STATIONARY RANDOM WALKS AND ISOTONIC REGRESSION

by

Ou Zhao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2008

Doctoral Committee:

Professor Michael B. Woodroofe, Chair
Professor Tailen Hsing
Professor Robert W. Keener
Assistant Professor Erhan Bayraktar

ACKNOWLEDGEMENTS

I would like to thank Michael Woodroofe for being my thesis advisor in the first place, and also among many other things, for seriously treating me as a collaborator. My first steps in mathematics were guided by him; from our collaborations I learned to appreciate his way of approaching mathematics—that is, always trying to be playful. Being Michael's student, I was led to many interesting topics—from operator theory to Markov chains, from ergodic theory to martingales, and from stationary processes to isotonic regression. I thank him for teaching me both probability and statistics, even some physics. His guidance has fundamentally changed my outlook.

Thanks to Tailen Hsing, Bob Keener, and my cognate member, Erhan Bayraktar, for serving on my thesis committee. Their support are very important for me to finish this dissertation. I am particularly grateful to Tailen during my job hunting process; his kind encouragements are indeed something I very much needed. Wei Biao Wu, from the University of Chicago, helped me in many different ways. I appreciate his great friendship.

The statistics department at the University of Michigan has been a supportive place for me to study in the past five years. Thanks to all the faculty members, staff, and my fellow students for creating the necessary environment. Last, but not least, I wish to thank my parents and my wife for their unconditional love and support. This dissertation is dedicated to them and my daughter, Ashley.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

This thesis makes some contributions to the study of stationary processes, with a view towards applications to time series analysis. Ever since the work of Gordin [15], martingale approximations have been a useful tool to study stationary random walks (random walks with stationary increments). One highlight of this dissertation is a useful necessary and sufficient condition for martingale approximations. This condition provides new insight unifying several recent developments on this topic in the literature. As an application of martingale approximations, we derive a fairly sharp sufficient condition for the law of the iterated logarithm, including the functional form, and an improvement of the conditional central limit theorem of Maxwell and Woodroofe [30]. For statistical applications, we consider the problem of estimating a monotone trend nonparametrically for time series data. The asymptotic distributions of isotonic estimators are analyzed, and the accuracy of the approximations are studied numerically. Estimation of the end point value is a main focus because of the practical importance and mathematical difficulty.

CHAPTER I

Introduction

This thesis makes some contributions to the study of stationary processes, with a view towards applications to time series analysis. The topic has a long history; but the treatment here is indeed modern, and results are nearly optimal. On the mathematical side, one novel feature is the consistent interplay between certain parts of ergodic theory and probability theory from an operator-theoretical point of view. The techniques employed to achieve this interplay are mostly from harmonic analysis and functional analysis, but otherwise quite down-to-earth. So I hope readers interested in ergodic theory and probability theory will find something valuable in this dissertation. On the more practical side, one nonparametric statistical application is considered, and the purpose here is to show by example how a theory for time series analysis might be pursued, based on the development of ergodic dynamical systems. We divide our treatment into five chapters, with the first chapter as the introduction here, then the rest addressing four different classes of problems. In this introduction, we shall briefly preview our main results chapter by chapter.

Let $\ldots, W_{-1}, W_0, W_1, \ldots$ be an ergodic, and (strictly) stationary Markov chain assuming values in a measurable space $(\mathcal{W}, \mathscr{B})$. The transition kernel and stationary distribution are denoted by Q and π, so $P[W_k \in B] = \pi(B)$ and $Q^n(w; B) =$

$P[W_{n+k} \in B | W_k = w]$ for all nonnegative integers n, k, and $B \in \mathscr{B}$. We shall also denote the Markov operator by Q,

$$Qf(w) = \int_{\mathcal{W}} f(z) Q(w; dz)$$

for any $f \in L^1(\mathcal{W}, \pi)$. Consider now functions defined on the state space, and in particular, we are interested in $g \in L^2_0(\pi)$, the space of square integrable functions for which $\int_{\mathcal{W}} g d\pi = 0$. Interest centers on

$$S_n = S_n(g) = g(W_1) + \cdots + g(W_n), \tag{1.1}$$

called *stationary random walks* (partial sums of stationary processes). It turns out that, to study (1.1), one can often introduce martingale approximations. These are of the form

$$S_n = M_n + R_n, \tag{1.2}$$

where M_n is a centered martingale with square integrable, stationary increments, and $\|R_n\| = o(\sqrt{n})$. Here $\| \cdot \| = \langle \cdot, \cdot \rangle$ denotes the norm in an L^2 space, which may vary from one usage to another. One prototype of (1.2) was used in Gordin and Lifsic [17], where they used *Poisson's equation*, $g = (I - Q)h$. The idea is that, given g and Q, one tries to solve for some $h \in L^2$. The consequence is then the following straightforward decomposition

$$S_n = \sum_{k=1}^{n} [h(W_k) - Qh(W_{k-1})] + Qh(W_0) - Qh(W_n).$$

So this assumes the form (1.2), with a negligible remainder term. From this phenomenon, one can easily speculate that more g should admit martingale approximations, if only $\|R_n\| = o(\sqrt{n})$ is required. Apparently, (1.2) is sufficient to deduce the central limit theorem (CLT); and even more so, one can derive a conditional version of the CLT (e.g., [26] and [30]), which is more useful in many real applications; for example, MCMC. Kipnis and Varadhan [26] made a significant contribution along this line by showing that, if the underling chain is reversible, then all that be needed to apply (1.2) is the following variance stability condition

$$\sigma^2 := \lim_{n\to\infty} \frac{E[S_n^2]}{n} \in [0, \infty). \tag{1.3}$$

It can be shown that, in the case of reversibility, (1.3) is equivalent to the solvability of a fractional version of Poisson's equation, $g = \sqrt{I-Q}h$. This is a special case (with index 1/2) of the so-called *Fractional Poisson's Equation*, which was systematically studied by Derriennic and Lin [11]. Taking advantage of the equivalence, Derriennic and Lin [10] were able to generalize the Kipnis-Varadhan theorem to Markov chains, whose transition operators are normal. To state their result, let Q^* denote the ajoint operator of Q, acting on the space $L^2(\pi)$. Then the normality of the chain means $QQ^* = Q^*Q$, which is a substantial generalization of the reversibility assumption. In [10] it is shown that, under the normality assumption, $g \in \sqrt{I-Q}L^2(\pi)$ is sufficient to imply the existence of (1.2). As applications, they considered random walks on compact abelian groups, which give rise to "normal" chains due to the commutativity of group operations. Motivated by the seminal work of Kipnis and Varadhan, but unaware of [10], Maxwell and Woodroofe [30] are able to drop the condition of reversibility, and derived (1.2) under their surprisingly elegant condition

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|E[S_n|W_1]\| < \infty. \tag{1.4}$$

This growth condition is optimal within logarithmic factors, as proved in [30]. It is also sufficient for the functional central limit theorem in a *best possible* way (e.g., Peligrad and Utev [32]).

Allowing for general dependence structure of the underlying chain and general state space has important implications for *general* stationary random walks. To see that, let $\dots, X_{-1}, X_0, X_1, \dots$ denote a centered, (strictly) stationary process. Then its partial sums $S_n = X_1 + X_2 + \cdots + X_n$ have the form (1.1), with $W_k = (\dots, X_{k-1}, X_k)$, and g just the first coordinate mapping. Representing in this way, we can work out Q and Q^*, but typically, $Q^*Q \neq QQ^*$. So normality is not generally satisfied. It can be shown that (1.4) is stronger than the condition $g \in \sqrt{I-Q}L^2(\pi)$; a natural question is then: can one drop the normality assumption in the theorem of Derriennic and Lin [10]? This question will be answered, along with a much more ambitious goal to derive a usable necessary and sufficient condition for (1.2). The important special case of co-isometries, $QQ^* = I$, will receive special attentions. For easy reference, we now state the main result of Chapter III, put

$$V_n g = \sum_{k=0}^{n-1} Q^k g,$$

so that $E(S_n|W_1) = V_n g(W_1)$. If a *martingale approximation* exists, then $\|V_n g\|^2 = E[E(S_n|W_1)^2] \leq 2E(M_1^2) + 2E(R_n^2) = o(n)$, and $\lim_{n\to\infty} E(S_n^2)/n = E(M_1^2)$. So for a given g, obvious necessary conditions for the existence of martingale approximations are that

$$\|V_n g\| = o(\sqrt{n}) \tag{1.5}$$

and

$$\|g\|_+^2 := \limsup_{n\to\infty} \frac{1}{n} E[S_n(g)^2] < \infty. \tag{1.6}$$

Let $\mathcal{L}$ denote the set of $g \in L_0^2(\pi)$ for which $\|g\|_+ < \infty$. Then $\mathcal{L}$ is a linear space, and $\|\cdot\|_+$ defines a pseudo norm on $\mathcal{L}$, called the *plus norm* in Chapter III. Moreover, Q maps $\mathcal{L}$ into itself. It will be shown that:

Theorem 1.1. *g admits a martingale approximation iff $\|V_n g\| = o(\sqrt{n})$, and*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \|Q^k g\|_+^2 = 0.$$

This result extends [26], [30] and [11] in a unified manner. For more details, we refer interested readers to Chapter III, or [47].

As we have already seen, in the case Poisson's equation can be solved, the martingale approximation can be derived with a rather negligible remainder term. This gives a hope to study the fluctuation behavior of (1.1) using martingale techniques. Indeed, in Chapter IV, we shall develop a set of techniques to prove a law of the iterated logarithm for stationary processes by slightly strengthening (1.4). To state the main result, let $\ell(\cdot)$ be a positive, nondecreasing, slowly varying function; and let

$$\ell^*(n) = \sum_{j=1}^{n} \frac{1}{j\ell(j)}.$$

Theorem 1.2. *If ℓ is a positive, slowly varying, nondecreasing function and*

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \sqrt{\ell(n)} \log(n) \|E(S_n|W_0)\| < \infty, \tag{1.7}$$

then

$$S_n = M_n + R_n \tag{1.8}$$

where M_n is a square integrable martingale with stationary increments and

$$\lim_{n\to\infty} \frac{R_n}{\sqrt{n\ell^*(n)}} = 0 \ w.p.1.$$

The law of the iterated logarithm can be easily derived as a corollary. For instance, if (1.7) holds with $\ell(n) = \log n$, then $\sigma^2 = \lim_{n\to\infty} E[S_n^2]/n$ exists, and

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = \sigma \quad w.p.1 \tag{1.9}$$

Other side products of Theorem 1.2 include *quenched* CLT, and invariance principles of Strassen's type. These are presented in Chapter IV, which is based on [48]. The proof employs a variety of techniques—perturbations of linear operators, Fourier analysis of renewal equations, and operator-theoretical ergodic theory. Whether the Maxwell-Woodroofe condition (1.4) be sufficient for (1.9) still remains open.

The study of stationary random walks has important applications to time series analysis. To illustrate one such usage, we consider a model for a time series which is thought to consist of a nondecreasing trend observed with stationary errors:

$$y_k = \mu_k + \varepsilon_k,\ \ k = 1, 2, \cdots$$

where $-\infty < \mu_1 < \mu_2 < \cdots$ and $\ldots\varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots$ is a strictly stationary sequence with mean 0 and finite variance. The global temperature anomalies, in Example 1.3, provide a particular example. If a segment of the series is observed, say $y_1, \cdots, y_n$, then isotonic methods suggest themselves for estimating the μ_k nonparametrically. The isotonic estimators may be described as

$$\hat{\mu}_k = \max_{i \le k} \min_{k \le j} \frac{y_i + \cdots + y_j}{j - i + 1}.$$

Example 1.3. Figure 1.1 plots the annual global temperature anomalies from 1850-2000 with the isotonic estimates of trend superimposed as a step function.

With the global warming data, there is special interest in estimating μ_n, the current temperature anomaly; and there isotonic methods encounter the *spiking problem*, described in Section 7.2 of [35] for a closely related problem of estimating monotone densities. We consider two methods for correcting this problem, the penalized estimators of [43] and the method of [28], both introduced for monotone densities. The former estimates μ_n by

$$\hat{\mu}_{p,n} = \max_{1 \le i \le n} \frac{y_i + \cdots + y_n - \beta_n}{n - i + 1},$$

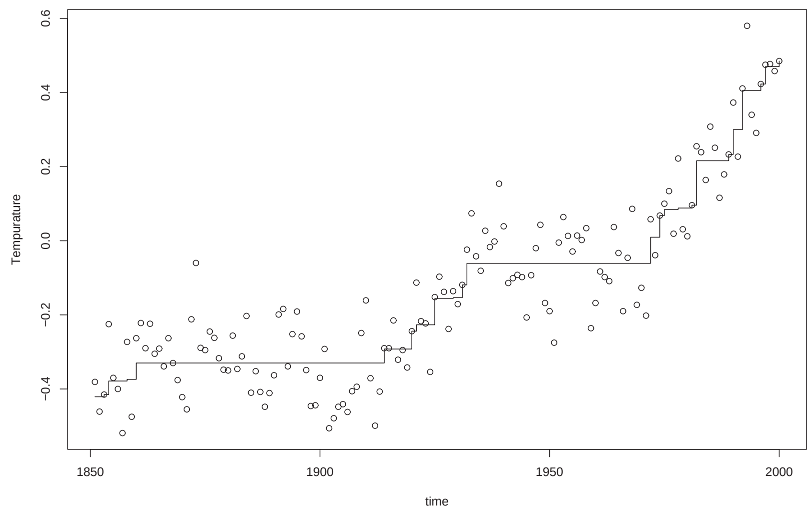

Figure 1.1: Global Temperature Anomalies

where $\beta_n > 0$ is a smoothing parameter, and the latter by

$$\tilde{\mu}_{b,n} = \hat{\mu}_{m_n},$$

where $m_n < n$ is another smoothing parameter. To analyze the behavior of these estimators, we shall need an important technical tool as developed in [32]. Stating it slightly specialized for our purpose, let $S_n = \varepsilon_1 + \cdots + \varepsilon_n$, $\mathcal{F}_n = \sigma(\ldots, \varepsilon_{n-1}, \varepsilon_n)$; define

$$B_n(t) = \frac{S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\, \varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}},$$

and let $\mathbb{B}$ denote a standard Brownian motion. Both B_n and $\mathbb{B}$ are regarded as random elements in $C[0,1]$. It can be shown that if

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|E[S_n|\mathcal{F}_0]\| < \infty, \tag{1.10}$$

then

$$\Gamma = \sum_{j=0}^{\infty} \left\| \frac{E[S_{2^j}|\mathcal{F}_0]}{2^{j/2}} \right\| < \infty,$$

and

$$E\left[\max_{1 \le k \le n} S_k^2\right] \le 6\big(E[\varepsilon_1^2] + \Gamma\big)^2 n, \tag{1.11}$$

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} E[S_n^2]$$

exists; moreover, B_n converges in distribution to $\sigma\mathbb{B}$.

The main results of Chapter II obtain the asymptotic distributions of estimation errors, properly normalized, for the three estimators described above. One of these results is well known for monotone regression with i.i.d. errors, and analogues of the others are known for the closely related problem of monotone density estimation. Interest here is in extending these results to allow for dependence. Other researchers have been interested in this question recently—notably Anevski and Hössjer [1]. Our results go beyond theirs in several ways: (1) we consider the boundary case—estimating μ_n, where the analysis of estimation errors pose particular mathematical challenges; (2) our results may be valid conditional on the past unobserved history of the series, this appealing fact is still under investigation, but with good progress; and (3) our conditions are weaker. Instead of the strong mixing condition, called (A9) in [1], we use a condition like (1.4) above, for the errors. One objective is to show by example how recent results on the central limit question for sums of stationary processes can be used to weaken mixing conditions in statistical applications.

Chapter V describes some problems pertaining to the conditional central limit theorem.

CHAPTER II

Estimating a Monotone Trend

2.1 Preliminaries

In order to model a monotone trend for a time series, such as the global warming data presented in the introduction, one may consider the following regression model. Given a nondecreasing function $\phi(\cdot)$ on the unit interval $[0,1]$, consider observations $y_1, \ldots, y_n$ from

$$y_k = \phi\left(\frac{k}{n}\right) + \varepsilon_k \qquad k = 1, \cdots, n, \tag{2.1}$$

where the errors are assumed to be ergodic, stationary with mean 0, finite second moments. For convenience we will sometimes be working with its two-sided extension, $\ldots \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots$. The problem is to nonparametrically estimate $\phi(\cdot)$, with particular focus on the boundary point $\phi(1)$.

To deal with the dependence among errors, we always assume (1.10) holds. This condition has been central to many recent developments concerning weak dependence, ever since its first appearance in [30]. Taking advantage of the shape assumption about the regression function, isotonic estimates are suggested. To describe them, let $\mu_k = \phi(k/n)$ and choose $\hat{\mu}_1, \ldots, \hat{\mu}_n$ to minimize

$$\sum_{i=1}^{n} (y_i - \mu_i)^2,$$

subject to the monotonicity constraint, $-\infty < \mu_1 \le \mu_2 \le \cdots \le \mu_n < \infty$. This is a well-known optimization problem (e.g., [35]), with unique solutions given by

$$\hat{\mu}_k = \max_{i \le k} \min_{k \le j} \frac{y_i + \cdots + y_j}{j - i + 1}. \tag{2.2}$$

Alternatively, letting Y_n denote the cumulative sum diagram, i.e.,

$$Y_n(t) = \frac{y_1 + \cdots + y_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) y_{\lfloor nt \rfloor + 1}}{n},$$

and $\tilde{Y}_n$ its greatest convex minorant, $\hat{\mu}_k$ is the left hand derivative of $\tilde{Y}_n$ evaluated at $t = k/n$. See Chapter 1 of [35] for background on isotonic methods.

Next, we define $\hat{\phi}_n(\cdot)$ to be the left-continuous step function on $[0,1]$, with values $\hat{\phi}_n(k/n) = \hat{\mu}_k$ at the knots k/n, $k = 1, \ldots, n$. The large sample behavior of $\hat{\phi}_n(t)$ is of great interest to us, because it is needed to set confidence intervals for $\phi(t)$. This will be studied for $t \in (0,1)$ in Section 2.2.

For the boundary value $\phi(1)$, the standard isotonic estimates suffer from the spiking problem. Here we propose two estimators, which will be shown to be consistent with convergence rate $n^{-1/3}$. The first one has its analogue in the context of density estimation, [28], it has the following form with boundary correction

$$\tilde{\mu}_{b,n} = \hat{\phi}_n(1 - \alpha n^{-\frac{1}{3}}), \tag{2.3}$$

where $\alpha > 0$ is a smoothing parameter. The other estimator has its prototype in [43]; it modifies the standard isotonic estimator, $\hat{\mu}_n$, by shrinking the size of it,

$$\hat{\mu}_{p,n} = \max_{1 \le j \le n} \frac{y_j + \cdots + y_n - \beta_n}{n - j + 1}, \tag{2.4}$$

where $\beta_n > 0$ is another smoothing parameter, depending on the sample size.

The aim of this chapter is to study the behavior of three estimators, (2.2),(2.3), and (2.4). Their limiting distributions are studied numerically, and the performance at the end points are compared through simulations.

2.2 Asymptotic Distributions

Isotonic estimators with boundary corrections. We first study (2.2) and (2.3), and start by introducing more notations. Define $\Phi_n(\cdot)$ to be the linear interpolation between its values $\Phi_n(k/n) = (\mu_1 + \cdots + \mu_k)/n$ at knots k/n, and $\Phi_n(0) = 0$; these give rise to continuous functions in $C[0,1]$. More generally, we shall denote by $C(K)$ the space of continuous functions on an arbitrary compact set $K \subset \mathbb{R}$. The topology on $C(K)$ will always be induced by the uniform metric ρ. Further let $S_n = \varepsilon_1 + \cdots + \varepsilon_n$, and define the partial sum process

$$B_n(t) = \frac{1}{\sqrt{n}}\big[S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\, \varepsilon_{\lfloor nt \rfloor + 1}\big] \tag{2.5}$$

on the unit interval. Then from (2.1), it is not difficult to check

$$Y_n(t) = \Phi_n(t) + \frac{1}{\sqrt{n}} B_n(t) \tag{2.6}$$

for $0 \le t \le 1$.

Put $\lambda_n = n^{-1/3}$, and take $\{t_n\}_1^\infty$ to be a sequence of numbers in $(0,1)$. We say $\{t_n\}$ is a *regular sequence* if $t_n \to t_0 \in (0,1]$, and, in the case $t_n \to 1$, $n^{1/3}(1-t_n) \to \alpha$ for some $\alpha > 0$. Next, introduce the process

$$Z_n(s) = n^{2/3}\big[Y_n(t_n + \lambda_n s) - Y_n(t_n) - \phi(t_n)\lambda_n s\big], \tag{2.7}$$

which is well-defined for s contained in

$$I_n := \big[-n^{1/3} t_n,\ n^{1/3}(1 - t_n)\big]\,. \tag{2.8}$$

Using (2.6) and letting $\Phi(t) := \int_0^t \phi(s) ds$, one can easily verify the following decomposition

$$\begin{aligned}
Z_n(s) &= n^{2/3}\left[\Phi(t_n+\lambda_n s)-\Phi(t_n)-\phi(t_n)\lambda_n s\right] \\
&\qquad +n^{1/6}\left[B_n(t_n+\lambda_n s)-B_n(t_n)\right] \\
&\qquad\qquad +n^{2/3}\big[R_n(t_n+\lambda_n s)-R_n(t_n)\big] \\
&=: \Psi_n(s)+W_n(s)+\Delta_n(s), \qquad (2.9)
\end{aligned}$$

where $R_n(\cdot) = \Phi_n(\cdot) - \Phi(\cdot)$; so, by the mean value theorem, $\sup_{t\in[0,1]}|R_n(t)| = O(1/n)$.

We will first prove a weak convergence result for $Z_n(\cdot)$. In it we suppose

(A1). $\phi \in C^1[0,1]$, and $\gamma_1 = \inf_{t\in(0,1]}\phi'(t) > 0$,

which is slightly stronger than necessity. To state the result, let $W(\cdot)$ denote a standard two-sided Brownian motion (starting from 0), and for $t_0 \in (0,1]$, define

$$Z(s) = \sigma W(s) + \frac{1}{2}\phi'(t_0)s^2.$$

Theorem 2.1. *Suppose (A1) and (1.10) hold; let $\{t_n\}$ be a regular sequence with $t_n \to t_0$; further, let K be any compact interval, $K_n = K \cap I_n$, and $K_\infty = \cup_{n\geq 1}K_n$. Then for Z_n, as defined in (2.7),*

$$Z_n(\cdot)|_{K_n} \Rightarrow Z(\cdot)|_{K_\infty},$$

in $C(K_\infty)$ (note $K_n = K_\infty$ for all large n).

Proof. Recall $Z_n(s) = \Psi_n(s)+W_n(s)+\Delta_n(s)$ for $s \in I_n = [-n^{1/3}t_n,\ n^{1/3}(1-t_n)]$. By the Taylor series expansion, as $n \to \infty$, we have

$$\Psi_n(s) \to \frac{1}{2}\phi'(t_0)s^2$$

uniformly for $s \in K_\infty$. It will be easy to observe $\sup_{s\in K_n}|\Delta_n(s)| = O(n^{-1/3})$, so it only remains to study the behavior of $W_n(s)$. First

$$W_n(s) = \frac{1}{\sqrt{n\lambda_n}}\big\{S_{\lfloor nt_n+n\lambda_n s\rfloor} + \langle nt_n + n\lambda_n s\rangle\ \varepsilon_{\lfloor nt_n+n\lambda_n s\rfloor+1} - S_{\lfloor nt_n\rfloor} - \langle nt_n\rangle\varepsilon_{\lfloor nt_n\rfloor+1}\big\},$$

where $\langle\cdot\rangle$ denotes the fractional part. Letting

$$W_{n,1}(s) = \frac{S_{\lfloor n\lambda_n s+\lfloor nt_n\rfloor\rfloor} - S_{\lfloor nt_n\rfloor}}{\sqrt{n\lambda_n}}, \quad W_{n,2}(s) = \frac{S_{\lfloor nt_n+n\lambda_n s\rfloor} - S_{\lfloor n\lambda_n s+\lfloor nt_n\rfloor\rfloor}}{\sqrt{n\lambda_n}}$$

and

$$W_{n,3}(s) = \frac{\langle nt_n + n\lambda_n s\rangle\ \varepsilon_{\lfloor nt_n+n\lambda_n s\rfloor+1} - \langle nt_n\rangle\ \varepsilon_{\lfloor nt_n\rfloor+1}}{\sqrt{n\lambda_n}},$$

then one can write $W_n(s) = W_{n,1}(s)+W_{n,2}(s)+W_{n,3}(s)$. For brevity, we only consider the case $K_\infty = [0,a]$ with some fixed positive a, more general K_∞ can be dealt with similarly. Starting with $W_{n,3}(s)$, we can observe

$$\sup_{0\le s\le a} \frac{\varepsilon_{\lfloor nt_n+n\lambda_n s\rfloor+1}}{\sqrt{n\lambda_n}} \le \max_{\lfloor nt_n\rfloor+1\le k\le\lfloor nt_n\rfloor+\lfloor n\lambda_n a\rfloor+3} \frac{|\varepsilon_k|}{\sqrt{n\lambda_n}},$$

then by stationarity,

$$\left\|\sup_{0\le s\le a} \frac{\varepsilon_{\lfloor nt_n+n\lambda_n s\rfloor+1}}{\sqrt{n\lambda_n}}\right\| \le \frac{1}{\sqrt{n\lambda_n}} \left\|\max_{1\le k\le\lfloor n\lambda_n a\rfloor+3} |\varepsilon_k|\right\|.$$

Using the fact that, for arbitrary $M > 0$,

$$\max_{1\le i\le n} \varepsilon_i^2 \le M + \sum_{i=1}^n \varepsilon_i^2 \mathbf{1}_{\{|\varepsilon_i|>\sqrt{M}\}}.$$

It follows, in view of the mean ergodic theorem,

$$\left\|\sup_{0\le s\le a} |W_{n,3}(s)|\right\| \to 0$$

as $n\to\infty$. That $\sup_{0\le s\le a} W_{n,2}(s) = o_p(1)$ follows along the similar line.

For $W_{n,1}(s)$, first by stationarity,

$$\frac{S_{\lfloor n\lambda_n s+\lfloor nt_n\rfloor\rfloor} - S_{\lfloor nt_n\rfloor}}{\sqrt{n\lambda_n}} \stackrel{d}{=} \frac{S_{\lfloor n\lambda_n s\rfloor}}{\sqrt{n\lambda_n}} \quad \text{in } D[0,a];$$

then by essentially adapting the method in [32], pp.807-809, one can show

$$W_{n,1}(\cdot)|_{K_n} \Rightarrow \sigma W(\cdot)|_{[0,a]}$$

in $D[0,a]$ with uniform topology. Observing the limiting process is continuous with probability one, the assertion of the theorem follows. $\square$

To introduce more notations, let f be a function defined on an interval $J \subset \mathbb{R}$, and let K be a subinterval contained in J. If f is bounded from below on K, we shall use $\widetilde{f|_K}$ to denote the greatest convex minorant (GCM) of the restriction of f to K. Viewing $Z_n(\cdot)$, introduced in (2.7), as a random function in $C(I_n)$, then for each fixed n, its GCM is given by

$$\tilde{Z}_n(s) = n^{2/3}\big[\tilde{Y}_n(t_n + \lambda_n s) - Y_n(t_n) - \phi(t_n)\lambda_n s\big].$$

By the chain rule, one immediately has the relation

$$n^{\frac{1}{3}}\left[\hat{\phi}_n(t_n) - \phi(t_n)\right] = \tilde{Z}_n'(0). \tag{2.10}$$

So the behavior of $\tilde{Z}_n'(0)$ will be of major interest, but there are difficulties analyzing it if we want to apply the continuous mapping theorem. One obvious complication is that I_n is expanding, the other concern is the continuity of the functional under consideration. For these purposes, we need some technical preparations. The first is simply a restatement of Lemmas 5.1 and 5.2 of [41].

Lemma 2.2. *Let f be a bounded continuous function on a closed interval I and let $\tilde{f}$ denote its greatest convex minorant. If $a, b \in I, a < b$, and*

$$f\left(\frac{a+b}{2}\right) < \frac{\tilde{f}(a) + \tilde{f}(b)}{2},$$

then $f(x) = \tilde{f}(x)$ for some $a \leq x \leq b$.

Let $a, b \in I$ and let f^* denote the greatest convex minorant of the restriction of f to $[a,b]$. If $a \leq x_0 < x_1 \leq b$, and $f(x_i) = \tilde{f}(x_i), i = 0, 1,$ then $\tilde{f} = f^*$ on $[x_0, x_1]$.

Lemma 2.3. *Suppose (1.10) holds and $\{t_n\}$ is regular, then for each fixed $\epsilon > 0$, there exists positive $M = M(\epsilon)$, such that*

$$\liminf_{n\to\infty} P\left\{\sup_{s\in I_n}\left[|W_n(s)| - \epsilon s^2\right] \le M\right\} \ge 1-\epsilon, \tag{2.11}$$

where $W_n(s)$ is defined in (2.9), and I_n in (2.8).

Proof. We only consider $\alpha_n = n^{1/3}(1-t_n) \to \infty$, then the case $\alpha_n \to \alpha > 0$ is almost trivial. Given $\epsilon > 0$, and in view of the weak convergence of W_n, one can choose M_1 large enough such that

$$\begin{aligned}
&\limsup_{n\to\infty} P\left\{|W_n(s)| \ge M_1 + \epsilon s^2,\ \exists\ s \in [-1,1]\right\}\\
&\qquad \le \limsup_{n\to\infty} P\left\{\sup_{-1\le s\le 1} \sigma|W(s)| \ge M_1\right\} \le \frac{\epsilon}{3}.
\end{aligned} \tag{2.12}$$

On the other hand, let κ_n be the biggest k for which $[2^k, -2^{k+1}] \subset I_n$, then

$$\begin{aligned}
P&\left\{|W_n(s)| \ge M + \epsilon s^2,\ \exists\ s \in I_n \cap [1,+\infty)\right\}\\
&\le \sum_{k=0}^{\kappa_n} P\left\{|W_n(s)| \ge M + \epsilon s^2,\ \exists\ s \in [2^k, 2^{k+1}]\right\}\\
&\quad + P\left\{|W_n(s)| \ge M + \epsilon s^2,\ \exists\ s \in [2^{\kappa_n+1}, \alpha_n]\right\}\\
&\le \sum_{k=0}^{\kappa_n} P\left\{\sup_{2^k\le s\le 2^{k+1}} |W_n(s)| \ge M + \epsilon 2^{2k}\right\}\\
&\quad + P\left\{\sup_{2^{\kappa_n+1}\le s\le \alpha_n} |W_n(s)| \ge M + \epsilon 2^{2\kappa_n+2}\right\}\\
&\le \sum_{k=0}^{\kappa_n+1} \frac{c(n\lambda_n 2^{k+1}+1)}{n^{2/3}(M+\epsilon 2^{2k})^2},
\end{aligned}$$

where c is a universal constant, not depending on M and ϵ; and the last inequality in the previous display follows from stationarity and (1.11). Hence, we can find M_2 so large that

$$\limsup_{n\to\infty} P\left\{|W_n(s)| \ge M_2 + \epsilon s^2,\ \exists\ s \in I_n \cap [1,+\infty)\right\} \le \frac{\epsilon}{3}. \tag{2.13}$$

Similarly one can show an analogue of (2.13) by restricting $s \in I_n \cap (-\infty, -1]$; thus, combining (2.12) and (2.13), choosing M large enough, assertion (2.11) follows.

Lemma 2.4. *Suppose (A1) holds, and $I_n \nearrow \mathbb{R}$; then for any $0 < \epsilon < \gamma_1/2$ and $M > 0$, there is a compact interval $K \subset (0, +\infty)$ such that for all large n,*

$$\left\{|W_n(s)| \leq M + \epsilon s^2,\ s \in I_n\right\} \subset \left\{Z_n(s) - \tilde{Z}_n(s) = 0 \text{ for some } s \in K\right\}; \tag{2.14}$$

where $\tilde{Z}_n(\cdot)$ is the GCM of $Z_n(\cdot)$ in (2.9). Similarly, (2.14) holds for some $K \subset (-\infty, 0)$.

Proof. By the assumption on $\phi'(\cdot)$ and applying Taylor's formula with remainder term to $\Psi_n(s)$ in (2.9), we can find $\gamma_2 > 0$ such that

$$\frac{\gamma_1}{2} s^2 \leq \Psi_n(s) \leq \gamma_2 s^2$$

for all $s \in I_n$. For fixed $0 < \epsilon < \gamma_1/2$, $M > 0$, let

$$A_n = \left\{|W_n(s)| \leq M + \epsilon s^2,\ s \in I_n\right\};$$

then on the events A_n with n sufficiently large,

$$\left(\frac{\gamma_1}{2} - \epsilon\right) s^2 - M - \delta \leq Z_n(s) \leq (\gamma_2 + \epsilon)\, s^2 + M + \delta \tag{2.15}$$

for $s \in I_n$, where $\delta \geq \sup_{s \in I_n} |\Delta_n(s)|$. It is easy to see (2.15) is also true with Z_n replaced by $\tilde{Z}_n$. Taking $0 < a < b < \infty$ and employing (2.15),

(2.16)
$$Z_n\left(\frac{a+b}{2}\right) - \left[\frac{\tilde{Z}_n(a) + \tilde{Z}_n(b)}{2}\right] \leq (\gamma_2 + \epsilon)\left(\frac{a+b}{2}\right)^2 - \left(\frac{\gamma_1}{2} - \epsilon\right)\frac{a^2+b^2}{2} + 2M + 2\delta.$$

Noticing that

$$\left(\frac{a+b}{2}\right)^2 - \frac{a^2+b^2}{2} = \frac{-(b-a)^2}{4},$$

and taking $\delta = 1$, say, then one can find a compact interval K containing a pair of a, b, for which

$$Z_n\left(\frac{a+b}{2}\right) - \left[\frac{\tilde{Z}_n(a) + \tilde{Z}_n(b)}{2}\right] < 0.$$

Applying Lemma 2.2, one can claim $Z_n(s) = \tilde{Z}_n(s)$ for some $s \in [a, b]$, and therefore, (2.14) holds. It is not difficult to see such a compact interval $K \subset (-\infty, 0)$ can be found similarly. $\square$

Theorem 2.5. *Suppose (1.10) and (A1) hold, and* $\{t_n\}$ *is regular; let* K_0 *be any compact interval containing* 0 *as an interior point, and let* Z_n^* *denote the GCM of the restriction of* $Z_n(\cdot)$ *to* $I_n \cap K_0$*, then*

$$\lim_{K_0 \nearrow \mathbb{R}} \limsup_{n\to\infty} P\left\{\tilde{Z}_n'(0) \neq Z_n^{*'}(0)\right\} = 0. \tag{2.17}$$

Proof. Recall $Z_n(\cdot)$ in (2.9), and by Lemma 2.2, we first make the observation,

$$\left\{\tilde{Z}_n'(0) \neq Z_n^{*'}(0)\right\} \subset \left\{\tilde{Z}(s) \neq Z_n(s),\ \forall\ s \in I_n \cap K_0^\circ \cap (-\infty, 0)\right\} \tag{2.18}$$

$$\cup \left\{\tilde{Z}_n(s) \neq Z_n(s),\ \forall\ s \in I_n \cap K_0^\circ \cap (0, \infty)\right\}; \tag{2.19}$$

and it suffices to show, given any $\epsilon > 0$, there exists $\mathcal{K}_0$, such that for all $K_0 \supset \mathcal{K}_0$,

$$\limsup_{n\to\infty} P\left\{\tilde{Z}_n'(0) \neq Z_n^{*'}(0)\right\} \leq \epsilon. \tag{2.20}$$

Consider the event in (2.19) now, we shall find $\mathcal{K}_0^+$, for which

$$\liminf_{n\to\infty} P\left\{\tilde{Z}_n(s) = Z_n(s),\ \exists\ s \in I_n \cap (\mathcal{K}_0^+)^\circ \cap (0, \infty)\right\} \geq 1 - \frac{\epsilon}{2}. \tag{2.21}$$

If $\alpha_n = n^{1/3}(1 - t_n) \to \alpha > 0$, this situation is simple because one can take $\mathcal{K}_0^+$ to be containing α as an interior point; then for all large n, it is not hard to see $\alpha_n \in \mathcal{K}_0^+$ and $\tilde{Z}_n(\alpha_n) = Z_n(\alpha_n)$.

Next suppose $\alpha_n \to \infty$, which means I_n is expanding to $\mathbb{R}$. There exists $M = M(\epsilon/2)$ by Lemma 2.3, such that

$$\liminf_{n\to\infty} P\left\{|W_n(s)| \le M + \frac{\epsilon s^2}{2},\ s \in I_n\right\} \ge 1 - \frac{\epsilon}{2}.$$

In view of Lemma 2.4, such a $\mathcal{K}_0^+$ will be easy to find to make (2.21) hold. Similarly, we can find $\mathcal{K}_0^-$ for which

$$\liminf_{n\to\infty} P\left\{Z_n(s) = \tilde{Z}_n(s),\ \exists\ s \in I_n \cap (K_0^-)^\circ \cap (-\infty, 0)\right\} \ge 1 - \frac{\epsilon}{2}.$$

Taking $K_0 \supset \mathcal{K}_0 = \mathcal{K}_0^+ \cup \mathcal{K}_0^-$, using relations (2.18) and (2.19), the assertion (2.20) follows. This completes our proof. □

Theorem 2.6. *Suppose (1.10) and (A1) hold, then for* $t \in (0,1)$,

$$n^{\frac{1}{3}}\left(\frac{\hat{\phi}_n(t) - \phi(t)}{\kappa}\right) \Rightarrow \underset{-\infty<s<\infty}{\arg\min}\ [W(s) + s^2], \tag{2.22}$$

where W *is a two-sided Brownian motion, and* $\kappa = \left[\frac{1}{2}\sigma^2\phi'(t)\right]^{\frac{1}{3}}$ *with* $\sigma^2 = \lim \frac{1}{n}E[S_n^2]$; *moreover,*

$$n^{\frac{1}{3}}\left[\tilde{\mu}_{b,n} - \phi(1)\right] \Rightarrow Z^{\#} - \alpha\phi'(1),$$

where

$$Z^{\#} = D\left\{T_{(-\infty,\alpha]}\big(\sigma W(s) + \frac{1}{2}\phi'(1)s^2\big)\right\}|_{s=0},$$

with D *denoting the left-derivative, and* $T_{(-\infty,\alpha]}(\cdot)$ *denoting the GCM of a function on* $(-\infty, \alpha]$.

Proof. We first recall the relation $n^{1/3}[\hat{\phi}_n(t_n) - \phi(t_n)] = \tilde{Z}_n'(0)$. Applying Theorem 2.5 in conjuction with Theorem 3.2 of [3], it is not hard to see the behavior of $\tilde{Z}_n'(0)$ is the same as $Z_n^{*'}(0)$. It will be easy to study $Z_n^{*'}(0)$ by first fixing K_0, and then letting $n \to \infty$. Take $t_n = t \in (0,1)$, in this case $I_n \nearrow \mathbb{R}$; let $K_0 = [-m, m]$, then using Marshall's lemma and properties of convex functions, one can apply the continuous mapping theorem to $Z_n^{*'}(0)$, with K_0 fixed. So, letting $n \to \infty$ and then $m \to \infty$, one can show

$$n^{\frac{1}{3}}\left[\hat{\phi}_n(t) - \phi(t)\right] \Rightarrow D\left\{T_{(-\infty,\infty)}(\sigma W(s) + \frac{1}{2}\phi'(t)s^2)\right\}|_{s=0}$$

Assertion (2.22) follows by a standard switching argument as in [19].

Similarly, when $n^{1/3}(1-t_n) \to \alpha > 0$, one can take $K_0 = [-m, \alpha]$, and establish the second assertion. $\square$

Comparison with Anevski and Hössjer [1]. The behavior of $\hat{\phi}_n(t)$, $t \in (0,1)$, has been considered in [1] under mixing conditions. These conditions are stronger than (1.10), as shown below.

Let $\mathcal{F}_n = \sigma(\ldots, \varepsilon_{n-1}, \varepsilon_n)$ and $\mathcal{G}_n = \sigma(\varepsilon_n, \varepsilon_{n+1}, \ldots)$, define the α-mixing coefficients

$$\alpha(n) = \sup_{A \in \mathcal{F}_0, B \in \mathcal{G}_n} |P(A \cap B) - P(A)P(B)|.$$

Further denote the L^p norm of a random variable X by $\|X\|_p = E[|X|^p]^{1/p}$.

Proposition 2.7. *Assume $E[\varepsilon_i^4] < \infty$, and*

$$\sum_{n=1}^{\infty} \alpha(n)^{\frac{1}{2}-\epsilon} < \infty$$

for some $\epsilon > 0$, then (1.10) holds.

Proof. Let $X \in L^4$ and $Y \in L^2$ be two random variables, which are measurable with respect to $\mathcal{G}_n$ and $\mathcal{F}_0$, respectively. Applying a mixing inequality (e.g., [21], Corollary A.2) with $p = 2q = 4$, we have

$$|E[XY]| \leq 8\|X\|_4\|Y\|_2\, \alpha(n)^{\frac{1}{4}}. \tag{2.23}$$

Hence using the identity

$$\|E[S_n|\mathcal{F}_0]\| = \sup_{Y \in \mathcal{F}_0, \|Y\| \leq 1} \int Y E[S_n|\mathcal{F}_0]\, dP = \sup_{Y \in \mathcal{F}_0, \|Y\| \leq 1} \sum_{k=1}^{n} E[\varepsilon_k Y],$$

in conjuction with (2.23), one can show

$$\sum_{n=1}^{\infty} n^{-3/2}\|E[S_n|\mathcal{F}_0]\| \leq 8\|\varepsilon_0\|_4 \sum_{n=1}^{\infty} n^{-3/2}\sum_{k=1}^{n}\alpha(k)^{\frac{1}{4}} \leq 24\|\varepsilon_0\|_4 \sum_{k=1}^{\infty}\frac{\alpha(k)^{\frac{1}{4}}}{\sqrt{k}}. \tag{2.24}$$

It is not hard to see this series is summable. Since for each given $\epsilon > 0$, we can take $q' < 2$ for which $\frac{1}{4}q' > \frac{1}{2} - \epsilon$, and $p' > 2$ for which $\frac{1}{p'} + \frac{1}{q'} = 1$, such that

$$\sum_{k=1}^{\infty}\frac{1}{\sqrt{k}}\,\alpha(k)^{\frac{1}{4}} \leq \left(\sum_{k=1}^{\infty} k^{-\frac{1}{2}p'}\right)^{\frac{1}{p'}}\left(\sum_{k=1}^{\infty}\alpha(k)^{\frac{1}{4}q'}\right)^{\frac{1}{q'}};$$

and this is finite by assumption. $\square$

Penalized Least Squares Estimators Let $\beta_n = \beta n^{1/3}$ for some fixed positive β; the penalized least squares estimator is given by

$$\hat{\mu}_{p,n} = \max_{1\leq j\leq n}\frac{y_j + \cdots + y_n - \beta_n}{n-j+1} = \max_{1\leq j\leq n}\frac{y_{n+1-j} + \cdots + y_n - \beta_n}{j}.$$

Then clearly

$$\hat{\mu}_{p,n} - \phi(1) = \max_{1\leq j\leq n}\frac{S_j^{(n)} + \Phi_j^{(n)} - \beta_n}{j},$$

where $\Phi_j^{(n)} = \mu_{n+1-j} - \phi(1) + \cdots + \mu_n - \phi(1)$ and $S_j^{(n)} = S_n - S_{n-j}$. Next, define

$$G_n(t) := S^{(n)}_{\lfloor n^{2/3}t\rfloor} + \Phi^{(n)}_{\lfloor n^{2/3}t\rfloor} - \langle n^{2/3}t\rangle\left[\varepsilon_{\lfloor n^{2/3}t\rfloor} + \phi\left(\frac{n+\lceil n^{2/3}t\rceil}{n}\right) - \phi(1)\right].$$

and

$$\Lambda_n := \sup_{t\in J_n}\frac{G_n(t) - \beta_n}{\lfloor n^{2/3}t\rfloor},$$

where $J_n = [n^{-2/3}, n^{1/3}]$. It is then not hard to verify $\hat{\mu}_{p,n} - \phi(1) = \Lambda_n$.

Lemma 2.8. *Suppose (1.10) and (A1) hold. Then for any compact interval $K \subset [0,\infty)$,*

$$n^{-\frac{1}{3}}G_n(\cdot)|_{K\cap J_n} \Rightarrow G(\cdot)|_K$$

in $C(K)$, where $G(t) = \sigma W(t) - \frac{1}{2}\phi'(1)t^2$ with $W(\cdot)$ denoting the standard two-sided Brownian motion.

Proof. First, by the mean value theorem,

$$n^{-\frac{1}{3}}\Phi^{(n)}_{\lfloor n^{2/3}t\rfloor} = n^{-\frac{1}{3}} \sum_{k=n+1+\lfloor n^{\frac{2}{3}}t\rfloor}^{n} \left[\phi\left(\frac{k}{n}\right) - \phi(1)\right] = -n^{-\frac{1}{3}} \sum_{k=n+1+\lfloor n^{\frac{2}{3}}t\rfloor}^{n} \phi'(t^*)\left(\frac{n-k}{n}\right)$$

for $t^* = t^*(k,n,t)$ which is contained in the interval$[(n+1+\lfloor n^{2/3}t\rfloor)/n, 1]$; hence,

$$n^{-\frac{1}{3}}\Phi^{(n)}_{\lfloor n^{2/3}t\rfloor} \to -\frac{1}{2}\phi'(1)t^2$$

uniformly for $t \in K$. Next, utilizing a similar truncation argument as in Theorem 2.1, it is straightforward to show

$$n^{-\frac{1}{3}}\left\|\sup_{t\in K\cap J_n} \varepsilon_{\lfloor n^{2/3}t\rfloor}\right\| \to 0$$

as $n \to \infty$. The assertion of the lemma then follows from Theorem 1 of [32]. $\square$

Lemma 2.9. *Suppose (1.10) and (A1) hold. Then given any $\epsilon > 0$ and $b > 0$, there exists δ_0 such that for all $0 < \delta \le \delta_0$,*

$$\limsup_{n\to\infty} P\left\{\sup_{1/\delta\le t\le n^{1/3}} \frac{n^{1/3}[G_n(t)-\beta_n]}{\lfloor n^{2/3}t\rfloor} > -b\right\} < \epsilon. \tag{2.25}$$

Proof. For any fixed $\delta > 0, b > 0$ and all large n, by simple calculations,

$$\begin{aligned} &P\left\{\sup_{1/\delta\le t\le n^{1/3}} \frac{n^{1/3}[G_n(t)-\beta_n]}{\lfloor n^{2/3}t\rfloor} > -b\right\} \\ &\quad\le P\left\{\frac{S_j^{(n)} + \Phi_j^{(n)} - \beta_n}{j} > -n^{-\frac{1}{3}}b,\ \exists\, n^{2/3}/\delta \le j \le n\right\} \\ &\quad= P\left\{S_j > -n^{-\frac{1}{3}}jb + \beta_n - \Phi_j^{(n)},\ \exists\, n^{2/3}/\delta \le j \le n\right\}; \end{aligned} \tag{2.26}$$

on the other hand, by the mean value theorem,

$$-\Phi_j^{(n)} \ge \sum_{k=n+1-j}^{n} \gamma_1\left(\frac{n-k}{n}\right) = \frac{\gamma_1(j^2-j)}{2n}.$$

So, over the range $j \ge n^{2/3}/\delta$, the leading term in (2.26) is given by j^2/n for large n (as long as δ is bounded from above). In particular, if $\delta < \gamma_1/(4b)$, then the probability in (2.26) is majorized by

$$P\left\{S_j > \frac{\gamma_1}{8}\frac{j^2}{n},\ \exists\, j \geq n^{2/3}/\delta\right\},$$

for all large n. To estimate this probability, take integer m for which $2^{m-1} < n^{2/3}/\delta \leq 2^m$, then

$$\begin{aligned}
&P\left\{S_j > \gamma^*\frac{j^2}{n},\ \exists\, j \geq \frac{n^{2/3}}{\delta}\right\}\\
&\quad\leq P\left\{S_j > \gamma^*\frac{j^2}{n}\ \exists\, j > 2^{m-1}\right\}\\
&\quad\leq \sum_{k=m}^{\infty} P\left\{S_j > \frac{\gamma^*}{n}(2^{k-1})^2,\ \exists\, 1 \leq j \leq 2^k\right\}\\
&\quad\leq \sum_{k=m}^{\infty} \frac{n^2}{\gamma^{*2}2^{4(k-1)}} E\left[\max_{1\leq j\leq 2^k} S_j^2\right]\\
&\quad\leq cn^2 2^{-3m}
\end{aligned}$$

for some positive constant c, after an application of the maximal inequality (1.11). Since $n^2 2^{-3m} \leq \delta^3$, the probability in the previous display can be made arbitrarily small by letting $\delta \to 0$. $\square$

Proposition 2.10. *Suppose (1.10) and (A1) hold, then*

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left\{\sup_{n^{-\frac{2}{3}}\leq t\leq n^{\frac{1}{3}}} \frac{G_n(t)-\beta_n}{\lfloor n^{2/3}t\rfloor} \neq \sup_{\delta\leq t\leq 1/\delta} \frac{G_n(t)-\beta_n}{\lfloor n^{2/3}t\rfloor}\right\} = 0. \tag{2.27}$$

Proof. It will be shown that the supremum of $n^{1/3}[G_n(t)-\beta_n]/\lfloor n^{2/3}t\rfloor$ is unlikely to be achieved on $[n^{-2/3},\delta]$ and $[1/\delta, n^{1/3}]$ when n is large. In view of Lemma 2.9, it suffices to show given any $\epsilon > 0$ and $b > 0$, then for all sufficiently small $\delta > 0$,

$$\limsup_{n\to\infty} P\left\{\sup_{n^{-\frac{2}{3}}\leq t\leq \delta} \frac{n^{\frac{1}{3}}[G_n(t)-\beta_n]}{\lfloor n^{2/3}t\rfloor} > -b\right\} < \epsilon. \tag{2.28}$$

To see that, applying Lemma 2.8 we have

$$
\begin{aligned}
P&\left\{n^{\frac{1}{3}}[G_n(t)-\beta_n] > -\lfloor n^{2/3}t\rfloor b,\ \exists\ n^{-2/3}\le t\le\delta\right\}\\
&\le P\left\{n^{\frac{1}{3}}[G_n(t)-\beta_n] > -b(n^{2/3}\delta+1),\ \exists\ n^{-2/3}\le t\le\delta\right\}\\
&= P\left\{n^{-\frac{1}{3}}G_n(t) > \beta-b(\delta+n^{-2/3}),\ \exists\ n^{-2/3}\le t\le\delta\right\}\\
&\le P\left\{\max_{0\le t\le\delta} n^{-\frac{1}{3}}G_n(t) > \beta-b(\delta+n^{-2/3})\right\}\\
&\to P\left\{\max_{0\le t\le\delta}\left[\sigma W(t)-\frac{1}{2}\phi'(1)t^2\right] > \beta-b\delta\right\}
\end{aligned}
$$

as $n\to\infty$. On the other hand, one can choose δ small enough such that $\beta-b\delta>\beta/2$, then

$$
P\left\{\max_{0\le t\le\delta}\left[\sigma W(t)-\frac{1}{2}\phi'(1)t^2\right] > \beta-b\delta\right\} \le P\left\{\max_{0\le t\le\delta} W(t) > \frac{\beta}{2\sigma}\right\} = 2\left[1-\Phi_{0,1}\left(\frac{\beta}{2\sigma\sqrt{\delta}}\right)\right],
$$

where $\Phi_{0,1}$ denotes the standard normal distribution. It is easy to see the probability in the previous display can be made arbitrarily small if δ is sufficiently close to 0, establishing (2.28). □

Theorem 2.11. *Under the hypotheses of Proposition 2.10,*

$$
n^{\frac{1}{3}}[\hat{\mu}_{p,n}-\phi(1)] \Rightarrow \sup_{t>0}\frac{\sigma W(t)-\beta-\frac{1}{2}\phi'(1)t^2}{t}, \tag{2.29}
$$

as $n\to\infty$.

Proof. Using the relation $\hat{\mu}_{p,n}-\phi(1)=\Lambda_n$ and Proposition 2.10, assertion (2.29) follows by an application of Theorem 3.2 of [3]. □

2.3 Numerical Studies

In this section, we report simulation studies to compare the performance of $\hat{\mu}_{p,n}$ and $\tilde{\mu}_{b,n}$. For simplicity, we choose the errors to be AR(1) with autoregressive parameter .25, and the regression functions to be either convex or concave near the right end point of the unit interval. In order to compute our estimates, there are issues of choosing smoothing parameters. Recall

$$\hat{\mu}_{p,n} = \max_{1 \le j \le n} \frac{y_j + \cdots + y_n - \beta n^{\frac{1}{3}}}{n - j + 1},$$

which depends on β; and for each fixed β,

$$n^{\frac{1}{3}}[\hat{\mu}_{p,n} - \phi(1)] \Rightarrow S_\beta(\sigma, \gamma) := \sup_{t>0} \frac{\sigma W(t) - \beta - \gamma t^2}{t}$$

where $\gamma = \phi'(1)/2$. Also we can recall $\tilde{\mu}_{b,n} = \hat{\phi}_n(1 - \alpha n^{-\frac{1}{3}})$, with limiting behavior

$$n^{\frac{1}{3}}[\tilde{\mu}_{b,n} - \phi(1)] \Rightarrow Z_\alpha(\sigma, \phi'(1)),$$

where

$$Z_\alpha(\sigma, \phi'(1)) = D\left\{T_{(-\infty,\alpha]}\big(\sigma W(s) + \frac{1}{2}\phi'(1)s^2\big)\right\}|_{s=0} - \alpha\phi'(1).$$

So, the choice of α has to be made to implement our estimation procedure. Here we choose α and β to minimize $E[Z_\alpha(\sigma, \phi'(1))^2]$ and $E[S_\beta(\sigma, \gamma)^2]$ respectively, supposing other parameters are given.

Moments of both $Z_\alpha(\sigma, \cdot)$ and $S_\beta(\sigma, \gamma)$ are apparently hard to get, but we can replace them by Monte Carlo estimates. To select α, we first generate two-sided Brownian paths using random walk approximations with step size 0.001, on the interval $[-2, \alpha]$. Then, based on each realization of discrete observations combining the drift term, we can compute the isotonic estimate corresponding to knot 0. Averaging over 1000 realizations gives us the Monte Carlo estimates. Similarly, we can select β, we refer interested readers to [38] for more details. Shown in Figure 2.1 is a picture suggesting the choice of α and β, when the regression function is chosen to be $\phi(x) = (2x - 1)^3 + 1$. In this case, $\sigma = 4/3, \gamma = 3 = \phi'(1)/2$, and the minima seem to be occuring at $\alpha \approx 0.17, \beta \approx 0.68$.

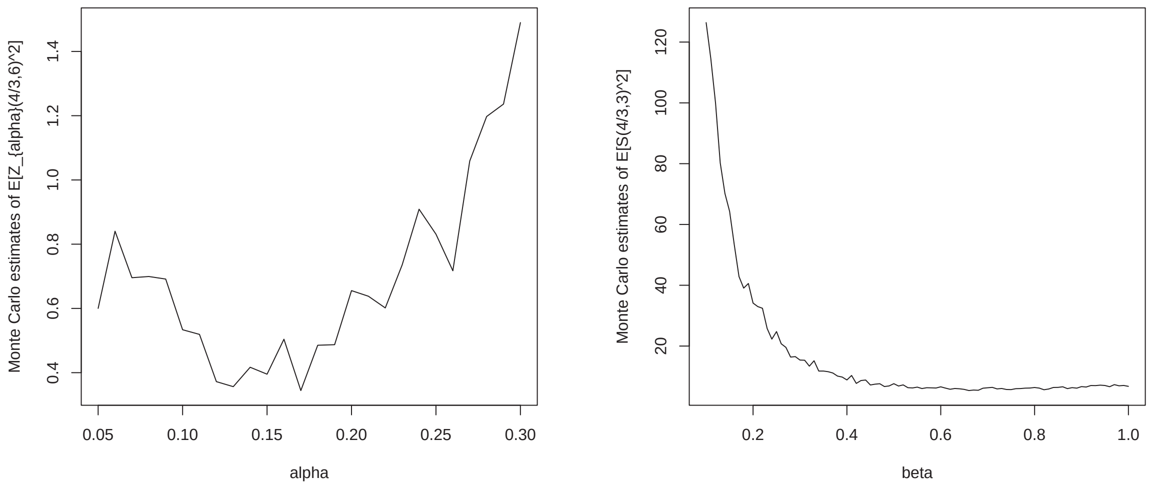

Figure 2.1: Choosing Smoothing Parameters for $\phi(x) = (2x-1)^3 + 1$

To compare the performance of estimators using mean squared error, we simulate 10,000 samples of sizes $n = 50, 100, 200$. For each sample, $\hat{\mu}_{p,n} - \phi(1)$ and $\tilde{\mu}_{b,n} - \phi(1)$ are computed with the suggested choices of α and β. Table 2.1 summarizes the comparison.

Table 2.1: $\phi(x) = (2x-1)^3 + 1$

	n	mean	var	mse
$\hat{\mu}_{p,n} - \phi(1)$	50	-0.34928	0.18234	0.30434
	100	-0.28044	0.13767	0.21632
	200	-0.21823	0.09680	0.14443
$\tilde{\mu}_{b,n} - \phi(1)$	50	-0.17655	0.23416	0.26533
	100	-0.16146	0.16457	0.19064
	200	-0.10302	0.11785	0.12847

Note: errors are AR(1) with autoregressive parameter .25

In the previous study, the regression function is convex near the end point. Next, we shall look at a scenario where the regression function is concave. Take $\phi(x) = \exp(13x-9)/(1+\exp(13x-9))$ and the autoregressive parameter still be .25, so $\sigma = 4/3$, and $\phi(1) = 0.982$, $\phi'(1) = 0.23$. Following a similar procedure as above, we are suggested to choose $\alpha \approx .35$ and $\beta \approx .28$; reported in Table 2.2 is a comparison of the performance. In both scenarios, the boundary corrected isotonic estimator, $\tilde{\mu}_{b,n}$, outperforms the penalized LSE. Overall, both estimators are reasonable by looking at the total mean squared errors, which are quite small. The Monte Carlo estimates of the means do not drop down significantly as sample size increases. This is partially due to the moderate sample sizes that we have chosen; but still, the biases are quite negligible.

Table 2.2: $\phi(x) = \exp(13x - 9)/(1 + \exp(13x - 9))$

	n	mean	var	mse
$\hat{\mu}_{p,n} - \phi(1)$	50	0.21352	0.24516	0.29075
	100	0.23784	0.18224	0.23881
	200	0.22535	0.13246	0.18324
$\tilde{\mu}_{b,n} - \phi(1)$	50	0.04325	0.12557	0.12744
	100	0.05203	0.10230	0.10501
	200	0.05124	0.05285	0.05548

Note: errors are AR(1) with parameter .25

Global temperature anomalies. There are $n = 150$ annualy observations for this time series in the period 1850-2000. One major thorny issue is to estimate $\phi'(1)$, the change rate of the underlying regression function at 1. We decided to fit an ordinary regression model with $x_i = i/n, i = 1, \ldots, n$, and y_i as the observations, using a second order polynomial. Based on the estimates, we are suggested $\hat{\phi}'(1) \approx 1.75$; for the residuals, we fit an ARMA model, and it seems an AR(1) gives the best fit. A 95% confidence interval for the autoregressive parameter is given by (0.17,0.39), the midpoint 0.28 is taken to be our estimate, so $\hat{\sigma} \approx 1.39$. Based on these estimates, we use the same criteria as before to choose α, β, and they are suggested as $\alpha = 0.3$, $\beta = 0.16$. Next, we study the distribution of $S_{0.16}(1.39, 0.875)$; Table 2.3 below presents the Monte Carlo estimates for the distribution function. The study of $Z_{0.3}(1.39, 1.75)$ is computationally very intensive, and we have not succeeded at this moment. These results may be used to set confidence intervals for $\phi(1)$.

Table 2.3: Monte Carlo estimates of $F(x) = P\{S_{.16}(1.39, 0.875) \leq x\}$

x	$\hat{F}(x)$	x	$\hat{F}(x)$	x	$\hat{F}(x)$	x	$\hat{F}(x)$
-2.9	0.0000	-1.4	0.0080	0.1	0.7335	1.6	0.9910
-2.8	0.0000	-1.3	0.0180	0.2	0.7795	1.7	0.9925
-2.7	0.0000	-1.2	0.0300	0.3	0.8215	1.8	0.9945
-2.6	0.0000	-1.1	0.0505	0.4	0.8530	1.9	0.9950
-2.5	0.0000	-1.0	0.0805	0.5	0.8900	2.0	0.9965
-2.4	0.0000	-0.9	0.1185	0.6	0.9155	2.1	0.9975
-2.3	0.0000	-0.8	0.1670	0.7	0.9310	2.2	0.9975
-2.2	0.0000	-0.7	0.2195	0.8	0.9420	2.3	0.9990
-2.1	0.0000	-0.6	0.2855	0.9	0.9525	2.4	0.9995
-2.0	0.0000	-0.5	0.3575	1.0	0.9630	2.5	0.9995
-1.9	0.0000	-0.4	0.4335	1.1	0.9735	2.6	0.9995
-1.8	0.0005	-0.3	0.4960	1.2	0.9790	2.7	0.9995
-1.7	0.0010	-0.2	0.5645	1.3	0.9835	2.8	1.0000
-1.6	0.0020	-0.1	0.6210	1.4	0.9880	2.9	1.0000
-1.5	0.0060	0.0	0.6800	1.5	0.9900	3.0	1.0000

Estimates are based on 2000 simulated data points.

CHAPTER III

Martingale Approximations

3.1 Main Results

Some notation is necessary to describe the results of this chapter. Let $\ldots W_{-1}, W_0$, $W_1, \ldots$ denote a stationary, ergodic Markov chain with values in a measurable space $\mathcal{W}$. The marginal distribution and transition function of the chain are denoted by π and Q; thus, $\pi\{B\} = P[W_n \in B]$ and $Q(w; B) = P[W_{n+1} \in B | W_n = w]$ for $w \in \mathcal{W}$ and measurable sets $B \subseteq \mathcal{W}$. In addition, Q denotes the operator, defined by

$$Qf(w) = \int_{\mathcal{W}} f(z)Q(w; dz) \ a.e. \ (\pi)$$

for $f \in L^1(\pi)$, and the iterates of Q are denoted by $Q^k = Q \circ \cdots \circ Q$ (k times). Thus, $Q^k f(w) = E[f(W_{n+k})|W_n = w]$ *a.e.* (π) for $f \in L^1(\pi)$. The probability space on which $\ldots, W_{-1}, W_0, W_1, \ldots$ are defined is denoted by $(\Omega, \mathcal{A}, P)$, and $\mathcal{F}_n = \sigma\{\ldots, W_{n-1}, W_n\}$. Finally, $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the norm and inner product in an L^2 space, which may vary from one usage to the next.

Observe that no stringent conditions, like Harris recurrence or even irreducibility, have been placed on the Markov chain. In particular, if $\ldots \xi_{-1}, \xi_0$, $\xi_1, \ldots$ are i.i.d. with common distribution ρ say, then the *shift process* $W_k = (\ldots \xi_{k-1}, \xi_k)$ satisfies the conditions placed on the chain with $\pi = \rho^{\mathbb{N}}$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$, and $Qg(w) = \int g(w, x)\rho\{dx\}$ for $g \in L^1(\pi)$. Shift processes abound in books on time series–for example, [6] and [39].

Next let $L_0^2(\pi)$ be the set of $g \in L^2(\pi)$ for which $\int_{\mathcal{W}} g d\pi = 0$; and, for $g \in L_0^2(\pi)$, consider stationary sequences of the form $X_k = g(W_k)$ and their sums $S_n = X_1 + \cdots + X_n$. Thus,

$$S_n = S_n(g) = g(W_1) + \cdots + g(W_n).$$

The question addressed here is the existence of a martingale $M_1, M_2, \ldots$ with respect to $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$ having stationary increments and a sequence of remainder terms $R_1, R_2, \ldots$ for which $\|R_n\| = o(\sqrt{n})$ and

$$S_n = M_n + R_n. \tag{3.1}$$

If (3.1) holds, we say that g admits a *martingale approximation.* Ever since the work of Gordin [15], martingale approximations have been an effective tool for studying the (conditional) central limit question and law of the iterated logarithm for stationary processes; see, for example, [7], [45], [8], [47], and their references for recent developments. The terminology here differs slightly from that of [45].

The sequence $X_k = g(W_k)$ is said to admit a *co-boundary* if there is a stationary sequence of martingale differences d_k and another stationary process Z_k for which

$$X_k = d_k + Z_k - Z_{k-1},$$

for all k, in which case $S_n = \tilde{M}_n + \tilde{R}_n$ with $\tilde{M}_n = d_1 + \cdots + d_n$ and $\tilde{R}_n = Z_n - Z_0$. Here $\tilde{M}_n$ is a martingale and $\tilde{R}_n$ is stochastically bounded, but does not necessarily satisfy $\|\tilde{R}_n\| = o(\sqrt{n})$. Conversely, a martingale approximation does not require R_n to be stochastically bounded. The relation between co-boundaries and martingale approximations is further clarified by the examples of [13].

Letting Q^* denote the adjoint of the restriction of Q to $L^2(\pi)$, so that $\langle Qf, g\rangle = \langle f, Q^*g\rangle$ for $f, g \in L^2(\pi)$, Q is said to be a co-isometry if $QQ^* = I$, in which case Q^*

is an isometry. Importantly, this condition is satisfied by shift processes. In Section 3.3, a convenient orthonormal basis for $L_0^2(\pi)$ is identified when Q is a co-isometry, and a simple necessary and sufficient condition for the existence of a martingale approximation is given in terms of the coefficients in the expansion of g with respect to this basis.

Returning to the main question, define

$$V_n g = \sum_{k=0}^{n-1} Q^k g,$$

so that $E(S_n|\mathcal{F}_1) = V_n g(W_1)$. If (3.1) holds, then $\|V_n g\|^2 = E[E(S_n|\mathcal{F}_1)^2] \leq 2E(M_1^2) + 2E(R_n^2) = o(n)$, and $\lim_{n\to\infty} E(S_n^2)/n = E(M_1^2)$. So, obvious necessary conditions for (3.1) are that

$$\|V_n g\| = o(\sqrt{n}) \tag{3.2}$$

and

$$\|g\|_+^2 := \limsup_{n\to\infty} \frac{1}{n} E[S_n(g)^2] < \infty. \tag{3.3}$$

Let $\mathcal{L}$ denote the set of $g \in L_0^2(\pi)$ for which $\|g\|_+ < \infty$. Then $\mathcal{L}$ is a linear space, and $\|\cdot\|_+$ is a pseudo norm on $\mathcal{L}$, called the *plus norm* below. Moreover, Q maps $\mathcal{L}$ into itself, since

$$S_n(g) = S_n(Qg) + \sum_{k=1}^{n}[g(W_k) - Qg(W_{k-1})] + Qg(W_0) - Qg(W_n); \tag{3.4}$$

and, therefore, $\|Qg\|_+ \leq \|g\|_+ + \sqrt{E\{[g(W_1) - Qg(W_0)]^2\}}$. In Section 3.4 it is shown that g admits a martingale approximation iff (3.2) holds and

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \|Q^k g\|_+^2 = 0.$$

These results are used in Section 3.5 to study the relationship between martingale approximations and solutions to the fractional Poisson equation, $g = \sqrt{(I-Q)}h$.

The relation between martingale approximations and the conditional central limit theorem is explored in Section 3.6 with special attention to superpositions of linear processes. Section 3.2 contains some preliminaries.

3.2 Preliminaries

In this section, upon exhibiting some preliminary facts, we establish a useful criterion for martingale approximations; and in particular, we show martingale approximations are unique. Let

$$\bar{V}_n = \frac{V_1 + \cdots + V_n}{n} = \sum_{k=0}^{n-1} \left(1 - \frac{k}{n}\right) Q^k.$$

Then

$$E[S_n(g)^2] = 2n\langle g, \bar{V}_n g\rangle - n\|g\|^2, \tag{3.5}$$

from [6], p. 219, and

$$\bar{V}_n = \bar{V}_n Q^i + V_i - \frac{1}{n} Q V_n V_i \tag{3.6}$$

for all $n \geq 1$, $i \geq 1$ by simple algebra and induction. Next, let π_1 denote the joint distribution of W_0 and W_1, define

$$H_n(w_0, w_1) = V_n g(w_1) - Q V_n g(w_0) \tag{3.7}$$

and

$$\bar{H}_n(w_0, w_1) = \bar{V}_n g(w_1) - Q \bar{V}_n g(w_0) \tag{3.8}$$

for $w_0, w_1 \in \mathcal{W}$. Then H_n and $\bar{H}_n$ are in $L^2(\pi_1)$.

Lemma 3.1. *If (3.2) holds, then* $S_k = M_{nk} + R_{nk}$ *where* $M_{nk} = \bar{H}_n(W_0, W_1) + \cdots + \bar{H}_n(W_{k-1}, W_k)$ *and* $\max_{k \leq n} \|R_{nk}\| = o(\sqrt{n})$.

Proof. The lemma is almost a special case of Theorem 1 of [45]. Using Equation (3.6) with $i = 1$,

$$R_{nk} = S_k - M_{nk} = Q\bar{V}_n g(W_0) - Q\bar{V}_n g(W_k) + \frac{1}{n} S_k(QV_n g),$$

from which it follows that $\max_{k \le n} \|R_{nk}\| \le 3 \max_{k \le n} \|V_k g\|$, which is $o(\sqrt{n})$ by (3.2). □

Of course, M_{nk} is a martingale in k for each n. The following proposition is closely related to Theorem 1 of [40].

Proposition 3.2. *$g \in L_0^2(\pi)$ admits a martingale approximation iff (3.2) holds and $\bar{H}_n$ converges to a limit H in $L^2(\pi_1)$, in which case*

$$M_n = M_n(g) := \sum_{k=1}^{n} H(W_{k-1}, W_k). \tag{3.9}$$

Consequently, martingale approximations are unique.

Proof. Suppose first that g admits a martingale approximation, $S_n = M_n + R_n$. Then (3.2) holds and $S_n = M_{nn} + R_{nn}$, where $\|R_{nn}\| = o(\sqrt{n})$, by Lemma 3.1. So,

$$nE\{[\bar{H}_n(W_0, W_1) - M_1]^2\} = E[(M_{nn} - M_n)^2] = E[(R_{nn} - R_n)^2] = o(n),$$

implying the convergence of $\bar{H}_n(W_0, W_1)$ in $L^2(P)$; and this is equivalent to the convergence of $\bar{H}_n$ in $L^2(\pi_1)$.

Conversely, if (3.2) holds and $\bar{H}_n$ converges to a limit H, say; we can let $M_n = H(W_0, W_1) + \cdots + H(W_{n-1}, W_n)$ and $R_n = S_n - M_n$. Then (3.1) holds, $R_n = M_{nn} - M_n + R_{nn}$, and $\|R_n\| \le \sqrt{n}\|\bar{H}_n - H\| + \|R_{nn}\| = o(\sqrt{n})$, establishing both the sufficiency and (3.9). That martingale approximations are unique is then clear. □

As a first use of Proposition 3.2, we shall recover Theorem 1 of [30]. To begin, let $g_n = V_n g$ and $\bar{g}_n = \bar{V}_n g$, recall the Maxwell-Woodroofe condition

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|V_n g\| < \infty. \tag{3.10}$$

Proposition 3.3. *Suppose (3.10) holds, then the conditions of Proposition 3.2 are verified, and therefore, g has a martingale approximation.*

To prove Proposition 3.3, we need two lemmas.

Lemma 3.4. *Let $\bar{H}_n$ be as defined in (3.8), then*

$$\|\bar{H}_n - \bar{H}_m\| \le \frac{1}{\sqrt{n}} \left(\|\bar{g}_m\| + \|\bar{g}_n\| + 2\|g_n\| + \frac{2n}{m}\|g_m\| \right). \tag{3.11}$$

PROOF By simple algebra, one has

$$\|\bar{H}_n - \bar{H}_m\|^2 = \|\bar{H}_n\|^2 + \|\bar{H}_m\|^2 - 2\langle \bar{H}_n, \bar{H}_m \rangle$$

and

$$\langle \bar{H}_n, \bar{H}_m \rangle = \langle \bar{g}_n, \bar{g}_m \rangle - \langle Q\bar{g}_n, Q\bar{g}_m \rangle.$$

Using the identity $Q\bar{V}_n = \bar{V}_n - I + \frac{1}{n}QV_n$ to expand $\langle Q\bar{g}_n, Q\bar{g}_m \rangle$ out, and by Cauchy-Schwarz inequality, we have

$$\|\bar{H}_n - \bar{H}_m\|^2 \le \left[\frac{\|g_n\|}{n} + \frac{\|g_m\|}{m} \right]^2 + 2\left[\frac{\|g_n\|}{n} + \frac{\|g_m\|}{m} \right] (\|\bar{g}_n\| + \|\bar{g}_m\|).$$

It is then easy to see

$$\|\bar{H}_n - \bar{H}_m\|^2 \le \left(\frac{1}{n^2} + \frac{1}{n} \right) \left(\|g_n\| + \frac{n}{m}\|g_m\| \right)^2 + \frac{1}{n}\left(\|\bar{g}_n\| + \|\bar{g}_m\| \right)^2,$$

establishing (3.11). $\square$

Lemma 3.5. *Fix $p > 1$, let $\ell(\cdot)$ be any nonnegative, slowly-varing function (at ∞); and let ν_n be a nonnegave, subadditive sequence. The following two conditions are equivalent:*

(i)

$$\sum_{n=1}^{\infty} n^{-p}\ell(n)\nu_n < \infty; \tag{3.12}$$

(ii)

$$\sum_{k=0}^{\infty} \frac{\ell(2^k)\bar{\nu}_{2^k}}{2^{k(p-1)}} < \infty, \tag{3.13}$$

where $\bar{\nu}_n = (\nu_1 + \nu_2 + \cdots + \nu_n)/n$.

PROOF. We first show (i)$\Rightarrow$ (ii) . Making a change of summation yields

$$\sum_{k=0}^{\infty} \frac{\ell(2^k)\bar{\nu}_{2^k}}{2^{k(p-1)}} = \sum_{k=0}^{\infty} \frac{\ell(2^k)}{2^{kp}} \left(\sum_{n=1}^{2^k} \nu_n \right) = \sum_{n=1}^{\infty} \left(\sum_{k=m_n}^{\infty} \frac{\ell(2^k)}{2^{kp}} \right) \nu_n,$$

where $m_n = \lceil \log_2 n \rceil$. Since $\ell(\cdot)$ is slowly-varing, one can find a positive integer k_0 for which, $\ell(2^k)/2^{kp}$ is nonincreasing in $k \geq k_0$. So, when $m_n \geq k_0 + 1$,

$$\sum_{k=m_n}^{\infty} \frac{\ell(2^k)}{2^{kp}} \leq \int_{m_n - 1}^{\infty} \frac{\ell(2^x)}{2^{xp}}\, dx \leq \frac{1}{\log 2} \int_{\frac{1}{2}n}^{\infty} \frac{\ell(y)}{y^{p+1}}\, dy \sim \frac{2^p}{p \log 2} \frac{\ell(n)}{n^p},$$

as $n \to \infty$, using Karamata's theorem [4, p. 27]. Hence, it follows that

$$\sum_{n=1}^{\infty} \left(\sum_{k=m_n}^{\infty} \ell(2^k) 2^{-kp} \right) \nu_n < \infty,$$

establishing (i) $\Rightarrow$ (ii).

To see (ii) $\Rightarrow$ (i), fix any integer $k \geq 0$, by subadditivity $\nu_n \leq \nu_k + \nu_{n-k}, k = 1, \ldots, n$, it is not hard to check

$$\nu_n \leq \sum_{j=0}^{k} \nu_{2^j}$$

for any integer $n \in [2^k, 2^{k+1} - 1]$. It then follows that

$$\sum_{n=1}^{\infty} \frac{\ell(n)\nu_n}{n^p} = \sum_{k=0}^{\infty} \sum_{n=2^k}^{2^{k+1}-1} n^{-p} \ell(n) \nu_n \leq \sum_{k=0}^{\infty} \sum_{n=2^k}^{2^{k+1}-1} n^{-p} \ell(n) \left(\sum_{j=0}^{k} \nu_{2^j} \right)$$

Choose n_0 large enough such that $n^{-p}\ell(n)$ is nonincreasing in $n \geq n_0$, then one can find K_0, such that for all $k \geq K_0$,

$$\sum_{n=2^k}^{2^{k+1}-1} n^{-p} \ell(n) \left(\sum_{j=0}^{k} \nu_{2^j} \right) \leq 2^{-k(p-1)} \ell(2^k) \left(\sum_{j=0}^{k} \nu_{2^j} \right).$$

Thus,

$$\sum_{k=0}^{\infty}\sum_{n=2^k}^{2^{k+1}-1} n^{-p}\ell(n)\left(\sum_{j=0}^{k}\nu_{2^j}\right) \ll \sum_{j=0}^{\infty}\nu_{2^j}\sum_{k=j}^{\infty}2^{-k(p-1)}\ell(2^k),$$

where $\ll$ is used in place of big O notation. Further by an integral test and Fatou's lemma,

$$\sum_{k=j}^{\infty}2^{-k(p-1)}\ell(2^k) = 2^{-j(p-1)}\sum_{k=0}^{\infty}2^{-k(p-1)}\ell(2^{k+j}) \ll 2^{-j(p-1)}\ell(2^j).$$

So (3.12) holds in view of $\nu_n \leq 2\bar{\nu}_{n-1} \leq 4\bar{\nu}_n, n = 2, \ldots$. □

Proof of Proposition 3.3. Let $\nu_n = \|g_n\|$, and $\bar{\nu}_n$ its Cesàro average. By Lemma 3.4, and using Lemma 3.5 with $p = 3/2$, $\ell(\cdot)$ =constant, we have

$$\sum_{k=1}^{\infty}\|\bar{H}_{2^k} - \bar{H}_{2^{k-1}}\| \leq \sum_{k=1}^{\infty}\frac{4}{\sqrt{2^k}}\left(\bar{\nu}_{2^{k-1}} + \bar{\nu}_{2^k} + \nu_{2^k} + \nu_{2^{k-1}}\right) < \infty.$$

On the other hand, by the subadditivity of ν_n, $\nu_n \leq 2\bar{\nu}_{n-1} \leq 4\bar{\nu}_n$ for $n \geq 2$; and also, $\bar{\nu}_m \leq 2\bar{\nu}_{2^k}$ for all $2^{k-1} < m \leq 2^k, k = 1, 2, \ldots$. It follows that

$$\begin{aligned}\max_{2^{k-1}<m\leq 2^k}\|\bar{H}_m - \bar{H}_{2^k}\| &\leq \max_{2^{k-1}<m\leq 2^k}\frac{1}{\sqrt{2^k}}\left[\|\bar{g}_m\| + \|\bar{g}_{2^k}\| + 2\|g_{2^k}\| + \frac{2^{k+1}}{m}\|g_m\|\right]\\ &\leq \frac{1}{\sqrt{2^k}}\left[39\bar{\nu}_{2^k} + 2\nu_{2^k}\right] \to 0\end{aligned}$$

as $k \to \infty$. Hence by Cauchy's rule, $\{\bar{H}_n\}$ is convergent in $L^2(\pi_1)$. That g admits a martingale approximation follows from Proposition 3.2. □

Proposition 3.2 is useful, but not so convenient to be used in concrete problems. Next, we will consider more specialized examples. The following corollary can be easily deduced and will be important later.

Corollary 3.6. *If g admits a martingale approximation, then so does $Q^k g$, and $M_1(Q^k g) = H(W_0, W_1) - H_k(W_0, W_1)$ with H_k as defined in (3.7).*

Proof. For $k = 1$, this follows directly from (3.4); and for $k = 2, 3, \ldots$, it follows by induction. □

As a second corollary, we may obtain necessary and sufficient conditions for linear processes; the following decomposition for a linear process has its roots in [24] and [45]. Let $\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots$ be i.i.d. random variables with mean 0 and unit variance; let $a_0, a_1, a_2, \ldots$ be a square summable sequence; and consider a causal linear process

$$X_j = \sum_{i=0}^{\infty} a_i \xi_{j-i} = \sum_{i \le j} a_{j-i} \xi_i. \tag{3.14}$$

Such a process is of the form $X_k = g(W_k)$, where $W_k = (\ldots, \xi_{k-1}, \xi_k)$. Letting $b_{-1} = 0$, $b_n = a_0 + \cdots + a_n$ for $n \ge 0$, and using (3.14),

$$S_n = \sum_{i \le 1} (b_{n-i} - b_{-i}) \xi_i + \sum_{i=0}^{n-2} b_i \xi_{n-i},$$

where the first term on the right is $E(S_n|W_1)$. It follows that

$$\|V_n g\|^2 = \|E(S_n|W_1)\|^2 = \sum_{i=-1}^{\infty} (b_{i+n} - b_i)^2,$$

also, $V_n g(W_1) - Q V_n g(W_0) = b_n \xi_1$, and $\bar{H}_n(W_0, W_1) = \bar{b}_n \xi_1$ with $\bar{b}_n = (b_1 + \cdots + b_n)/n$. Thus, for a linear process, (3.2) specializes to

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=-1}^{\infty} (b_{i+n} - b_i)^2 = 0. \tag{3.15}$$

Corollary 3.7. *For the linear process defined in (3.14), the following are equivalent:*

(a) *There is a martingale approximation.*

(b) *(3.15) holds and $\bar{b}_n$ converges.*

(c) *(3.15) holds and $\bar{b}_n^2$ converges.*

PROOF. In this case $\|\bar{H}_n - \bar{H}_m\|^2 = (\bar{b}_n - \bar{b}_m)^2$. Hence, (a) and (b) are equivalent by Proposition 3.2. It is clear that (b) implies (c) and remains only to show that (c) imples (b). If $\bar{b}_n^2$ converges, but $\bar{b}_n$ does not, then $\bar{b}_n$ would have to oscillate between two values, there would be a positive ϵ for which $|\bar{b}_{n+1} - \bar{b}_n| \ge \epsilon$ infinitely often; but this is impossible, since $\bar{b}_{n+1} - \bar{b}_n = (b_{n+1} - \bar{b}_n)/(n+1)$ and $b_n = O(\sqrt{n})$, as $a_0, a_1, \ldots$ are square summable. □

In the next section, we show how to extend this example from linear functions of shift processes to measurable ones with mean 0 and finite variance.

3.3 Co-Isometries

We suppose throughout this section that the chain has a trivial left tail field and that Q is a co-isometry; that is,

$$\lim_{n\to\infty} \|Q^n f\| = 0 \qquad \text{and} \qquad QQ^* = I. \tag{3.16}$$

for all $f \in L_0^2(\pi)$. We also suppose $L_0^2(\pi)$ is separable. These conditions are satisfied, for example, by (one-sided) shift processes.

With a view towards later examples, we work with $\mathscr{L}_0^2(\pi)$, the space of complex valued, square integrable functions with mean 0 under π. Then (3.16) is still valid for this space if we extend the definition of Q to the imaginary part.

Let $\mathcal{H}$ denote a closed linear subspace of $\mathscr{L}_0^2(\pi)$ that is invariant under both Q and Q^*; restrict Q and Q^* to $\mathcal{H}$; and let $\mathcal{K} = Q^*\mathcal{H}$. Then Q^* is an isometry from $\mathcal{H}$ onto $\mathcal{K}$, since $\langle Q^*f, Q^*g\rangle = \langle f, QQ^*g\rangle = \langle f, g\rangle$ for $f, g \in \mathcal{H}$. This is the origin of the term "co-isometry." Moreover,

$$Q^*h(W_1) = h(W_0) \ w.p.1 \tag{3.17}$$

for any $h \in L_0^2(\pi)$, since $E[Q^*h(W_1)h(W_0)] = \langle QQ^*h, h\rangle = \|h\|^2$ by conditioning on W_0, and therefore, $E\{[Q^*h(W_1) - h(W_0)]^2\} = \|Q^*h\|^2 - 2\langle QQ^*h, h\rangle + \|h\|^2 = 0$. It then can be easily checked (3.17) also holds for $h \in \mathscr{L}_0^2(\pi)$.

Lemma 3.8. *$\mathcal{K}$ is a closed, proper linear subspace of $\mathcal{H}$; and $\cap_{j=0}^{\infty} Q^{*j}\mathcal{H} = \{0\}$.*

Proof. That $\mathcal{K}$ is closed is clear, since Q^* is an isometry; and that $\mathcal{K}$ is proper follows from $\cap_{j=0}^{\infty} Q^{*j}\mathcal{H} = \{0\}$. So, it suffices to establish the latter. If $f \in \cap_{j=0}^{\infty} Q^{*j}\mathcal{H}$, then there are $h_0, h_1, \ldots \in \mathcal{H}$ for which $f = Q^{*j}h_j$ with each j. In this case, $\|h_j\| = \|f\|$, since Q^* is an isometry, $h_j = Q^j Q^{*j} h_j = Q^j f$, and $\lim_{j\to\infty} \|Q^j f\| = 0$. So, $\|f\| = 0$, establishing the lemma. $\square$

Next, let $\mathcal{K}^\perp = \{f \in \mathcal{H} : \langle f, h\rangle = 0 \text{ for all } h \in \mathcal{K}\}$. Then $\mathcal{K}^\perp = \{g \in \mathcal{H} : Qg = 0\}$, since $\langle Q^* f, g\rangle = \langle f, Qg\rangle = 0$ for all $f \in \mathcal{H}$ iff $Qg = 0$; and Q^*Q is the projection operator onto $\mathcal{K}$, since $(Q^*Q)^2 = Q^*Q$ and $Q(I - Q^*Q) = 0$. Let $\mathcal{E}_0 = \{e_j : j \in J\}$ be an orthonormal basis for $\mathcal{K}^\perp$, let $\mathcal{E}_i = Q^{*i}\mathcal{E}_0$ and $\mathcal{E} = \cup_{i\geq 0}\mathcal{E}_i$.

Lemma 3.9. *$\mathcal{E}$ is an orthonormal basis for $\mathcal{H}$.*

Proof. $\mathcal{E}_i$ consists of orthonormal elements for each $i \geq 0$, since Q^* is an isometry; for any $f \in \mathcal{E}_i$ and $f' \in \mathcal{E}_{i'}$, where $i < i'$, there are $e, e' \in \mathcal{E}_0$ for which $f = Q^{*i}e$ and $f' = Q^{*i'}e'$, in which case $\langle f, f'\rangle = \langle Q^{*i}e, Q^{*i'}e'\rangle = \langle Q^{i'-i}e, e'\rangle = 0$, since $Qe = 0$. Finally, if $f \perp \mathcal{E}_0$, then $f \in \mathcal{K}$ and $f = Q^*h_1$ for some $h_1 \in \mathcal{H}$. If also, $f \perp Q^*\mathcal{E}_0$, then $Qf \perp \mathcal{E}_0$, $Qf = Q^*h_2$ for some $h_2 \in \mathcal{H}$, and $f = Q^*Qf = Q^{*2}h_2$. Continuing, we find that if $f \perp \mathcal{E}$, then $f \in Q^{*j}\mathcal{H}$ for all j, and completeness follows from Lemma 3.8. $\square$

Now write $e_{i,j} = Q^{*i}e_j$, so that $\mathcal{E}_i = \{e_{i,j} : j \in J\}$, and let $\mathcal{H}_j = \text{span}(e_{i,j} : i \geq 0)$, the closed linear span of $\{e_{i,j} : i \geq 0\}$. Then $Q\mathcal{H}_j = \mathcal{H}_j$ for each j, and $\mathcal{H} = \oplus_{j\in J}\mathcal{H}_j$. In the language of [9, 16], the $\mathcal{H}_j$, $j \in J$, are an orthogonal invariant splitting of $\mathcal{H}$. Then, any $g \in \mathcal{H}$ may be written as $g = \sum_{j\in J}\sum_{i=0}^{\infty} c_{i,j}e_{i,j}$, where $c_{i,j}$ are square summable. Let $b_{n,j} = c_{0,j} + \cdots + c_{n-1,j}$, $\bar{b}_{n,j} = (b_{1,j} + \cdots + b_{n,j})/n$ and regard $\boldsymbol{b}_n = (b_{n,j} : j \in J)$ and $\bar{\boldsymbol{b}}_n = (\bar{b}_{n,j} : j \in J)$ as elements of $\ell^2(J)$.

Theorem 3.10. *$g \in L_0^2(\pi)$ admits a martingale approximation iff $\bar{\boldsymbol{b}}_n$ converges in*

$\ell^2(J)$, and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{\infty}\|\boldsymbol{b}_{i+n}-\boldsymbol{b}_i\|^2=0. \tag{3.18}$$

PROOF. We take $\mathcal{H}=\mathscr{L}_0^2(\pi)$. Since $Qe_{i,j}=QQ^{*i}e_j=0$ if $i=0$ and $e_{i-1,j}$ if $i\geq 1$, $Qg=\sum_{j\in J}\sum_{i=1}^{\infty}c_{i,j}e_{i-1,j}$,

$$Q^kg=\sum_{i=k}^{\infty}\sum_{j\in J}c_{i,j}e_{i-k,j}=\sum_{i=0}^{\infty}\sum_{j\in J}c_{i+k,j}e_{i,j}\ ,$$

$$V_ng=\sum_{i=0}^{\infty}\sum_{j\in J}(b_{i+n,j}-b_{i,j})e_{i,j}$$

and

$$\|V_ng\|^2=\sum_{i=0}^{\infty}\sum_{j\in J}|b_{i+n,j}-b_{i,j}|^2=\sum_{i=0}^{\infty}\|\boldsymbol{b}_{i+n}-\boldsymbol{b}_i\|^2.$$

So (3.18) is just (3.2), specialized to the present context.

Next $Q^*Qg=\sum_{j\in J}\sum_{i=1}^{\infty}c_{i,j}e_{i,j}$, so that from (3.17),

$$g(W_1)-Qg(W_0)=[g-Q^*Qg](W_1)=\sum_{j\in J}c_{0,j}e_{0,j}(W_1),$$

$$\bar{H}_n(W_0,W_1)=\sum_{k=0}^{n-1}\left(1-\frac{k}{n}\right)\left[Q^kg(W_1)-Q^{k+1}g(W_0)\right]=\sum_{j\in J}\bar{b}_{n,j}e_{0,j}(W_1)$$

and

$$\|\bar{H}_n-\bar{H}_m\|=\|\bar{\boldsymbol{b}}_n-\bar{\boldsymbol{b}}_m\|.$$

The theorem now follows directly from Proposition 3.2. □

Example 3.11 (*Bernoulli Shifts*). The one-sided Bernoulli shift process is defined by

$$W_k=\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^{j+1}\xi_{k-j},$$

where $\dots\xi_{-1},\xi_0,\xi_1,\dots$ are i.i.d. random variables taking the values 0 and 1 with probability 1/2 each. The state space $\mathcal{W}$ is the unit interval, the marginal distribution

π is the uniform distribution, $Qg(w) = \frac{1}{2}[g(\frac{1}{2}w) + g(\frac{1}{2}w + \frac{1}{2})]$, and $Q^*g(w) = g(2w)$ for *a.e.* $w \in \mathcal{W}$ and $g \in L^1(\pi)$ with the convention that g is continued periodically. For this example, any $g \in \mathscr{L}_0^2(\pi)$ has a Fourier expansion

$$g = \sum_{r \neq 0} c_r e_r, \tag{3.19}$$

where $e_r(w) = e^{2\pi\imath rw}$ and c_r, $r \in \mathbb{Z}$, are square summable. Then $Qe_r = 0$ or $e_{\frac{1}{2}r}$ accordingly as r is odd or even, and $Q^*e_r = e_{2r}$ for all r. With $\mathcal{H} = \mathscr{L}_0^2(\pi)$, it follows that $\mathcal{K}$, respectively $\mathcal{K}^\perp$, consists of all functions g for which $c_r = 0$ for odd, respectively even, r. Thus, $\mathcal{E}_0 = \mathrm{span}(e_r : r \in \mathrm{Odd})$, and $\mathcal{E}_i = \mathrm{span}(e_{r2^i} : r \in \mathrm{Odd})$, and there is an invariant splitting with $e_{i,j} = e_{j2^i}$. Necessary and sufficient conditions for the existence of a martingale approximation can be read from Theorem 3.10. See [42] for more on the Fourier analysis of Bernoulli shifts.

Example 3.12 (*Lebesgue Shifts*)**.** By a (one-sided) Lebesgue shift, we mean the Markov chain $W_k = (\ldots U_{k-1}, U_k)$ where $\ldots U_{-1}, U_0, U_1, \ldots$ are independent uniformly distributed random variables over $[0, 1)$, in which case $\mathcal{W} = [0, 1)^{\mathbb{N}}$ and $\pi = \lambda^{\mathbb{N}}$, where λ is the uniform distribution. Lebesgue shifts are similar to Bernoulli shifts. Let Γ denote the set of sequences $j = (j_0, j_1, \ldots) \in \mathbb{Z}^{\mathbb{N}}$ for which $j_i = 0$ for all but finite number of i. Then, letting $j \cdot w = j_0 w_0 + j_1 w_{-1} + \cdots$ and $e_j(w) = e^{2\pi\imath j \cdot w}$ for $w = (\ldots w_{-1}, w_0) \in [0, 1)^{\mathbb{N}}$ and $j \in \Gamma$, any $g \in \mathscr{L}_0^2(\pi)$ has a Fourier expansion,

$$g(w) = \sum_{j \in \Gamma} c_j e_j$$

where c_j are square summable. Next, let $J = \{j \in \Gamma : j_0 \neq 0\}$. Then, since

$$Qe_j(w) = \left[\int_0^1 e^{2\pi\imath j_0 u} du\right] \prod_{i=1}^{\infty} \exp(2\pi\imath j_i w_{-i+1}),$$

$\mathcal{E}_0 = \{e_k : k \in J\}$ is an orthonormal basis for $\mathcal{K}^\perp$ (with $\mathcal{H} = \mathscr{L}_0^2(\pi)$ and $\mathcal{K} = Q^*\mathcal{H}$). Define $\psi : \Gamma \to \Gamma$ by $\psi(j) = (0, j_0, j_1, \ldots)$, then it is not difficult to check $Q^*e_k =$

$e_{\psi(k)}$, $Q^{*i}e_k = e_{\psi^i(k)}$, where ψ^i is the composition of ψ with itself i times. Necessary and sufficient conditions can be read from Theorem 3.10.

Example 3.13 (*Superlinear Processes*)**.** Let $\xi_{i,j}, i \in \mathbb{Z},\ j \in \mathbb{N}$, be independent random variables, all having mean 0 and bounded variances, for which $\ldots \xi_{-1,j}, \xi_{0,j}, \xi_{1,j}, \ldots$ are identically distributed for each j, and let $c_{ij}, i \in \mathbb{Z}, j \in \mathbb{N}$, be a square summable array. Then

$$X_k = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} c_{i,j}\xi_{k-i,j} \tag{3.20}$$

converges $w.p.1$ and in mean square for each k and defines a stationary process. Letting $\boldsymbol{\xi}_i = (\xi_{i,0}, \xi_{i,1}, \ldots)$, X_k is of the form $X_k = g(W_k)$, where $W_k = (\ldots, \boldsymbol{\xi}_{k-1}, \boldsymbol{\xi}_k)$ is a shift process. Next, letting $\mathcal{H} = \operatorname{span}(\xi_{i,j} : i \le 0,\ j \ge 0)$, one finds easily that there is an invariant splitting with $e_{i,j} = \xi_{-i,j}$ for $i, j \ge 0$. Necessary and sufficient conditions for the existence of a martingale approximation can again be read from Theorem 3.10.

3.4 The Plus Norm

To study the plus norm, we first recall the definition $\|g\|_+^2 = \limsup_{n\to\infty} E[S_n(g)^2]/n$. The following example serves as a simple illustration.

Example 3.14. If Q is a co-isometry and the chain has a trivial left tail field, we may write $g = \sum_{j\in J}\sum_{i=0}^{\infty} c_{i,j}e_{i,j}$, as in Section 3.3, and $\bar{H}_n(W_0, W_1) = \sum_{j\in J}\bar{b}_{n,j}e_{0,j}(W_1)$, as in the proof of Theorem 3.10. So, if (3.2) holds, $E(S_n^2) = nE[\bar{H}_n^2(W_0, W_1)] + o(n) = n\|\bar{\boldsymbol{b}}_n\|^2 + o(n)$, and $\|g\|_+^2 = \limsup_{n\to\infty}\|\bar{\boldsymbol{b}}_n\|^2$.

The main result of this section is that g admits a martingale approximation iff $\|V_n g\| = o(\sqrt{n})$ and $\sum_{k=1}^{m}\|Q^k g\|_+^2 = o(m)$. The following two lemmas are needed; their proofs are given after the proof of Theorem 3.17.

Lemma 3.15. *If $g \in L_0^2(\pi)$ and (3.2) holds, then*

$$\lim_{n\to\infty}\left[\|\bar{H}_n - \bar{H}_m\|^2 - \frac{2}{m}\langle \bar{V}_n g, QV_m g\rangle\right] = -\frac{2}{m}\langle \bar{V}_m g, QV_m g\rangle - \left\|\frac{QV_m g}{m}\right\|^2.$$

Lemma 3.16. *If $g \in L_0^2(\pi)$ and $\|g\|_+ < \infty$, then*

$$\limsup_{n\to\infty}\ \langle \bar{V}_n g, QV_m g\rangle \le \frac{1}{2}\sum_{k=1}^{m}\|Q^k g\|_+^2 + \frac{1}{2}\|QV_m g\|^2 + \langle g, V_m Q g\rangle;$$

and if g admits a martingale approximation, then the limit exists and there is equality.

Theorem 3.17. *g admits a martingale approximation iff (3.2) holds and*

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m}\|Q^k g\|_+^2 = 0. \tag{3.21}$$

Proof of Theorem 3.17. Suppose first that g admits a martingale approximation. Then $\|V_n g\| = o(\sqrt{n})$ and $\lim_{m\to\infty}[\lim_{n\to\infty}\|\bar{H}_n - \bar{H}_m\|^2] = 0$ by Proposition 3.2. Next, by Lemmas 3.15 and 3.16,

$$\begin{aligned}\lim_{n\to\infty}\|\bar{H}_n - \bar{H}_m\|^2 &= \lim_{n\to\infty}\frac{2}{m}\langle \bar{V}_n g, QV_m g\rangle - \left[\frac{2}{m}\langle \bar{V}_m g, QV_m g\rangle + \left\|\frac{QV_m g}{m}\right\|^2\right] \\ &= \frac{1}{m}\sum_{k=1}^{m}\|Q^k g\|_+^2 + \frac{1}{m}\|QV_m g\|^2 + \frac{2}{m}\langle g, QV_m g\rangle \\ &\qquad - \frac{2}{m}\langle \bar{V}_m g, QV_m g\rangle - \left\|\frac{QV_m g}{m}\right\|^2.\end{aligned}$$

Since $\|V_m g\| = o(\sqrt{m})$, the last four terms on the right approach 0 as $m \to \infty$, and, therefore, so does the first. This establishes the necessity of (3.21).

Next suppose that (3.2) and (3.21) hold, then $\lim_{m\to\infty}[\limsup_{n\to\infty}\|\bar{H}_n - \bar{H}_m\|^2] = 0$, by Lemmas 3.15 and 3.16. It follows easily that, $\sup_{n\ge 1}\|\bar{H}_n\| < \infty$, which implies $\bar{H}_1, \bar{H}_2, \ldots$ is weakly compact in $L^2(\pi_1)$. Let H^* denote any weak limit point of $\bar{H}_1, \bar{H}_2, , \ldots$. Then $\|H^* - \bar{H}_m\| \le \limsup_{n\to\infty}\|\bar{H}_n - \bar{H}_m\|$ for each m (cf. [12], p. 68). Thus, $\lim_{m\to\infty}\|\bar{H}_m - H^*\| = 0$ from which the converse follows from Proposition 3.2.

□

Proof of Lemma 3.15. To begin, write

$$\begin{aligned}\|\bar{H}_n - \bar{H}_m\|^2 &= \|(\bar{V}_n - \bar{V}_m)g\|^2 - \|Q(\bar{V}_n - \bar{V}_m)g\|^2 \\ &= \langle (I+Q)(\bar{V}_n - \bar{V}_m)g, (I-Q)(\bar{V}_n - \bar{V}_m)g\rangle \\ &= 2\left\langle (\bar{V}_n - \bar{V}_m)g, \left(\frac{QV_m g}{m} - \frac{QV_n g}{n}\right)\right\rangle - \left\|\frac{QV_m g}{m} - \frac{QV_n g}{n}\right\|^2;\end{aligned}$$

and when the first term in the last line is expanded, it becomes

$$\frac{2}{m}\langle \bar{V}_n g, QV_m g\rangle - \frac{2}{n}\langle \bar{V}_n g, QV_n g\rangle - \frac{2}{m}\langle \bar{V}_m g, QV_m g\rangle + \frac{2}{n}\langle \bar{V}_m g, QV_n g\rangle.$$

The lemma now follows directly from (3.2) and the mean ergodic theorem, which implies that all those terms multiplied by $1/n$ approach 0 as $n \to \infty$. $\square$

Proof of Lemma 3.16. Writing

$$\langle \bar{V}_n g, QV_m g\rangle = \sum_{k=1}^{m} \langle \bar{V}_n g, Q^k g\rangle,$$

and using (3.6), then

$$\langle \bar{V}_n g, QV_m g\rangle = \sum_{k=1}^{m} \left[\langle \bar{V}_n Q^k g, Q^k g\rangle + \langle V_k g, Q^k g\rangle - \frac{1}{n}\langle QV_n V_k g, Q^k g\rangle\right].$$

Here

$$\begin{aligned}\sum_{k=1}^{m} \langle V_k g, Q^k g\rangle &= \sum_{k=1}^{m}\sum_{j=1}^{k} \langle Q^{j-1} g, Q^k g\rangle \\ &= \frac{1}{2}\sum_{k=1}^{m}\sum_{j=1}^{m} \langle Q^j g, Q^k g\rangle - \frac{1}{2}\sum_{j=1}^{m} \|Q^j g\|^2 + \langle g, V_m Q g\rangle \\ &= \frac{1}{2}\|V_m Q g\|^2 - \frac{1}{2}\sum_{j=1}^{m} \|Q^j g\|^2 + \langle g, V_m Q g\rangle.\end{aligned}$$

Combining terms together,

$$\begin{aligned}\langle \bar{V}_n g, QV_m g\rangle \quad = \quad & \frac{1}{2}\sum_{k=1}^{m} \left[2\langle \bar{V}_n Q^k g, Q^k g\rangle - \|Q^k g\|^2\right] \\ & + \frac{1}{2}\|QV_m g\|^2 + \langle g, V_m Q g\rangle - \sum_{k=1}^{m} \frac{1}{n}\langle QV_n V_k g, Q^k g\rangle.\end{aligned}$$

The first assertion follows directly from (3.2) and (3.5). So does the second; for if g admits a martingale approximation, then the limit exists in the definition of $\|Q^k g\|_+$.

□

3.5 The Fractional Poisson Equation

It is possible to attach a meaning to the symbol $\sqrt{I-Q}$ by replacing t with Q in the series expansion of $\sqrt{1-t}$. The definition may be written

$$\sqrt{I-Q} = I - \sum_{k=1}^{\infty} \beta_k Q^k,$$

where $\beta_k = (-1)^{k-1}\binom{1/2}{k}$ and the series converges in the operator norm, since $\beta_k \sim 1/(2\sqrt{\pi}k^{3/2})$ as $k \to \infty$. A function $h \in L_0^2(\pi)$ is said to solve the *fractional Poisson equation* (for g) if $g = \sqrt{(I-Q)}h$. The relation between the existence of a solution to the fractional Poisson equation and the existence of a martingale approximation is considered in this section for co-isometries and normal operators ($QQ^* = Q^*Q$).

Lemma 3.18. *If $g \in \sqrt{(I-Q)}L_0^2(\pi)$, then $\|V_n g\| = o(\sqrt{n})$; and if $g = \sqrt{(I-Q)}h = \sqrt{(I-Q^*)}h^*$, then $\|g\|_+^2 = \langle (I+Q)h, h^*\rangle$.*

PROOF. Observe that $(I-Q^k)V_n = (I-Q^n)V_k$. So, if $g = \sqrt{(I-Q)}h$, then $V_n g = \sum_{k=0}^{\infty} \beta_k (I - Q^{k\vee n})V_{k\wedge n}h$, where $\wedge$ ($\vee$) denotes minimum (maximum). Using the mean ergodic theorem, $\|V_n h\| = o(n)$, then

$$\|V_n g\| \le 2\sum_{k=0}^{\infty} \beta_k \|V_{k\wedge n}h\| = 2\sum_{k=0}^{\infty} \beta_k o(k\wedge n) = o(\sqrt{n}),$$

establishing the first assertion. If, in addition, $g = \sqrt{(I-Q^*)}h^*$, then $\|g\|^2 = \langle (I-Q)h, h^*\rangle$, and

$$\langle \bar{V}_n g, g\rangle = \langle (I-Q)\bar{V}_n h, h^*\rangle = \langle h, h^*\rangle - \frac{1}{n}\langle QV_n h, h^*\rangle \to \langle h, h^*\rangle,$$

using the mean ergodic theorem again in the final step. Thus, in view of (3.5), $\|g\|_+^2 = \lim_{n\to\infty}[2\langle \bar{V}_n g, g\rangle - \|g\|^2] = \langle (I+Q)h, h^*\rangle$; similar calculations also appear in [11]. □

Normal Operators. As an interesting generalization of [26] in the reversible case, it is known, [5, 10, 18], that if Q is a normal operator and there is a solution to the fractional Poisson equation, then g admits a martingale approximation. This result can be easily deduced from our Theorem 3.17. Recall that if R is any bounded normal operator on a Hilbert Space $\mathcal{H}$, then $\sqrt{I-R}$ and $\sqrt{I-R^*}$ have the same range (cf. [10], Lemma 2).

Proposition 3.19. *Suppose that Q is normal, then any $g \in \sqrt{(I-Q)}L_0^2(\pi)$ admits a martingale approximation.*

Proof. If $g \in \sqrt{(I-Q)}L_0^2(\pi)$, then (3.2) follows from Lemma 3.18, and it suffices to establish (3.21). Since the ranges of $\sqrt{(I-Q)}$ and $\sqrt{(I-Q^*)}$ are the same, there are $h, h^* \in L_0^2(\pi)$ for which $g = \sqrt{(I-Q)}h = \sqrt{(I-Q^*)}h^*$. Then $Q^k g = \sqrt{(I-Q)}Q^k h = \sqrt{(I-Q^*)}Q^k h^*$, so that $\|Q^k g\|_+^2 = \langle (I+Q)Q^k h, Q^k h^*\rangle$. Thus, letting $R = Q^*Q$, $\|Q^k g\|_+^2 = \langle (I+Q)h, R^k h^*\rangle$, and it is necessary to show

$$\lim_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} \langle (I+Q)h, R^k h^*\rangle = 0. \tag{3.22}$$

To see this let $\mathcal{R}$ be the closure of $(I-R)L_0^2(\pi)$. Then $\mathcal{R}^\perp$ consists of all f for which $Rf = f$, and Q, Q^*, and R map both $\mathcal{R}$ and $\mathcal{R}^\perp$ into themselves. Write $h = h_1 + h_2$ with $h_1 \in \mathcal{R}$, $h_2 \in \mathcal{R}^\perp$, and let $g_i = \sqrt{(I-Q)}h_i$. Then $g_1 \in \mathcal{R}$ and $g_2 \in \mathcal{R}^\perp$, since Q maps $\mathcal{R}$ and $\mathcal{R}^\perp$ into themselves. Next, write $h^* = h_1^* + h_2^*$ with $h_1^* \in \mathcal{R}$, $h_2^* \in \mathcal{R}^\perp$, then $g_i = \sqrt{(I-Q^*)}h_i^*$ by the uniqueness of direct sum decomposition of g. Returning to (3.22), we have

$$\begin{aligned}\langle (I+Q)h, R^k h^*\rangle &= \langle (I+Q)h_1, R^k h_1^*\rangle + \langle (I+Q)h_2, h_2^*\rangle \\ &= \langle (I+Q)h_1, R^k h_1^*\rangle + \|g_2\|_+^2\end{aligned}$$

by orthogonality and Lemma 3.18. It will be first shown that $\|g_2\|_+ = 0$; to see it, note $Rg_2 = g_2$, then

$$\begin{aligned}\|V_n g_2\|^2 &= 2\sum_{j=0}^{n-1}\sum_{k=j}^{n-1}\langle Q^j g_2, Q^k g_2\rangle - \sum_{j=0}^{n-1}\|Q^j g_2\|^2 \\ &= 2\sum_{j=0}^{n-1}\langle g_2, V_{n-j} g_2\rangle - n\|g_2\|^2 = n\left[2\langle g_2, \bar{V}_n g_2\rangle - \|g_2\|^2\right],\end{aligned}$$

thus, $\|g_2\|_+ = 0$ follows from (3.5) and Lemma 3.18. That (3.22) holds when $h_1^* \in (I-R)L_0^2(\pi)$ is clear by forming a telescoping sum, and the boundary case then follows by approximation. □

Co-isometries. The existence of a solution to the fractional Poisson equation does not imply the existence of a martingale approximation for co-isometries. Here is a simple example.

Example 3.20. Let $\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots$ be i.i.d. with mean 0 and unit variance; consider the shift process $W_k = (\ldots, \xi_{k-1}, \xi_k)$. For $j \geq 0$, let $a_j = 1/[\sqrt{(j+1)}\log(j+2)]$ and define h by

$$h(W_0) = \sum_{j=0}^{\infty} a_j \xi_{-j},$$

so that $h(W_k)$ is a linear process. Then $g = \sqrt{(I-Q)}h$ admits a solution to the fractional Poisson equation, and

$$g(W_0) = \sum_{k=1}^{\infty}\beta_k (I - Q^k) h(W_0) = \sum_{j=0}^{\infty} c_j \xi_{-j}$$

with

$$c_j = \sum_{k=1}^{\infty}\beta_k (a_j - a_{j+k}),$$

after some straightforward calculation. Observe that $a_j - a_{j+k} \geq 0$ for all $j \geq 0$ and $k \geq 1$, and that $a_{j+k} \leq 3a_j/4$ for all $k \geq j+1$ and all $j \geq 0$. So,

$$c_j \geq \frac{1}{4}a_j \sum_{k=j+1}^{\infty} \beta_k \geq \left(\frac{a_j}{9\sqrt{j}}\right)$$

for all sufficiently large j. Therefore, $b_n = c_0 + \cdots + c_n \to \infty$, and also, its Cesàro average $\bar{b}_n \to \infty$ as $n \to \infty$. No martingale approximation can exist.

However, the existence of solutions to both the forward and backward fractional Poisson equations, does imply the existence of a martingale approximation.

Proposition 3.21. *Suppose Q is a co-isometry and the chain has a trivial left tail field, and if $g \in \sqrt{(I-Q)}L_0^2(\pi) \cap \sqrt{(I-Q^*)}L_0^2(\pi)$, then g admits a martingale approximation.*

Proof. As in Section 3.3, we can take $\mathcal{H} = \mathscr{L}_0^2(\pi)$, and there is an orthogonal invariant splitting, $\mathcal{H} = \oplus_{j\in J}\mathcal{H}_j$. Let $g = \sqrt{(I-Q)}h$ for some $h \in \mathcal{H}$, $g = \sum_{j\in J} g_j$, and $h = \sum_{j\in J} h_j$ with $g_j, h_j \in \mathcal{H}_j$ for all j. Clearly $g = \sum_{j\in J}\sqrt{(I-Q)}h_j$ and, therefore, $g_j = \sqrt{(I-Q)}h_j$, by taking the projection on each $\mathcal{H}_j$. Similarly, $g = \sqrt{(I-Q^*)}h^*$, where $h^* = \sum_{j\in J} h_j^*$ with $h_j^* \in \mathcal{H}_j$, and $g_j = \sqrt{(I-Q^*)}h_j^*$ for each j. It then follows easily from Lemma 3.18 and Example 3.14 that $\lim_{n\to\infty} |\bar{b}_{n,j}|^2 = \|g_j\|_+^2 = \langle (I+Q)h_j, h_j^*\rangle$ exists for each j and that $\lim_{n\to\infty} \|\bar{\boldsymbol{b}}_n\|^2 = \|g\|_+^2 = \langle (I+Q)h, h^*\rangle$ exist. It then follows from (the proof of) Corollary 3.7 that $b_j = \lim_{n\to\infty} \bar{b}_{n,j}$ exists for each j, so that $\bar{\boldsymbol{b}}_n$ converges weakly to $\boldsymbol{b} = (b_j : j \in J)$. So, to show convergence of $\bar{\boldsymbol{b}}_n$ in the norm of $\ell^2(J)$ and, therefore, the existence of a martingale aproximation, it suffices to show that $\lim_{n\to\infty} \|\bar{\boldsymbol{b}}_n\|^2 = \|\bar{\boldsymbol{b}}\|^2$; and this follows easily from Lemma 3.18 which implies

$$\lim_{n\to\infty} \|\bar{\boldsymbol{b}}_n\|^2 = \langle (I+Q)h, h^*\rangle = \sum_{j\in J}\langle (I+Q)h_j, h_j^*\rangle = \sum_{j\in J} |b_j|^2 = \|\boldsymbol{b}\|^2. \quad \square$$

3.6 The CCLT for Superlinear Processes

Let F_n denote the conditional distribution function of $S_n/\sqrt{n}$ given W_0,

$$F_n(w; z) = P\left[\frac{S_n}{\sqrt{n}} \leq z | W_0 = w\right].$$

We will say that the *conditional central limit theorem* (CCLT) holds (with a $\sqrt{n}$ normalization) iff

$$\lim_{n\to\infty} \frac{E[S_n^2]}{n} = \kappa^2 \in [0, \infty)$$

and

$$\lim_{n\to\infty} \int_{\mathcal{W}} d[\Phi_\kappa, F_n(w; \cdot)]\pi\{dw\} = 0,$$

where Φ_κ denotes the normal distribution function with mean 0 and standard deviation κ, and d is the Lévy metric or any other bounded metric that metrizes weak convergence of distribution functions.

It is clear that the existence of a martingale approximation implies the CCLT; see, for example, [30]. It is also clear, for simple linear process as defined in (3.14), CCLT necessarily requires the existence of martingale approximation. However, in general, the converse is not true as shown in the example below. To proceed as in Example 3.13, let F_j be the common distribution function of $\xi_{i,j}, i = \ldots -1, 0, 1, \ldots$ and suppose that the F_j have mean 0 and bounded variances. Recall the notation $b_{n,j} = c_{0,j} + \cdots + c_{n-1,j}$ and $\bar{b}_{n,j} = (b_{1,j} + \cdots + b_{n,j})/n$ and that $\boldsymbol{b}_n = (b_{n,1}, b_{n,2}, \ldots)$ and $\bar{\boldsymbol{b}}_n$ may be regarded as elements of $\ell^2(\mathbb{N})$.

Example 3.22 (*superlinear process revisited).* Consider a superlinear process, defined in (3.20), with $c_{i,j} = 0$ for all $j \geq 2$, $b_{n,0} = \cos(\sqrt{\log n}), b_{n,1} = \sin(\sqrt{\log n})$, and $c_{0,j} = c_{1,j} = 0$ for $j = 0, 1$. Then $c_{n,j} = b_{n,j} - b_{n-1,j} = O(1/(n\sqrt{\log n}))$ for $j = 0, 1$. So, the process is well-defined. If F_0 and F_1 both have mean 0 and unit variance, then the CCLT holds, but martingale approximation does not exist. To see this, first observe that for any $\delta > 0$,

$$\begin{aligned}\sum_{k=0}^{\infty}(b_{k+n,0}-b_{k,0})^2 &\leq \left(\sum_{k\leq n\delta}+\sum_{k>n\delta}\right)[\cos(\sqrt{\log(n+k)})-\cos(\sqrt{\log k})]^2\\ &\leq 4n\delta+\sum_{k>n\delta}\left(\frac{n}{2k\sqrt{\log k}}\right)^2,\end{aligned}$$

so that $\sum_{k=0}^{\infty}(b_{k+n,0}-b_{k,0})^2 = o(n)$, and similarly, $\sum_{k=0}^{\infty}(b_{k+n,1}-b_{k,1})^2 = o(n)$. So, $\|V_n g\|^2 = o(n)$. Next, for any $\epsilon > 0$,

$$\begin{aligned}\bar{b}_{n,0}-b_{n,0} &= \frac{1}{n}\sum_{k=1}^{n}[\cos(\sqrt{\log k})-\cos(\sqrt{\log n})]\\ &\leq \frac{1}{n}\sum_{k\leq n\epsilon}[\cos(\sqrt{\log k})-\cos(\sqrt{\log n})]+\frac{1}{n}\sum_{n\epsilon<k\leq n}[\sqrt{\log n}-\sqrt{\log k}]\\ &\leq 2\epsilon+\frac{1}{n}\sum_{n\epsilon<k\leq n}\frac{n-k}{2k\sqrt{\log k}}\leq 2\epsilon+\frac{1}{n}(n-n\epsilon)\frac{1}{2\epsilon\sqrt{\log(n\epsilon)}}\end{aligned}$$

for all large n. It follows that $\bar{b}_{n,0}-b_{n,0} = o(1)$. Similarly $\bar{b}_{n,1}-b_{n,1} = o(1)$, and therefore, $\bar{b}_{n,0}^2+\bar{b}_{n,1}^2 \to 1$. So, applying Theorem 2 of [45], CCLT holds; but martingale approximation does not exist since $\bar{b}_{n,j}$ does not converge for $j = 0, 1$. $\square$

Next, we investigate some partial converses for superlinear processes.

Theorem 3.23. *If the CCLT holds for all choices $F_1, F_2, \ldots$ with means 0 and unit variances, then $\bar{\boldsymbol{b}}_n$ is pre-compact in $\ell^2(\mathbb{N})$; and if the CCLT holds for all $F_1, F_2, \ldots$ with means 0 and bounded variances, then $\bar{\boldsymbol{b}}_n$ converges in $\ell^2(\mathbb{N})$.*

Proof. If the CCLT holds, then (3.2) holds by Corollary 1 of [30]. So, by Lemma 3.1, $S_n = M_{nn} + R_{nn}$, where $\|R_{nn}\| = o(\sqrt{n})$ and

$$M_{nn} = \sum_{j=1}^{\infty}\bar{b}_{n,j}\zeta_{n,j},$$

where $\zeta_{n,j} = \xi_{1,j} + \ldots + \xi_{n,j}$. So, if the CCLT holds for any choice of $F_1, F_2, \ldots$ with means 0 and unit variances, then $\lim_{n\to\infty}\|\bar{\boldsymbol{b}}_n\|^2 = \kappa^2$. In particular, $\bar{\boldsymbol{b}}_n,\ n \geq 1$, are bounded and, therefore, weakly pre-compact. To show pre-compactness, it therefore suffices to show that any weak limit point is a strong limit point. Let $\boldsymbol{b} \in \ell^2(\mathbb{N})$ be an arbitrary weak limit point and let $\mathbb{N}_0$ be a subsequence for which $\lim_{n\in\mathbb{N}_0} \bar{\boldsymbol{b}}_n = \boldsymbol{b}$. Then $\lim_{n\in\mathbb{N}_0} \bar{b}_{n,j} = b_j$ for all j, and

$$\lim_{n\in\mathbb{N}_0} \sum_{j=1}^{j_n} [\bar{b}_{n,j} - b_j]^2 = 0$$

for some subsequence $j_n \to \infty$. By thining the subsequence $\mathbb{N}_0$, if necessary, we may suppose that $j_n, n \in \mathbb{N}_0$ are strictly increasing. There is a strictly decreasing sequence $1 > q_1 > q_2, \ldots$ for which $\lim_{n\in\mathbb{N}_0} nq_{j_n} = 0$. Let $p_j = q_j - q_{j+1}$ and let F_j be the distribution which assigns mass $\frac{1}{2}p_j$ to $\pm 1/\sqrt{p_j}$ and mass $1 - p_j$ to 0. With this choice of $F_1, F_2, \ldots$, let

$$\tilde{M}_{n,n} = \sum_{j=1}^{j_n} \bar{b}_{n,j}\zeta_{n,j}.$$

Then $P[\zeta_{n,j} \neq 0] \leq np_j$, and

$$P[M_{n,n} \neq \tilde{M}_{n,n}] \leq nq_{j_n} \to 0$$

as $n \to \infty$ in $\mathbb{N}_0$. So, $\tilde{M}_{n,n}/\sqrt{n}$ has a limiting normal distribution with mean 0 and variance κ^2 and, therefore,

$$\liminf_{n\in\mathbb{N}_0} \sum_{j=1}^{j_n} \bar{b}_{n,j}^2 = \liminf_{n\in\mathbb{N}_0} \frac{1}{n} E(\tilde{M}_{n,n}^2) \geq \kappa^2,$$

and

$$\lim_{n\in\mathbb{N}_0} \sum_{j=j_n+1}^{\infty} \bar{b}_{n,j}^2 = 0.$$

It follows easily that $\lim_{n\in\mathbb{N}_0} \bar{\boldsymbol{b}}_n = \boldsymbol{b}$ in $\ell^2(\mathbb{N})$, and since $\boldsymbol{b}$ was an arbitrary weak limit point, this establishes the first assertion.

The second assertion is now immediate. Setting all of the variances but one to zero, shows that $\lim_{n\to\infty} \bar{b}_{n,j}^2$ exists for a fixed j, in which case $\lim_{n\to\infty} \bar{b}_{n,j}$ exists, since $|b_{n+1,j} - \bar{b}_{n,j}| = O(\sqrt{n})$, as in the proof of Corollary 3.7. It then follows that $\bar{\boldsymbol{b}}_n$ converges weakly, from which the assertion follows since $\bar{\boldsymbol{b}}_n,\ n \geq 1$, are pre-compact.

□

CHAPTER IV

Law of the Iterated Logarithm

4.1 Introduction

Let $\ldots X_{-1}, X_0, X_1, \ldots$ denote a centered, square integrable, (strictly) stationary and ergodic process, defined on a probability space $(\Omega, \mathcal{A}, P)$, with partial sums denoted by $S_n = X_1 + \cdots + X_n$. The main question addressed is the Law of the Iterated Logarithm: under what conditions is

$$(4.1) \qquad \limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log_2 n}} = \sigma \ w.p.1$$

for some $0 \le \sigma < \infty$, where $\log_2 n = \log\log n$. Of course, (4.1) holds if the X_i are independent, by the classic work of Hartman and Wintner [22], and more generally– for example, [37], [23], and [34]. Here we employ an approach which has been used recently in the study of the central limit question for stationary processes, martingale approximations.

As in Maxwell and Woodroofe [30], it is convenient to suppose that X_k is of the form $X_k = g(W_k)$, where $\ldots W_{-1}, W_0, W_1, \ldots$ is a stationary, ergodic Markov chain. The state space, transition function, and (common) marginal distribution are denoted by $\mathcal{W}$, Q, and π; thus, $\pi(B) = P[W_n \in B]$, and

$$Qf(w) = E[f(W_{n+1})|W_n = w]$$

for a.e. $w \in \mathcal{W}$, measurable $B \subseteq \mathcal{W}$, and $f \in L^1(\pi)$. The iterates of Q are denoted by Q^k. It is also convenient to suppose that the probability space Ω is endowed with an ergodic, measure preserving transformation θ for which $W_k \circ \theta = W_{k+1}$ for all k. Neither convenience entails any loss of generality, since we may let the probability space be $\mathbb{R}^{\mathbb{Z}}$, X_k be the coordinate functions, $W_k = (\ldots X_{k-1}, X_k)$, and θ be the shift transformation. Some other choices of W_k are considered in the examples.

Let $\|\cdot\|$ denote the norm in $L^2(P)$, $\mathcal{F}_k = \sigma(\ldots, W_{k-1}, W_k)$, and recall the main result of [30]; if

$$\sum_{n=1}^{\infty} n^{-3/2} \|E(S_n|\mathcal{F}_0)\| < \infty, \tag{4.2}$$

then

$$\sigma^2 := \lim_{n\to\infty} \frac{1}{n} E(S_n^2) \tag{4.3}$$

exists and is finite, and

$$S_n = M_n + R_n, \tag{4.4}$$

where M_n is a square integrable martingale with ergodic, stationary increments, and $||R_n|| = o(\sqrt{n})$. It is shown in [30] that if (4.2) holds, then the conditional distributions of $S_n/\sqrt{n}$, given $\mathcal{F}_0$ converge *in probability* to the normal distribution with mean 0 and variance σ^2 (See their Corollary 1). It can also be shown that (4.2) is *best possible* through Peligrad and Utev [32].

To state the main result of this chapter, let ℓ be a positive, nondecreasing and slowly varying (at ∞) function and let

$$\ell^*(n) = \sum_{j=1}^{n} \frac{1}{j\ell(j)}.$$

Theorem 4.1. *If ℓ is a positive, slowly varying, nondecreasing function and*

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \sqrt{\ell(n)} \log(n) \| E(S_n | \mathcal{F}_0) \| < \infty, \tag{4.5}$$

then

$$\lim_{n \to \infty} \frac{R_n}{\sqrt{n \ell^*(n)}} = 0 \ w.p.1.$$

Corollary 4.2. *If (4.5) holds with $\ell(n) = 1 \vee \log(n)$, then (4.1) holds.*

Proof. In this case $\ell^*(n) \sim \log_2 n$, so that $R_n/\sqrt{n \log_2 n} \to 0$ as $n \to \infty$, and

$$\limsup_{n \to \infty} \frac{S_n}{\sqrt{2n \log_2 n}} = \limsup_{n \to \infty} \frac{M_n}{\sqrt{2n \log_2 n}}$$

both *w.p.1.* The corollary now follows from the Law of the Iterated Logarithm of martingales–for example, Stout [37]. $\square$

The next corollary strengthens the conclusion of [30] from convergence in probability to convergence *w.p.1*, under a slightly stronger hypothesis. Kipnis and Varadhan [26] call this an important question in a closely related context (see their Remark 1.7). Let F_n denote a regular conditional distribution function for $S_n/\sqrt{n}$ given $\mathcal{F}_0$, so that

$$F_n(\omega; z) = P[\frac{S_n}{\sqrt{n}} \leq z | \mathcal{F}_0](\omega)$$

for $\omega \in \Omega$ and $-\infty < z < \infty$; and Φ_σ denote the normal distribution with mean 0 and variance σ^2.

Corollary 4.3. *If (4.5) holds with some ℓ for which $1/[n\ell(n)]$ is summable, then $F_n(\omega; \cdot)$ converges weakly to Φ_σ for a.e. ω.*

Proof. Let G_n be a regular conditional distribution for $M_n/\sqrt{n}$ given $\mathcal{F}_0$. Then $G_n(\omega; \cdot)$ converges weakly to Φ_σ for *a.e.* ω, essentially by the Martingale Central

Limit Theorem, applied conditionally given $\mathcal{F}_0$. See [30] for the details. Moreover, $P[\lim_{n\to\infty} R_n/\sqrt{n} = 0|\mathcal{F}_0] = 1$ $w.p.1$, since $P[\lim_{n\to\infty} R_n/\sqrt{n} = 0] = 1$, by Theorem 4.1. The corollary follows easily. $\square$

A major contribution of this chapter is to obtain a simple, general sufficient condition (4.5) for the LIL. Our results differ from those of Arcones [2], for example, by not requiring normality, and those of Rio [34] by not requiring strong mixing. In [29] Lai and Stout have a quite general result for strongly dependent variables. Their results require a condition on the moment generating function of the delayed partial sums, and only cover the upper half of LIL. Yokoyama [46] also uses martingale approximation in a similar setting to ours. His results require a martingale approximation, as in (4.4) and bounds on higher moments of the remainder term.

The rest of this chapter is organized as follows. The proof of Theorem 4.1 is outlined in Section 4.2, with supporting details in Sections 4.3 and 4.4. Invariance principles are considered in Section 4.5, and examples in Section 4.6.

4.2 Outline of the Proof

In this section we give an outline of the proof for the main result. Let

$$h_\epsilon = \sum_{k=1}^{\infty} \frac{Q^{k-1}g}{(1+\epsilon)^k} \tag{4.6}$$

and $H_\epsilon(w_0, w_1) = h_\epsilon(w_1) - Qh_\epsilon(w_0)$. Thus $H_\epsilon \in L^2(\pi_1)$, where π_1 denotes the joint distribution of W_0 and W_1. In [30] it is shown that if (4.2) holds, then $H := \lim_{\epsilon\downarrow 0} H_\epsilon$ exists in $L^2(\pi_1)$ and that (4.4) holds with $M_n = H(W_0, W_1) + \cdots + H(W_{n-1}, W_n)$. Letting $\xi_k = g(W_k) - H(W_{k-1}, W_k)$ leaves

$$R_n = \sum_{k=1}^{n} \xi_k = \sum_{k=1}^{n} \xi_0 \circ \theta^k \tag{4.7}$$

in (4.4).

For appropriately chosen $\beta_k \sim c/\sqrt{k^3\ell(k)}$ (see (4.12), below), the series

$$B(z) = \sum_{k=1}^{\infty} \beta_k z^k \tag{4.8}$$

converges for all complex $|z| \leq 1$, is analytic in $|z| < 1$, $B(1) = 1$, and $|1 - B(z)| > 0$ for $z \neq 1$. Letting T be the operator on $L^2(P)$ defined by $T\eta = \eta \circ \theta$, it is also true that $B(T)$ converges in the operator norm. Thus,

$$B(T)\eta = \sum_{k=1}^{\infty} \beta_k T^k \eta = \sum_{k=1}^{\infty} \beta_k \eta \circ \theta^k. \tag{4.9}$$

With this notation, there are two main steps to the proof. It is first shown that in (4.7), $\xi_0 \in [I - B(T)]L^2(P)$, the range of $I - B(T)$, so that $\xi_0 = \eta_0 - B(T)\eta_0$ for some $\eta_0 \in L^2(P)$. It is then shown that for any $\xi \in [I - B(T)]L^2(P)$, with probability one,

$$\lim_{n\to\infty} \frac{1}{\sqrt{n\ell^*(n)}} \sum_{k=1}^{n} T^k \xi = 0.$$

The broad brush strokes follow Derriennic and Lin [11], but with complications. Formally, the solution to the equation $\xi_0 = \eta_0 - B(T)\eta_0$ is $\eta_0 = A(T)\xi_0$, where

$$A(z) = \frac{1}{1 - B(z)} = \sum_{k=0}^{\infty} \alpha_k z^k, \tag{4.10}$$

but there are technicalities in attaching a meaning to $A(T)\xi_0$.

4.3 Fourier Analysis

The Size of R_n. The first item of business is to estimate the size of $\|R_n\|$. Here and below, the symbol $\|\cdot\|$ is used more generally to denote the norm in an L^2 space, which may vary from one usage to the next.

Lemma 4.4. *Let $\delta_j = 2^{-j}$. If (4.5) holds, then*

$$\sum_{j=1}^{\infty} j\sqrt{\ell(2^j)}\sqrt{\delta_j}\|h_{\delta_j}\| < \infty,$$

where (now) $\|\cdot\|$ denotes the norm in $L^2(\pi)$.

PROOF. Let $V_n g = g + Qg + \cdots + Q^{n-1}g$, so that $V_n g(w) = E[S_n|W_1 = w]$ and $\|V_n g\| \le 2||X_0|| + ||E(S_n|\mathcal{F}_0)||$. Then, rearranging terms in (4.6),

$$\|h_{\delta_j}\| \le \delta_j \sum_{n=1}^{\infty} \frac{\|V_n g\|}{(1+\delta_j)^n},$$

and

$$\sum_{j=1}^{\infty} j\sqrt{\ell(2^j)}\sqrt{\delta_j}\|h_{\delta_j}\| \le \sum_{n=1}^{\infty}\Big[\sum_{j=1}^{\infty} \frac{j\sqrt{\ell(2^j)\delta_j^3}}{(1+\delta_j)^n}\Big]\|V_n g\|.$$

Comparing the inner sum to an integral for any fixed integer $n \ge 0$, then

$$\sum_{j=1}^{\infty} \frac{j\sqrt{\ell(2^j)\delta_j^3}}{(1+\delta_j)^n} \le \log_2(e)\int_0^1 \frac{\sqrt{t\ell(2/t)}\log(2/t)}{(1+\frac{1}{2}t)^n}\,dt.$$

By a change of variables and the dominated convergence theorem, using Potter's bound (cf. [4], page 25) to supply a dominating function, the integral on the right hand side of last inequality is just

$$\frac{1}{\sqrt{n^3}}\int_0^n \sqrt{t\ell(\frac{2n}{t})}\log\left(\frac{2n}{t}\right)\left(1+\frac{t}{2n}\right)^{-n} dt \sim \frac{\sqrt{\ell(n)}\log(n)}{\sqrt{n^3}}\int_0^{\infty}\sqrt{t}e^{-\frac{1}{2}t}dt,$$

from which the lemma follows. $\square$

Proposition 4.5. *If (4.5) holds, then*

$$\lim_{n\to\infty}\sqrt{\ell(n)}\,\frac{\|R_n\|}{\sqrt{n}} = 0 \qquad \text{and} \qquad \sum_{n=1}^{\infty}\sqrt{\frac{\ell(n)}{n^3}}\,\|R_n\| < \infty. \tag{4.11}$$

PROOF. Let $H_\epsilon(w_0, w_1) = h_\epsilon(w_1) - Qh_\epsilon(w_0)$, and $M_n(\epsilon) = H_\epsilon(W_0, W_1) + \cdots + H_\epsilon(W_{n-1}, W_n)$. Then, it is shown in [30] that $S_n = M_n(\epsilon) + R_n(\epsilon)$ for each $\epsilon > 0$ with $R_n(\epsilon) = \epsilon S_n(h_\epsilon) + Qh_\epsilon(W_0) - Qh_\epsilon(W_n)$ and $S_n(h_\epsilon) = h_\epsilon(W_1) + \cdots + h_\epsilon(W_n)$. So,

$$R_n = M_n(\epsilon) - M_n + \epsilon S_n(h_\epsilon) + Qh_\epsilon(W_0) - Qh_\epsilon(W_n)$$

and

$$\|R_n\| \le \|M_n(\epsilon) - M_n\| + (n\epsilon + 2)\|h_\epsilon\| \le \sqrt{n}\|H_\epsilon - H\| + (n\epsilon + 2)\|h_\epsilon\|.$$

Now let $\epsilon_n = 2^{-k_n}$, where $2^{k_n-1} \le n < 2^{k_n}$. Then $1/(2n) \le \epsilon_n = \delta_{k_n} \le 1/n$, and $\|H_{\delta_{j+1}} - H_{\delta_j}\| \le 4\sqrt{\delta_j}\|h_{\delta_j}\|$, by Lemma 2 of [30],

$$\|R_n\| \le \sqrt{n}\sum_{j=k_n}^{\infty} \|H_{\delta_{j+1}} - H_{\delta_j}\| + 3\|h_{\delta_{k_n}}\| \le 10\sqrt{n}\sum_{j=k_n}^{\infty}\sqrt{\delta_j}\|h_{\delta_j}\|.$$

Since $k_n \le j$ implies $n < 2^j$, and so

$$\sum_{k_n \le j} \frac{\sqrt{\ell(n)}}{n} \le \sqrt{\ell(2^j)}\sum_{n<2^j}\frac{1}{n} \le 2j\sqrt{\ell(2^j)},$$

then we derive

$$\begin{aligned}\sum_{n=1}^{\infty}\sqrt{\frac{\ell(n)}{n^3}}\|R_n\| &\le 10\sum_{j=1}^{\infty}\Big[\sum_{k_n\le j}\frac{\sqrt{\ell(n)}}{n}\Big]\sqrt{\delta_j}\|h_{\delta_j}\| \\ &\le 20\sum_{j=1}^{\infty}\sqrt{\ell(2^j)}j\sqrt{\delta_j}\|h_{\delta_j}\|,\end{aligned}$$

which is finite by the previous lemma. Thus, the series in (4.11) converges. That $\sqrt{\ell(n)}\|R_n\|/\sqrt{n} \to 0$ then follows from the sub-additivity of $\|R_n\|$; $\|R_{m+n}\| \le \|R_m\| + \|R_n\|$. Since $\|R_n\| \le \|R_k\| + \|R_{n-k}\|$ for all $k = 1, \cdots n-1$, and therefore,

$$\sqrt{\frac{\ell(n)}{n}}\|R_n\| \le 6\sqrt{\frac{\ell(n)}{n^3}}\sum_{\frac{1}{4}n\le k\le\frac{3}{4}n}\|R_k\| \le 6\sum_{\frac{1}{4}n\le k\le\frac{3}{4}n}\sqrt{\frac{\ell(k)}{k^3}}\|R_k\|$$

for all sufficiently large n, and this approaches 0 as already shown. □

The Size of α_n. Let

$$\beta_k = \frac{c}{k}\sum_{n=k}^{\infty}\frac{1}{\sqrt{n^3\ell(n)}} \tag{4.12}$$

where c is chosen so that $\beta_1 + \beta_2 + \cdots = 1$. Then, $B(z) = \sum_{k=1}^{\infty}\beta_k z^k$ converges for all $|z| \le 1$ in (4.8) and $\mathcal{R}B(z) < 1$ for all $z \ne 1$, so that $A(z)$ is well-defined in (4.10) for all $|z| \le 1$, except $z = 1$. Observe that $A(z)[1 - B(z)] = 1$ and, therefore,

$$\alpha_n = \sum_{k=1}^{n}\beta_k\alpha_{n-k} \tag{4.13}$$

for $n \geq 1$ and $\alpha_0 = 1$. Let

$$b(t) = B(e^{it}) = \sum_{k=1}^{\infty} \beta_k e^{ikt} \tag{4.14}$$

for $-\pi < t \leq \pi$.

Proposition 4.6. *b is twice differentiable on $-\pi < t \neq 0 < \pi$, $|1 - b(t)| \sim \kappa_0 \sqrt{|t|}/\sqrt{\ell(1/|t|)}$, and*

$$|b'(t)| \sim \frac{2c\sqrt{\pi}}{\sqrt{|t|\ell(1/|t|)}}, \qquad |b''(t)| \sim \frac{\kappa_2}{\sqrt{|t|^3 \ell(1/|t|)}} \tag{4.15}$$

as $t \to 0$, where $\kappa_0 \neq 0$ and κ_2 are constants (identified) in the proof.

Proof. Clearly (4.14) is absolutely convergent, b is continuous, and $b(0) = 1$. By Theorem 2.6 of Zygmund [49, p. 4], the formal expression for the derivative

$$b'(t) = i \sum_{k=1}^{\infty} \Big[\sum_{n=k}^{\infty} \frac{c}{\sqrt{n^3 \ell(n)}} \Big] e^{ikt} \tag{4.16}$$

converges uniformly on $\epsilon \leq |t| \leq \pi$ for any $\epsilon > 0$, and therefore, is the derivative of b. By Theorem 4.3.2 of [4, p. 207],

$$|b'(t)| \sim \frac{2c\sqrt{\pi}}{\sqrt{|t|\ell(1/|t|)}}$$

as $t \to 0$. So, $|1 - b(t)| \sim 4c\sqrt{\pi|t|}/\sqrt{\ell(1/|t|)}$. Reversing the order of summation in (4.16) (which can be justified by truncating the outer sum at K and letting $K \to \infty$) gives us,

$$b'(t) = i \sum_{n=1}^{\infty} \Big[\sum_{k=1}^{n} e^{ikt} \Big] \frac{c}{\sqrt{n^3 \ell(n)}} = \frac{e^{it}}{1 - e^{it}} \sum_{n=1}^{\infty} (1 - e^{int}) \frac{ic}{\sqrt{n^3 \ell(n)}} = f(t) g(t),$$

where $f(t) = e^{it}/(1 - e^{it})$ is continuously differentiable on $-\pi < t \neq 0 < \pi$, and g is continuous. As above,

$$g'(t) = \sum_{n=1}^{\infty} e^{int} \frac{c}{\sqrt{n \ell(n)}}$$

converges uniformly on $\epsilon \le |t| \le \pi$ and

$$|g'(t)| \sim c\sqrt{\pi}\frac{1}{\sqrt{|t|\ell(1/|t|)}}$$

as $t \to 0$. Hence, b is twice continuously differentiable on $-\pi < t \neq 0 < \pi$, and the second relationship in (4.15) follows from $b''(t) = f'(t)g(t) + f(t)g'(t) = f(t)g'(t) + [ib'(t)/(1-e^{it})]$ and symmetry. $\square$

In (4.10), $A(z)$ is defined for all $|z| \le 1$, except $z = 1$. Let $a(t) = A(e^{it})$ for $-\pi < t \neq 0 < \pi$, then one can derive the following properties.

Corollary 4.7. *a is twice differentiable on $0 < |t| < \pi$, and*

$$|a'(t)| \sim \frac{1}{8c\sqrt{\pi}}\frac{\sqrt{\ell(1/|t|)}}{\sqrt{|t|^3}}, \qquad \text{and} \qquad |a''(t)| = O\left(\frac{\sqrt{\ell(1/|t|)}}{\sqrt{|t|^5}}\right)$$

as $t \to 0$.

Proof. This follows directly from (4.10) and Proposition 4.6. $\square$

Proposition 4.8. *Let α_n be the coefficients of $A(z)$, then $0 < \alpha_n \le 1$ for all $n \ge 0$ and*

$$\alpha_n - \alpha_{n+1} = O\left(\frac{\sqrt{\ell(n)}}{\sqrt{n^3}}\right)$$

as $n \to \infty$.

Proof. The first assertion follows easily from (4.13) and induction. By Proposition 4.6, a is absolutely integrable, so that $2\pi\alpha_n = \int_{-\pi}^{\pi} e^{-int}a(t)dt$, and then

$$\alpha_n - \alpha_{n+1} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-int}a_*(t)dt,$$

where $a_*(t) = [1-e^{-it}]a(t)$. Both $a_*'(s)$ and $sa_*''(s)$ are integrable over $(-\pi, \pi]$. Hence, integration by parts (twice) is justified and yields

$$\alpha_n - \alpha_{n+1} = \frac{1}{2\pi in}\int_{-\pi}^{\pi} e^{-int}a_*'(t)dt = \frac{1}{2\pi n^2}\int_{-\pi}^{\pi}[1-e^{-int}]a_*''(t)dt.$$

By Corollary 4.7, there is a C for which $|a''_*(t)| \le C\sqrt{\ell(1/|t|)/|t|^3}$ for all $0 < |t| \le \pi$. So

$$\begin{aligned}
|\alpha_n - \alpha_{n+1}| &= \frac{1}{2\pi n^3}\left|\int_{-\pi n}^{\pi n} [1-e^{-it}]a''_*(\frac{t}{n})\ dt\ \right| \\
&\le \frac{C}{2\pi n^3}\int_{-\pi n}^{\pi n} |1-e^{-it}|\sqrt{\frac{n^3}{|t|^3}\ell(\frac{n}{|t|})}\ dt \\
&\sim \frac{C}{2\pi}\sqrt{\frac{\ell(n)}{n^3}}\int_{-\infty}^{\infty} |1-e^{-it}|\frac{dt}{\sqrt{|t|^3}},
\end{aligned}$$

using Potter's theorem again and monotonicity of ℓ. This establishes the proposition.

$\square$

Existence of η_0. We need the following fact which is easily deduced from Lemma 1.3 of Krengel [27, p. 4]: Let $L^2_0(P)$ be the set of $\eta \in L^2(P)$ with mean 0; if θ is ergodic, then $[I-T]L^2_0(P)$ is dense in $L^2_0(P)$. Recall the defintion of ξ_0 in (4.7) and expression for $B(T)$ in (4.9); observe that $\xi_0 \in L^2_0(P)$; and let $A_N(T) = \sum_{n=0}^N \alpha_n T^n$ and $U_n = T + \cdots + T^n$.

Proposition 4.9. *If (4.5) is satisfied, then* $\eta_0 = \lim_{N\to\infty} A_N(T)\xi_0$ *exists in* $L^2(P)$*, and* $\xi_0 = [I - B(T)]\eta_0$*.*

Proof. From (4.7), we have $U_n\xi_0 = R_n$. Then, summing by parts,

$$A_N(T)\xi_0 = \xi_0 + \alpha_N R_N + \sum_{n=1}^{N-1}(\alpha_n - \alpha_{n+1})R_n.$$

In view of Propositions 4.5 and 4.8 and Karamata's theorem, the sum converges in $L^2(P)$ and $\alpha_N R_N \to 0$.

For the second assertion, let $\eta_N = A_N(T)\xi_0$. Then, rearranging terms and using

(4.13),

$$
\begin{aligned}
B(T)\eta_N &= \sum_{k=1}^{\infty}\beta_k \sum_{j=0}^{N}\alpha_j T^{j+k}\xi_0 \\
&= \sum_{m=1}^{N}\alpha_m T^m \xi_0 + \sum_{m=N+1}^{\infty}\big[\sum_{j=0}^{N}\alpha_j\beta_{m-j}\big]T^m\xi_0 \\
&= \eta_N - \xi_0 + C_N(T)\xi_0
\end{aligned}
$$

where $C_N(T) := I - [I - B(T)]A_N(T)$. So, it suffices to show that $\|C_N(T)\xi_0\| \to 0$. For this, first observe that, replacing T by z in the definition of $C_N(T)$, $1 - C_N(z) = [1 - B(z)]A_N(z)$. Then $C_N(1) = 1$ and the coefficients of $C_N(z)$ are all positive, so that $\|C_N(T)\|_{op} \leq 1$, where $\|\cdot\|_{op}$ stands for operator norm. So, it suffices to show that $\|C_N(T)\xi\| \to 0$ for all $\xi \in [I - T]L_0^2(P)$, a dense subset of $L_0^2(P)$. This is easy: for if $\xi = \psi - T\psi$, then

$$C_N(T)\xi = \sum_{j=0}^{N}\alpha_j\left[\beta_{N+1-j}T^{N+1}\psi + \sum_{m=N+1}^{\infty}(\beta_{m+1-j} - \beta_{m-j})T_m\psi\right]$$

and

$$\|C_N(T)\xi\| \leq 2\|\psi\|\sum_{j=0}^{N}\alpha_j\beta_{N+1-j} \to 0$$

as $N \to \infty$ by (4.13) and Proposition 4.8. $\square$

4.4 Ergodic Theory

Some preparation is necessary for the second step. First, for any $\eta \in L^2(P)$, $\eta^* := \sup_{n\geq 1} U_n|\eta|/n \in L^2(P)$ by the Dominated Ergodic Theorem (see, for example, Krengel [27, p. 52]). We will also use the following fact:

$$E\left(\sqrt{(\eta^2)^*}\,\right) \leq 2||\eta||, \tag{4.17}$$

whose proof is essentially an application of the Maximal Ergodic Theorem [33, Corollary 2.2] to $(\eta^2)^*$.

The proof of Theorem 4.1 will be completed by proving:

Theorem 4.10. *If $\xi \in [I - B(T)]L^2(P)$, then*

$$\lim_{n\to\infty} \frac{U_n \xi}{\sqrt{n\ell^*(n)}} = 0 \ w.p.1.$$

Proof. By assumption, there is an $\eta \in L^2(P)$ for which, $\xi = \eta - B(T)\eta = \sum_{k=1}^{\infty} \beta_k[\eta - T^k\eta]$, and there is no loss of generality in supposing that $\eta \in L_0^2(P)$. Observe that $|T^k\eta|^p = T^k(|\eta|^p)$ for any integer $k \geq 0$ and real $p > 0$, and write

$$U_n\xi = I_n\eta + II_n\eta,$$

where

$$I_n\eta = \sum_{k=1}^{n} \beta_k U_n[\eta - T^k\eta],$$

and

$$II_n\eta = \sum_{k=n+1}^{\infty} \beta_k U_n[\eta - T^k\eta].$$

If $k > n$, then $|U_n(\eta - T^k\eta)| \leq |U_n\eta| + |U_nT^k\eta| \leq [\eta^* + T^k\eta^*]n$. So,

$$|II_n\eta| \leq n \sum_{k=n+1}^{\infty} \beta_k[\eta^* + T^k\eta^*].$$

Here

$$\sum_{k=n+1}^{\infty} \beta_k T^k\eta^* \leq \sum_{k=n+1}^{\infty} \Delta\beta_k U_k\eta^* \leq \sum_{k=n+1}^{\infty} k\Delta\beta_k\eta^{**},$$

where $\Delta\beta_k = \beta_k - \beta_{k+1}$ and $\eta^{**} = \sup_{k\geq 1} U_k\eta^*/k$. Observing that

$$\sum_{k=n+1}^{\infty} (\beta_k + k\Delta\beta_k) = n\beta_{n+1} + 2\sum_{k=n+1}^{\infty} \beta_k,$$

Thus,

$$|II_n\eta| \leq n(\eta^* \vee \eta^{**})\Big[\sum_{k=n+1}^{\infty} \beta_k + \sum_{k=n+1}^{\infty} k\Delta\beta_k\Big] = (\eta^* \vee \eta^{**}) \times O\left(\sqrt{\frac{n}{\ell(n)}}\right),$$

and

$$\lim_{n\to\infty} \frac{II_n\eta}{\sqrt{n\ell^*(n)}} = 0 \ w.p.1. \tag{4.18}$$

Similarly, for $k \le n$, $U_n\eta - U_nT^k\eta = U_k\eta - U_kT^n\eta$, then

$$I_n\eta = \sum_{k=1}^{n} \beta_k U_k \eta - \sum_{k=1}^{n} \beta_k U_k T^n \eta.$$

Letting $\gamma_j = \sum_{k=j}^{\infty} \beta_k$ and recalling (4.12), we have

$$\sum_{j=1}^{n} \gamma_j^2 \sim (4c)^2 \left(\sum_{j=1}^{n} \frac{1}{j\ell(j)} \right) = (4c)^2 \ell^*(n),$$

and

$$|I_n\eta| \le \sum_{k=1}^{n} \beta_k \sum_{j=1}^{k} \left[T^j|\eta| + T^{j+n}|\eta|\right] \le \sum_{j=1}^{n} \gamma_j \left[T^j|\eta| + T^{j+n}|\eta|\right]$$

$$\le \sqrt{\sum_{j=1}^{n} \gamma_j^2} \times \sqrt{2 \times \sum_{j=1}^{2n} T^j \eta^2}.$$

Using (4.17), there exists a constant $C > 0$, such that

$$E\left(\sup_n \frac{|I_n\eta|}{\sqrt{n\ell^*(n)}} \right) \le C||\eta||,$$

where C doesn't depend on η. Hence, to show

$$\lim_{n\to\infty} \frac{I_n\eta}{\sqrt{n\ell^*(n)}} = 0 \ w.p.1 \qquad (4.19)$$

for each $\eta \in L_0^2(P)$, one only needs to consider $\eta \in (I - T)L_0^2(P)$, a dense subset in $L_0^2(P)$, and this is easy. If $\eta = \phi - T\phi$ for some $\phi \in L_0^2(P)$, then $U_kT^n\eta = T^{n+1}\phi - T^{k+n+1}\phi$ for $1 \le k \le n$, so that

$$|I_n\eta| \le |T \sum_{k=1}^{n} \beta_k(\phi - T^k\phi)| + |T^{n+1} \sum_{k=1}^{n} \beta_k(\phi - T^k\phi)| \le T\tilde{\phi} + T^{n+1}\tilde{\phi},$$

where

$$\tilde{\phi} = \sum_{k=1}^{\infty} \beta_k |\phi - T^k\phi| \in L^2(P).$$

Since $\tilde{\phi} \in L^2(P)$, $\lim_{n\to\infty} T^{n+1}\tilde{\phi}/\sqrt{n} = 0 \ w.p.1$ by an easy application of the Borel-Cantelli lemmas, and therefore, $\lim_{n\to\infty} I_n\eta/\sqrt{n\ell^*(n)} = 0 \ w.p.1$. The theorem now follows by combining (4.18) and (4.19). $\square$

4.5 Invariance Principles

Let $C[0,1]$ be the space of all real-valued continuous functions on $[0,1]$, endowed with the metric

$$\rho(x,y) = \sup_{0\le t\le 1} |x(t) - y(t)|,$$

where $x, y \in C[0,1]$. For any $\nu \ge 0$, let K_ν denote the set of absolutely continuous functions $x \in C[0,1]$ such that $x(0) = 0$ and

$$\int_0^1 [x'(t)]^2 \, dt \le \nu^2.$$

Set $S_0 = M_0 = 0$, define sequences of random functions $\{\theta_n(\cdot)\}$ and $\{\zeta_n(\cdot)\}$ respectively by

$$\begin{aligned} \theta_n(t) &= \frac{S_k + (nt-k)X_{k+1}}{\sqrt{2n\log_2(n)}} \\ \zeta_n(t) &= \frac{M_k + (nt-k)(M_{k+1} - M_k)}{\sqrt{2n\log_2(n)}} \end{aligned}$$

for $k \le nt \le k+1, k = 0, 1, \cdots, n-1$. Then $\theta_n, \zeta_n \in C[0,1]$.

Corollary 4.11. *If the hypothesis in Corollary 4.2 holds, then w.p.1, $\{\theta_n\}_{n\ge 3}$ are relatively compact in $C[0,1]$, and the set of limit points is K_σ.*

PROOF. Under the hypothesis, (4.3) and (4.4) hold, then

$$\rho(\theta_n, \zeta_n) \le \max_{k\le n} \frac{|R_k|}{\sqrt{2n\log_2(n)}} \to 0 \ w.p.1,$$

which implies that θ_n and ζ_n have the same limit points; and the limit points of ζ_n are known to be K_σ $w.p.1$ (see, for example, Heyde and Scott [23], Corollary 2). $\square$

Let

$$\mathbb{B}_n(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}$$

for $0 \leq t < 1, \mathbb{B}_n(1) = \mathbb{B}_n(1-)$, where $\lfloor \cdot \rfloor$ denotes the integer part. Then $\mathbb{B}_n \in D[0,1]$, the space of cadlag functions as described in Chapter 3 of Billingsley [3]. Let F_n denote a regular conditional distribution for $\mathbb{B}_n$ given $\mathcal{F}_0$, so that $F_n(\omega; B) = P[\mathbb{B}_n \in B|\mathcal{F}_0](\omega)$ for Borel sets $B \subseteq D[0,1]$; and let Φ_σ denote the distribution of $\sigma\mathbb{B}$, where $\mathbb{B}$ is a standard Brownian motion. Let Δ denote the Prokhorov metric on $D[0,1]$ (cf. [3], page 72).

Corollary 4.12. *If the hypothesis in Corollary 4.3 holds, then*

$$\lim_{n\to\infty} \Delta[F_n(\omega;\cdot), \Phi_\sigma] = 0 \qquad a.e.\ \omega \tag{4.20}$$

PROOF. For $S_n = M_n + R_n$, let $M_n^*(t) = M_{\lfloor nt \rfloor}/\sqrt{n}, 0 \leq t < 1$ and $M_n^*(1) = M_n^*(1-)$. Let G_n denote a regular conditional distribution for the random element M_n^* given $\mathcal{F}_0$. Then $G_n(\omega;\cdot)$ converges to Φ_σ for a.e. ω (P), by verifying Theorem 2.5 of Durrett and Resnick [14] in view of the mean ergodic theorem. Under the hypothesis of Corollary 4.3, $\max_{1\leq k\leq n} |R_k|/\sqrt{n} \to 0$ $w.p.1$, and therefore,

$$\rho(M_n^*, \mathbb{B}_n) = \sup_{0\leq t\leq 1} |M_n^*(t) - \mathbb{B}_n(t)| \to 0 \qquad w.p.1.$$

(4.20) follows. □

4.6 Examples

In this section, we illustrate our conditions by considering linear processes, additive functionals of a Bernoulli shift, and ρ-mixing processes.

Linear processes. Let $\ldots \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ be an ergodic stationary martingale difference sequence with common mean 0 and variance 1. Define a linear process

$$X_k = \sum_{j=0}^{\infty} a_j \epsilon_{k-j},$$

where $a_0, a_1, \ldots$ is a square summable sequence, and observe that X_k is of the form $g(W_k)$ with $W_k = (\ldots, \epsilon_{k-1}, \epsilon_k)$.

Proposition 4.13. *Suppose* $a_n = O\left[1/(nL(n))\right]$*, where* $L(\cdot)$ *is a positive, non-decreasing, slowly varying function. If*

$$\sum_{n=2}^{\infty} \frac{\log^{\alpha}(n)}{nL(n)} < \infty \tag{4.21}$$

with $\alpha = 3/2$*, then (4.5) holds with* $\ell(n) = 1 \vee \log(n)$ *and, thus the conclusions to Corollaries 4.2 and 4.11; Furthermore, if (4.21) holds with some* $\alpha > 3/2$*, then also the conclusions to Corollaries 4.3 and 4.12.*

PROOF. Letting $s_{j,n} = a_{j+1} + \cdots + a_{j+n}$, straight forward calculations yield that

$$\|E[S_n|\mathcal{F}_0]\|^2 = \sum_{j=0}^{\infty} s_{j,n}^2.$$

If $j \geq 3$, then

$$|s_{j,n}| \leq \frac{C}{L(j)} \int_j^{j+n} \frac{1}{x}\, dx \leq \frac{C}{L(j)} \log(1 + \frac{n}{j})$$

for some constant $C > 0$, and therefore,

$$\begin{aligned}\sum_{j=3}^{\infty} s_{j,n}^2 &\leq C^2 \int_2^{\infty} \frac{1}{L^2(x)} \log^2(1 + \frac{n}{x})\, dx \\ &= nC^2 \int_0^{n/2} \frac{1}{L^2(n/t)} \frac{\log^2(1+t)}{t^2}\, dt = O\left[\frac{n}{L^2(n)}\right],\end{aligned}$$

where the last step follows from the dominated convergence theorem, using Potter's bound to supply the dominating function, or by Fatou's lemma. It is then easily verified that $\|E(S_n|\mathcal{F}_0)\| = O[\sqrt{n}/L(n)]$, and the proposition is an immediate consequence. $\square$

REMARK 1. If $L(n) \sim \log^{\beta}(n)$, then (4.21) requires $\beta > 5/2$. This is similar to, but not strictly comparable with, the results of Yokoyama [46], who required finite moments of order $p > 2$ and $\beta \geq 1 + (2/p)$.

Additive Functionals of the Bernoulli Shift. Now consider a Bernoulli shift process, say

$$W_k = \sum_{j=1}^{\infty} \frac{1}{2^j} \epsilon_{k-j+1},$$

where $\ldots \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ are i.i.d. random variables that take the values 0 and 1 with probability 1/2 each. Then $\mathcal{W} = [0,1]$, π is the uniform distribution, and

$$Qf(w) = \frac{1}{2}\left[f\left(\frac{w}{2}\right) + f\left(\frac{1+w}{2}\right)\right]$$

for $f \in L^1$. Next, consider a stationary process of the form $X_k = g(W_k)$, where g is square integrable with respect to π and has mean 0. In this case, it is possible to relate (4.5) to a weak regularity condition on g.

Proposition 4.14. *If*

$$\int_0^1 \int_0^1 \frac{[g(x)-g(y)]^2}{|x-y|} \log^{\frac{5}{2}+\delta}\left[\log\left(\frac{1}{|x-y|}\right)\right] dxdy < \infty \tag{4.22}$$

for some $\delta > 0$, then the conclusions to Corollaries 4.3 and 4.12 hold, and so also those of Corollaries 4.2 and 4.11.

Proof (*sketched*). The proof involves showing that (4.22) implies (4.5), for which, $\ell(n)$ can be chosen such that $\ell^*(n)$ remains bounded. The details are similar to the proof of Proposition 3 of Maxwell and Woodroofe [30], and will be omitted. □

ρ-mixing processes. Our condition (4.5) can be checked when a mixing rate is available for a ρ-mixing process, see [31, pp. 4-5] for a definition.

Corollary 4.15. *Let $\rho(n)$ be the ρ-mixing coefficients of a centered, square integrable, stationary process $(X_k)_{k\in\mathbb{Z}}$. If $\rho(n) = O(\log^{\gamma} n)$ for some $\gamma > 5/2$, as $n \to \infty$, then (4.1) holds.*

Proof (*outline*). Let $S_n = X_1 + \cdots + X_n$ and $h(x) = (1 \vee \log x)^{3/2}$. By a similar argument as in [31, p. 15], one can easily show that, for some constant $C > 0$,

$$\sum_{r=0}^{\infty} \frac{h(2^r)||E(S_{2^r}|\mathcal{F}_0)||}{2^{r/2}} \leq C \sum_{j=0}^{\infty} h(2^j)\rho(2^j) < \infty.$$

Since $||E(S_n|\mathcal{F}_0)||$ is sub-additive, it's then straightforward to argue as in Lemma 2.7 of [32], that

$$\sum_{n=1}^{\infty} \frac{h(n)||E(S_n|\mathcal{F}_0)||}{n^{3/2}} < \infty.$$

Therefore, (4.1) holds by Corollary 4.2. $\square$

Remark 2. Shao [36] showed that LIL holds when $\rho(n) = O(\log^{\gamma} n)$ for some $\gamma > 1$, but through a completely different approach.

CHAPTER V

Conditional Central Limit Theorem

5.1 The Problem

This is the only chapter where there is no theorem yet. Following the notations in Chapters III and IV, let $W_0, W_1, \ldots$ be an ergodic and strictly stationary Markov chain with measurable state space $(\mathcal{W}, \mathscr{B})$, and let π, Q denote the invariant distribution and transition kernel. Consider $g \in L_0^2(\mathcal{W}, \pi)$, the space of square-integrable functions with mean 0 under π; and let

$$S_n(g) := g(W_1) + \cdots + g(W_n).$$

The purpose of this chapter is to further pursue a study of conditional central limit questions with σ_n normalization, where $\sigma_n = \sigma_n(g) := \|S_n(g)\|$, and $\|\cdot\|$ stands for L^2 norm. Varying g, the standard deviation, $\sigma_n(g)$, may exhibit diffent kinds of behavior; see an example below. The linear case, $\sigma_n^2 \sim n\kappa$ for some $\kappa \geq 0$, has been well understood (e.g., [30, 47]); but the general case, including sublinear and superlinear, requires further investigations. To state the conditional central limit theorem (CCLT), let $S_n^* := S_n/\sigma_n$, and let F_n denote the conditional distribution function

$$F_n(w; z) := P(S_n^* \leq z | W_0 = w).$$

The CCLT asserts that

$$\lim_{n\to\infty} \int_{\mathcal{W}} \Delta[\Phi, F_n(w;\cdot)]\pi(dw) = 0, \tag{5.1}$$

where Φ is the standard normal distribution, and Δ denotes the Lévy metric which metrizes the weak convergence in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

It is shown in Wu and Woodroofe [45] that, if (5.1) holds, then necessarily $\nu_n = o(\sigma_n)$, which entails $\sigma_n^2 = n\ell(n)$ for some slowly varying $\ell(\cdot)$. Our aim here is to find a simple and usable criterion for (5.1). The importance of CCLT has been discussed in [7] and [45]. We recall here, CCLT will gurantee CLT, but not vice versa (cf. Example 1 of [45]). CCLT also bears relevance to MCMC, since (5.1) implies the asymptotic normality of S_n^* even when the chain starts at certain different distribution other than π.

5.2 Reversible Markov Chains

In the context of reversible Markov chains, it has been a well-known result, due to Kipnis and Varadhan [26], that if

$$\lim_{n\to\infty} \frac{E[S_n^2(g)]}{n} \to \sigma^2 \in [0, \infty),$$

then (5.1) holds. A natural question is then, to what extent, can the result be extended to $\sigma_n^2 = n\ell(n)$? To study this question, we first present a lemma, which points some direction along the line in [45].

Lemma 5.1. *Suppose the chain is reversible, i.e., $Q = Q^*$, then $\sigma_n^2 = n\ell(n)$ for some slowly-varying function $\ell(\cdot)$ iff $\|V_n g\| = o(\sigma_n)$.*

PROOF. One direction has been shown in [45]; now let us look at the other direction, assuming $\sigma_n^2 = n\ell(n)$. The following calculation is straightforward:

$$\begin{aligned}
\|V_n g\|^2 &= \sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\langle g, Q^{k+j} g\rangle \\
&= \sum_{i=0}^{n-1}(i+1)\langle g, Q^i g\rangle + \sum_{i=n}^{2n-2}(2n-1-i)\langle g, Q^i g\rangle \\
&= \sum_{i=0}^{2n-2}(2n-1-i)\langle g, Q^i g\rangle - 2\sum_{i=0}^{n-1}(n-1-i)\langle g, Q^i g\rangle \\
&= \frac{1}{2}\left[\sigma_{2n-1}^2 + (2n-1)\|g\|^2\right] - \left[\sigma_{n-1}^2 + (n-1)\|g\|^2\right] \\
&= \frac{1}{2}\sigma_{2n-1}^2 - \sigma_{n-1}^2 + \frac{1}{2}\|g\|^2.
\end{aligned}$$

The assertion of the lemma then follows easily. $\square$

Lemma 5.1 assures martingale approximations in triangular array under the single condition $\sigma_n^2 = n\ell(n)$, but it is still unclear whether (5.1) holds. Dalibor Volný (personal communications) indicated an example for which, the variances $\sigma_n^2 = n\ell(n)$, growing nonlinearly, and the CLT fails. So some additional conditions may be needed. Just to illustrate the approach, as developed in [45], is not as effective as we have expected, we shall now study $H_n/\sqrt{\ell(n)}$, where $H_n = \bar{V}_n g(w_1) - Q\bar{V}_n g(w_0)$. One can show $H_n/\sqrt{\ell(n)}$ is not Cauchy when $\ell(n) \to \infty$. First, by simple algebra,

$$\left\|\frac{H_n}{\sqrt{\ell(n)}} - \frac{H_m}{\sqrt{\ell(m)}}\right\|^2 = \frac{1}{\ell(n)}\|H_n\|^2 + \frac{1}{\ell(m)}\|H_m\|^2 - \frac{2}{\sqrt{\ell(m)\ell(n)}}\langle H_m, H_n\rangle,$$

where

$$\begin{aligned}
\langle H_m, H_n\rangle &= \langle \bar{V}_n g(w_1) - Q\bar{V}_n g(w_0), \bar{V}_m g(w_1) - Q\bar{V}_m g(w_0)\rangle \\
&= \langle \bar{V}_n g, \bar{V}_m g\rangle - \langle Q\bar{V}_n g, Q\bar{V}_m g\rangle \\
&= \langle \bar{V}_n g, \bar{V}_m g\rangle - \langle Q^2\bar{V}_n g, \bar{V}_m g\rangle \\
&= \langle (I - Q^2)\bar{V}_n g, \bar{V}_m g\rangle \\
&= \langle (V_2 - \frac{1}{n}QV_nV_2)g, \bar{V}_m g\rangle
\end{aligned}$$

then for any fixed m,

$$\lim_{n\to\infty}\left\|\frac{H_n}{\sqrt{\ell(n)}}-\frac{H_m}{\sqrt{\ell(m)}}\right\|^2=1+\frac{1}{\ell(m)}\|H_m\|^2,$$

in view of the mean ergodic theorem. Letting $m\to\infty$,

$$\lim_{m\to\infty}\left(\limsup_{n\to\infty}\left\|\frac{H_n}{\sqrt{\ell(n)}}-\frac{H_m}{\sqrt{\ell(m)}}\right\|^2\right)=2\neq 0.$$

So even in the context of reversible chains, more tools will be needed.

The Metropolis-Hastings algorithm. We shall construct Markov chains using the Metropolis-Hastings algorithm. Let

$$W_n=\begin{cases} W_{n-1} & \text{if } U_n\le p(W_{n-1})\\ Y_n & \text{o.w.}\end{cases}$$

where U_n are i.i.d. U(0,1), Y_n are i.i.d. with symmetric density $f(y)$, and $p(x)$ is some symmetric function with range $[0,1)$. To make the chain stationary, one can choose the common distribution π for W_n to be with density

$$\pi(x)=\frac{c_0 f(x)}{1-p(x)},$$

where c_0 is a normalizing constant. It is easy to check that the chain is also ergodic.

The transition kernel of the chain is given by

$$Q(x;dy)=p(x)\delta_x\{dy\}+(1-p(x))F\{dy\} \tag{5.2}$$

where $\delta_x(\cdot)$ is the Dirac measure putting unit mass at x, and F is the distribution corresponding to the density $f(y)$. Then for any F-integrable function $g:\mathbb{R}\to\mathbb{R}$,

$$Qg(x)=p(x)g(x)+(1-p(x))\kappa_g$$

where $\kappa_g=\int g(y)\ dF=E[g(Y_1)]$. Thus, in particular, when g is odd,

$$Qg(x)=p(x)g(x).$$

Now let us look at the functional $g(x) = x$ which is an odd function, then

$$Qg(x) = p(x)x, \quad Q^2 g(x) = (p(x))^2 x, \quad \cdots, \; Q^n g(x) = (p(x))^n x.$$

Specializing $p(x) = \exp(-1/|x|)$, then obviously $p(x) \in [0, 1)$ for any $x \in \mathbb{R}$. It is worth observing here $h(x) := x/(1 - p(x))$ formally solves Poisson's equation

$$x = g(x) = (I - Q)h(x),$$

but $h \notin L^2(\pi)$. Further, let the density of F be

$$f(x) = \frac{1 - p(x)}{\kappa_0(1 + |x|)^4} \sim \frac{1}{\kappa_0 |x|^5},$$

then

$$\pi(x) = \frac{3}{2(1 + |x|)^4}.$$

It follows that

$$\langle g, Q^n g\rangle = \int_{\mathbb{R}} [p(x)]^n x^2 \pi(x)\; dx = \int_{\mathbb{R}} e^{-\frac{n}{|x|}} \frac{x^2\; dx}{(1 + |x|)^4} = \int_{\mathbb{R}} e^{-\frac{1}{|y|}} \frac{n^3 y^2}{(1 + n|y|)^4}\; dy.$$

Using

$$\exp\left(-\frac{1}{|y|}\right) \frac{y^2}{(\frac{1}{n} + |y|)^4} \le \exp\left(-\frac{1}{|y|}\right) \frac{1}{y^2},$$

and applying the dominated convergence theorem, it can be shown that

$$\langle g, Q^n g\rangle \sim C\frac{1}{n}$$

for some constant $C > 0$. Further, one can show

$$\frac{\sigma_n^2}{n} = 2\sum_{k=0}^{n-1}(1 - \frac{k}{n})\langle g, Q^k g\rangle - \|g\|^2 \sim 2C\sum_{k=0}^{n-1}\frac{1}{k} + \text{constant} \sim 2C\log(n).$$

Thus, $\sigma_n^2 = n\ell(n)$ with $\ell(n) \sim \gamma \log n$ for some constant $\gamma > 0$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Anevski, D. and Hössjer, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874-1930.

[2] Arcones, M. A. (1999). The law of the iterated logarithm over a stationary Gaussian sequence of random vectors. *J. Theor. Prob.* **12** 615-641.

[3] Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley, New York.

[4] Bingham, N.H., Goldie, C.M. and Tuegels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[5] Borodin, A. N. and Ibragimov, I. A. (1994). *Limit theorems for functionals of random walks.* Trudy Mat. Inst. Steklov. **195.** Translated into English: Amer. Math. Soc. (1995) (Ch. 4, Sections 7-9) .

[6] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer, New York.

[7] Dedecker, J. and Merlevède, F. (2002). Necessary and sufficient conditions for the conditional central limit theorem. *Ann. Probab.* **30** 1044-1081.

[8] Dedecker, J., Merlevède, F. and Volný, D. (2007). On the weak invariance principle for non adapted sequences under projective criteria. *J. Theoret. Probab.* **20** 971-1004.

[9] Denker, M. and Gordin, M. I. (1998). The central limit theorem for random perturbations of rotations. *Probab. Theory Related Fields* **111** 1-16.

[10] Derriennic, Y. and Lin, M. (1996). Sur le théorème limite central de Kipnis et Varadhan pour les chaînes réversibles ou normales. *C. R. Acad. Sci. Paris Sér. I Math.* **323** 1053-1057.

[11] Derriennic, Y. and Lin, M. (2001). The central limit theorem for Markov chains with normal transition operators, started at a point. *Probab. Theory Related Fields* **119** 508-528.

[12] Dunford, N. and Schwartz, J. (1958). *Linear Operators.* Vol. I. Interscience, New York.

[13] Durieu, O. and Volný, D. (2007). Comparison between criteria leading to the weak invariance principle. *Ann. Inst. H. Poincaré Probab. Statist.* **44** 324-340.

[14] Durrett, R. and Resnick, S. (1978). Functional limit theorems for dependent variables. *Ann. Probab.* **6** 829-846.

[15] Gordin, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10** 1174-1176.

[16] Gordin, M. I. and Holzmann, H. (2004). The central limit theorem for stationary Markov chains under invariant splittings. *Stoch. Dyn.* **4** 15-30.

[17] Gordin, M. and Lifšic, B.A. (1978). Central limit theorems for stationary Markov processes. *Dokl. Akad. Nauk. SSSR* **239** 766-767.

[18] Gordin, M. and Lifšic, B.A. (1981). A remark about a Markov process with normal transition operator. In Third Vilnius Conference on Probability and Statistics, Akad. Nauk Litovsk, Vilnius. **1** 147-148 (in Russian).

[19] Groeneboom, P. (1985). Estimating a monotone density. In *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer* (Edited by L. LeCam and R. Olshen) **II** 539-555. Wadsworth, Belmont, CA.

[20] Groeneboom, P. and Wellner, J. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388-400.

[21] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

[22] Hartman, P. and Wintner, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.* **63** 169-176.

[23] Heyde, C.C. and Scott D.J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probab.* **1** 428-436.

[24] Ho, H.C. and Hsing, T. (1997) Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636-1669.

[25] Kim J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191-219.

[26] Kipnis, C. and Varadhan, S.R.S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1-19.

[27] Krengel, U. (1985). *Ergodic Theorems.* de Gruyter, Berlin.

[28] Kulikov, V. N. and Lopuhaä, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34** 742-768.

[29] Lai, T.L. and Stout, W.F. (1980). Limit theorems for sums of dependent random variables. *Z. Wahrsch. Verw. Gebiete* **51** 1-14.

[30] Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28** 713-724.

[31] Merlevède, F., Peligrad, M. and Utev, S. (2006). Recent advances in invariance principles for stationary sequences. *Probab. Surv.* **3** 1-36.

[32] Peligrad, M. and Utev, S. (2005). A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.* **33** 798-815.

[33] Petersen, K. (1983). *Ergodic Theory.* Cambridge University Press, Cambridge.

[34] Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.* **23** 1188-1203.

[35] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, Chichester.

[36] Shao, Q. M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stochastic Process. Appl.* **48** 319-334.

[37] Stout, W. F. (1970). The Hartman-Wintner law of the iterated logarithm for martingales. *Ann. Math. Statist.* **41** 2158-2160.

[38] Sun, J. and Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Plann. Inference* **52** 143-159.

[39] Tong, H. (1990). *Non-linear time series: A dynamical system approach.* Oxford University Press.

[40] Volný, D. (1993). Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic Process. Appl.* **44** 41-74.

[41] Wang, X. and Woodroofe, M. (2007). A Kiefer-Wolfowitz comparison theorem for Wicksell's problem. *Ann. Statist.* **35** 1559-1575.

[42] Woodroofe, M. (1992). A central limit theorem for functions of a Markov chain with applications to shifts. *Stochastic Process. Appl.* **41** 33-44.

[43] Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statist. Sinica* **3** 501-515.

[44] Wright, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443-448.

[45] Wu, W. B. and Woodroofe, M. (2004). Martingale approximations for sums of stationary processes. *Ann. Probab.* **32** 1674-1690.

[46] Yokoyama, R. (1995). On the central limit theorem and law of the iterated logarithm for stationary processes with applications to linear processes. *Stochastic Process. Appl.* **59** 343-351.

[47] Zhao, O. and Woodroofe, M. (2008). Law of the iterated logarithm for stationary processes. *Ann. Probab.* **36** 127-142.

[48] Zhao, O. and Woodroofe, M. (2008). On martingale approximations. *Ann. Appl. Probab.* to appear.

[49] Zygmund, A. (2003). *Trigonometric Series.* Paperback with a foreword by Robert Fefferman, Cambridge University Press, Cambridge.