0) + f(X_0)\right)^2\right].$$

Proof. We want to show that $Z_n := \displaystyle\sum_{j=0}^{\infty} \mathbb{E}\left(f(X_{n+j})|\mathcal{F}_n\right)$ is well-defined and square-integrable for all $n \in \mathbb{N} \cup \{0\}$.
If Z_0 can be defined, then we can define Z_n using the shift operator T, via $Z_n(\omega) = Z_0(T^n\omega)$. Thus we only need to show that Z_0 is well-defined and square-integrable.
For each $n = 0, 1, 2, \ldots$, define $Q_n f$ by

$$(Q_n f)(x) := \mathbb{E}\left(f(X_n)|X_0 = x\right) \quad \text{for all } x \in X.$$

Then

$$(Q_n f)(X_0) = \mathbb{E}\left(f(X_n)|\mathcal{F}_0\right).$$

We claim that $Q_n(Q_m)f = Q_{m+n}f$, for any $n, m \in \mathbb{N} \cup \{0\}$.
To prove this claim, let $m, n \in \mathbb{N}$ and let $\hat{X}_k$ be a random variable with the same distribution as X_k for each $k \in \mathbb{N} \cup \{0\}$. Then

$$\begin{aligned}(Q_n(Q_m)f)(X_0) &= \mathbb{E}\left(Q_m f(X_n)|\mathcal{F}_0\right) \\ &= \mathbb{E}\left(\mathbb{E}\left[f(\hat{X}_m)|\hat{X}_0 = X_n\right]|\mathcal{F}_0\right).\end{aligned}$$

Note that starting a Markov chain from state X_n and letting it run for time m is equivalent to considering a Markov chain which is in state X_n at time n and letting it run up to time $m + n$. Thus

$$\mathbb{E}\left[f(\hat{X}_m)|\hat{X}_0 = X_n\right] = \mathbb{E}\left[f(X_{m+n})|\mathcal{F}_n\right],$$

and so

$$\begin{aligned}(Q_n(Q_m)f)(X_0) &= \mathbb{E}\left(\mathbb{E}\left[f(X_{m+n})|\mathcal{F}_n\right]|\mathcal{F}_0\right) \\ &= \mathbb{E}\left[f(X_{m+n})|\mathcal{F}_0\right] \quad \text{by the tower property (1.3.1)} \\ &= Q_{m+n},\end{aligned}$$

as claimed.
Therefore $Q_n f = Q_{n-1}(Q_1 f) = Q_{n-2}(Q_1^2 f) = \cdots = Q_1^n f$.
Fix $x \in X$. Then

$$(Q_1 f)(x) = \mathbb{E}\left(f(X_1)|X_0 = x\right) = \sum_{y \in X} f(y)\Pi(x,y)dy = (\Pi f)(x).$$

So we have that $Q_1 = \Pi$ and thus, whenever the sum converges,

$$\sum_{j=0}^{\infty} \mathbb{E}\left(f(X_j)|\mathcal{F}_0\right) = \sum_{j=0}^{\infty} (Q_j f)(X_0) = \sum_{j=0}^{\infty} (Q_1^j f)(X_0) = \sum_{j=0}^{\infty} (\Pi^j f)(X_0).$$

Moreover, if the sum does converge,

$$\sum_{j=0}^{\infty} (\Pi^j f)(X_0) = \left([I - \Pi]^{-1} f\right)(X_0),$$

since

$$[I - \Pi] \sum_{j=0}^{\infty} \Pi^j f = \sum_{j=0}^{\infty} \Pi^j f - \sum_{j=0}^{\infty} \Pi^{j+1} f = \Pi^0 f = f.$$

As in the hypothesis of the theorem, suppose that $\exists U \in \mathcal{L}^2$ such that $[I - \Pi]\, U = f$. Then

$$Z_0 = \sum_{j=0}^{\infty} \mathbb{E}\left(f(X_j)|\mathcal{F}_0\right) = \left([I - \Pi]^{-1} f\right)(X_0) = U(X_0) \tag{3.2.8}$$

converges and we see that $Z_n \in \mathcal{L}^2$ with

$$Z_n = \sum_{j=0}^{\infty} \mathbb{E}\left(f(X_{n+j})|\mathcal{F}_0\right) = U(X_n), \qquad \forall n = 0, 1, 2, \dots .$$

We now appeal to the previous lemma which gives us that, under the assumptions of the theorem, we have a central limit theorem:

$$\frac{\sum_{j=1}^{n} f(X_j)}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma^2\right)$, for some $\sigma > 0$.
To complete the proof of the theorem, we now need to calculate the variance σ^2.
Using our notation from earlier, we have that $f(X_n) = \xi_{n+1} + \eta_{n+1}$, where

$$\xi_{n+1} = U(X_{n+1}) - U(X_n) + f(X_n) \text{ is a martingale difference,}$$

as shown in Lemma 3.2.3, and

$$\mathbb{E}\left[\left(\sum_{j=1}^{n} \eta_j\right)^2\right] = o(n). \tag{3.2.9}$$

Then we have

$$
\begin{aligned}
\operatorname{Var}\left(\frac{\sum_{j=1}^{n} f(X_j)}{\sqrt{n}}\right) &= \frac{1}{n}\operatorname{Var}\left(\sum_{j=1}^{n}\xi_{j+1} + \sum_{j=1}^{n}\eta_{j+1}\right) \\
&= \frac{1}{n}\left(\operatorname{Var}\left[\sum_{j=1}^{n}\xi_{j+1}\right] + \operatorname{Var}\left[\sum_{j=1}^{n}\eta_{j+1}\right]\right. \\
&\qquad \left. + 2\operatorname{Cov}\left[\sum_{i=1}^{n}\xi_{i+1}, \sum_{j=1}^{n}\eta_{j+1}\right]\right).
\end{aligned}
$$

Note that all of the random variables ξ_j, η_j, $j = 1, 2, \ldots$, have mean zero. Therefore

$$
\operatorname{Var}\left[\sum_{j=1}^{n}\eta_{j+1}\right] = \mathbb{E}\left[\Big(\sum_{j=1}^{n}\eta_{j+1}\Big)^2\right] = o(n) \text{ by (3.2.9)}
$$

and

$$
\begin{aligned}
\operatorname{Var}\left[\sum_{j=1}^{n}\xi_{j+1}\right] &= \mathbb{E}\left[\Big(\sum_{j=1}^{n}\xi_{j+1}\Big)^2\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[\xi_{j+1}^2\right],
\end{aligned}
$$

since for any $i < j$,

$$
\operatorname{Cov}\left[\xi_i, \xi_j\right] = \mathbb{E}\left[\xi_i\xi_j\right] = \mathbb{E}\left(\mathbb{E}\left[\xi_i\xi_j|\mathcal{F}_i\right]\right) = \mathbb{E}\left(\xi_i\mathbb{E}\left[\xi_j|\mathcal{F}_i\right]\right) = 0,
$$

as $\{\xi_k\}$ is a martingale difference, and similarly for $i > j$.
Then, by stationarity, we see that

$$
\sum_{j=1}^{n}\mathbb{E}\left[\xi_{j+1}^2\right] = n\mathbb{E}\left[\xi_0^2\right].
$$

Putting these together, we get

$$
\operatorname{Var}\left(\frac{\sum_{j=1}^{n} f(X_j)}{\sqrt{n}}\right) = \frac{n\mathbb{E}\left[\xi_0^2\right]}{n} + \frac{o(n)}{n} + \frac{2}{n}\operatorname{Cov}\left[\sum_{i=1}^{\infty}\xi_{i+1}, \sum_{j=1}^{\infty}\eta_{j+1}\right].
$$

By the Cauchy-Schwarz inequality for covariance (1.1.6), we have that

$$
\begin{aligned}
0 \le \frac{2}{n}\left|\operatorname{Cov}\left[\sum_{i=1}^{\infty}\xi_{i+1}, \sum_{j=1}^{\infty}\eta_{j+1}\right]\right| &\le \frac{2}{n}\sqrt{\operatorname{Var}\left[\sum_{j=1}^{\infty}\xi_{j+1}\right]}\sqrt{\operatorname{Var}\left[\sum_{j=1}^{\infty}\eta_{j+1}\right]} \\
&= \frac{2}{n}\sqrt{n\mathbb{E}\left[\xi_0^2\right]}\sqrt{o(n)} \\
&= 2\sqrt{\mathbb{E}\left[\xi_0^2\right]}\sqrt{\frac{o(n)}{n}} \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

Hence

$$\operatorname{Var}\left(\frac{\sum_{j=1}^{n} f(X_j)}{\sqrt{n}}\right) \xrightarrow{n\to\infty} \mathbb{E}\left[\xi_0^2\right]$$
$$= \mathbb{E}^{\mathbb{P}_\mu}\left[\left(U(X_1) - U(X_0) + f(X_0)\right)^2\right].$$

That is

$$\sigma^2 = \mathbb{E}^{\mathbb{P}_\mu}\left[\left(U(X_1) - U(X_0) + f(X_0)\right)^2\right].$$

□

3.3 The CLT for Markov Chains II

We now restrict ourselves to Markov chains on a finite state space and present an alternative approach to proving a central limit theorem for stationary ergodic Markov chains. This method allows us to arrive at the result more directly, but we get a less general result here. What follows is adapted from handwritten notes by Bálint Tóth [13].

Let X be a finite set and let $(\Omega, \mathcal{F})$ be the space of sequences which take values in X, with the product σ-field.
Let $\{X_n\}$ be a stationary, irreducible, aperiodic Markov chain on $(\Omega, \mathcal{F})$, adapted to a filtration $(\mathcal{F}_n)$, with transition probabilities given by Π and state space X. Because we are working on a finite state space, we can apply Theorem 2.5.4. This tells us that $\{X_n\}$ is ergodic, since it is a stationary, irreducible and aperiodic Markov chain on a finite state space.
Define $\mathbb{P}$ to be the stationary ergodic probability measure on $(\Omega, \mathcal{F})$ which describes the time-evolution of the process $\{X_n\}$.
Let $f \in \mathcal{L}^2(X, \mathcal{B}, \mu)$ and suppose that f has mean zero under the unique stationary distribution μ; i.e. $\sum_{x\in X} \mu(x) f(x) = 0$.
Define

$$S_n := \sum_{k=0}^{n-1} f(X_k).$$

The main result of this section is the following central limit theorem:

Theorem 3.3.1.
$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$
where $Z \sim \mathcal{N}(0, \sigma_1^2)$ *for some* $\sigma_1^2 > 0$.

To prove this, we are going to prove a central limit theorem for another quantity which approximates S_n, as in [13], and we will see that our desired result then follows.
Set $X_0 = x_0$ deterministically.
We are going to use the fact that the times between consecutive returns to x_0 and the behaviour of the chain in these time intervals are i.i.d.. We define

several random variables which we are going to work with.
Define the return times to x_0 inductively by

$$R_0 := 0$$
$$R_{n+1} := \min\{k > R_n : X_k = x_0\}.$$

Then R_n is the n^{th} return time to x_0 after time 0 for each $n \in \mathbb{N}$.
Next we define the length of the time interval between returns as

$$T_n := R_n - R_{n-1}, \quad \forall n \in \mathbb{N}$$

and the sum of the function of the Markov chain in each of these time intervals by

$$Y_n := \sum_{k=R_{n-1}}^{R_n - 1} f(X_k).$$

By the Markov property (1.2.1), we see that $(Y_n, T_n)_{n\in\mathbb{N}}$ are i.i.d..
We also have that all exponential moments of Y_n and T_n are finite, as we are in a finite state space.
Clearly $\mathbb{E}(Y_n) = 0$. Let us set $\sigma_0^2 = \mathbb{E}\left(Y_n^2\right)$ and $b = \mathbb{E}\left(T_n\right)$.
Let the total number of visits to x_0 (including the one at time 0) *before* time n be given by

$$\nu_n := \min\{k : R_k \geq n\} \ \forall n \in \{0, 1, 2, \dots\}.$$

Equivalently,

$$\nu_n = \min\left\{k : \sum_{l=1}^{k} T_l \geq n\right\}.$$

Define $U_n := R_{\nu_n} = \min\{m \geq n : X_m = x_0\}$, so that U_n is the time of the first return to x_0 after time n. Then

$$S_n = \sum_{k=0}^{U_n - 1} f(X_k) - \sum_{k=n}^{U_n - 1} f(X_k).$$

Remark 3.3.1. As stated in [13], we can show that $\sum_{k=n}^{U_n-1} f(X_k)$ is stochastically bounded, by a random variable independent of n. The idea of the proof is that this sum is bounded by the sum of $|f|$ taken between any two consecutive return times. We do not provide details of this here.

We also have that

$$S_{U_n} = \sum_{k=0}^{U_n - 1} f(X_k) = \sum_{k=1}^{\nu_n} Y_k = \tilde{S}_{\nu_n},$$

where we define

$$\tilde{S}_m := \sum_{k=1}^{m} Y_k.$$

The proof of the main result will follow immediately from the central limit theorem below. We next turn our attention to proving this theorem, which is the main focus of [13].

Theorem 3.3.2. *For* $\tilde{S}_m := \sum_{k=1}^{m} Y_k$,

$$\frac{\tilde{S}_{\nu_n}}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \frac{\sigma_0^2}{b}\right)$.

Proof. Define a function $\psi : \mathbb{R} \to \mathbb{R}$ by

$$\exp\left\{\psi(\lambda)\right\} := \mathbb{E}\left(\exp\left\{\lambda Y_1\right\}\right).$$

We claim that

$$\psi(\lambda) = \frac{1}{2}\sigma_0^2\lambda^2 + O(\lambda^3) \quad \text{as } \lambda \to 0.$$

Note first that, by the series representation of the exponential function,

$$\begin{aligned}\exp\left\{\frac{1}{2}\sigma_0^2\lambda^2 + O(\lambda^3)\right\} &= 1 + \left[\frac{1}{2}\sigma_0^2\lambda^2 + O(\lambda^3)\right] + \frac{1}{2!}\left[\frac{1}{2}\sigma_0^2\lambda^2 + O(\lambda^3)\right]^2 + \cdots \\ &= 1 + \frac{1}{2}\sigma_0^2\lambda^2 + O(\lambda^3), \text{ as } \lambda \to 0.\end{aligned}$$

Now

$$\begin{aligned}\exp\left\{\psi(\lambda)\right\} = \mathbb{E}\left(\exp\left\{\lambda Y_1\right\}\right) &= \mathbb{E}\left[1 + \lambda Y_1 + \frac{1}{2}\lambda^2 Y_1^2 + O(\lambda^3)\right] \\ &= 1 + 0 + \frac{1}{2}\lambda^2\mathbb{E}(Y_1^2) + O(\lambda^3) \\ &= 1 + \frac{1}{2}\lambda^2\sigma_0^2 + O(\lambda^3) \\ &= \exp\left\{\frac{1}{2}\sigma_0^2\lambda^2 + O(\lambda^3)\right\},\end{aligned}$$

and taking logarithms proves our claim.

Next we claim that $\exp\left\{\lambda\tilde{S}_k - \psi(\lambda)k\right\}$ is a martingale.

Since we have finite exponential moments, the following calculation is all that is needed to prove this:

$$\begin{aligned}&\mathbb{E}\left[\exp\left\{\lambda\tilde{S}_{k+1} - \psi(\lambda)\,(k+1)\right\}|\mathcal{F}_k\right] \\ &\qquad = \mathbb{E}\left[\exp\left\{\lambda\tilde{S}_k - \psi(\lambda)k\right\}\exp\left\{\lambda Y_{k+1} - \psi(\lambda)\right\}|\mathcal{F}_k\right] \\ &\qquad = \exp\left\{\lambda\tilde{S}_k - \psi(\lambda)k\right\}\mathbb{E}\left[\exp\left\{\lambda Y_{k+1}\right\}\right]\exp\left\{-\psi(\lambda)\right\} \\ &\qquad = \exp\left\{\lambda\tilde{S}_k - \psi(\lambda)k\right\}\exp\left\{\psi(\lambda)\right\}\exp\left\{-\psi(\lambda)\right\} \\ &\qquad = \exp\left\{\lambda\tilde{S}_k - \psi(\lambda)k\right\}.\end{aligned}$$

Note that $\mathbb{E}\left[\exp\left\{\lambda\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}\right\}\right]$ is the moment generating function of $\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}$ and $\exp\left\{\frac{\lambda^2}{2}\frac{\sigma_0^2}{b}\right\}$ is the moment generating function of a normal random variable with mean zero and variance $\frac{\sigma_0^2}{b}$.
We want to show that $\mathbb{E}\left[\exp\left\{\lambda\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}\right\}\right] \xrightarrow{n\to\infty} \exp\left\{\frac{\lambda^2}{2}\frac{\sigma_0^2}{b}\right\}$. Then Theorem 1.1.3 will give us the result of this theorem.

Let us fix n. Then ν_n is a stopping time and by the optional stopping theorem (Theorem 1.3.3), for any $\theta \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left\{\theta\tilde{S}_{\nu_n} - \psi(\theta)\nu_n\right\}\right] = \mathbb{E}\left[\exp\left\{\theta\tilde{S}_0 - 0\right\}\right] = e^0 = 1. \tag{3.3.2}$$

Therefore

$$\begin{aligned}
\mathbb{E}\left[\exp\left\{\lambda\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}\right\}\right] &= \mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\nu_n\right\}\exp\left\{\psi\left(\frac{\lambda}{\sqrt{n}}\right)\nu_n\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\nu_n\right\}\exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\}\right] \\
&\quad + \mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\nu_n\right\}\right. \\
&\qquad \left.\cdot\left(\exp\left\{n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\} - \exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\}\right)\right] \\
&= \exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\} + \mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\nu_n\right\}\right. \\
&\qquad \left.\cdot\left(\exp\left\{n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\} - \exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\}\right)\right],
\end{aligned}$$

where the last line follows from (3.3.2) with $\theta = \frac{\lambda}{\sqrt{n}}$.
Call

$$\begin{aligned}
E_n := \mathbb{E}&\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\nu_n\right\}\right. \\
&\left.\cdot\left(\exp\left\{n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\} - \exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\}\right)\right],
\end{aligned}$$

so that

$$\mathbb{E}\left[\exp\left\{\lambda\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}\right\}\right] = \exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\} + E_n.$$

We will show that E_n is an error term which tends to 0 as $n \to \infty$.

We can write

$$\begin{aligned}
E_n &= \mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\exp\left\{\left[\frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right) - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\right]\nu_n\right\}\cdot\right.\\
&\qquad\left.\cdot\left(\exp\left\{n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\} - \exp\left\{\frac{\sigma_0^2\lambda^2}{2b}\right\}\right)\right]\\
&= e^{\frac{\sigma_0^2\lambda^2}{2b}}\mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\right.\\
&\qquad\qquad\cdot\exp\left\{\left[\frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right) - \psi\left(\frac{\lambda}{\sqrt{n}}\right)\right]\nu_n\right\}\\
&\qquad\qquad\left.\cdot\left(\exp\left\{n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n} - \frac{\sigma_0^2\lambda^2}{2b}\right\} - 1\right)\right]\\
&= e^{\frac{\sigma_0^2\lambda^2}{2b}}\mathbb{E}\left[\exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\exp\left\{\frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\right.\\
&\qquad\qquad\left.\cdot\left(\exp\left\{-\frac{\sigma_0^2\lambda^2}{2b}\right\} - \exp\left\{-n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\}\right)\right].
\end{aligned}$$

We are going to use the Cauchy-Schwarz inequality (1.1.5) to bound this error. Note that for any random variables $X, Y, Z \in \mathcal{L}^2$, with $Y^2, Z^2 \in \mathcal{L}^2$,

$$\begin{aligned}
(\mathbb{E}[XYZ])^4 &\le \left(\mathbb{E}\left[X^2\right]\right)^2\left(\mathbb{E}\left[Y^2Z^2\right]\right)^2 \quad \text{by Cauchy-Schwarz}\\
&\le \left(\mathbb{E}\left[X^2\right]\right)^2\mathbb{E}\left[\left(Y^2\right)^2\right]\mathbb{E}\left[\left(Z^2\right)^2\right] \quad \text{applying Cauchy-Schwarz again}\\
&= \left(\mathbb{E}\left[X^2\right]\right)^2\mathbb{E}\left[Y^4\right]\mathbb{E}\left[Z^4\right].
\end{aligned} \tag{3.3.3}$$

Define the following random variables:

$$\begin{aligned}
E_{1,n} &:= \exp\left\{\frac{\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\\
E_{2,n} &:= \exp\left\{\frac{1}{2}\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\\
E_{3,n} &:= \exp\left\{-\frac{\sigma_0^2\lambda^2}{2b}\right\} - \exp\left\{-n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\}.
\end{aligned}$$

Then

$$E_n = e^{\frac{\sigma_0^2\lambda^2}{2b}}\mathbb{E}\left[E_{1,n}\cdot E_{2,n}\cdot E_{3,n}\right]$$

and it is easy to check that we can apply the inequality (3.3.3) to get

$$E_n^4 \le e^{\frac{2\sigma_0^2\lambda^2}{b}}\left(\mathbb{E}\left[E_{1,n}^2\right]\right)^2\mathbb{E}\left[E_{2,n}^4\right]\mathbb{E}\left[E_{3,n}^4\right].$$

We can now look at each term in the product individually. The constant $e^{\frac{\sigma_0^2\lambda^2}{2b}}$ plays no role here.

Using (3.3.2), we see that

$$\begin{aligned}\mathbb{E}\left[E_{1,n}^2\right] &= \mathbb{E}\left[\exp\left\{\frac{2\lambda}{\sqrt{n}}\tilde{S}_{\nu_n} - \psi\left(\frac{2\lambda}{\sqrt{n}}\right)\nu_n\right\}\right] \\ &= 1.\end{aligned}$$

In order to estimate the expectation of the other random variables in the error term, we claim that we have a strong law of large numbers for ν_n:

$$\frac{\nu_n}{n} \xrightarrow{n\to\infty} \frac{1}{b} \quad \text{a.s..} \tag{3.3.5}$$

We prove this claim as follows. Since the $\{T_j\}$ are i.i.d.,

$$\frac{1}{J}\sum_{j=1}^{J} T_j \xrightarrow{J\to\infty} \mathbb{E}(T_1) = b \quad \text{a.s.,}$$

by the strong law of large numbers for i.i.d. random variables.
Note that we have the following equivalence, for any $J \in \mathbb{N}$:

$$\frac{1}{J}\sum_{j=1}^{J} T_j \geq \frac{n}{J} \quad \Leftrightarrow \quad \frac{J}{n} \geq \frac{\nu_n}{n}.$$

Fix $\varepsilon > 0$ and define $J_n := \left\lfloor \frac{n}{b+\varepsilon} \right\rfloor$. Then $J_n \xrightarrow{n\to\infty} \infty$, so

$$\frac{1}{J_n}\sum_{j=1}^{J_n} T_j \xrightarrow{n\to\infty} b \qquad \text{a.s..}$$

But $\frac{n}{J_n} > b$, so the event $\left\{\frac{1}{J_n}\sum_{j=1}^{J_n} T_j \geq \frac{n}{J_n}\right\} = \left\{\frac{J_n}{n} \geq \frac{\nu_n}{n}\right\}$ occurs for only finitely many n almost surely.
Therefore $\exists N \in \mathbb{N}$ such that $\forall n > N$,

$$\frac{\nu_n}{n} > \frac{1}{n}\left\lfloor \frac{n}{b+\varepsilon} \right\rfloor \qquad \text{a.s..}$$

Hence

$$\liminf_{n\to\infty} \frac{\nu_n}{n} \geq \frac{1}{b+\varepsilon} \qquad \text{a.s..}$$

Now define $\tilde{J}_n := \left\lfloor \frac{n}{b-\varepsilon} \right\rfloor$. Again we have

$$\frac{1}{\tilde{J}_n}\sum_{j=1}^{\tilde{J}_n} T_j \xrightarrow{n\to\infty} b \qquad \text{a.s..}$$

But $\frac{n}{\tilde{J}_n} < b$, so the event $\left\{ \frac{1}{\tilde{J}} \sum_{j=1}^{\tilde{J}_n} T_j < \frac{n}{\tilde{J}_n} \right\} = \left\{ \frac{\tilde{J}_n}{n} < \frac{\nu_n}{n} \right\}$ occurs for only finitely many n almost surely.
Therefore $\exists M \in \mathbb{N}$ such that $\forall n > M$,

$$\frac{\nu_n}{n} < \frac{1}{n} \left\lfloor \frac{n}{b - \varepsilon} \right\rfloor \qquad \text{a.s..}$$

Hence

$$\limsup_{n\to\infty} \frac{\nu_n}{n} \le \frac{1}{b - \varepsilon} \qquad \text{a.s..}$$

We now have

$$\frac{1}{b+\varepsilon} \le \liminf_{n\to\infty} \frac{\nu_n}{n} \le \limsup_{n\to\infty} \frac{\nu_n}{n} \le \frac{1}{b-\varepsilon} \qquad \text{a.s.,}$$

so by taking the limit $\varepsilon \to 0$, we have that the desired limit exists and we have proved our claim that

$$\lim_{n\to\infty} \frac{\nu_n}{n} \xrightarrow{n\to\infty} \frac{1}{b} \qquad \text{a.s..}$$

We can now see that

$$\begin{aligned}
\mathbb{E}\left[E_{2,n}^4\right] &= \mathbb{E}\left[\exp\left\{2n\psi\left(\frac{2\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{2n\left(\frac{1}{2}\frac{\sigma_0^2(2\lambda)^2}{n} + O\left(\frac{8\lambda^3}{n^{\frac{3}{2}}}\right)\right)\frac{\nu_n}{n}\right\}\right] \\
&\xrightarrow{n\to\infty} e^{\frac{4\lambda^2\sigma_0^2}{b}},
\end{aligned}$$

by the strong law of large numbers for ν_n (3.3.5) and bounded convergence (Theorem 1.1.5). By the same reasoning,

$$\begin{aligned}
\mathbb{E}\left[E_{3,n}^4\right] &= \mathbb{E}\left[\left(\exp\left\{-\frac{\lambda^2\sigma_0^2}{2b}\right\} - \exp\left\{-n\psi\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\nu_n}{n}\right\}\right)^4\right] \\
&= \mathbb{E}\left[\left(\exp\left\{-\frac{\lambda^2\sigma_0^2}{2b}\right\} - \exp\left\{-n\left(\frac{1}{2}\frac{\sigma_0^2\lambda^2}{n} + O\left(\frac{\lambda^3}{n^{\frac{3}{2}}}\right)\right)\frac{\nu_n}{n}\right\}\right)^4\right] \\
&\xrightarrow{n\to\infty} 0.
\end{aligned}$$

So

$$0 \le \lim_{n\to\infty} E_n^4 \le e^{\frac{\sigma_0^2\lambda^2}{2b}} \cdot 1 \cdot e^{\frac{4\lambda^2\sigma_0^2}{b}} \cdot 0 = 0.$$

Thus $E_n \to 0$ as $n \to \infty$, and so we have

$$\mathbb{E}\left[\exp\left\{\lambda\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}\right\}\right] \xrightarrow{n\to\infty} \exp\left\{\frac{\lambda^2}{2}\frac{\sigma_0^2}{b}\right\}.$$

Appealing to Theorem 1.1.3, we have that $\frac{\tilde{S}_{\nu_n}}{\sqrt{n}}$ converges in distribution to a mean zero normal random variable with variance $\frac{\sigma_0^2}{b}$, as required. □

The proof of the main theorem is now straight forward.

Proof of Theorem 3.3.1. Let Z be a random variable such that $Z \sim \mathcal{N}\left(0, \frac{\sigma_0^2}{b}\right)$. We have that

$$\frac{S_n}{\sqrt{n}} = \frac{\tilde{S}_{\nu_n}}{\sqrt{n}} - \frac{1}{\sqrt{n}} \sum_{k=n}^{U_n - 1} f(X_k),$$

by the definitions of S and $\tilde{S}$. But by Remark 3.3.1, the sum in the second term is stochastically bounded by a quantity which is independent of n, so

$$\frac{1}{\sqrt{n}} \sum_{k=n}^{U_n - 1} f(X_k) \xrightarrow{\mathcal{D}} 0.$$

Also, by the previous theorem,

$$\frac{\tilde{S}_{\nu_n}}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z.$$

Hence we have

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

as required. $\square$

We have proved that we have a central limit theorem and we now wish to calculate the variance $\frac{\sigma_0^2}{b}$ of the limiting distribution, again following Tóth in [13]. The following remark is a key observation to facilitate the calculation in [13].

Remark 3.3.2. Since f has mean zero, $\exists g$ such that $f = (I - \Pi)\, g$. In fact, $g = \sum_{n=0}^{\infty} \Pi^n f$.

Proof. Suppose that $g = \sum_{n=0}^{\infty} \Pi^n f$ converges. Then

$$(I - \Pi)\, g = (I - \Pi) \sum_{n=0}^{\infty} \Pi^n f = \sum_{n=0}^{\infty} \Pi^n f - \sum_{n=0}^{\infty} \Pi^{n+1} f = \Pi^0 f = f.$$

So we just need to show the convergence of the infinite sum. However, it is well-known that, when Π is the transition matrix of an irreducible aperiodic Markov chain on a finite state space, $\Pi^n f$ converges exponentially, so the infinite sum converges. $\square$

Theorem 3.3.3. *If we define an inner product on the space of functions on X by*

$$\langle a_1, a_2 \rangle := \sum_{x \in X} a_1(x) a_2(x) \mu(x),$$

for any functions a_1, a_2*, then*

$$\frac{\sigma_0^2}{b} = 2\langle f, g\rangle - \langle f, f\rangle,$$

where $f = (I - \Pi)g$*.*

Proof. By the above remark,

$$\begin{aligned}
S_n &= \sum_{k=0}^{n-1} f(X_k) \\
&= \sum_{k=0}^{n-1} [g(X_k) - \Pi g(X_k)] \\
&= g(X_0) - g(X_n) + \sum_{k=1}^{n} [g(X_k) - \Pi g(X_{k-1})].
\end{aligned}$$

Define

$$M_n := \sum_{k=1}^{n} [g(X_k) - \Pi g(X_{k-1})].$$

We claim that M_n is a martingale. To prove this we note that

$$\mathbb{E}\left(g(X_{n+1})|\mathcal{F}_n\right) = \Pi g(X_n),$$

by definition of the transition probability Π. Thus

$$\begin{aligned}
\mathbb{E}\left(M_{n+1}|\mathcal{F}_n\right) &= \mathbb{E}\left(\sum_{k=1}^{n+1} [g(X_k) - \Pi g(X_{k-1})]\,|\mathcal{F}_n\right) \\
&= \mathbb{E}\left(\sum_{k=1}^{n} [g(X_k) - \Pi g(X_{k-1})]\,|\mathcal{F}_n\right) \\
&\quad + \mathbb{E}\left(g(X_{n+1})|\mathcal{F}_n\right) - \mathbb{E}\left(\Pi g(X_n)|\mathcal{F}_n\right) \\
&= M_n + \Pi g(X_n) - \Pi g(X_n) \\
&= M_n,
\end{aligned}$$

and so M_n is a martingale.
Let $T = T_1$. Then

$$S_T = \sum_{k=0}^{T_1-1} f(X_k) = \sum_{k=0}^{R_1-1} f(X_k) = Y_1.$$

T is a stopping time and $S_T = g(X_0) - g(X_T) + M_T = g(x_0) - g(x_0) + M_T = M_T$.
So by optional stopping (Theorem 1.3.3),

$$\mathbb{E}(S_T) = \mathbb{E}(M_T) = \mathbb{E}(M_0) = 0.$$

Now define another martingale

$$N_n := M_n^2 - \sum_{k=0}^{n-1} \left[\Pi g^2 - (\Pi g)^2\right](X_k).$$

We confirm that this is indeed a martingale by the following calculations. By definition of M_n and N_n,

$$\mathbb{E}\left(N_{n+1}|\mathcal{F}_n\right) - N_n = \mathbb{E}\left(M_{n+1}^2|\mathcal{F}_n\right) - M_n^2 - \Pi g^2(X_n) + \left[\Pi g(X_n)\right]^2.$$

Now note that

$$\begin{aligned}
\mathbb{E}\left(M_{n+1}^2|\mathcal{F}_n\right) - M_n^2 &= \mathbb{E}\left(M_{n+1}^2 - M_n^2|\mathcal{F}_n\right) \\
&= \mathbb{E}\left(\left[(M_{n+1} - M_n) + M_n\right]^2 - M_n^2|\mathcal{F}_n\right) \\
&= \mathbb{E}\left((M_{n+1} - M_n)^2|\mathcal{F}_n\right) + 2\mathbb{E}\left((M_{n+1} - M_n)M_n|\mathcal{F}_n\right) \\
&= \mathbb{E}\left((M_{n+1} - M_n)^2|\mathcal{F}_n\right) + 2M_n\mathbb{E}\left(M_{n+1}|\mathcal{F}_n\right) - 2M_n^2 \\
&= \mathbb{E}\left((M_{n+1} - M_n)^2|\mathcal{F}_n\right) + 2M_n^2 - 2M_n^2 \\
&\qquad \text{since } M_n \text{ is a martingale} \\
&= \mathbb{E}\left((M_{n+1} - M_n)^2|\mathcal{F}_n\right).
\end{aligned}$$

We then see that

$$\mathbb{E}\left(N_{n+1}|\mathcal{F}_n\right) - N_n = 0,$$

since

$$\begin{aligned}
\mathbb{E}\left((M_{n+1} - M_n)^2|\mathcal{F}_n\right) &= \mathbb{E}\left(\left[g(X_{n+1}) - \Pi g(X_n)\right]^2|\mathcal{F}_n\right) \\
&= \mathbb{E}\left(g(X_{n+1})^2|\mathcal{F}_n\right) - 2\Pi g(X_n)\mathbb{E}\left(g(X_{n+1})|\mathcal{F}_n\right) \\
&\quad + \mathbb{E}\left(\left[\Pi g(X_n)\right]^2|\mathcal{F}_n\right) \\
&= \Pi g^2(X_n) - 2\Pi g(X_n)\Pi g(X_n) + \left[\Pi g(X_n)\right]^2 \\
&= \Pi g^2(X_n) - \left[\Pi g(X_n)\right]^2.
\end{aligned}$$

Thus N_n is a martingale, as claimed.
Using this we can calculate

$$\begin{aligned}
\mathbb{E}(S_T^2) &= \mathbb{E}(M_T^2) \\
&= \mathbb{E}\left(N_T + \sum_{k=0}^{T-1}\left[\Pi g^2 - (\Pi g)^2\right](X_k)\right) \\
&= \mathbb{E}(N_0) + \mathbb{E}\left(\sum_{k=0}^{T-1}\left[\Pi g^2 - (\Pi g)^2\right](X_k)\right) \\
&\qquad \text{by optional stopping (Theorem 1.3.3)} \\
&= \mathbb{E}\left(\sum_{k=0}^{T-1}\left[\Pi g^2 - (\Pi g)^2\right](X_k)\right).
\end{aligned}$$

We claim that

$$\mathbb{E}\left(\sum_{k=0}^{T-1}\left[\Pi g^2-(\Pi g)^2\right](X_k)\right)=\sum_{x\in X}\frac{\mu(x)}{\mu(x_0)}\left[\Pi g^2(x)-(\Pi g(x))^2\right].$$

Define $F(x):=\mathbb{E}\left(\sum_{n=1}^{T}\mathbb{1}\left\{X_n=x\right\}\right)$.

We will show that $F(x)=\sum_{y\in X}F(y)\Pi(y,x)$. Then, by uniqueness of the stationary distribution (Theorem 1.2.1), we must have $F=C\mu$ for some constant C.
We have that

$$\begin{aligned}
F(x)&=\mathbb{E}\left(\sum_{n=1}^{\infty}\mathbb{1}\left\{X_n=x;T\geq n\right\}\right)\\
&=\sum_{n=1}^{\infty}\mathbb{P}\left(\mathbb{1}\left\{X_n=x;T\geq n\right\}\right)\\
&=\sum_{n=1}^{\infty}\sum_{y\sim x}\mathbb{P}\left(\mathbb{1}\left\{X_n=x;X_{n-1}=y;T\geq n\right\}\right),\ \text{where } y\sim x\Leftrightarrow\Pi(y,x)>0\\
&=\Pi(0,x)+\sum_{n=2}^{\infty}\sum_{\substack{y\sim x\\ y\neq x_0}}\mathbb{P}\left(\mathbb{1}\left\{X_n=x;X_{n-1}=y;T\geq n\right\}\right)\\
&=\Pi(x_0,x)+\sum_{n=2}^{\infty}\sum_{\substack{y\sim x\\ y\neq x_0}}\mathbb{P}\left(\mathbb{1}\left\{X_n=x;X_{n-1}=y;T\geq n-1\right\}\right),
\end{aligned}$$

where the last line can be explained by considering the following two cases.
Suppose first that $x=x_0$. Then

$$\begin{aligned}
\left\{X_n=x;X_{n-1}=y;T\geq n\right\}&=\left\{X_n=x;X_{n-1}=y;T=n\right\}\\
&=\left\{X_n=x;X_{n-1}=y;T\geq n-1\right\}\ \text{since } y\neq x_0.
\end{aligned}$$

Now suppose $x\neq x_0$. Then clearly

$$\left\{X_n=x;X_{n-1}=y;T\geq n\right\}\subseteq\left\{X_n=x;X_{n-1}=y;T\geq n-1\right\}$$

and, since $y\neq x_0$,

$$\left\{X_n=x;X_{n-1}=y;T\geq n-1\right\}\subseteq\left\{X_n=x;X_{n-1}=y;T\geq n\right\},$$

giving the required equality.
We next note that

$$\mathbb{P}\left(X_n=x|X_{n-1}=y;T\geq n-1\right)=\Pi(y,x).$$

Then we have

$$
\begin{aligned}
F(x) &= \Pi(x_0,x) + \sum_{n=2}^{\infty}\sum_{\substack{y\sim x\\ y\neq x_0}} \mathbb{P}\left(\mathbb{1}\left\{X_n = x; X_{n-1} = y; T \geq n-1\right\}\right)\\
&= \Pi(x_0,x)\\
&\quad + \sum_{n=2}^{\infty}\sum_{\substack{y\sim x\\ y\neq x_0}} \mathbb{P}\left(X_n = x | X_{n-1} = y; T \geq n-1\right)\mathbb{P}\left(X_{n-1} = y; T \geq n-1\right)\\
&= \Pi(x_0,x) + \sum_{\substack{y\sim x\\ y\neq x_0}} \Pi(y,x)\sum_{n=2}^{\infty}\mathbb{P}\left(X_{n-1} = y; T \geq n-1\right)\\
&= \Pi(x_0,x) + \sum_{\substack{y\sim x\\ y\neq x_0}} \Pi(y,x)\sum_{n=1}^{\infty}\mathbb{P}\left(X_n = y; T \geq n\right).
\end{aligned}
$$

But, since

$$
\sum_{n=1}^{\infty}\mathbb{P}\left(X_n = x_0; T \geq n\right) = \sum_{n=1}^{\infty}\mathbb{P}\left(T = n\right) = 1,
$$

we get that

$$
\begin{aligned}
F(x) &= \Pi(x_0,x) + \sum_{\substack{y\sim x\\ y\neq x_0}} \Pi(y,x)\sum_{n=1}^{\infty}\mathbb{P}\left(X_n = y; T \geq n\right)\\
&= \sum_{y\sim x} \Pi(y,x)\sum_{n=1}^{\infty}\mathbb{P}\left(X_n = y; T \geq n\right)\\
&= \sum_{y\sim x} \Pi(y,x)F(y).
\end{aligned}
$$

Therefore $\exists C \in \mathbb{R}$ such that $F(x) = C\mu(x)$, $\forall x \in X$, by uniqueness of the stationary distribution.
We have that $F(x_0) = 1$, so $C = \frac{1}{\mu(x_0)}$ and

$$
F(x) = \frac{\mu(x)}{\mu(x_0)} \quad \text{for all } x \in X.
$$

Our claim follows by noting that $\sum_{x \in X} \mathbb{1}\{X_k = x\} = 1$ for any k. This gives

$$\begin{aligned}
\mathbb{E}\left(\sum_{k=0}^{T-1}\left[\Pi g^2 - (\Pi g)^2\right](X_k)\right) &= \mathbb{E}\left(\sum_{k=0}^{T-1}\sum_{x \in X}\mathbb{1}\{X_k = x\}\left[\Pi g^2 - (\Pi g)^2\right](X_k)\right) \\
&= \mathbb{E}\left(\sum_{x \in X}\left[\Pi g^2 - (\Pi g)^2\right](x)\sum_{k=0}^{T-1}\mathbb{1}\{X_k = x\}\right) \\
&= \sum_{x \in X}\left[\Pi g^2(x) - (\Pi g(x))^2\right]\mathbb{E}\left(\sum_{k=0}^{T-1}\mathbb{1}\{X_k = x\}\right) \\
&= \sum_{x \in X}\left[\Pi g^2(x) - (\Pi g(x))^2\right]F(x) \\
&= \sum_{x \in X}\frac{\mu(x)}{\mu(x_0)}\left[\Pi g^2(x) - (\Pi g(x))^2\right].
\end{aligned}$$

Next, we calculate

$$\begin{aligned}
\sum_{x \in X}\mu(x)\Pi g^2(x) &= \sum_{y \in X}\sum_{x \in X}\mu(x)\Pi(x,y)g^2(y) \\
&= \sum_{y \in X}\mu(y)g^2(y) \quad \text{by stationarity} \\
&= \langle g, g\rangle \quad \text{by definition of the inner product.}
\end{aligned}$$

If we take $\|\cdot\|$ to be the norm induced by the inner product $\langle\cdot,\cdot\rangle$, then we have

$$\sum_{x \in X}\frac{\mu(x)}{\mu(x_0)}\left[\Pi g^2(x) - (\Pi g(x))^2\right] = \frac{1}{\mu(x_0)}\left(\|g\|^2 - \|\Pi g\|^2\right).$$

We now see that

$$\begin{aligned}
\frac{1}{\mu(x_0)} = \sum_{x \in X}\frac{\mu(x)}{\mu(x_0)} = \sum_{x \in X}F(x) &= \sum_{x \in X}\mathbb{E}\left(\sum_{n=1}^{T}\mathbb{1}\{X_n = x\}\right) \\
&= \mathbb{E}\left(\sum_{n=1}^{T}\sum_{x \in X}\mathbb{1}\{X_n = x\}\right) = \mathbb{E}(T) = b.
\end{aligned}$$

So we have shown that

$$\begin{aligned}
\mathbb{E}(S_T^2) &= \mathbb{E}\left(\sum_{k=0}^{T-1}\left[\Pi g^2 - (\Pi g)^2\right](X_k)\right) \\
&= \sum_{x \in X}\frac{\mu(x)}{\mu(x_0)}\left[\Pi g^2(x) - (\Pi g(x))^2\right] \\
&= b\left(\|g\|^2 - \|\Pi g\|^2\right).
\end{aligned}$$

In fact,

$$\mathbb{E}(S_T^2) = b\sigma_1^2,$$

where

$$\sigma_1^2 := 2\langle f, g\rangle - \langle f, f\rangle.$$

We see this by the following calculation, noting that $f = g - \Pi g$.

$$\begin{aligned}\|g\|^2 - \|\Pi g\|^2 &= \langle g, g\rangle - \langle \Pi g, \Pi g\rangle \\ &= \langle g - \Pi g, g\rangle + \langle \Pi g, g\rangle - \langle \Pi g, \Pi g\rangle \\ &= \langle f, g\rangle + \langle \Pi g, g - \Pi g\rangle \\ &= \langle f, g\rangle - \langle -\Pi g, g - \Pi g\rangle \\ &= \langle f, g\rangle - \langle g - \Pi g, g - \Pi g\rangle + \langle g, g - \Pi g\rangle \\ &= \langle f, g\rangle - \langle f, f\rangle + \langle g, f\rangle \\ &= 2\langle f, g\rangle - \langle f, f\rangle.\end{aligned}$$

But $\sigma_0^2 = \mathbb{E}(Y_1^2) = \mathbb{E}(S_T^2) = b\sigma_1^2$. So

$$\frac{\sigma_0^2}{b} = \sigma_1^2,$$

as required. □

We have shown in Theorems 3.3.1 and 3.3.3 that

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma_1^2\right)$, and

$$\sigma_1^2 = 2\langle f, g\rangle - \langle f, f\rangle.$$

We have now proved a central limit theorem for Markov chains via two different methods, under different conditions, one following Varadhan in [15] and the other Tóth in [13]. In each case we derived a formula for the variance of the limiting distribution. We now provide a calculation to check that the two expressions for the variance agree, in the case where both theorems are valid.

Remark 3.3.3. Take $\{X_n\}$ to be a stationary, irreducible, aperiodic Markov chain with finite state space X, and let $f \in \mathcal{L}^2$ have mean zero under the unique stationary distribution μ.
Then we have that $\exists g$ such that $f = (I - \Pi)g$, and

$$\frac{\sum_{j=1}^{n} f(X_j)}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where Theorem 3.2.2 asserts that $Z \sim \mathcal{N}\left(0, \sigma^2\right)$, with

$$\sigma^2 = \mathbb{E}^{\mathbb{P}_\mu}\left[\left(g(X_1) - g(X_0) + f(X_0)\right)^2\right],$$

and Theorems 3.3.1 and 3.3.3 assert that $Z \sim \mathcal{N}\left(0, \sigma_1^2\right)$, with

$$\sigma_1^2 = 2\langle f, g\rangle - \langle f, f\rangle .$$

We claim that $\sigma^2 = \sigma_1^2$.

Proof. To prove this equality, we express both σ^2 and σ_1^2 in terms of g only and we will see that we arrive at the same expression.
Let us first look at σ^2. We see that

$$\begin{aligned}
\sigma^2 &= \sum_{x,y\in X} \mu(x)\Pi(x,y)\left[g(y) - g(x) + f(x)\right]^2 \\
&= \sum_{y\in X} g(y)^2 \sum_{x\in X} \mu(x)\Pi(x,y) \\
&\qquad + \sum_{x\in X} \mu(x)\left[f(x)^2 + g(x)^2 - 2f(x)g(x)\right] \sum_{y\in X} \Pi(x,y) \\
&\quad + 2\sum_{x,y\in X} \mu(x)\Pi(x,y)g(y)\left[f(x) - g(x)\right] \\
&= \sum_{y\in X} \mu(y)g(y)^2 + \sum_{x\in X} \mu(x)\left[f(x)^2 + g(x)^2 - 2f(x)g(x)\right] \\
&\quad + 2\sum_{x,y\in X} \mu(x)\Pi(x,y)g(y)\left[f(x) - g(x)\right].
\end{aligned} \tag{3.3.6}$$

Now we use the relation $f = (I - \Pi)g$ to find that

$$\begin{aligned}
&\sum_{x\in X} \mu(x)\left[f(x)^2 + g(x)^2 - 2f(x)g(x)\right] \\
&= \sum_{x\in X} \mu(x)\Bigg[\Big(g(x) - \sum_{y\in X} \Pi(x,y)g(y)\Big)^2 + g(x)^2 \\
&\qquad\qquad -2\Big(g(x) - \sum_{y\in X} \Pi(x,y)g(y)\Big)g(x)\Bigg] \\
&= \sum_{x\in X} \mu(x)\Bigg[g(x)^2 - 2\sum_{y\in X} \Pi(x,y)g(x)g(y) + \sum_{y\in X} \Pi(x,y)g(y)\sum_{z\in X} \Pi(x,z)g(z) \\
&\qquad\qquad +g(x)^2 - 2g(x)^2 + 2\sum_{y\in X} \Pi(x,y)g(x)g(y)\Bigg] \\
&= \sum_{x\in X} \mu(x)\sum_{y\in X} \Pi(x,y)g(y)\sum_{z\in X} \Pi(x,z)g(z).
\end{aligned}$$

and

$$
\begin{aligned}
&2 \sum_{x,y\in X} \mu(x)\Pi(x,y)g(y)\left[f(x)-g(x)\right] \\
&= \sum_{x,y\in X} \mu(x)\Pi(x,y)g(y)\left[g(x)-\sum_{z\in X}\Pi(x,z)g(z)-g(x)\right] \\
&= -2\sum_{x\in X}\mu(x)\sum_{y\in X}\Pi(x,y)g(y)\sum_{z\in X}\Pi(x,z)g(z).
\end{aligned}
$$

Substituting these expressions into (3.3.6), we get

$$
\begin{aligned}
\sigma^2 &= \sum_{x\in X}\mu(x)g(x)^2 + \sum_{x\in X}\mu(x)\sum_{y\in X}\Pi(x,y)g(y)\sum_{z\in X}\Pi(x,z)g(z) \\
&\quad - 2\sum_{x\in X}\mu(x)\sum_{y\in X}\Pi(x,y)g(y)\sum_{z\in X}\Pi(x,z)g(z) \\
&= \sum_{x\in X}\mu(x)g(x)^2 - \sum_{x\in X}\mu(x)\sum_{y\in X}\Pi(x,y)g(y)\sum_{z\in X}\Pi(x,z)g(z).
\end{aligned}
$$

We now consider σ_1^2. Recalling the definition $\langle a,b\rangle = \sum_{x\in X} a(x)b(x)\mu(x)$, we have

$$
\begin{aligned}
\sigma_1^2 &= 2\sum_{x\in X} f(x)g(x)\mu(x) - \sum_{x\in X} f(x)^2\mu(x) \\
&= 2\sum_{x\in X} g(x)\left[g(x)-\sum_{y\in X}\Pi(x,y)g(y)\right]\mu(x) \\
&\quad - \sum_{x\in X}\left[g(x)-\sum_{y\in X}\Pi(x,y)g(y)\right]^2\mu(x) \\
&= 2\sum_{x\in X}\mu(x)g(x)^2 - 2\sum_{x,y\in X}\mu(x)\Pi(x,y)g(x)g(y) \\
&\quad - \sum_{x\in X}\mu(x)g(x)^2 + 2\sum_{x,y\in X}\mu(x)\Pi(x,y)g(x)g(y) \\
&\quad - \sum_{x\in X}\mu(x)\sum_{y\in X}\Pi(x,y)g(y)\sum_{z\in X}\Pi(x,z)g(z) \\
&= \sum_{x\in X}\mu(x)g(x)^2 - \sum_{x\in X}\mu(x)\sum_{y\in X}\Pi(x,y)g(y)\sum_{z\in X}\Pi(x,z)g(z).
\end{aligned}
$$

Thus we have verified that $\sigma^2 = \sigma_1^2$. $\square$

Chapter 4

Applications of the Markov Chain CLT

In this chapter, we are going to give some examples of how one can apply the theory which we have studied in this report.

4.1 Simple Random Walks on a Torus

Our first two examples concern a simple random walk. This process represents a very simple physical situation: we have a particle which jumps either to the left or right in each time step with given probabilities. However, we would not be able to say anything about this using the well-known central limit theorem for i.i.d. random variables (Theorem 3 of Chapter III, Section 3 in [11]).
Random walks are studied in many texts, and more discussion of this topic can be found in [3] and [6], for example. We adapt some of the definitions and results from Section 3.9 of [6], where simple random walks on the integers are treated.
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $K \in \mathbb{N}$ and consider a torus with K sites; i.e. a line in one dimension with K discrete points labelled $0, 1, \ldots, K-1$ such that site $K-1$ neighbours site 0, as shown in Figure 4.1.

4.1.1 The simple symmetric random walk

We consider a simple symmetric random walk $\{X_n\}$ on the torus, which we define as follows, adapting Grimmett and Stirzaker's definition of a simple symmetric random walk on $\mathbb{Z}$ in [6].
Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. random variables such that $\mathbb{P}(Z_i = \pm 1) = \frac{1}{2}$.

Let $X_0 = 0$ and, for $n \in \mathbb{N}$, define $X_n = \displaystyle\sum_{i=1}^{n} Z_i \mod K$.

It is easy to see that $\{X_n\}$ is a Markov chain, by adapting an argument from Section 3.1 of [6], and that the chain has transition matrix Π, whose entries are given by $\Pi_{i,j} = \Pi(i,j)$, with

$$\begin{aligned} \Pi(i,i+1) = \Pi(i,i-1) &= \frac{1}{2} \quad \text{for } i = 1,2,\cdots,K-2 \\ \Pi(0,1) = \Pi(0,K-1) &= \frac{1}{2} \\ \Pi(K-1,K-2) = \Pi(K-1,0) &= \frac{1}{2}, \end{aligned}$$

and all other entries zero.

The physical interpretation of this random walk is that we consider a particle constrained to jump between K sites on a torus. The particle starts in state 0 and in each discrete time step the particle jumps one site to the left or one site to the right with equal probability. We can see this in Figure 4.1, with $p = q = \frac{1}{2}$.

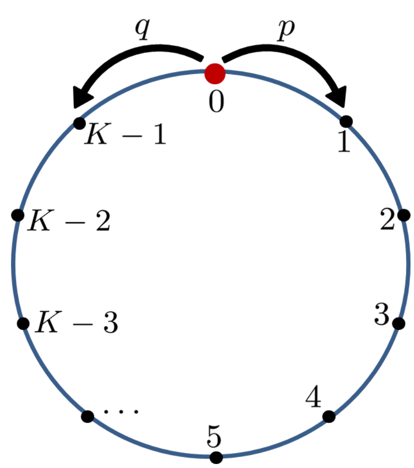

Figure 4.1: We have created a figure to demonstrate a simple random walk on a torus: The particle in red starts at site 0 and, in each time step, it jumps to the right with probability p, or to the left with probability q, where $p+q=1$. The case $p = q = \frac{1}{2}$ corresponds to the simple symmetric random walk in Section 4.1.1 and the case $p \neq q$ corresponds to an asymmetric random walk, which is treated in Section 4.1.2.

Example 4.1.1. We are going to consider the **local time** at zero of the Markov chain up to time n, which we define as

$$T_K(n) := \sum_{k=0}^{n} \mathbb{1}_0(X_k).$$

The local time measures the amount of time the Markov chain spends in state 0 up to time n.

We will prove a central limit theorem for this quantity, using Theorem 3.2.2. In order to apply this theorem, the first condition which we need to check is that the Markov chain is ergodic.
To do this, we will show that $\{X_n\}$ has a unique stationary distribution μ. Then, by Remark 2.5.1, we will have that $\{X_n\}$ is ergodic.
One can easily calculate that if a probability measure μ on $\{0, 1, \ldots, K-1\}$ satisfies

$$\mu(i) = \sum_{j=0}^{K-1} \Pi(j,i)\mu(j) \qquad \forall i = 0, 1, \ldots, K-1,$$

then

$$\mu(i) = \frac{1}{K} \qquad \forall i \in \{0, 1, \ldots, K-1\}. \tag{4.1.2}$$

So μ as defined above is the unique stationary distribution for $\{X_n\}$ and thus the Markov chain is ergodic.
Now recall that, by Theorem 3.2.2, we have a central limit theorem for $f(X_n)$ if f has mean zero, $f \in \mathcal{L}^2$ and we can solve the equation $(I - \Pi)U = f$ for some $U \in \mathcal{L}^2$.
Recalling the definition of T_K, we choose to set f to be the indicator function centred so that it has mean zero under the stationary distribution μ of the Markov chain; that is

$$\begin{aligned} & f(0) = 1 - \frac{1}{K} \\ \text{and} \qquad & f(i) = \quad -\frac{1}{K} \qquad \forall i = 1, \cdots, K-1. \end{aligned}$$

Then

$$T_K(n) - \mathbb{E}[T_K(n)] = \sum_{k=0}^{n} f(X_k).$$

We claim that

$$\frac{T_K(n) - \mathbb{E}[T_K(n)]}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma^2\right)$, with $\sigma^2 = \dfrac{(K-1)(K-2)}{3K^2}$.

Proof. We look for a solution to the equation $(I - \Pi)U = f$ and see that U must satisfy the following system of equations:

$$U(0) - \frac{1}{2}U(1) - \frac{1}{2}U(K-1) = 1 - \frac{1}{K} \tag{4.1.4a}$$

$$-\frac{1}{2}U(i-1) + U(i) - \frac{1}{2}U(i+1) = -\frac{1}{K} \qquad \text{for } 1 \leq i \leq K-2 \tag{4.1.4b}$$

$$-\frac{1}{2}U(0) - \frac{1}{2}U(K-2) + U(K-1) = -\frac{1}{K}. \tag{4.1.4c}$$

The general solution to the homogeneous form of equation (4.1.4b) is $U(i) = A + Bi$, and a particular solution to the inhomogeneous equation is

$U(i) = \dfrac{i^2}{K}$.
Therefore, the general solution to (4.1.4b) is

$$U(i) = A + Bi + \frac{i^2}{K}.$$

We see that the following solution satisfies the required system of equations:

$$U(i) = \frac{i^2}{K} - i \qquad \forall i = 0, 1, \ldots, K-1.$$

Clearly $\displaystyle\sum_{i=0}^{K-1} [U(i)]^2 \mu(i) < \infty$, so that $U \in \mathcal{L}^2$. Therefore, by Theorem 3.2.2, we have a central limit theorem

$$\frac{T_K(n) - \mathbb{E}[T_K(n)]}{\sqrt{n}} = \frac{\sum_{k=1}^{n} f(X_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}(0, \sigma^2)$, for $\sigma^2 = \mathbb{E}^{\mathbb{P}_\mu}\left[(U(X_1) - U(X_0) + f(X_0))^2\right]$.
We now need to calculate the variance, as follows.
Define $G(i,j) := (U(j) - U(i) + f(i))^2$. Then

$$\begin{aligned}
\sigma^2 &= \mathbb{E}^{\mathbb{P}_\mu}\left[(U(X_1) - U(X_0) + f(X_0))^2\right] \\
&= \sum_{i,j} \mu(i)\Pi(i,j)G(i,j) \\
&= \frac{1}{K}\sum_{i,j} \Pi(i,j)G(i,j).
\end{aligned}$$

Substituting in the values of Π gives

$$\begin{aligned}
\sigma^2 = \frac{1}{2}\frac{1}{K}\Bigg\{ &\sum_{i=1}^{K-2} [G(i,i+1) + G(i,i-1)] \\
&+ G(0,1) + G(0,K-1) + G(K-1,K-2) + G(K-1,0) \Bigg\}.
\end{aligned}$$

We then substitute in our definition of G and simplify, resulting in

$$\sigma^2 = \frac{1}{K}\left\{ \sum_{i=1}^{K-2} \left[1 - \frac{2i}{K}\right]^2 + \left[1 - \frac{2}{K}\right]^2 \right\}.$$

We can simplify this further by noting $\left[1 - \dfrac{2}{K}\right]^2 = \left[1 - \dfrac{2(K-1)}{K}\right]^2$, so that

$$\begin{aligned}
\sigma^2 &= \frac{1}{K}\sum_{i=1}^{K-1} \left[1 - \frac{2i}{K}\right]^2 \\
&= \sum_{i=1}^{K-1} \frac{1}{K} - \frac{4}{K^2}\sum_{i=1}^{K-1} i + \frac{4}{K^3}\sum_{i=1}^{K-1} i^2.
\end{aligned}$$

Then, using standard summation formulae, we get

$$\begin{aligned}
\sigma^2 &= \frac{K-1}{K} - \frac{4}{K^2}\frac{K(K-1)}{2} + \frac{4}{K^3}\frac{(K-1)K(2K-1)}{6} \\
&= \frac{K-1}{K^2}\left(K - 2K + \frac{2}{3}(2K-1)\right) \\
&= \frac{(K-1)(K-2)}{3K^2}.
\end{aligned}$$

We now have a closed form for the variance, and we have shown that

$$\frac{T_K(n) - \mathbb{E}[T_K(n)]}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma^2\right)$, with

$$\sigma^2 = \frac{(K-1)(K-2)}{3K^2}.$$

□

Remark 4.1.1. We see that as $K \to \infty$ we have the limit $\sigma^2 \to \frac{1}{3}$.
That is, as the number of sites on the torus increases to infinity, the limiting distribution above tends to $\mathcal{N}\left(0, \frac{1}{3}\right)$.
It is of interest to ask what happens if we switch the order in which we take the limits, first letting $K \to \infty$, then taking the time $n \to \infty$. This scenario corresponds to looking for a central limit theorem for the local time of a simple symmetric random walk on the integers.
It is known that in this case the limiting distribution is actually the distribution of $|\xi|$, where $\xi \sim \mathcal{N}(0,1)$, as proved on slide 12 of [14]. This distribution on the non-negative real numbers is quite different from the distribution $\mathcal{N}\left(0, \frac{1}{3}\right)$, which is non-zero on the whole real line.
Thus we cannot switch the order in which the limits are taken without affecting the result.

4.1.2 Asymmetric random walks

In the previous section, we showed that we have a central limit theorem for the local time of a simple symmetric random walk on a torus. We can treat the asymmetric case in the same way.
Let $p, q \in (0,1)$, $p \neq q$. Let $\{X_n\}$ now be the simple random walk defined as follows, where again we adapt the definition of a simple random walk on the integers from [6].
Let $Z_1, Z_2, \dots$ be a sequence of i.i.d. random variables such that $\mathbb{P}(Z_i = 1) = p$ and $\mathbb{P}(Z_i = -1) = q$, for all $i \in \mathbb{N}$.
As in the symmetric case, let $X_0 = 0$ and, for $n \in \mathbb{N}$, define $X_n = \sum_{i=1}^{n} Z_i \mod K$.

Then, as before, $\{X_n\}$ is a Markov chain with transition matrix Π, whose entries are now given by $\Pi_{i,j} = \Pi(i,j)$, with

$$\begin{aligned} \Pi(i,i+1) &= p \qquad \text{for } i = 0,1,\dots,K-2 \\ \Pi(K-1,0) &= p \\ \Pi(i,i-1) &= q \qquad \text{for } i = 1,2,\dots,K-1 \\ \Pi(0,K-1) &= q, \end{aligned}$$

and all other entries zero.

This random walk represents a similar physical situation to the symmetric case, but now the particle jumps to the right with probability p and to the left with probability q, where these probabilities are unequal. This is also shown in Figure 4.1.

Example 4.1.2. Again we consider the local time at zero,

$$T_K(n) = \sum_{k=0}^{n} \mathbb{1}_0(X_k).$$

It is straight forward to check that the Markov chain has a unique stationary distribution, μ, which is the same as in (4.1.2) for the symmetric case:

$$\mu(i) = \frac{1}{K} \qquad \forall i = 0,1,\dots K-1.$$

Thus the chain is ergodic, by Remark 2.5.1.

As before, we set f to be the function defined by

$$\begin{aligned} & f(0) = 1 - \frac{1}{K} \\ \text{and} \qquad & f(i) = \quad -\frac{1}{K} \qquad \forall i = 1,\cdots,K-1, \end{aligned}$$

so that $T_K(n) - \mathbb{E}[T_K(n)] = \sum_{k=0}^{n} f(X_k)$.

We claim that

$$\frac{T_K(n) - \mathbb{E}[T_K(n)]}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0,\sigma^2\right)$ for some $\sigma > 0$.

Proof. To prove this, we will use Theorem 3.2.2 once again. Thus, we want to find a solution to the equation $(I - \Pi)U = f$. We see that such a solution U must satisfy the system of equations below:

$$U(0) - pU(1) - qU(K-1) = 1 - \frac{1}{K}, \tag{4.1.7a}$$

$$-qU(i-1) + U(i) - pU(i+1) = -\frac{1}{K}, \qquad \text{for } 1 \le i \le K-2, \tag{4.1.7b}$$

$$-qU(0) - pU(K-2) + U(K-1) = -\frac{1}{K}. \tag{4.1.7c}$$

We find that the general solution to (4.1.7b) is

$$U(i) = A + B\left(\frac{q}{p}\right)^i + \frac{i}{K(p-q)}.$$

After some manipulation of the equations, it turns out that the system of equations is solved by the function U defined by

$$U(i) = \frac{p^{K-i}q^i}{(p-q)\left(p^K - q^K\right)} + \frac{i}{K(p-q)},$$

for all $i = 0, 1, \ldots, K-1$.

It is clear that $U \in \mathcal{L}^2$; i.e. $\displaystyle\sum_{i=0}^{K-1} [U(i)]^2 \mu(i) < \infty$.

Thus, by Theorem 3.2.2, we have a central limit theorem

$$\frac{T_K(n) - \mathbb{E}[T_K(n)]}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \sigma^2\right)$ for

$$\sigma^2 = \mathbb{E}^{\mathbb{P}_\mu}\left[\left(U(X_1) - U(X_0) + f(X_0)\right)^2\right].$$

We can now compute the variance, in a similar way to our calculation in the symmetric case in Section 4.1.1. The actual form of σ^2 is complicated, so we only describe its asymptotic behaviour here. We find that σ^2 is asymptotically of the order $\frac{1}{K}$. There is a coefficient which blows up as $p \to \frac{1}{2}$, so this agrees with our calculations for the symmetric case, in Remark 4.1.1, where we see that the variance is asymptotically constant. $\square$

4.2 A Queuing Problem

Another Markov chain which we can study with the tools from Chapter 3 is the one suggested by Varadhan in Exercise 6.15 of [15]. Let $\{X_n\}$ be a Markov chain which takes values on the non-negative integers and has transition probabilities given by

$$\Pi(x,y) = \begin{cases} \frac{1}{2} & \text{for } x = y \geq 0 \\ \frac{1-\delta}{4} & \text{for } y = x+1,\ x \geq 1 \\ \frac{1+\delta}{4} & \text{for } y = x-1,\ x \geq 1 \\ \frac{1}{2} & \text{for } x = 0,\ y = 1, \end{cases}$$

for some $\delta \in (0,1)$.

This Markov chain could model the length of a queue, where in each discrete time step there are three possibilities when the queue is not empty: with probability $\frac{1}{2}$ the queue remains unchanged, with probability greater than $\frac{1}{4}$ one person leaves the queue, or with probability less than $\frac{1}{4}$ one person joins the queue. When the queue is empty, the queue either remains empty or grows to one with equal probabilities.

Example 4.2.1. An interesting quantity to consider is the amount of time for which the queue is empty; i.e. the number of times k at which $X_k = 0$. Therefore, as in the previous examples, we will look for a central limit theorem for the local time of the Markov chain at zero, defined by

$$T(n) = \sum_{k=0}^{n} \mathbb{1}_0(X_k).$$

We claim that

$$\frac{T(n) - \mathbb{E}^{\mathbb{P}_\mu}\left[T(n)\right]}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \dfrac{2-\delta}{(1+\delta)^2}\right)$.

Proof. As in our previous two examples, we wish to apply Theorem 3.2.2. We first need to show that $\{X_n\}$ is ergodic and, in order to do this, we prove that the Markov chain has a unique stationary distribution.
We need to find μ such that, for all $x \in \mathbb{N} \cup \{0\}$, $\mu(x) = \sum_{y \in \mathbb{N}\cup\{0\}} \Pi(y,x)\mu(y)$; that is, μ must satisfy

$$\mu(0) = \frac{1}{2}\mu(0) + \frac{1+\delta}{4}\mu(1) \tag{4.2.1a}$$

$$\mu(1) = \frac{1}{2}\mu(0) + \frac{1}{2}\mu(1) + \frac{1+\delta}{4}\mu(2) \tag{4.2.1b}$$

$$\mu(x) = \frac{1-\delta}{4}\mu(x-1) + \frac{1}{2}\mu(x) + \frac{1+\delta}{4}\mu(x+1), \qquad \text{for } x = 2, 3, \ldots. \tag{4.2.1c}$$

We also have the constraint

$$\sum_{x=0}^{\infty} \mu(x) = 1.$$

We find that the general solution to (4.2.1c) is

$$\mu(x) = A + B\left(\frac{1-\delta}{1+\delta}\right)^x, \qquad \text{for } x = 2, 3, \ldots,$$

and the constraint (4.2) gives us that $A = 0$.
We can then see that we have the following unique solution to the system of equations (4.2.1) and (4.2):

$$\mu(x) = \begin{cases} \frac{\delta}{1+\delta} & \text{for } x = 0, \\ 2\delta\frac{(1-\delta)^{x-1}}{(1+\delta)^{x+1}} & \text{for } x = 1, 2, \ldots. \end{cases}$$

Therefore, by Remark 2.5.1, $\{X_n\}$ is ergodic. Let us set f to be the centred indicator function on the state 0,

$$\begin{aligned} f &= \mathbb{1}_0 - \mathbb{E}^\mu\left[\mathbb{1}_0\right] \\ &= \mathbb{1}_0 - \mu(0) \\ &= \mathbb{1}_0 - \frac{\delta}{1+\delta}. \end{aligned}$$

That is

$$f(x) = \begin{cases} \frac{1}{1+\delta} & \text{for } x = 0 \\ -\frac{\delta}{1+\delta} & \text{for } x = 1, 2, \ldots . \end{cases}$$

Then $T(n) - \mathbb{E}[T(n)] = \sum_{k=0}^{n} f(X_k)$.

We next need to look for a solution U to the equation $(I - \Pi)U = f$, so that we can apply Theorem 3.2.2. U must satisfy the system of equations

$$\sum_{y=0}^{\infty} (\delta_{0,y} - \Pi(0, y))\, U(y) = \frac{1}{1+\delta} \qquad \text{and}$$

$$\sum_{y=0}^{\infty} (\delta_{x,y} - \Pi(x, y))\, U(y) = -\frac{\delta}{1+\delta} \qquad \text{for } x = 1, 2, \ldots,$$

which is equivalent to

$$\frac{1}{2}U(0) - \frac{1}{2}U(1) = \frac{1}{1+\delta} \qquad \text{and}$$

$$\frac{1}{2}U(x) - \frac{1+\delta}{4}U(x-1) - \frac{1-\delta}{4}U(x+1) = -\frac{\delta}{1+\delta} \qquad \text{for } x = 1, 2, \ldots .$$

One can show that the solution to this system of equations is

$$U(x) = -\frac{2x}{1+\delta} \qquad \text{for all } x = 0, 1, 2, \ldots .$$

We have that $U \in \mathcal{L}^2$, since

$$\sum_{x=0}^{\infty} [U(x)]^2 \mu(x) = 8\delta(1+\delta)^4 \sum_{x=1}^{\infty} x^2 \left(\frac{1-\delta}{1+\delta}\right)^{x-1} < \infty,$$

recalling that $\delta \in (0, 1)$.

Thus, by Theorem 3.2.2, we have a central limit theorem for the local time at zero:

$$\frac{T(n) - \mathbb{E}[T(n)]}{\sqrt{n}} = \frac{\sum_{k=0}^{n} f(X_k)}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}(0, \sigma^2)$ and the variance can be calculated using the now familiar formula

$$\sigma^2 = \mathbb{E}^{\mathbb{P}_\mu}\left[(U(X_1) - U(X_0) + f(X_0))^2\right].$$

After some computation, the variance comes out as

$$\sigma^2 = \frac{2-\delta}{(1+\delta)^2}.$$

We have therefore shown that we have

$$\frac{T(n) - \mathbb{E}^{\mathbb{P}_\mu}[T(n)]}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

where $Z \sim \mathcal{N}\left(0, \frac{2-\delta}{(1+\delta)^2}\right)$. □

Conclusion

In Chapter 3, our most general result (Theorem 3.2.2) showed that there is a central limit theorem for $\sum_{k=1}^{n} \frac{f(X_k)}{\sqrt{n}}$ when $\{X_n\}$ is an ergodic Markov chain with transition matrix Π, under some conditions. Namely, we required that the chain has a countable state space in $\mathbb{R}$ and f is a mean zero square-integrable function such that $(I - \Pi)U = f$, for some $U \in \mathcal{L}^2$. Several extensions of this result have been proved, and perhaps the most notable of these is the 1984 paper of Kipnis and Varadhan [7].

In the examples of simple random walks on a torus which we discussed in Section 4.1, we were modelling a situation of a single particle moving around K sites on a torus. We were able to deal with this example using our Theorem 3.2.2. If we suppose that we are in a similar physical situation, but with some of the sites on the torus occupied by other particles and our jumping particle only able to move to unoccupied sites, then we can no longer handle this situation with the tools that we have studied in this report.

The situation described above is an example of what is known as a simple exclusion process. These processes are explained further by Komorowski, Landim and Olla in Part II of [8]. In [7], the authors extend the Markov chain central limit theorem and apply their results to symmetric simple exclusion processes, under some conditions. Further extensions to this work have been achieved more recently; for example, in a 2004 paper [2], Bernardin considers asymmetric simple exclusion processes in one and two dimensions. More results on the topic of simple exclusion processes are collected in Part II of the 2012 book [8].

Another distinct application area of central limit theorems for Markov chains is Markov chain Monte Carlo. This subject is explained in Gamerman and Lopes' book [4], and here we outline some of the main ideas of the subject, summarising material from that text. Monte Carlo integration is a method for estimating an integral by using a random sample. The idea is that if we have a large random sample, for which a law of large numbers applies, and we take the sample mean of a function, then this quantity approximates the integral of the function with respect to the distribution of the sample. If we have a central limit theorem for the random sample, then we can control the error of the approximation.

For more complicated and high-dimensional distributions, it can be difficult to generate a random sample which has a given distribution. Markov chain Monte Carlo methods provide a solution to this problem. If we can construct a Markov chain which converges to its stationary distribution, and whose stationary distribution is equal to the distribution in which we are interested, then at large times, the random variables in the chain are approximately distributed according to the distribution of interest. If, moreover, we have a law of large numbers and a central limit theorem for the Markov chain, then we can take values of the chain after it has been running for a specified large time and use these as a random sample in Monte Carlo integration.

Markov chain Monte Carlo is frequently applied in statistics, most often to perform inference on the posterior distributions in Bayesian models, as explained in [4]. A good example of this is given in [12]. Markov chain Monte Carlo methods also have wider applications. For example, in [5] the authors use Markov chain Monte Carlo to estimate coefficients, which cannot be explicitly computed, in the theory of elliptic partial differential equations.

These are just a few examples of how the work which we have presented in this report can be extended and applied to diverse areas of mathematics.

Bibliography

[1] Robert G. Bartle, *The elements of integration and Lebesgue measure*, Wiley Classics library ed., John Wiley and Sons, Inc., 1995.

[2] Cédric Bernardin, *Fluctuations in the occupation time of a site in the asymmetric simple exclusion process*, The Annals of Probability **32** (2004), no. 1B, 855–879.

[3] Rick Durrett, *Probability: Theory and examples*, fourth ed., Cambridge University Press, 2010.

[4] Dani Gamerman and Hedibert F. Lopes, *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*, second ed., Chapman and Hall/CRC, 2006.

[5] Antoine Gloria and Jean-Christophe Mourrat, *Quantitative version of the Kipnis-Varadhan theorem and Monte Carlo approximation of homogenized coefficients*, The Annals of Applied Probability **23** (2013), no. 4, 1544–1583.

[6] Geoffrey R. Grimmett and David R. Stirzaker, *Probability and random processes*, third ed., Oxford University Press, 2001.

[7] C. Kipnis and S.R.S. Varadhan, *Central limit theorem for additive functionals of reversible markov processes and applications to simple exclusions*, Communications in Mathematical Physics **104** (1986), 1–19.

[8] Tomasz Komorowski, Claudio Landim, and Stefano Olla, *Fluctuations in Markov processes: Time symmetry and martingale approximation*, A Series of Comprehensive Studies in Mathematics, vol. 345, Springer, 2012.

[9] Erwin Kreyszig, *Introductory functional analysis with applications*, Wiley Classics library ed., John Wiley and Sons, Inc., 1989.

[10] Sheldon Ross, *A first course in probability*, ninth ed., Pearson Education Limited, 2014.

[11] A. N. Shiryaev, *Probability*, second ed., Graduate Texts in Mathematics, vol. 95, Springer, 1989.

[12] D.J. Spiegelhalter, N.G. Best, W.R. Gilks, and H. Inskip, *Hepatitis B: a case study in MCMC methods*, Markov Chain Monte Carlo in Practice (W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds.), Chapman and Hall, 1996, pp. 21–44.

[13] Bálint Tóth, *Central limit theorem for finite Markov chains - Doeblin's method: Reduction to i.i.d.*, Unpublished handwritten notes.

[14] ______, *Limit theorems of probability theory*, `http://www.math.bme.hu/~balint/oktatas/hetnet/jegyzet/ltp3_wbh_slides.pdf`, 2014, Accessed: 29 April 2015.

[15] S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes, vol. 7, American Mathematical Society, 2001.

[16] David Williams, *Probability with martingales*, Cambridge University Press, 1991.