# Marcinkiewicz Strong Law of Large Numbers for Products of Long Range Dependent and Heavy Tailed Linear Processes

by

Sounak Paul

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Mathematics

Department of Mathematical and Statistical Sciences
University of Alberta

# Abstract

Classical methods of inference are often rendered inapplicable while dealing with data exhibiting heavy tails, which gives rise to infinite variance and frequent extremes, and long memory, which induces inertia in the data. In this thesis, we develop the Marcinkiewicz Strong Law of Large Numbers, $\lim_{n\to\infty} n^{-\frac{1}{p}} \sum_{k=1}^{n} (d_k - d) = 0$ a.s. with $p \in (1,2)$, for products $d_k = \prod_{r=1}^{s} x_k^{(r)}$, $s \in \mathbb{N}$, where each $x_k^{(r)} = \sum_{l=-\infty}^{\infty} c_{k-l}^{(r)} \xi_l^{(r)}$ is a two-sided univariate linear process for $1 \le r \le s$, with coefficients $\{c_l^{(r)}\}_{l\in\mathbb{Z}}$ and i.i.d. zero-mean innovations $\{\xi_l^{(r)}\}_{l\in\mathbb{Z}}$ respectively. The decay of the coefficients $c_l^{(r)}$ as $|l| \to \infty$, can be slow enough that $\{x_k^{(r)}\}$ can have long memory while $\{d_k\}$ can have heavy tails. The aim of this thesis is to handle the long-range dependence and heavy-tailedness for $\{d_k\}$ simultaneously, and to prove a decoupling property that shows the convergence rate is dictated by the worst of long-range dependence and heavy tails, and not their combination.

# Preface

I intend to publish my entire thesis with my supervisor Professor Michael Kouritzin. My contribution was concept formation and mathematical analysis behind the proofs of lemmas and theorems, while the key ideas belonged to my supervisor.

*To my family*

*For everything they have taught me.*

# Acknowledgements

Firstly, I would like to thank my supervisor, Professor Michael Kouritzin, for his priceless guidance which helped me gradually transform from a student to a researcher. He has always listened to my questions and thoughts patiently, and has been very eager to help, both in person and via email. He has been a pillar of support, without whom this thesis would never had come to fruition.

Special thanks to Professors Ivan Mizera, Yaozhong Hu, Alexander Litvak and Bin Han, for teaching me mathematics and supporting me through my Master's degree. They helped me develop interest in an eclectic mix of areas encompassing analysis, probability, statistics and machine learning.

I would also like to express my sincere gratitude to my parents and my sister for always being there for me. I shall remain indebted to them forever.

Finally, I would like to gratefully acknowledge that my research has been supported by the University of Alberta Master's Scholarship, Pundit RD Sharma Memorial Scholarship, Josephine M. Mitchell Scholarship, and a NSERC Discovery Grant.

# Contents

# List of Figures

# Chapter 1

# Introduction

With today's internet of things, big data has become abundant and huge opportunities await those who can effectively mine it. However, this data, especially in finance, econometrics, networks, machine learning, signal processing, and environmental science, often posseses heavy-tails and long memory (see [8], [13], [30], and [35]). Data exhibiting this combination of heavy-tails (HT) and long-range dependence (LRD) can often be modeled well by linear processes but is lethal for most classical statistics. Recently, certain covariance estimators and stochastic approximation algorithms have been shown capable of handling this kind of data. In particular, Marcinkiewicz strong laws of large numbers (MSLLN) were established for showing polynomial rates of convergence (see [17], [20] and [29]).

## 1.1 Basic definitions

We first discuss the concept of MSLLN. Loeve [19] (Case 4 of Theorem A, Section 17) provides the following statement of the Marcinkiewicz-Zygmund strong law of large numbers.

**Theorem 1.1** (Marcinkiewicz-Zygmund Strong Law of Large Numbers)**.** *Let* $\{X_n\}_{n\in\mathbb{Z}}$ *be a sequence of i.i.d. random variables, and let* $0 < p < 2$*. Then,* $E\left[|X_1|^p\right] < \infty$ *if and only if*

$$\lim_{n\to\infty} n^{-\frac{1}{p}} \sum_{k=1}^{n} (X_k - c) = 0 \quad a.s.\ , \quad where \ \ c = \begin{cases} 0, & p < 1 \\ E|X_1|, & p \geq 1 \end{cases} . \tag{1.1}$$

More generally, for a stationary time series $\{X_n\}_{n\in\mathbb{Z}}$ with given conditions on $\{X_n\}$, any result regarding the almost sure convergence of $n^{-\frac{1}{p}}\sum_{k=1}^{n}(X_n - c)$ for some constant $c$ and some $p \in (0, 2)$, is known as a Marcinkiewicz-Zygmund strong law, or simply a Marcinkiewicz strong law of order $p$.

Next we move to heavy-tailedness. Foss et al. [6] (Definition 2.1) called a distribution $F$ heavy tailed if $\int_{-\infty}^{\infty} e^{\lambda x} F(dx) = \infty,\ \forall\ \lambda > 0$. They went on to show in their Theorem 2.6, that $F$ is heavy-tailed if and only if its tail is not bounded by any exponentially decreasing function. We use the following weaker definition throughout our paper, which was also used by Kouritzin and Sadeghi [17], that basically says that the tails decay like $x^{-\beta}$ for some real number $\beta$.

**Definition 1.2** (Heavy Tails)**.** A random variable $X$ is said to be heavy tailed, if

$$\beta \;=\; \sup \left\{ q \in \mathbb{R} : \sup_{x \geq 0} \; x^q P\left(|X| > x\right) < \infty \right\} \;\; < \;\infty \;, \tag{1.2}$$

and $\beta$ will be called the heavy tail coefficient of $X$. Notice that if $\beta > 1$, then $X$ will have finite expectation, and if $\beta > 2$, $X$ will have finite variance. The smaller the value of $\beta$, the heavier the tail of $X$.

Now, we shall discuss long-range dependence. Pipiras and Taqqu [27] (Chapter 2) provides five non-equivalent conditions for LRD, and explored relations among them in detail. We shall use their first condition as the definition of long-range dependence.

**Definition 1.3** (Long-range Dependence)**.** The time series $X = \{X_n\}_{n\in\mathbb{Z}}$ is said to be long-range dependent, if can be represented as $X_n = \sum_{l=-\infty}^{\infty} c_{n-l}\xi_l$, such that $\{\xi_l\}_{l\in\mathbb{Z}}$ are i.i.d. zero-mean random variables with finite variance, and $\{c_l\}_{l\in\mathbb{Z}}$ satisfies

$$c_l = \frac{L(l)}{|l|^{\sigma}} \qquad \text{for some} \;\; \sigma \in \left(\frac{1}{2}, 1\right) , \tag{1.3}$$

for some function $L$ slowly varying at infinity, i.e. $L$ is eventually positive, and $\lim_{x\to\infty} \frac{L(ax)}{L(x)} = 1, \;\; \forall\; a > 0.$

Pipiras and Taqqu [27] explains that Definition 1.3 implies that the autocovariance function of the time series $X$, i.e. $\gamma_X(k) = E[X_0 X_k]$ , will be equal to $k^{1-2\sigma}\overline{L}(k)$, where $\overline{L}$ is another slowly varying function at infinity, and that these autocovariances are not absolutely summable.

## 1.2 History of LRD and HT

A detailed history of long-range dependence can be found in Graves et al. [7]. Indication of long memory in environmental and hydrological time series drew a lot of attention in the mid-twentieth century (see [10],[11],[23] and [24]). While trying to find the ideal height of a dam that can be constructed on the Nile river, H.E. Hurst looked at its river flow data through a new statistic that he defined, i.e. the *rescaled range* statistic, or the $R/S$ statistic (see [10] and [11]). Samorodnitsky [32] defined the $R/S$ statistic as follows.

**Definition 1.4** (R/S Statistic)**.** Let $\{X_k\}_{k\in\mathbb{N}}$ be a sequence of identical random variables, with a non-degenerate distribution, and $S_n$ denote the $n$th partial sum $X_1 + X_2 + \ldots + X_n$. Then,

$$\frac{R}{S}(n) = \frac{R}{S}(X_1, X_2, \ldots, X_n) = \frac{\max\limits_{1\leq k\leq n}\{S_k - \frac{k}{n}S_n\} - \min\limits_{1\leq k\leq n}\{S_k - \frac{k}{n}S_n\}}{\sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \frac{1}{n}S_n)^2}} \ . \tag{1.4}$$

Hurst [10] observed from the river flow data, that $R/S(n)$ was proportional to $n^{0.72}$, instead of $\sqrt{n}$, as would have been the case under the assumption of $\{X_k\}_{k\in\mathbb{N}}$ being independent and Gaussian (see Feller [5]). This unintuitive discrepancy came to be known as the *Hurst phenomenon.*

Several mathematicians and hydrologists attempted to explain the Hurst phenomenon by assuming non-normality of the marginal distribution (see Moran [26]) or non-stationarity (see Potter [28]), but failed to come up with a practical explanation. There were cases when the Hurst phenomenon could be observed in stationary data as well. A general consensus gradually arose that a combination of transience and autocorrelation effects was the probable culprit in those cases (see Wallis and O'Connell [38]).

Motivated by the results of Mandelbrot [22] on erratic behavior of noises in certain solids, Mandelbrot and Van Ness [23] introduced and studied concepts of Fractional Brownian Motion (FBM) and Fractional Gaussian Noise (FGN) and laid the groundwork to study the Hurst phenomenon. Mandelbrot and Wallis [24] used the names *Noah effect* for heavy tails and *Joseph effect* for LRD and showed that models exhibiting self-similarity accounted well for Hurst's findings. Self-similarity is a property seen in many long range dependent time series, and also forms the basis for fractals. Pipiras and Taqqu [27] defined self-similarity as follows.

**Definition 1.5** (Self-similarity)**.** A stochastic process $\{X_t\}_{t\in\mathbb{R}}$ is called self-similar if there exists $H > 0$, such that $\forall\ c > 0$, we have

$$\{X_{ct}\}_{t\in\mathbb{R}} \stackrel{d}{=} c^H\{X_t\}_{t\in\mathbb{R}} \ , \tag{1.5}$$

where $\stackrel{d}{=}$ means that the processes agree on all finite-dimensional distributions. The parameter $H$ is called the Hurst parameter, or self-similarity parameter.

Hosking [9] laid the foundation for the class of $ARFIMA$ (Autoregressive fractionally integrated moving average) models, which are now the most extensively used models for simulating long-range dependence. Due to its widespread usage, we present the definition of $ARFIMA$. The following definition of $ARFIMA$ can be found in Pipiras and Taqqu [27], although they used the term $FARIMA$ to mean $ARFIMA$.

**Definition 1.6** ($ARFIMA$)**.** Let $\{X_n\}_{n\in\mathbb{Z}}$ be a time series, $\{\epsilon_n\}_{n\in\mathbb{Z}}$ are zero-mean noise terms with finite variance, $I$ be the identity operator, and $B$ be the backshift operator, i.e. $BX_n = X_{n-1}$, $\forall\, n \in \mathbb{Z}$ . Let $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \ldots + \theta_p z^q$ be complex valued polynomials of degree $p$ and $q$ respectively, and assume that they don't have common zeros and that $\phi$ has no roots on the unit circle. Then, for $\frac{1}{2} > d \in \mathbb{R}$, $\{X_n\}_{n\in\mathbb{Z}}$ is called an $ARFIMA(p, d, q)$ series if

$$X_n \quad = \quad \phi^{-1}(B)\; \theta(B)\; (I - B)^{-d}\; \epsilon_n \quad = \quad \sum_{l=-\infty}^{\infty} c_l \epsilon_{n-l}, \quad \forall\, n \in \mathbb{Z}, \tag{1.6}$$

where $\{c_l\}_{l\in\mathbb{Z}}$ are coefficients of the Laurent expansion, $\phi^{-1}(z)\; \theta(z)\; (1 - z)^{-d}$ $= \sum_{k\in\mathbb{Z}} c_k z^k$ .

Pipiras and Taqqu [27] (Proposition 2.4.11) proved that when $0 < d < \frac{1}{2}$ and $\phi$ (from Definition 1.6) has all roots outside the unit circle, $ARFIMA(p, d, q)$ is long-range dependent. They also mentioned that the operator $(I - B)^{-d}$

is responsible for the long memory characteristics of the series.

A survey of covariance methods, $R/S$ analysis, and FGNs can be found in Mandelbrot [21]. Today, long-range dependence frequently comes up in fluid flow data (see [13], [30]), network traffic (see [8], [12]), finance and stock markets (see [35]), and is often accompanied by heavy tails.

Autocovariance estimation under LRD and HT conditions is currently a field of great importance, owing to the widespread use of autocovariance functions (see [2], [16], and [39]). Davis and Resnick [2] gave limit theorems for sample covariances of linear processes whose innovations are i.i.d. with regularly varying tail probabilities. Kouritzin [16] studied strong Gaussian approximations for

$$S_{\lfloor t \rfloor} = \sum_{n=1}^{\lfloor t \rfloor} [x_n y_n - E(x_n y_n)]$$

where $\{x_n\}_{n \in \mathbb{N}}$ and $\{y_n\}_{n \in \mathbb{N}}$ are causal linear processes with finite fourth moments, and with independent innovations. Wu and Min [39] considered the asymptotic behavior of sample covariances of linear processes with weakly dependent innovations, and provided a central limit theorem for the same. Wu and Min [40] also studied the asymptotic behavior of sample covariances of long range dependent linear processes, and provided central as well as non-central limit theorems for the same.

## 1.3 Goal of thesis

Consider the following model. Let $\left\{ \mathbf{x}_k = \left( x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(s)} \right) \right\}_{k \in \mathbb{Z}}$ be $R^s$-valued random vectors such that

$$x_k^{(r)} = \sum_{l=-\infty}^{\infty} c_{k-l}^{(r)} \xi_l^{(r)}, \qquad \forall\, 1 \le r \le s, \tag{1.7}$$

are two-sided linear processes, where $\left\{ \left( \xi_l^{(1)}, \xi_l^{(2)}, \ldots, \xi_l^{(s)} \right) \right\}_{l \in \mathbb{Z}}$ (called innovations) are independent and identically distributed $R^s$-valued random vectors, $E\left[ \left| \xi_1^{(r)} \right|^{s \vee 2} \right] < \infty$, and $\left\{ \mathbf{c}_l = \left( c_l^{(1)}, c_l^{(2)}, \ldots, c_l^{(s)} \right) \right\}_{l \in \mathbb{Z}}$ are coefficients satisfying $\sup_{l \in \mathbb{Z}} |l|^{\sigma_r} \left| c_l^{(r)} \right| < \infty$ for some $\sigma_r \in \left( \frac{1}{2}, 1 \right]$. $\sigma_r > \frac{1}{2}$ along with $E\left[ \left| \xi_1^{(r)} \right|^2 \right] < \infty$, ensure the almost sure convergence of the series in (1.7) due to the Khinchin-Kolmogorov Theorem (Theorem 2 of Shiryaev [33], Chapter 4, Section 2), since $E\left( c_{k-l}^{(r)} \xi_l^{(r)} \right) = 0$, and

$$\sum_{l \in \mathbb{Z}} E\left[ \left( c_{k-l}^{(r)} \xi_l^{(r)} \right)^2 \right] = E\left[ \left( \xi_0^{(r)} \right)^2 \right] \sum_{l \in \mathbb{Z}} \left( c_{k-l}^{(r)} \right)^2 < \infty \ ,$$

satisfy the conditions of the theorem. $\sigma_r > \frac{1}{2}$ also guarantees the stationarity of $x_k^{(r)}$. Alternatively, Samorodnitsky [31], Theorem 1.4.1, gives detailed conditions that ensures the existence of (1.7), and mentions that the series converges *unconditionally*, i.e. the series converges to the same sum for any deterministic permutation of its terms.

As per the condition on the coefficients, $\{c_l\}_{l \in \mathbb{Z}}$ may decay slowly enough that $\{x_k\}$ has long memory. In that case, $d_k = \prod_{r=1}^{s} x_k^{(r)}$ is said to possess long-range dependence as well. Based on how the innovations $\{\xi_k^{(r)}\}$ depend on each other, $d_k$ may also possess heavy tails. This gives rise to a few challenges. Since the linear processes we will deal with, will be two sided, we have to care about both the past and the future. Presence of long memory indicates absence of strong mixing, and heavy-tails give rise to infinite variance which makes use of moments impossible without truncation or other techniques.

Very few MSSLN results have been explored for the combination of LRD and HT data. Louhichi and Soulier [20] gave a MSSLN for linear processes where the innovations are linear symmetric $\alpha$-stable processes, and with coefficients $\{c_i\}_{i\in\mathbb{Z}}$ satisfying $\sum_{i=-\infty}^{\infty} |c_i|^s < \infty$ for some $1 \leq s < \alpha$. Rio [29] explored MSSLN results for a strongly mixing sequence $\{X_n\}_{n\in\mathbb{Z}}$ assuming conditions on the mixing rate function and the quantile function of $|X_0|$. Kouritzin and Sadeghi [17] gave a MSLLN for the outer product of two-sided linear processes exhibiting both long memory and heavy tailedness.

In this thesis, we shall generalize Theorem 3 of Kouritzin and Sadeghi [17], to prove a MSSLN for a general product of linear processes in lieu of the outer product, assuming the conditions necessary for its existence. Our goal is to find a bound $\chi$ (a function of the LRD coefficients $\sigma_r$ and HT coefficients $\alpha_i$ in (1.7)), such that

$$\lim_{n\to\infty} n^{-\frac{1}{p}} \sum_{k=1}^{n} (d_k - d) = 0 \quad \text{a.s.} \qquad \forall\, p < \chi \,, \tag{1.8}$$

where $d_k = \prod_{r=1}^{s} x_k^{(r)}$ and $d = E(d_k)$. We shall make further assumptions like $\sup_{l\in\mathbb{Z}} |l|^{\sigma_r}|c_l^{(r)}| < \infty$, $\sigma_r \in \left(\frac{1}{2}, 1\right]$ (which allows for the presence of LRD), and

$$\max_{\pi\in\Pi_s} \max_{1\le i\le \left\lfloor \frac{s-1}{2} \right\rfloor} \sup_{t\ge 0} t^{\alpha_i} P\left( \prod_{r\in\{\pi(1),\ldots,\pi(s-i)\}} \left|\xi_1^{(r)}\right| > t \right) < \infty \, .$$

for some $\alpha_0 > 1$, $\alpha_i = \frac{s}{s-i}\alpha_0$ for $i \in \left\{1, 2, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor \right\}$, and $\Pi_s$ denotes the collection of permutations of $\{1, 2, \ldots, s\}$. This assumption basically deals with the dependence between the tails of the innovations in a general manner, and allows for the presence of heavy tails. It can be seen in Corollary 3.6, that the conditions become much simpler while dealing with $s$ copies of the same linear process. We also provide a corresponding multivariate generalization, Theorem 3.7.

Kouritzin and Sadeghi [17] prove (1.8) for $\chi = 2 \wedge \alpha_0 \wedge \frac{1}{2-\sigma_1-\sigma_2}$ in the outer product case, i.e. $s = 2$. This shows that the rate of convergence is dictated by the worst of the LRD condition $p < \frac{1}{2-\sigma_1-\sigma_2}$ and the heavy-tail condition $p < (\alpha_0 \wedge 2)$, but not the combination. This implies that when $\alpha_0 < 2$ and $\alpha_0 = \frac{1}{2-\sigma_1-\sigma_2}$, a bifurcation takes place due to the structure of $d_k$ while considering outer products. Partitioning $d_k$ into diagonal and off-diagonal terms, we see that the off-diagonal sum $\sum_{l_1\neq l_2} c_{k-l_1}^{(1)} c_{k-l_2}^{(2)} \xi_{l_1}^{(1)} \xi_{l_2}^{(2)}$ does not have heavy tails when $\alpha_0 > 1$, and that LRD is absent in the diagonal sum $\sum_{l\in\mathbb{Z}} c_{k-l}^{(1)} c_{k-l}^{(2)} \xi_l^{(1)} \xi_l^{(2)}$ because $\sigma_1 + \sigma_2 > 1$. In this thesis, we will prove a similar *decoupling* result for general $s$, though the reason does not remain as simple as in the outer product setting.

# Chapter 2

# Applications

## 2.1 Detection and estimation of Long Memory and Heavy Tails

Detecting the presence of long memory and heavy tails in data, and measuring their intensity has been a much investigated problem, with a plethora of tests and methods developed over the last few decades. Some estimation methods are also known to falsely detect the presence of long-range dependence under certain conditions. Montanari et al. [25] compared the performance of various estimation methods to determine the best one to be used in the presence of periodicity. Dette et al. [3] developed tests for short versus long memory in non-stationary time series using $ARFIMA$ (see Definition 1.6). Since non-stationarity is often mistaken for LRD, Torre et al. [37] evaluated the performance of ARFIMA models for detecting and measuring LRD, since ARFIMA has a tendency of falsely detecting LRD even in short-range dependent data. Smith [34] used simulations to estimate the power of five tests for detecting heavy-tails in distributions, where the null hypothesis was that the distribution was normal, and alternative hypotheses was that it belonged to the class of symmetric stable Paretian distributions.

It is evident that most methods to detect or estimate LRD or HT in data are quite complex. In particular, the problem of estimating the intensity of both LRD and HT simultaneously in given data, is almost untouched. We can use Corollary 3.6 to formulate a simple test which could estimate the long-range dependence coefficient $\sigma$ (from (1.3)), and heavy tails coefficient $\beta$ (from (1.2)). Note that $\beta = s\alpha$, where $s$ and $\alpha$ are as in Corollary 3.6, neither of which are known for a given model. We will make the following assumptions.

- The data provided to us can be modeled using two-sided linear processes, and that the underlying distribution of the innovations, has zero mean and finite variance.
- The MSLLN rates in Theorem 3.1 (and by extension, Corollary 3.6) given in Chapter 3, are optimal in a polynomial sense when $s$ is odd or when $s = 2$ (for other values of $s$, note that (3.5) refutes optimality).

For clarification, by optimality in polynomial sense, we mean that it is not possible to achieve polynomially better rates than the ones given by the MSLLN results. The following method will be based completely on Corollary 3.6.

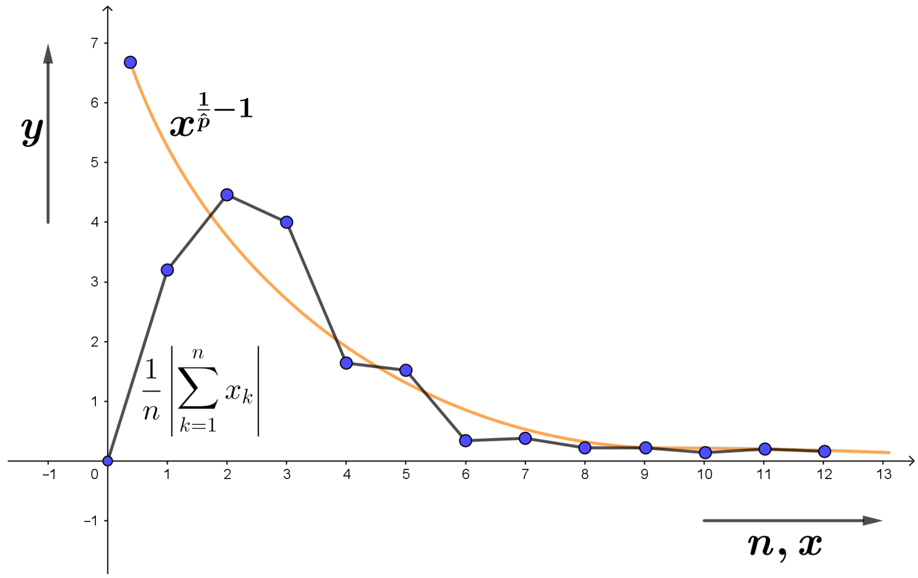


Figure 2.1: Using plots of $\frac{1}{n}\left|\sum_{k=1}^{n} x_k\right|$ and $x^{\frac{1}{p}-1}$ to estimate $\hat{p}$

Suppose the sample size of the data given to us is $m$ (which is supposedly very large), and the data is $x_1, x_2, \ldots, x_m$. Subtracting the sample mean from each of the data points does not affect it's LRD or HT properties, thus without loss of generality, we assume that the sample mean of th dataset is 0. Next, we plot $\frac{1}{n}\left|\sum_{k=1}^{n} x_k\right|$ for $n \in \{1, 2, \ldots, m\}$, and find $\hat{p}$, the highest value of $p$ such that the plot of $\frac{1}{n}\left|\sum_{k=1}^{n} x_k\right|$ asymptotically lies below the curve $y = x^{\frac{1}{p}-1}$. The absolute value sign is being added because $\frac{1}{n}\left|\sum_{k=1}^{n} x_k\right|$ is non-negative, thus helping us to find $\hat{p}$ better (see Figure 2.1 for an example). Alternatively, we can use the *bisection method* to estimate $\sigma$. First take $p = 2$, and look at the convergence of $n^{-\frac{1}{p}} \sum_{k=1}^{n} x_k$ as $n \to \infty$. If the sequence converges to 0, that means the data is not long range dependent. If the sequence does not converge, we take $p = \frac{3}{2}$ and check convergence of the same expression again. If the sequence converges to 0, we check with $p = \frac{7}{4}$, otherwise we check with $p = \frac{5}{4}$, and so on. Repeating this multiple times, we can estimate $\hat{p}$, the optimum value of $p$ for which $\lim_{n\to\infty} n^{-\frac{1}{p}} \sum_{k=1}^{n} x_k = 0$. Due to the optimality assumption on the MSLLN result in Corollary 3.6, we see that,

$$\hat{p} = \frac{2}{3 - 2\hat{\sigma}} \, , \tag{2.1}$$

where $\hat{\sigma}$ is our estimator for $\sigma$. This way, we can estimate the long memory coefficient for a given dataset.

Once we have found $\hat{\sigma}$, we proceed to estimate $\hat{\beta}$. This time we take $p = 1$, and look at the convergence of $n^{-\frac{1}{p}} \sum_{k=1}^{n} \left( x_k^s - \overline{x_k^s} \right)$ as $n \to \infty$, one by one for $s \in \{3, 5, 7, \ldots\}$. Here, $\overline{x_k^s}$ refers to the mean of the $s$th power of all the data points, and is a proxy for $E\left[(x_k)^s\right]$.

- **Case 1:** Suppose the sequence $\frac{1}{n} \sum_{k=1}^{n} \left( x_k^s - \overline{x_k^s} \right)$ does not converge for $s = 2$. This would refute our assumption of finite variance, and would imply that $\beta < 2$. This means that the amount of heavy tails in the data would be too large to detect by our method.
- **Case 2:** It might happen that even after checking for a large number of values of $s \in \{2\} \cup \{3, 5, 7, \ldots\}$, the sequence $\frac{1}{n} \sum_{k=1}^{n} \left( x_k^s - \overline{x_k^s} \right)$ as $n \to \infty$ always converges to 0. In that case, we must conclude that the data has very little or no heavy tails. We can still find a bound for the amount of heavy tails, because if the sequence does not converge for $s = s_0$, then $\beta > s_0$. This means

that the tails of the innovations are lighter than $x^{-s_0}$.

- **Case 3:** Suppose we find that $s_0 \in \{2\} \cup \{3, 5, 7, \ldots\}$ is the highest value of $s$ for which the sequence $\frac{1}{n}\sum_{k=1}^{n}\left(x_k^s - \overline{x_k^s}\right)$ convergences to 0. Then, we plot $\frac{1}{n}\left|\sum_{k=1}^{n}\left(x_k^{s_0} - \overline{x_k^{s_0}}\right)\right|$ for $n \in \{1, 2, \ldots, m\}$, and find $\hat{p}$, the highest value of $p$ such that the curve $y = x^{\frac{1}{p}-1}$ asymptotically lies below the plot of $\frac{1}{n}\left|\sum_{k=1}^{n}\left(x_k^{s_0} - \overline{x_k^{s_0}}\right)\right|$. Due to our optimality assumption, $\hat{p}$ is the estimate of the optimum value of $p$ for which $\lim_{n\to\infty} n^{-\frac{1}{p}}\left|\sum_{k=1}^{n}\left(x_k^{s_0} - \overline{x_k^{s_0}}\right)\right| = 0$. From Corollary (3.6), we get that

$$\hat{p} = \begin{cases} 2 \wedge \hat{\alpha} \wedge \frac{1}{2-2\hat{\sigma}}, & s_0 = 2 \\ \hat{\alpha} \wedge \frac{2}{3-2\hat{\sigma}}, & s_0 \in \{3, 5, 7, \ldots\} \end{cases}. \tag{2.2}$$

where $\hat{\sigma}$ is the LRD coefficient as estimated by (2.1). Now, when $s_0 = 2$, we can only estimate $\hat{\alpha}$ if $\alpha < 2 \wedge \frac{1}{2-2\hat{\sigma}}$, since in that case, $\hat{\alpha} = \hat{p}$. Similarly, when $s_0 > 2$, then we can only estimate $\hat{\alpha}$ if $\alpha < \frac{2}{3-2\hat{\sigma}}$, and again get that $\hat{\alpha} = \hat{p}$. Our estimate for $\hat{\beta}$ then becomes $\hat{\beta} = \hat{p}s_0$.

Thus, we see that in many cases, we can estimate both $\sigma$ and $\beta$. While $\sigma$ can be estimated in all cases, $\beta$ cannot be estimated if it is either less than 2 or very large (though we can still bound it as tightly as we want). Also, if there is a lot of long memory in the data (i.e. $\sigma$ is very close to $\frac{1}{2}$), that would make $\frac{1}{2-2\hat{\sigma}}$ and $\frac{2}{3-2\hat{\sigma}}$ very close to 1. Then, $\alpha$ itself would have to be very close to 1, to enable us to estimate $\beta$. Thus, this application of Corollary 3.6 to detect and estimate $\sigma$ and $\beta$, raises some possible problems for future research as well. One problem is proving optimality of the the MSSLN rates in Theorem 3.1, for odd $s$, and $s = 2$. Another direction for further research, could be to estimate $\sigma$ and $\beta$ simultaneously when $\beta < 2$, i.e. the underlying distribution of the innovations has infinite variance.

## 2.2 Stochastic Approximation

Stochastic approximation (SA) algorithms are widely used in adaptive filtering, optimization, signals and systems, machine learning and pattern recognition (see [1]). SA algorithms are used to iteratively produce estimates of a parameter vector in a model. The estimates are updated recursively at every step such that they converge to the true value of the parameter. Thus, their almost sure rates of convergence and invariance principles have been topics of a lot of research (see [4], [14], [15] and [18]).

One of the most used linear stochastic approximation algorithms applies to the following model:

$$y_{k+1} = z_k^T h + \epsilon_k \ , \qquad \forall\ k \in \mathbb{N} \ , \tag{2.3}$$

where $\{y_k\}_{k\in\mathbb{N}\setminus\{1\}}$ and $\{z_k\}_{k\in\mathbb{N}}$ are $\mathbb{R}$-valued and $\mathbb{R}^d$-valued stochastic processes respectively, $\{\epsilon_k\}_{k\in\mathbb{N}}$ are noise terms, and $h \in \mathbb{R}^d$ is an unknown parameter vector. We often want to find the value of $h$ that minimizes $E\left[(y_{k+1} - z^T h)^2\right]$. Under the assumptions of second-order stationarity, and the existence and positive definiteness of $E(z_k z_k^T)$, Kouritzin and Sadeghi [18] mentioned that the required value of $h$ is $E(z_k z_k^T)^{-1} E(y_{k+1} z_k)$. But since $E(z_k z_k^T)$ and $E(y_{k+1} z_k)$ are often unknown, we often use the SA algorithm

$$h_{k+1} = h_k + \mu_k (b_k - A_k h_k) \ , \tag{2.4}$$

where $A_k = z_k z_k^T, \ b_k = y_{k+1} z_k$ and $\mu_k$ is the step-size of the $k$th step. Kouritzin [14, 15] and Kouritzin and Sadeghi [18] studied the convergence of (2.4) under various conditions, taking $\mu_k = \frac{1}{k^\chi}$ for some $\chi \in (0, 1]$.

Kouritzin and Sadeghi [17] (Theorem 5) proved an important result dealing with the almost sure rate of convergence of (2.4) under LRD and HT conditions, by combining Theorem 4 of [17] with Corollary 3 of [18]. However, due to Remark 3.2, we should instead use Theorem 3.1 of our paper (with $s = 2$), and proceed the same way as in the proof of Theorem 5 of [17].

# Chapter 3

# Results

The following theorem is concerned with the rate of convergence of products of long range dependent and heavy tailed univariate linear processes. Motivations and general comments about the assumptions in this theorem are provided in Remarks 3.2-3.5.

**Theorem 3.1.** *Let* $s \in \mathbb{N}$, $\alpha_0 > 1$, $\alpha_i = \frac{s}{s-i}\alpha_0$ *for* $i \in \left\{1, 2, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}$, *and* $\Pi_s$ *denote the collection of permutations of* $\{1, 2, \ldots, s\}$.
*Let* $\left\{\left(\xi_l^{(1)}, \xi_l^{(2)}, \ldots, \xi_l^{(s)}\right)\right\}_{l \in \mathbb{Z}}$ *be i.i.d.* $\mathbb{R}^s$*-valued zero-mean random vectors such that the following hold,*

$$E\left[\left|\xi_1^{(r)}\right|^{s \vee 2}\right] < \infty \qquad \forall\ 1 \le r \le s, \tag{3.1}$$

$$\max_{\pi\in\Pi_s}\ \max_{1\le i\le\left\lfloor\frac{s-1}{2}\right\rfloor}\ \sup_{t\ge0}\ t^{\alpha_i}P\left(\prod_{r\in\{\pi(1),\ldots,\pi(s-i)\}}\left|\xi_1^{(r)}\right|>t\right)<\infty\ . \tag{3.2}$$

*Moreover, let constants* $\left\{\left(c_l^{(1)},c_l^{(2)},\ldots,c_l^{(s)}\right)\right\}_{l\in\mathbb{Z}}$ *satisfy*

$$\sup_{l\in\mathbb{Z}}\ |l|^{\sigma_r}\left|c_l^{(r)}\right|<\infty\qquad \textit{for some } \sigma_r\in\left(\frac{1}{2},1\right],\quad \forall\ 1\le r\le s. \tag{3.3}$$

*For* $1\le r\le s$, $k\in\mathbb{N}$, *define* $x_k^{(r)}=\sum_{l=-\infty}^{\infty}c_{k-l}^{(r)}\xi_l^{(r)}$, $d_k=\prod_{r=1}^{s}x_k^{(r)}$, *and* $d=E(d_k)$. *Then,* $\lim_{n\to\infty}n^{-\frac{1}{p}}\sum_{k=1}^{n}(d_k-d)=0$ *a.s. for*

$$p\quad<\quad\begin{cases}\frac{2}{3-2\sigma_1}, & s=1\\ 2\wedge\alpha_0\wedge\frac{1}{2-\sigma_1-\sigma_2}, & s=2\\ \alpha_0\wedge\frac{2}{3-2\min_{1\le i\le s}\{\sigma_i\}}, & s>2\end{cases}\quad. \tag{3.4}$$

*Furthermore, if* $\xi_1^{(1)}=\xi_1^{(2)}=\ldots=\xi_1^{(s)}$ *and* $\xi_1^{(1)}$ *is a symmetric random variable, and* $s$ *is even, then the constraint for (3.4) can be relaxed to*

$$p\quad<\quad 2\wedge\alpha_0\wedge\frac{1}{2-\min_{1\le i<j\le s}\{\sigma_i+\sigma_j\}}\ . \tag{3.5}$$

**Remark 3.2.** Taking $s=2$ in Theorem 3.1 gives us Kouritzin and Sadeghi [17] (Theorem 3) as a corollary. There is a minor miscalculation in the second-last line (Line 17) of Page 362 of Kouritzin and Sadeghi [17]. The term $\sum_{l=j+1}^{k+T}c_{j-l}c_{k-l}$ in Line 16 was erroneously taken to be smaller than $(j-k)^{-2\sigma}T^{2-2\sigma}$, which should actually have been $(j-k)^{1-2\sigma}$ instead. This miscalculation can be corrected by applying Lemma 4.1 (with $\gamma = \sigma$) in Section 4.2 of our paper, to Line 15 of [17], to obtain their results. Also, Kouritzin and Sadeghi [17] (Remark 2), mention that the constraints for handling LRD and those for HT *decouple*, which they explain through the structure of the terms $d_k$. This decoupling phenomenon is observed in our proof as well.

**Remark 3.3.** $\sigma_r \in \left(\frac{1}{2}, 1\right]$ allows for the presence of long memory in $x_k^{(r)}$ (see Definition 1.3). (3.1) implies that $E\left(\prod_{r=1}^{s}\left|\xi_1^{(r)}\right|\right) < \infty$ (ensuring the existence of $d = E(d_k)$), and that the second moment of $\xi_1^{(r)}$ is finite, ensuring the convergence of the series $x_k^{(r)} = \sum_{l\in\mathbb{Z}} c_{k-l}^{(r)}\xi_l^{(r)}$. The tail bound in (3.2) allows for the second moment of the product of more than $\frac{s}{2}$ of the $\xi_1^{(r)}$'s to be infinite, giving rise to heavy tails in corresponding sums. The condition $\alpha_i = \frac{s}{s-i}\alpha_0$ is motivated by the case when $\xi_1^{(1)} = \ldots = \xi_1^{(s)} = \xi_1$, where the tail condition $\sup_{t\geq 0}\ t^{\alpha_0}P\left(|\xi_1|^s > t\right) < \infty$ implies that $\sup_{t\geq 0}\ t^{\frac{s}{s-i}\alpha_0}P\left(|\xi_1|^{s-i} > t\right) < \infty$ .

**Remark 3.4.** Since $\sigma_r \in (\frac{1}{2}, 1]$, we can find an $\epsilon > 0$ such that $\sigma_r - \epsilon \in \left(\frac{1}{2}, 1\right)$. Similarly, since $\alpha_i \in (1,\infty)$, we can find $\bar{\epsilon} > 0$ such that $\alpha_i - \bar{\epsilon} \in (1,2)\cup(2,\infty)$. It can be checked that (3.2,3.3) also hold for $\alpha_i - \bar{\epsilon}$ and $\sigma_r - \epsilon$ instead of $\alpha_i$ and $\sigma_r$ respectively. Thus, proving that $\lim_{n\to\infty} n^{-\frac{1}{p}}\sum_{k=1}^{n}(d_k - d) = 0$ a.s, for $p < 2\wedge(\alpha_0 - \bar{\epsilon})\wedge\frac{2}{3-2\min_{1\leq i\leq s}\{\sigma_i\}+2\epsilon}$ will imply (3.4) as well, since $\epsilon, \bar{\epsilon} > 0$ were arbitrary but fixed. Similar result will hold for (3.5). Therefore, it suffices to assume that $\sigma_r \in (\frac{1}{2}, 1)$, and $\alpha_i \in (1,2)\cup(2,\infty)$.

**Remark 3.5.** Note that (3.3) implies that $\left|c_l^{(r)}\right| \ll \begin{cases} 1 & l = 0 \\ |l|^{-\sigma_r} & l \neq 0 \end{cases}$. The proof of the general case only differs cosmetically from the notationally simpler case where $\xi_l^{(1)} = \xi_l^{(2)} = \ldots = \xi_l^{(s)} = \xi_l$, and $\sigma_1 = \sigma_2 = \ldots = \sigma_s = \sigma$, which means that we can further assume that $c_l^{(1)} = c_l^{(2)} = \ldots = c_l^{(s)} = c_l$, where $|c_l| \ll \begin{cases} 1 & l = 0 \\ |l|^{-\sigma} & l \neq 0 \end{cases}$. We only provide the proof of this later case.

When $x_1^{(1)} = x_1^{(2)} = \ldots = x_1^{(s)}$, several of the conditions in Theorem 3.1 merge, and we get a simple yet important corollary with very useful applications.

**Corollary 3.6.** *Let $s \in \mathbb{N}$, and $\{\xi_l\}_{l\in\mathbb{Z}}$ be i.i.d. zero-mean random variables with finite variance, such that $\sup_{t\geq 0} t^{\alpha} P\left(|\xi_1|^s > t\right) < \infty$ for some $\alpha > 1$, and let $\{c_l\}_{l\in\mathbb{Z}}$ satisfy $\sup_{l\in\mathbb{Z}} |l|^{\sigma} |c_l| < \infty$ for some $\sigma \in \left(\frac{1}{2}, 1\right]$. For $k \in \mathbb{N}$, define $x_k = \sum\limits_{l=-\infty}^{\infty} c_{k-l}\xi_l$, $d_k = (x_k)^s$, and $d = E(d_k)$. Then, $\lim\limits_{n\to\infty} n^{-\frac{1}{p}} \sum\limits_{k=1}^{n} (d_k - d) = 0$ a.s, for*

$$p < \begin{cases} \frac{2}{3-2\sigma}, & s = 1 \\ 2 \wedge \alpha \wedge \frac{1}{2-2\sigma}, & s = 2 \\ \alpha \wedge \frac{2}{3-2\sigma}, & s > 2 \end{cases} . \tag{3.6}$$

*Furthermore, if $\xi_1$ is a symmetric random variable, and $s$ is even, then the constraint for (3.6) can be relaxed to $p < 2 \wedge \alpha \wedge \frac{1}{2-2\sigma}$ .*

We can now extract a multivariate version of Theorem 3.1, analogous to Theorem 4 of Kouritzin and Sadeghi [17]. Refer to Notation List (Section 4.1) for some of the notation used in this theorem.

**Theorem 3.7.** *Let $s \in \mathbb{N}$, $\alpha_0 > 1$, $\alpha_i = \frac{s}{s-i}\alpha_0$ for $1 \leq i \leq \left\lfloor \frac{s-1}{2} \right\rfloor$, and $\Pi_s$*

*denote the collection of permutations of* $\{1,2,\ldots,s\}$.

*Let* $\left\{\left(\Xi_l^{(1)},\Xi_l^{(2)},\ldots,\Xi_l^{(s)}\right)\right\}_{l\in\mathbb{Z}}$ *be i.i.d. zero-mean random vectors in* $\mathbb{R}^{m\times s}$, *such that* $E\left[\left\|\Xi_1^{(r)}\right\|_F^{s\vee 2}\right]<\infty,\quad \forall\ 1\le r\le s$, *and*

$$\max_{\pi\in\Pi_s}\ \max_{1\le i\le\left\lfloor\frac{s-1}{2}\right\rfloor}\ \sup_{t\ge 0}\ t^{\alpha_i}P\left(\prod_{r\in\{\pi(1),\ldots,\pi(s-i)\}}\left\|\Xi_1^{(r)}\right\|_F>t\right)<\infty\ .$$

*Moreover, let* $\mathbb{R}^{d\times m}$*-valued matrices* $\left\{\left(C_l^{(1)},C_l^{(2)},\ldots,C_l^{(s)}\right)\right\}_{l\in\mathbb{Z}}$ *satisfy* $\sup_{l\in\mathbb{Z}}\ |l|^{\sigma_r}\left\|C_l^{(r)}\right\|_F<\infty$ *, for some* $\sigma_r\in\left(\frac{1}{2},1\right]$. *For* $1\le r\le s,\ k\in\mathbb{Z}$, *define* $X_k^{(r)}=\sum_{l=-\infty}^{\infty}C_{k-l}^{(r)}\Xi_l^{(r)}$, $\quad D_k=\bigotimes_{r=1}^{s}X_k^{(r)}$ *(the tensor product of* $X_k^{(1)},\ldots,X_k^{(s)}$*), and* $D=E(D_k)$. *Then,* $\lim_{n\to\infty}n^{-\frac{1}{p}}\sum_{k=1}^{n}(D_k-D)=0$ *a.s, for the values of p as in (3.4).*

This theorem follows from linearity of limits and Theorem 3.1.

# Chapter 4

# Proofs

## 4.1 Notation and Conventions

- $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, where $\mathbb{N}$ denotes the set of natural numbers.
- $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$.
- $|x|$ is Euclidean distance of $x \in \mathbb{R}^d$, with $d \in \mathbb{N}$.
- $\mathbf{1}_A$ is the indicator function of the event A, i.e. 1 if A occurs, or else 0.
- $|S|$ is the cardinality of the set $S$.
- $\bigotimes_{r=1}^{n} v^{(r)}$ denotes the tensor product of vectors $v^{(r)} \in \mathbb{R}^d$, $1 \leq r \leq n$, $d \in \mathbb{N}$.
- $\|A\|_F$ is the Frobenius norm of $A$, i.e. $\sqrt{trace(A^T A)}$ for any matrix $A \in \mathbb{R}^{m \times n}$, where $m, n \in \mathbb{N}$.
- $\|X\|_p = [E(X^p)]^{\frac{1}{p}}$ for any non-negative random variable $X$, and $p > 0$.
- $\lfloor c \rfloor = \max\{n \in \mathbb{N}_0 : n \leq c\}$ and $\lceil c \rceil = \min\{n \in \mathbb{N}_0 : n \geq c\}$ $\forall$ $c \geq 0$ .
- $a_{i,k} \overset{i}{\ll} b_{i,k}$ means that for each $k$, $\exists$ $c_k > 0$ that does not depend upon $i$

such that $|a_{i,k}| \le c_k |b_{i,k}|$ for all $i, k$ (also used in [15] and [17]).

- $$l_{n,\beta}(x) = \begin{cases} x^{n(1-2\beta)+1}, & \beta < \frac{n+1}{2n} \\ \log(x), & \beta = \frac{n+1}{2n} \\ 1, & \beta > \frac{n+1}{2n} \end{cases}, \ \forall \ n \in \mathbb{N} \text{ and } \beta \in \mathbb{R}.$$
- We shall follow the convention, that when $\{f_r\}_{r\in\mathbb{Z}}$ is a sequence of functions or constants, and $a, b \in \mathbb{N}_0$ such that $a > b$, then $\prod_{r=a}^{b} f_r = 1$.

## 4.2 Important Lemmas

We first present two lemmas on which Theorems 3.1 and 3.7 will rely.

**Lemma 4.1.** *For $j, k \in \mathbb{Z}$, $j \neq k$ and $\gamma > \frac{1}{2}$, we have,*

$$\sum_{\substack{l=-\infty \\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\gamma}|k-l|^{-\gamma} \overset{j,k}{\ll} \begin{cases} |j-k|^{1-2\gamma}, & \gamma \in \left(\frac{1}{2}, 1\right) \\ |j-k|^{-1}\ln(|j-k|), & \gamma = 1 \\ |j-k|^{-\gamma}, & \gamma > 1 \end{cases}.$$

*Proof.* Without loss of generality, we assume that $j > k$. When $\gamma \in \left(\frac{1}{2}, 1\right)$, using symmetry, integral approximation, and successive substitutions $t = k - l$ and $s = \frac{t}{j-k}$, we get

$$\sum_{\substack{l=-\infty \\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\gamma}|k-l|^{-\gamma}$$

$$\overset{j,k}{\ll} \sum_{l=-\infty}^{k-1} (j-l)^{-\gamma}(k-l)^{-\gamma} + \sum_{l=k+1}^{j-1} (j-l)^{-\gamma}(k-l)^{-\gamma}$$

$$
\begin{aligned}
&\overset{j,k}{\ll} \int_0^{\infty} (j-k+t)^{-\gamma} t^{-\gamma}\, dt \quad + \int_0^{j-k} (j-k-t)^{-\gamma} t^{-\gamma}\, dt \\
&\overset{j,k}{\ll} (j-k)^{1-2\gamma} \left( \int_0^{\infty} (1+s)^{-\gamma} s^{-\gamma}\, ds \quad + \int_0^1 (1-s)^{-\gamma} s^{-\gamma}\, ds \right). \qquad (4.1)
\end{aligned}
$$

Since $\gamma \in \left(\frac{1}{2}, 1\right)$, notice that $\int_0^1 (1+s)^{-\gamma} s^{-\gamma}\, ds \leq \int_0^1 (1-s)^{-\gamma} s^{-\gamma}\, ds = B(1-\gamma, 1-\gamma)$, which is the beta function evaluated at $(1-\gamma, 1-\gamma)$. Therefore, we get from (4.1), that

$$
\begin{aligned}
&\sum_{\substack{l=-\infty \\ l \notin \{j,k\}}}^{\infty} |j-l|^{-\gamma} |k-l|^{-\gamma} \\
&\overset{j,k}{\ll} (j-k)^{1-2\gamma} \left( \int_1^{\infty} (1+s)^{-\gamma} s^{-\gamma}\, ds \; + \; 2 \int_0^1 (1-s)^{-\gamma} s^{-\gamma}\, ds \right) \\
&\overset{j,k}{\ll} (j-k)^{1-2\gamma}. \qquad (4.2)
\end{aligned}
$$

Next, we consider the case where $\gamma = 1$.

$$
\begin{aligned}
&\sum_{\substack{l=-\infty \\ l \notin \{j,k\}}}^{\infty} |j-l|^{-1} |k-l|^{-1} \\
&\overset{j,k}{\ll} \sum_{l=-\infty}^{k-1} (j-l)^{-1} (k-l)^{-1} \quad + \sum_{l=k+1}^{j-1} (j-l)^{-1} (l-k)^{-1} \\
&= (j-k)^{-1} \left( \sum_{l=-\infty}^{k-1} \left[ (k-l)^{-1} - (j-l)^{-1} \right] \; + \sum_{l=k+1}^{j-1} \left[ (l-k)^{-1} + (j-l)^{-1} \right] \right) \\
&= (j-k)^{-1} \left( \sum_{l=1}^{j-k} l^{-1} \quad + \quad 2 \sum_{l=1}^{j-k-1} l^{-1} \right) \\
&\overset{j,k}{\ll} (j-k)^{-1} \log(j-k). \qquad (4.3)
\end{aligned}
$$

Finally we consider the case where $\gamma > 1$. Using symmetry, and summability of the sequence $\{|l|^{-\gamma}\}_{l\in\mathbb{Z}}$, we have

$$\begin{aligned}
\sum_{\substack{l=-\infty\\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\gamma}|k-l|^{-\gamma} \quad &\overset{j,k}{\ll} \sum_{\substack{l=-\infty\\ l\neq k}}^{\left\lfloor\frac{j+k}{2}\right\rfloor} (j-l)^{-\gamma}\,|k-l|^{-\gamma} \\
&\overset{j,k}{\ll} \left(j-\left\lfloor\frac{j+k}{2}\right\rfloor\right)^{-\gamma} \sum_{\substack{l=-\infty\\ l\neq k}}^{\left\lfloor\frac{j+k}{2}\right\rfloor} |k-l|^{-\gamma} \\
&\overset{j,k}{\ll} (j-k)^{-\gamma} \qquad (4.4)
\end{aligned}$$

From (4.2, 4.3) and (4.4), the proof of the lemma is complete. □

**Lemma 4.2.** *For $j,k\in\mathbb{Z}$, $j\neq k$ and $\gamma\in\left(\frac{1}{2},1\right)$, we have,*

$$\sum_{\substack{l=-\infty\\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\gamma}|k-l|^{-2\gamma} \qquad \overset{j,k}{\ll} \qquad |j-k|^{-\gamma}\ .$$

*Proof.* Without loss of generality, we assume that $j>k$. Then, we have

$$\begin{aligned}
&\sum_{\substack{l=-\infty\\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\gamma}|k-l|^{-2\gamma} \\
\overset{j,k}{\ll}\ & \sum_{\substack{l=-\infty\\ l\neq k}}^{\left\lfloor\frac{j+k}{2}\right\rfloor} |j-l|^{-\gamma}|k-l|^{-2\gamma} \ + \sum_{\substack{l=\left\lceil\frac{j+k}{2}\right\rceil\\ l\neq j}}^{\infty} |j-l|^{-\gamma}|k-l|^{-2\gamma} \\
\overset{j,k}{\ll}\ & |j-k|^{-\gamma} \sum_{\substack{l=-\infty\\ l\notin\{j,k\}}}^{\infty} |k-l|^{-2\gamma} \ +\ |j-k|^{-\gamma} \sum_{\substack{l=-\infty\\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\gamma}|k-l|^{-\gamma} \\
\overset{j,k}{\ll}\ & |j-k|^{-\gamma}, \qquad (4.5)
\end{aligned}$$

by Lemma 4.1. This concludes the proof of the lemma. □

The following lemma will help reduce the calculations dealing with heavy tails.

**Lemma 4.3.** *Let* $z > 1$, $n_r = 2^r \ \forall\, r \in \mathbb{N}$, *and* $\{X_n\}_{n\in\mathbb{N}}$ *be random variables such that* $E\big[|X_n|^z\big] < \infty$. *Then we have,*

$$E^{\frac{1}{z}}\left[\sup_{n_r \le n < n_{r+1}} |X_n - E(X_n)|^z\right] \overset{r}{\ll} E^{\frac{1}{z}}\left[\left|\sup_{n_r \le n < n_{r+1}} |X_n|\right|^z\right].$$

*Proof.* By Triangle Inequality, Minkowski's Inequality and Jensen's Inequality, we have,

$$\begin{aligned}
& E^{\frac{1}{z}}\left[\sup_{n_r \le n < n_{r+1}} |X_n - E(X_n)|^z\right] \\
\le\ & E^{\frac{1}{z}}\left[\left|\sup_{n_r \le n < n_{r+1}} |X_n|\right|^z\right] + E^{\frac{1}{z}}\left[\left|\sup_{n_r \le n < n_{r+1}} |E(X_n)|\right|^z\right] \\
\le\ & E^{\frac{1}{z}}\left[\left|\sup_{n_r \le n < n_{r+1}} |X_n|\right|^z\right] + E\left[\left|\sup_{n_r \le n < n_{r+1}} |X_n|\right|\right] \\
\le\ & 2\, E^{\frac{1}{z}}\left[\left|\sup_{n_r \le n < n_{r+1}} |X_n|\right|^z\right],
\end{aligned} \tag{4.6}$$

which concludes the proof of the lemma. □

Samorodnitsky [31] (Theorem 1.4.1), proved a general result providing sufficient conditions for the convergence of $x_k^{(r)}$ (in Theorem 3.1) and the existence of $E\left[\left(x_k^{(r)}\right)^s\right]$ when $s \in \mathbb{R}^+$, using Marcinkiewicz-Zygmund inequalities and induction. We will prove the special case of that theorem when $s \in \mathbb{N}$, using less complicated machinery.

**Lemma 4.4.** *Let $s \in \mathbb{N}$ and $\{\xi_l\}_{l\in\mathbb{Z}}$ be i.i.d. zero-mean random variables such that $E\left[|\xi_1|^{s\vee 2}\right] < \infty$, and $\{c_l\}_{l\in\mathbb{Z}}$ satisfy $\sup_{l\in\mathbb{Z}} |l|^\sigma |c_l| < \infty$, for some $\sigma \in \left(\frac{1}{2}, 1\right)$. Then,*

$$\sup_{k\in\mathbb{Z}} E\left(\left|\sum_{l=-\infty}^{\infty} c_l \xi_{k-l}\right|^s\right) < \infty \; . \tag{4.7}$$

*Proof.* By the i.i.d. nature of $\{\xi_k\}_{k\in\mathbb{Z}}$, $E\left(\left|\sum_{l=-\infty}^{\infty} c_l\xi_{k-l}\right|^s\right) = E\left(\left|\sum_{l=-\infty}^{\infty} c_l\xi_l\right|^s\right)$ for all $k$. Hence it suffices to prove the finiteness of $E\left(\left|\sum_{l=-\infty}^{\infty} c_l\xi_l\right|^s\right)$.
$\left(\sum_{l=-\infty}^{\infty} c_l\xi_{k-l}\right)^s$ can be broken up into sums based on the combinations of subscripts of $\xi$'s that are equal. That is, as sums of

$$s(q,\lambda_q) \;=\; \sum_{l_1,l_2,\ldots,l_q\in\mathbb{Z},\; distinct} \left(\prod_{r=1}^{q} c_{l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \xi_{k-l_r}^{a_r}\right), \tag{4.8}$$

where $q$ ranges over $\{1,2,\ldots,s\}$, and $\lambda_q = (a_1, a_2, \ldots, a_q)$ ranges over partitions of $s$ of length $q$ (we call an $r$-tuple of natural numbers $(b_1, b_2, \ldots, b_r)$ a partition of a natural number $n$ of length $r$, if $b_1 + \ldots + b_r = n$ and $b_1 \geq b_2 \geq \ldots \geq b_r \geq 1$). By Minkowski's Inequality,

$$\left|\sum_{l=-\infty}^{\infty} c_l\xi_{k-l}\right|^s \;\leq\; \sum_{q=1}^{s}\sum_{\lambda_q} |s(q,\lambda_q)| \; ,$$

so (4.7) follows from $E|s(q,\lambda_q)| < \infty$, which we prove below.
Let $v = \#\{1 \leq p \leq q : \; a_p = 1\}$. Hence, $\sum_m c_m^{a_r} < \infty$ for $1 \leq r \leq q-v$. If

$v < q$, by (4.8) and Jensen's Inequality, we have

$$\begin{aligned}
& E\left|s(q,\lambda_q)\right| \\
\leq\ & E\left|\sum_{l_1=-\infty}^{\infty} \left|c_{l_1}^{a_1}\right| \left|\sum_{l_2\in\mathbb{Z}\setminus\{l_1\}} \cdots \sum_{l_q\in\mathbb{Z}\setminus\{l_1,\ldots,l_{q-1}\}} \left(\prod_{r=2}^{q} c_{l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \xi_{l_r}^{a_r}\right)\right|\right| \\
=\ & \sum_{m=-\infty}^{\infty} \left|c_m^{a_1}\right| E\left|\sum_{l_1=-\infty}^{\infty} \frac{\left|c_{l_1}^{a_1}\right|}{\sum_m \left|c_m^{a_1}\right|} \left|\sum_{l_2\in\mathbb{Z}\setminus\{l_1\}} \cdots \sum_{l_q\in\mathbb{Z}\setminus\{l_1,\ldots,l_{q-1}\}} \left(\prod_{r=2}^{q} c_{l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \xi_{l_r}^{a_r}\right)\right|\right| \\
\leq\ & \sum_{l_1=-\infty}^{\infty} \left|c_{l_1}^{a_1}\right| E\left|\sum_{l_2\in\mathbb{Z}\setminus\{l_1\}} \cdots \sum_{l_q\in\mathbb{Z}\setminus\{l_1,\ldots,l_{q-1}\}} \left(\prod_{r=2}^{q} c_{l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \xi_{l_r}^{a_r}\right)\right| . \qquad (4.9)
\end{aligned}$$

Similarly, we can recursively bring summations over $l_2, \ldots, l_{q-v}$, out of the expectation in (4.9), and use the independence of $\xi$'s to get that

$$E|s(q,\lambda_q)| \ \leq \sum_{l_1,\ldots,l_{q-v}\in\mathbb{Z},\ distinct} \left(\prod_{r=1}^{q-v} \left|c_{l_r}^{a_r}\right|\right)\left(\prod_{r=1}^{q-v} E\left|\xi_{l_r}^{a_r}\right|\right) E(\phi_v) \ , \qquad (4.10)$$

where $\phi_v = \left|\sum_{l_{q-v+1}\in\mathbb{Z}\setminus\{l_1,\ldots,l_{q-v}\}} \cdots \sum_{l_q\in\mathbb{Z}\setminus\{l_1,\ldots,l_{q-1}\}} \left(\prod_{r=q-v+1}^{q} c_{l_r}\xi_{k-l_r}\right)\right|$, when $v \neq 0$, and 1 when $v = 0$. Note that (4.10) holds when $v = q$ as well, since none of the $l_r$'s need to be brought out of the expectation. Now, since $\{\xi_k\}_{k\in\mathbb{Z}}$

are i.i.d. and zero-mean,

$$
\begin{aligned}
E\left[|\phi_v|^2\right] &= E\left(\sum_{\substack{l_{q-v+1}\neq l_{q-v+2}\neq \ \dots \ \neq l_q \\ l_{q-v+1},\dots,l_q\notin\{l_1,\dots,l_{q-1}\}}} \prod_{r=q-v+1}^{q} \left(c_{l_r}^2 \xi_{l_r}^2\right)\right) \\
&= \left[E\left(\xi_1^2\right)\right]^v \sum_{\substack{l_{q-v+1}\neq l_{q-v+2}\neq \ \dots \ \neq l_q \\ l_{q-v+1},\dots,l_q\notin\{l_1,\dots,l_{q-1}\}}} \left(\prod_{r=q-v+1}^{q} c_{l_r}^2\right) \\
&\leq \left[E\left(\xi_1^2\right)\right]^v \sum_{l_{q-v+1}=-\infty}^{\infty} \sum_{l_{q-v+2}=-\infty}^{\infty} \dots \sum_{l_q=-\infty}^{\infty} \left(\prod_{r=q-v+1}^{q} c_{l_r}^2\right) < \infty. \qquad (4.11)
\end{aligned}
$$

Noting that $a_r \geq 2$ for $1 \leq r \leq q - v$, $E\left[|\xi_1|^s\right] < \infty$ , and using (4.11) on (4.10), we get that

$$
\begin{aligned}
E|s(q,\lambda_q)| &\ll \left(\prod_{r=1}^{q-v} E\left|\xi_1^{a_r}\right|\right) \sum_{l_1,\dots,l_{q-v}\in\mathbb{Z},\ distinct} \left(\prod_{r=1}^{q-v} \left|c_{l_r}^{a_r}\right|\right) \\
&\ll \sum_{l_1\in\mathbb{Z}} \sum_{l_2\in\mathbb{Z}} \dots \sum_{l_{q-v}\in\mathbb{Z}} \left(\prod_{r=1}^{q-v} \left|c_{l_r}^{a_r}\right|\right) < \infty .
\end{aligned}
$$

The proof of the lemma is complete. □

R.J. Serfling generalized a fundamental maximal inequality for orthogonal random variables, and is proven in Theorem 2.4.1 of Stout [36]. We conclude this section by presenting a simplified version of that generalization, which we shall use throughout the rest of this thesis.

**Theorem 4.5** (Serfling's Generalization). *Let* $\{Z_k\}_{k\in\mathbb{N}}$ *be a time series with* *finite second moments, and $f$ be a super-additive function on $\mathbb{N}$, such that*

$$\begin{aligned} f(a)+f(b) &\leq f(a+b) \qquad \forall\ a,b \in \mathbb{N} \\ E\left[\left(\sum_{i=o+1}^{n} Z_i\right)^2\right] &\leq f(n-o) \qquad \forall\ o<n \in \mathbb{N}_0\ . \end{aligned}$$

*Then, for $n_r = 2^r,\ r \in \mathbb{N}_0,$ and $o, n \in \mathbb{N},$ we have*

$$E\left[\max_{n_r \leq o < n < n_{r+1}} \left(\sum_{i=o+1}^{n} Z_i\right)^2\right] \overset{r}{\ll} r^2 f(n_r)\ . \tag{4.12}$$

## 4.3 Light Tailed Case of Theorem 3.1

The following calculation will explain why we consider the case $\alpha_0 > 2$ in (3.2) to be non-heavy tailed, and the case $\alpha_0 \in (1,2]$ to have possible heavy tails. If $\alpha_0 > 2$, then $\alpha_i = \frac{s}{s-i}\alpha_0 > 2$ for $i \in \{0,1,\ldots,\lfloor\frac{s-1}{2}\rfloor\}$. When $\pi$ is a permutation of $\{1,2,\ldots,s\}$, we see from (3.2), that

$$\begin{aligned} & E\left[\prod_{r\in\{\pi(1),\ldots,\pi(s-i)\}} \left|\xi_1^{(r)}\right|^2\right] \\ = \ & 2\int_0^\infty tP\left(\prod_{r\in\{\pi(1),\ldots,\pi(s-i)\}} \left|\xi_1^{(r)}\right| > t\right)\ dt \\ \ll \ & 2\int_0^1 1\ dt +\ 2\int_1^\infty t^{1-\alpha_i}\ dt \\ \ll \ & 2+\frac{2}{\alpha_i - 2} \ < \ \infty\ , \qquad \forall\ 0 \leq i \leq \left\lfloor\frac{s-1}{2}\right\rfloor . \end{aligned} \tag{4.13}$$

We conclude that $E\left[\prod_{r=1}^{s}\left(1+\left(\xi_1^{(r)}\right)^2\right)\right] < \infty$, which precludes heavy tails. When all the $\xi_1^{(r)}$'s are equal, to say $\xi_1$, we see that

$$E\left[|\xi_1|^{2s}\right] = 2\int_0^\infty tP\left(|\xi_1|^s > t\right)\,dt \ll 2 + 2\int_1^\infty t^{1-\alpha_0}\,dt < \infty\,.$$

Thus, keeping Remarks 3.4 and 3.5 in mind, we first present a theorem that handles long range dependence under the condition $\alpha_0 > 2$.

**Theorem 4.6.** *Let $s \in \mathbb{N}$ and $\{\xi_l\}_{l\in\mathbb{Z}}$ be i.i.d. zero-mean random variables such that $E\left[(\xi_1)^{2s}\right] < \infty$, and $\{c_l\}_{l\in\mathbb{Z}}$ satisfy $\sup_{l\in\mathbb{Z}} |l|^\sigma |c_l| < \infty$, for some $\sigma \in \left(\frac{1}{2}, 1\right)$. For $k \in \mathbb{Z}$, define $x_k = \sum_{l=-\infty}^{\infty} c_{k-l}\xi_l$, $d_k = (x_k)^s$, and $d = E(d_k)$. Then, $\lim_{n\to\infty} n^{-\frac{1}{p}} \sum_{k=1}^{n} (d_k - d) = 0$ a.s. for*

$$p < \begin{cases} 2 \wedge \frac{1}{2-2\sigma}, & s = 2 \\ \frac{2}{3-2\sigma}, & s \neq 2 \end{cases}. \tag{4.14}$$

*Furthermore, if $\xi_1$ is symmetric, and $s$ is even, then the constraint for (4.14) can be relaxed to*

$$p < 2 \wedge \frac{1}{2-2\sigma}\,. \tag{4.15}$$

*Proof.* By expanding the expressions for $d_k$ and $d$, we get that,

$$\sum_{k=1}^{n}(d_k - d) = \sum_{k=1}^{n}\sum_{l_1=-\infty}^{\infty}\cdots\sum_{l_s=-\infty}^{\infty}\left(\prod_{r=1}^{s} c_{k-l_r}\right)\left(\prod_{r=1}^{s}\xi_{l_r} - E\left(\prod_{r=1}^{s}\xi_{l_r}\right)\right).$$

This expression for $\sum_{k=1}^{n}(d_k - d)$ can be broken up in several sums based on the combinations of subscripts of $\xi$'s that are equal. That is, $\sum_{k=1}^{n}(d_k - d)$ can be seen as the sum of

$$S_n(q,\lambda_q) = \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left(\prod_{r=1}^{q} c_{k-l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \xi_{l_r}^{a_r} - E\left(\prod_{r=1}^{q} \xi_{l_r}^{a_r}\right)\right). \quad (4.16)$$

where $q$ ranges over $\{1,2,\ldots,s\}$, and $\lambda_q = (a_1, a_2, \ldots, a_q)$ satisfies $a_1 + \ldots + a_q = s$ and $a_1 \geq a_2 \geq \ldots \geq a_q \geq 1$.

Before we bound the second moment of $S_n(q,\lambda_q)$, we shall consider an analogous summation, $Y_{o,n,\delta}^{\lambda_q}$, but with more general random variables $\psi_l^{(r)}$ instead of $\xi_l^{a_r}$. For $q \in \mathbb{N},\ v \in \{1,2,\ldots,q\}$, and $\delta \geq 1$, we define $\{(\psi_l^{(1)},\ldots,\psi_l^{(q)})\}_{l\in\mathbb{Z}}$ to be i.i.d $\mathbb{R}^q$-valued random vectors, such that

$$\begin{cases} E\left(\psi_1^{(r)}\right) & \ll \mathbf{1}_{\{1\leq r\leq q-v\}}, \\ E\left[\left(\psi_1^{(r)}\right)^2\right] & \ll \delta\mathbf{1}_{\{r=1\}} + \mathbf{1}_{\{r\neq 1\}}, \end{cases} \qquad \forall\ 1 \leq r \leq q\ ,$$

and for $o < n \in \mathbb{N}_0$, and fixed $\lambda_q = (a_1, a_2, \ldots, a_q)$, we define

$$Y_{o,n,\delta}^{\lambda_q} = \sum_{k=o+1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left(\prod_{r=1}^{q} c_{k-l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \psi_{l_r}^{(r)} - E\left(\prod_{r=1}^{q} \psi_{l_r}^{(r)}\right)\right).$$

The moments of $\psi_l^{(r)}$ have been assumed in such a manner that allows for the applicability of this result in a variety of cases. The means of $\psi_l^{(q-v+1)}, \psi_l^{(q-v+2)}, \ldots, \psi_l^{(q)}$ have been assumed to be 0 keeping in mind that $\xi_l^{a_r}$ has zero-mean when $a_r = 1$, and the second moment of $\psi_l^{(1)}$ has been kept arbitrary to allow for substitution of different random variables like $\xi_l^{a_r}$, $U[-1,1]$, or truncated versions of random variables exhibiting heavy tails, which may have different second moments.

### 4.3.1 Bounding covariance of $\prod_{r=1}^{q} \psi_{l_r}^{(r)}$

We first give the following definitions.

**Definition 4.7.** For $q \in \mathbb{N}$, $v \in \{1, 2, \ldots, q\}$, let the sets $V_r = V_r^{v,q}$ for $1 \leq r \leq 6$, be such that $V_1, V_2, V_3$ partition $\{q - v + 1, \ldots, q\}$, and $V_4, V_5, V_6$ partition $\{1, \ldots, q - v\}$. Define a matching function $\nu = \nu^{q,v}(V_2, V_3, V_4, V_5)$, given by

$$\nu : V_2 \cup V_3 \cup V_4 \cup V_5 \ \to \ \{1, \ldots, q\},$$

such that $\nu$ is injective, $\nu(V_2 \cup V_4) \subseteq \{q - v + 1, \ldots, q\}$, and $\nu(V_3 \cup V_5) \subseteq \{1, \ldots, q - v\}$. For ease of notation, we further define $W_1 = W_1^{q,v}(\nu) = \{q - v + 1, \ldots, q\} \setminus \nu(V_2 \cup V_4)$, $W_r = W_r^{q,v}(\nu) = \nu(V_r)$ for $2 \leq r \leq 5$, and $W_6 = W_6^{q,v}(\nu) = \{1, \ldots, q - v\} \setminus \nu(V_3 \cup V_5)$.

**Remark 4.8.** In Definition 4.7, observe that $|V_1| + \ldots + |V_6| = |W_1| + |\nu(V_2)| + \ldots + |\nu(V_5)| + |W_6| = q$. Also, since $V_1, V_2, V_3$ partition $\{q-v+1, \ldots, q\}$, as do $W_1, \nu(V_2), \nu(V_4)$, we get that $|V_1| + |V_2| + |V_3| = |W_1| + |\nu(V_2)| + |\nu(V_4)| = v$. Similarly, $|V_4| + |V_5| + |V_6| = |\nu(V_3)| + |\nu(V_5)| + |W_6| = q - v$. Finally, due to injectivity of $\nu$, we have $|\nu(V_r)| = |V_r|$ for $2 \leq r \leq 5$.

**Definition 4.9.** Let $q \in \mathbb{N}$, $v \in \{1, 2, \ldots, q\}$, and $\Delta = \Delta_q$ be the set of all tuples in $\mathbb{Z}^q$ with distinct elements, i.e. $(l_1, \ldots, l_q) \in \Delta$ satisfies $l_i \neq l_j$ for all

$1 \leq i < j \leq q$. For sets $V_1, ..., V_6$ and matching function $\nu$ as in Definition 4.7, partition $\Delta \times \Delta$ into the sets

$$\begin{aligned} & \Delta \times \Delta(V_1, ..., V_6, \nu) \\ = & \ \{\big((l_1, \ldots, l_q), (m_1, \ldots, m_q)\big) \in \Delta \times \Delta : \ l_r = m_{\nu(r)}, \ \forall \ r \in V_2 \cup V_3 \cup V_4 \cup V_5\}. \end{aligned}$$

The following lemma bounds the covariance of $\prod_{r=1}^{q} \psi_{l_r}^{(r)}$.

**Lemma 4.10.** *Let* $q \in \mathbb{N}, \ v \in \{1, 2, \ldots, q\}, \ \delta \geq 1,$ *and* $\{(\psi_l^{(1)}, \ldots, \psi_l^{(q)})\}_{l \in \mathbb{Z}}$ *be i.i.d* $\mathbb{R}^q$*-valued random vectors, such that*

$$\begin{cases} E\left(\psi_1^{(r)}\right) & \ll \ \boldsymbol{1}_{\{1 \leq r \leq q-v\}}, \\ E\left[\left(\psi_1^{(r)}\right)^2\right] & \ll \ \delta \boldsymbol{1}_{\{r=1\}} + \boldsymbol{1}_{\{r \neq 1\}}, \end{cases} \qquad \forall \ \ 1 \leq r \leq q \ . \tag{4.17}$$

*For the same* $q$*,* $v$ *as in (4.17), let* $\big((l_1, l_2, \ldots, l_q), (m_1, m_2, \ldots, m_q)\big) \in \Delta \times \Delta(V_1, ..., V_6, \nu)$ *(from Definition 4.9).*

*Then,*

$$\begin{aligned} & \left| E\left(\prod_{r=1}^{q} (\psi_{l_r}^{(r)} \psi_{m_r}^{(r)})\right) - E\left(\prod_{r=1}^{q} \psi_{l_r}^{(r)}\right) E\left(\prod_{r=1}^{q} \psi_{m_r}^{(r)}\right) \right| \\ \overset{\delta}{\ll} & \begin{cases} 0, & |V_1| > 0 \ \ \textit{or} \ \ |W_1| > 0 \ \ \textit{or} \ \ |V_6| = q, \\ 1, & 0 < |V_6| < q, \ |V_1| = |V_4| = |V_5| = |W_1| = 0, \\ \delta, & \textit{otherwise.} \end{cases} \end{aligned} \tag{4.18}$$

*Proof.* When $V_1 \cup V_2 \cup V_3 \neq \phi$, due to the independence of $\psi$'s with different subscripts, and the zero-mean property of $\psi_{l_r}^{(r)}$ for $r \in V_1 \cup V_2 \cup V_3$ in (4.17), we have

$$E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right) = E\left(\prod_{r\in V_4\cup V_5\cup V_6}\psi_{l_r}^{(r)}\right)\left(\prod_{r\in V_1\cup V_2\cup V_3}E\left(\psi_{l_r}^{(r)}\right)\right) = 0\ .$$

Similarly, when $W_1\cup\nu(V_2)\cup\nu(V_4)\neq\phi$, we get that $E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)=0$. Hence, when $V_1\cup V_2\cup V_3\neq\phi$ or $W_1\cup\nu(V_2)\cup\nu(V_4)\neq\phi$, we get that

$$E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right) = 0\ . \tag{4.19}$$

**Case 1:** $|V_1|>0$ or $|W_1|>0$ or $|V_6|=q$.

$|V_1|>0$ implies that $V_1\neq\phi$, and $|W_1|>0$ implies that $W_1\neq\phi$, hence (4.19) holds in this case. When $V_1\neq\phi$, we see from Definition 4.7, that for all $r\in V_1$, $l_r\neq m_j$ for all $1\leq j\leq q$. Hence, due to the independence of $\psi$'s with different subscripts, and the zero-mean property of $\psi_{l_r}^{(r)}$ for $r\in V_1$, we get that

$$E\left(\prod_{r=1}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) = E\left(\prod_{r\in\{1,\ldots,q\}\setminus V_1}\psi_{l_r}^{(r)}\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\prod_{r\in V_1}E\left(\psi_{l_r}^{(r)}\right) = 0\ . \tag{4.20}$$

Similarly, (4.20) holds when $W_1\neq\phi$. Thus, when $|V_1|>0$ or $|W_1|>0$, from (4.19) and (4.20), we get that

$$\left|E\left(\prod_{r=1}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right| = 0\ . \tag{4.21}$$

When $|V_6|=q$, we must have $v=0$ and none of the $l$'s are equal to any of the $m$'s, i.e. $\{l_1,\ldots,l_q\}\cap\{m_1,\ldots,m_q\}=\phi$. In that scenario, due to the independence of $\psi_{l_r}^{(r)}$'s with $\psi_{m_r}^{(r)}$'s, (4.21) holds as well.

**Case 2:** $0 < |V_6| < q,\ |W_1| = |V_1| = |V_4| = |V_5| = 0.$

In this case we will show that $l_1 \notin \{m_1, \ldots, m_q\}$ and $m_1 \notin \{l_1, \ldots, l_q\}$. From Remark 4.8, note that $|V_4| + |V_5| + |V_6| = q - v$, hence $0 < |V_6| < q$ along with $|V_4| = |V_5| = 0$ implies that $0 < v < q$. Since $v$ is the cardinality of $V_1 \cup V_2 \cup V_3$, this means that $\{1, \ldots, q\} \neq V_1 \cup V_2 \cup V_3 \neq \phi$, and (4.19) also holds in this case.

From Remark 4.8, using injectivity of $\nu$, we get that $|V_1| + |V_2| + |V_3| = |W_1| + |V_2| + |V_4|$. Thus, $|V_1| = |W_1| = 0$ implies that $|V_3| = |V_4|$. Also, $v < q$ implies that $q - v \geq 1$, hence $1 \in V_4 \cup V_5 \cup V_6$ and $1 \in \nu(V_3) \cup \nu(V_5) \cup W_6$. Further, $|V_3| = |V_4| = |V_5| = 0$ ensures that $1 \in V_6$ and $1 \in W_6$. This means that $l_1 \notin \{m_1, \ldots, m_q\}$ and $m_1 \notin \{l_1, \ldots, l_q\}$. Hence, due to independence of $\psi$'s with unequal subscripts, Cauchy-Schwartz inequality, and (4.17), we get

$$\begin{aligned} E\left(\prod_{r=1}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) &= E\left(\psi_{l_1}^{(1)}\right) E\left(\psi_{m_1}^{(1)}\right) E\left(\prod_{r=2}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) \\ &\leq E\left(\psi_{l_1}^{(1)}\right) E\left(\psi_{m_1}^{(1)}\right) \sqrt{\prod_{r=2}^{q} E\left[\left(\psi_{l_r}^{(r)}\right)^2\right] \prod_{r=2}^{q} E\left[\left(\psi_{m_r}^{(r)}\right)^2\right]} \\ &\overset{\delta}{\ll} 1\ . \end{aligned} \tag{4.22}$$

From (4.19) and (4.22), we get that

$$\left| E\left(\prod_{r=1}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right) E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right| \overset{\delta}{\ll} 1\ . \tag{4.23}$$

**Case 3:** None of the above.

For all other cases, we will get various bounds, and we will show that the worst of them is $\delta$. Due to the independence of $\psi$'s with different subscripts, Cauchy-Schwartz inequality, and the fact that $E\left[\left(\psi_1^{(r)}\right)^2\right] \ll \delta$ (from (4.17)), we get that

$$
\begin{aligned}
E\left(\prod_{r=1}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) &\leq \sqrt{\prod_{r=1}^{q} E\left[\left(\psi_{l_r}^{(r)}\right)^2\right]\prod_{r=1}^{q} E\left[\left(\psi_{m_r}^{(r)}\right)^2\right]} \\
&\overset{\delta}{\ll} \sqrt{\delta^2\prod_{r=2}^{q} E\left[\left(\psi_{l_r}^{(r)}\right)^2\right]\prod_{r=2}^{q} E\left[\left(\psi_{m_r}^{(r)}\right)^2\right]} \\
&\overset{\delta}{\ll} \delta\ . \qquad (4.24)
\end{aligned}
$$

We also see that $E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right) \overset{\delta}{\ll} 1$, due to independence of $\psi$'s with different subscripts, so using (4.24) and Triangle Inequality, we get that

$$
\left|E\left(\prod_{r=1}^{q}(\psi_{l_r}^{(r)}\psi_{m_r}^{(r)})\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right| \overset{\delta}{\ll} \delta + 1 \overset{\delta}{\ll} \delta\ . \qquad (4.25)
$$

Lemma 4.10 follows from (4.21, 4.23) and (4.25). □

### 4.3.2 Bounding second moment of $Y_{o,n,\delta}^{\lambda_q}$

We now present a lemma that bounds the second moment of the difference in partial sums of a general expression which we will use not only to bound the second moment of $S_n(q,\lambda_q)$, but also later on to handle heavy tails. Here, we will work with a given fixed partition of $s$, i.e. $\lambda_q = (a_1, a_2, \ldots, a_q)$.

**Lemma 4.11.** *Let $o < n \in \mathbb{N}_0$, $s \in \mathbb{N}$, $\delta \geq 1$, $\lambda_q = (a_1, a_2, \ldots, a_q)$ satisfies $a_1 + \ldots + a_q = s$ and $a_1 \geq a_2 \geq \ldots \geq a_q \geq 1$, and $v = \#\{1 \leq r \leq q : a_r = 1\}$. Let $\{c_l\}_{l\in\mathbb{Z}}$ satisfy $\sup_{l\in\mathbb{Z}} |l|^\sigma |c_l| < \infty$, for some $\sigma \in \left(\frac{1}{2}, 1\right)$, and $\{(\psi_l^{(1)}, \ldots, \psi_l^{(q)})\}_{l\in\mathbb{Z}}$ be i.i.d $\mathbb{R}^q$-valued random vectors, such that*

$$\begin{cases} E\left(\psi_1^{(r)}\right) & \ll \boldsymbol{1}_{\{1\leq r\leq q-v\}}, \\ E\left[\left(\psi_1^{(r)}\right)^2\right] & \ll \delta \boldsymbol{1}_{\{r=1\}} + \boldsymbol{1}_{\{r\neq 1\}}, \end{cases} \qquad \forall \;\; 1 \leq r \leq q \;. \tag{4.26}$$

*Define,*

$$Y_{o,n,\delta}^{\lambda_q} = \sum_{k=o+1}^{n} \sum_{l_1\neq l_2\neq \ldots \neq l_q} \left(\prod_{r=1}^{q} c_{k-l_r}^{a_r}\right)\left(\prod_{r=1}^{q} \psi_{l_r}^{(r)} - E\left(\prod_{r=1}^{q} \psi_{l_r}^{(r)}\right)\right).$$

*Then* $$E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right] \overset{o,n,\delta}{\ll} \begin{cases} \delta\ (n-o), & a_q \geq 2 \\ \delta\ (n-o)\ l_{s,\sigma}(n-o), & a_1 = 1 \\ (\delta\ (n-o)) \ \vee\ ((n-o)\ l_{1,\sigma}(n-o)), & a_q = 1,\ a_1 \geq 2 \end{cases},$$

*where $l_{s,\sigma}$ is defined in the Notation List (Section 4.1). Furthermore, if $s$ is even and $E\left(\psi_1^{(r)}\right) = 0$ whenever $a_r$ is odd, then this bound can be tightened* *to*

$$E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right] \overset{o,n,\delta}{\ll} (\delta\ (n-o))\ \vee\ ((n-o)\ l_{2,\sigma}(n-o)).$$

*when* $a_q = 1$ *and* $a_1 \geq 2$.

*Proof.* We first bound the second moment of $Y_{o,n,\delta}^{\lambda_q}$.

$$\begin{aligned}
& E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right] \\
= & \sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\sum_{l_1\neq l_2\neq\ldots\neq l_q}\sum_{m_1\neq m_2\neq\ldots\neq m_q}\left(\prod_{r=1}^{q} c_{j-m_r}^{a_r} c_{k-l_r}^{a_r}\right) \\
& \left[E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\psi_{m_r}^{(r)}\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right] \\
\leq & \sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\sum_{l_1\neq l_2\neq\ldots\neq l_q}\sum_{m_1\neq m_2\neq\ldots\neq m_q}\left(\prod_{r=1}^{q}\left|c_{j-m_r}^{a_r}\right|\left|c_{k-l_r}^{a_r}\right|\right) \\
& \left|E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\psi_{m_r}^{(r)}\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right|. \qquad (4.27)
\end{aligned}$$

Notice that the summation in (4.27) is over $\Delta \times \Delta$ (from Definition 4.9). Based on $q$ and $v = \#\{1 \leq r \leq q : a_r = 1\}$, we can partition $\Delta \times \Delta$ into the sets $\Delta \times \Delta(V_1, ..., V_6, \nu)$. For sets $V_1, \ldots, V_6$ and matching function $\nu$ as in Definition 4.7, define

$$\begin{aligned}
S(V_1,\ldots,V_6,\nu) = & \sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\sum_{\substack{((l_1,\ldots,l_q),(m_1,\ldots,m_q)) \\ \in\Delta\times\Delta(V_1,\ldots,V_6,\nu)}}\left(\prod_{r=1}^{q}\left|c_{j-m_r}^{a_r}\right|\left|c_{k-l_r}^{a_r}\right|\right) \\
& \left|E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\psi_{m_r}^{(r)}\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right|. \qquad (4.28)
\end{aligned}$$

Using the fact that there can only be a finite number of possibilities for $V_1, \ldots, V_6$ and $\nu$, we get from (4.27) and (4.28), that

$$E\left[(Y^{\lambda_q}_{o,n,\delta})^2\right] \overset{o,n,\delta}{\ll} \max_{V_1,\ldots,V_6,\ \nu} S(V_1, \ldots, V_6, \nu) \ . \tag{4.29}$$

Observe that when $|V_1| > 0$ or $|W_1| > 0$, $S(V_1, \ldots, V_6, \nu) = 0$ according to Lemma 4.10, and need not be considered in (4.29). Hence we assume that $|V_1| = |W_1| = 0$. From Remark 4.8, recall that $|V_1| + |V_2| + |V_3| = |W_1| + |\nu(V_2)| + |\nu(V_4)| = v$. Due to injectivity of $\nu$, we have $|\nu(V_r)| = |V_r|$ for $2 \leq r \leq 5$, so when $|V_1| = |W_1| = 0$, we get our second observation, i.e. $|V_3| = |V_4|$. Similarly, since $|V_1| + \ldots + |V_6| = |W_1| + |\nu(V_2)| + \ldots + |\nu(V_5)| + |W_6| = q$, using $|V_1| = |W_1| = 0$, we get that $|V_6| = |W_6|$. Hence, we only need to consider $S(V_1, \ldots, V_6, \nu)$, where

$$\begin{cases} |V_1| = |W_1| = 0, \\ |V_3| = |V_4| \,, \\ |V_6| = |W_6| \,. \end{cases} \tag{4.30}$$

We now fix sets $V_1, \ldots, V_6$ and matching function $\nu$, from Definition 4.9, satisfying (4.30). Combining the bound of Lemma 4.10 with our observations in (4.30), we define

$$\rho_{u_2,\ldots,u_6} = \begin{cases} 1, & 0 < u_6 < q, \ u_4 = u_5 = 0 \\ \delta, & \text{otherwise} \end{cases} \ , \tag{4.31}$$

which we will use to bound $S(V_1, \ldots, V_6, \nu)$ below. Using (4.28) and (4.31), we first group the coefficients according to $V_1,\ldots,V_6$, and $\nu$, to get that

$$
\begin{aligned}
&S(V_1,\ldots,V_6,\nu)\\
\overset{o,n,\delta}{\ll}\ &\sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\sum_{\substack{((l_1,\ldots,l_q),(m_1,\ldots,m_q))\\ \in\Delta\times\Delta(V_1,\ldots,V_6,\nu)}}\left(\prod_{r=1}^{q}\left|c_{j-m_r}^{a_r}\right|\left|c_{k-l_r}^{a_r}\right|\right)\rho_{|V_2|,\ldots,|V_6|}\\
\overset{o,n,\delta}{\ll}\ &\rho_{|V_2|,\ldots,|V_6|}\sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\sum_{\substack{((l_1,\ldots,l_q),(m_1,\ldots,m_q))\\ \in\Delta\times\Delta(V_1,\ldots,V_6,\nu)}}\left(\prod_{r\in W_6}|c_{j-m_r}^{a_r}|\right)\\
&\left(\prod_{r\in V_6}|c_{k-l_r}^{a_r}|\right)\left(\prod_{r\in V_5}|c_{j-m_{\nu(r)}}^{a_{\nu(r)}}||c_{k-l_r}^{a_r}|\right)\left(\prod_{r\in V_4}|c_{j-m_{\nu(r)}}^{a_{\nu(r)}}||c_{k-l_r}^{a_r}|\right)\\
&\left(\prod_{r\in V_3}|c_{j-m_{\nu(r)}}^{a_{\nu(r)}}||c_{k-l_r}^{a_r}|\right)\left(\prod_{r\in V_2}|c_{j-m_{\nu(r)}}^{a_{\nu(r)}}||c_{k-l_r}^{a_r}|\right). \qquad (4.32)
\end{aligned}
$$

Note that $a_r \geq 2$ (hence $c_l^{a_r} \leq c_l^2$ ) for $r \in V_4\cup V_5\cup V_6\cup W_6$, and $a_r = 1$ for $r \in V_2\cup V_3$. Next, for $r \in V_2\cup V_3\cup V_4\cup V_5$, we use $l_r = m_{\nu(r)}$ in (4.32), then bring in the summations and extend them over all integers, to get

$$
\begin{aligned}
&S(V_1,\ldots,V_6,\nu)\\
\overset{o,n,\delta}{\ll}\ &\rho_{|V_2|,\ldots,|V_6|}\sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\left(\prod_{r\in W_6}\sum_{m_r=-\infty}^{\infty}|c_{j-m_r}^2|\right)\left(\prod_{r\in V_6}\sum_{l_r=-\infty}^{\infty}|c_{k-l_r}^2|\right)\\
&\left(\prod_{r\in V_5}\sum_{l_r=-\infty}^{\infty}|c_{j-l_r}^2||c_{k-l_r}^2|\right)\left(\prod_{r\in V_4}\sum_{l_r=-\infty}^{\infty}|c_{j-l_r}||c_{k-l_r}^2|\right)\\
&\left(\prod_{r\in V_3}\sum_{l_r=-\infty}^{\infty}|c_{j-l_r}^2||c_{k-l_r}|\right)\left(\prod_{r\in V_2}\sum_{l_r=-\infty}^{\infty}|c_{j-l_r}||c_{k-l_r}|\right)\\
\overset{o,n,\delta}{\ll}\ &\rho_{|V_2|,\ldots,|V_6|}\sum_{k=o+1}^{n}\sum_{j=o+1}^{n}\left(\sum_{m=-\infty}^{\infty}|c_{j-m}^2|\right)^{|W_6|}\left(\sum_{l=-\infty}^{\infty}|c_{j-l}^2|\right)^{|V_6|}\left(\sum_{l=-\infty}^{\infty}|c_{j-l}^2||c_{k-l}^2|\right)^{|V_5|}\\
&\left(\sum_{l=-\infty}^{\infty}|c_{j-l}||c_{k-l}^2|\right)^{|V_4|}\left(\sum_{l=-\infty}^{\infty}|c_{j-l}^2||c_{k-l}|\right)^{|V_3|}\left(\sum_{l=-\infty}^{\infty}|c_{j-l}||c_{k-l}|\right)^{|V_2|}. \qquad (4.33)
\end{aligned}
$$

Applying Lemma 4.1 with $\gamma = \sigma,\ 2\sigma$ and Lemma 4.2 with $\gamma = \sigma$, we have

$$\sum_{l=-\infty}^{\infty} |c_{j-l}^2||c_{k-l}^2| \overset{o,n,\delta}{\ll} \begin{cases} 1 + \sum_{\substack{l=-\infty \\ l\neq j}}^{\infty} |j-l|^{-4\sigma}, & j = k \\ \sum_{\substack{l=-\infty \\ l\notin\{j,k\}}}^{\infty} |j-l|^{-2\sigma}|k-l|^{-2\sigma} + |j-k|^{-2\sigma}, & j \neq k \end{cases}$$
$$\overset{o,n,\delta}{\ll} \begin{cases} 1, & j = k \\ |j-k|^{-2\sigma}, & j \neq k \end{cases} \tag{4.34}$$

$$\sum_{l=-\infty}^{\infty} |c_{j-l}||c_{k-l}^2| \overset{o,n,\delta}{\ll} \begin{cases} 1 + \sum_{\substack{l=-\infty \\ l\neq j}}^{\infty} |j-l|^{-3\sigma}, & j = k \\ \sum_{\substack{l=-\infty \\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\sigma}|k-l|^{-2\sigma} + |j-k|^{-\sigma}, & j \neq k \end{cases}$$
$$\overset{o,n,\delta}{\ll} \begin{cases} 1, & j = k \\ |j-k|^{-\sigma}, & j \neq k \end{cases} \tag{4.35}$$

$$\sum_{l=-\infty}^{\infty} |c_{j-l}||c_{k-l}| \overset{o,n,\delta}{\ll} \begin{cases} 1 + \sum_{\substack{l=-\infty \\ l\neq j}}^{\infty} |j-l|^{-2\sigma}, & j = k \\ \sum_{\substack{l=-\infty \\ l\notin\{j,k\}}}^{\infty} |j-l|^{-\sigma}|k-l|^{-\sigma} + |j-k|^{-\sigma}, & j \neq k \end{cases}$$
$$\overset{o,n,\delta}{\ll} \begin{cases} 1, & j = k \\ |j-k|^{1-2\sigma}, & j \neq k \end{cases} \tag{4.36}$$

Using (4.30,4.33,4.34,4.35) and (4.36), and the summability of $|c_l^2|$ over integers, we get that

$$
\begin{aligned}
& S(V_1,\ldots,V_6,\nu) \\
\overset{o,n,\delta}{\ll} & \rho_{|V_2|,\ldots,|V_6|} \sum_{k=o+1}^{n} \left( 1 \ + \sum_{\substack{j=o+1 \\ j\neq k}}^{n} |j-k|^{-2\sigma|V_5|} |j-k|^{-(|V_3|+|V_4|)\sigma} |j-k|^{(1-2\sigma)|V_2|} \right) \\
\overset{o,n,\delta}{\ll} & \rho_{|V_2|,\ldots,|V_6|} \sum_{k=o+1}^{n} \left( 1 \ + \sum_{\substack{j=o+1 \\ j\neq k}}^{n} |j-k|^{|V_2|-2(|V_2|+|V_3|+|V_5|)\sigma} \right). \qquad (4.37)
\end{aligned}
$$

(4.37) provides a bound for $S(V_1,\ldots,V_6,\nu)$ in terms of the cardinalities $|V_2|,\ldots|V_6|$. However, depending on the given partition $\lambda_q = (a_1, a_2, \ldots, a_q)$, the value of $v$ can be different, thus putting constraints on $V_2,\ldots,V_6$. We shall use (4.29) and (4.37) to bound the second moment of $Y^{\lambda_q}_{o,n,\delta}$.

**Case 1:** $a_q \geq 2$.

In this case, we see that $a_r \neq 1,\ \forall\ 1 \leq r \leq q$. Thus, Definition 4.7 gives us that $|V_2| = |V_3| = 0$. Also from (4.30), $|V_3| = |V_4|$ gives us that $|V_4| = 0$. If further, $|V_5| = 0$, then we will have $|V_6| = q$ (since $|V_2| + \ldots + |V_6| = q$). So by Lemma 4.10, we see that

$$
\left| E\left( \prod_{r=1}^{q} (\psi^{(r)}_{l_r} \psi^{(r)}_{m_r}) \right) - E\left( \prod_{r=1}^{q} \psi^{(r)}_{l_r} \right) E\left( \prod_{r=1}^{q} \psi^{(r)}_{m_r} \right) \right| = 0
$$

and hence $S(V_1,\ldots,V_6,\nu) = 0$. Since we need not consider cases where

$S(V_1,\ldots,V_6,\nu)=0$, we assume that $|V_5|\geq 1$.

Thus, we have $|V_2|=|V_3|=0,\ |V_5|\geq 1$, and get that $\rho_{0,0,0,|V_5|,|V_6|}=\delta$ (from (4.31)), and that $|V_2|-2(|V_2|+|V_3|+|V_5|)\sigma<-1$ (since $\sigma\in\left(\frac{1}{2},1\right)$). From (4.29,4.37), we get that

$$\begin{aligned} E\left[(Y^{\lambda_q}_{o,n,\delta})^2\right] &\overset{o,n,\delta}{\ll} \max_{|V_5|\geq 1,\ |V_6|}\left(\rho_{0,0,0,|V_5|,|V_6|}\right)\sum_{k=o+1}^{n}\left(1 \ + \sum_{\substack{j=o+1\\ j\neq k}}^{n}|j-k|^{-2|V_5|\sigma}\right) \\ &= \delta\ (n-o)\ . \end{aligned} \tag{4.38}$$

**Case 2:** $a_1=1$.

In this case, we see that $a_r=1,\ \forall\ 1\leq r\leq q$. Thus, Definition 4.7 gives us that $|V_4|=|V_5|=|V_6|=0$. Also from (4.30), $|V_3|=|V_4|$ gives us that $|V_3|=0$ and $|V_2|=q$. Since in this permutation, all the elements are 1, so $q=s$, and hence $|V_2|=s$.

Thus, we have $|V_3|=|V_5|=0,\ |V_2|=s$, and get that $\rho_{s,0,0,0,0}=\delta$ (from (4.31)), and that $|V_2|-2(|V_2|+|V_3|+|V_5|)\sigma=(1-2\sigma)s$. From (4.29,4.37), we get that

$$\begin{aligned} E\left[(Y^{\lambda_q}_{o,n,\delta})^2\right] &\overset{o,n,\delta}{\ll} (\rho_{s,0,0,0,0})\sum_{k=o+1}^{n}\left(1 \ + \sum_{\substack{j=o+1\\ j\neq k}}^{n}|j-k|^{(1-2\sigma)s}\right) \\ &\overset{o,n,\delta}{\ll} \delta\ (n-o)\ \ l_{s,\sigma}(n-o)\ , \end{aligned} \tag{4.39}$$

where $l_{s,\sigma}$ is from the Notation list in Section 4.1.

**Case 3:** $a_1 \geq 2,\ a_q = 1$.

In this case, we see from Definition 4.7, that $0 < |V_2| + |V_3| < q$ and $0 < |V_4| + |V_5| + |V_6| < q$.

First, assume that $|V_3| = |V_5| = 0$. Since from (4.30), we have $|V_3| = |V_4|$, thus we get that $|V_4| = 0$, and $|V_2|, |V_6| \in \{1, 2, \ldots, q-1\}$. So we have $\rho_{|V_2|,0,0,0,|V_6|} = 1$ (from (4.31)), and that $|V_2| - 2(|V_2| + |V_3| + |V_5|)\sigma = (1-2\sigma)|V_2|$, hence using (4.29,4.37), we get that

$$E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right] \overset{o,n,\delta}{\ll} \max_{|V_2|,|V_6| \in \{1,2,\ldots,q-1\}} \left(\rho_{|V_2|,0,0,0,|V_6|}\right) \sum_{k=o+1}^{n} \left( 1 + \sum_{\substack{j=o+1 \\ j \neq k}}^{n} |j-k|^{(1-2\sigma)|V_2|} \right)$$

$$\overset{o,n,\delta}{\ll} \max_{|V_2| \in \{1,2,\ldots,q-1\}} (n-o) \ \ l_{|V_2|,\sigma}(n-o) \ . \tag{4.40}$$

From Notation List in Section 4.1, $l_{|V_2|,\sigma}(n-o) = \begin{cases} (n-o)^{|V_2|(1-2\sigma)+1}, & \sigma < \frac{|V_2|+1}{2|V_2|} \\ \log(n-o), & \sigma = \frac{|V_2|+1}{2|V_2|} \\ 1, & \sigma > \frac{|V_2|+1}{2|V_2|} \end{cases}$.

Since $(1-2\sigma) < 0,\ \ n-o \geq 1$, and $\frac{|V_2|+1}{2|V_2|}$ decreases as $|V_2|$ increases, observe that $l_{|V_2|,\sigma}(n-o)$ is a non-increasing function of $|V_2| \in \{1, 2, \ldots, v\}$. Thus, we take $|V_2| = 1$ in (4.40) to bound the left hand side, and get

$$E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right] \overset{o,n,\delta}{\ll} (n-o) \ \ l_{1,\sigma}(n-o) \ .$$

For all other values of $|V_3|$ and $|V_5|$, we have $|V_2| - 2(|V_2| + |V_3| + |V_5|)\sigma < -1$ (since $\sigma \in \left(\frac{1}{2}, 1\right)$), and $\rho_{|V_2|,\ldots,|V_6|} \leq \delta$ (from (4.31)). Thus, we get from

(4.29,4.37), that

$$E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right] \overset{o,n,\delta}{\ll} \max_{|V_2|,\ldots,|V_6|}\left(\rho_{|V_2|,\ldots,|V_6|}\right)\sum_{k=o+1}^{n}\left(1+\sum_{\substack{j=o+1\\ j\neq k}}^{n}|j-k|^{|V_2|-2(|V_2|+|V_3|+|V_5|)\sigma}\right)$$
$$\overset{o,n,\delta}{\ll} \delta\,(n-o)\,. \tag{4.41}$$

**Case 4:** $a_1 \geq 2,\ a_q = 1$, $s$ is even, and $E\left(\psi_1^{(r)}\right) = 0$ whenever $a_r$ is odd.

Under these new conditions, we will show that it is possible to tighten the bound for $E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right]$ in (4.40). We had taken $|V_2| = 1$ to bound $E\left[(Y_{o,n,\delta}^{\lambda_q})^2\right]$ in (4.40) in Case 3, under the assumption that $|V_3| = |V_4| = |V_5| = 0$ and $|V_2|, |V_6| \in \{1, 2, \ldots, q-1\}$.

Further, when $|V_2| = 1$, it means that $\psi_{l_q}^{(q)}$ and $\psi_{m_q}^{(q)}$ are the only two $\psi$'s with zero-mean, and that they must be matched. This gives us that $\nu(q) = q$, $|V_6| = q-1$ and that $V_1 \cup V_2 \cup V_3 \neq \phi$. So, we apply (4.19) and the independence of $\psi$'s with different subscripts, to the definition of $S(V_1, \ldots, V_6, \nu)$ in (4.28), and get that

$$\left|E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\psi_{m_r}^{(r)}\right) - E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\right)E\left(\prod_{r=1}^{q}\psi_{m_r}^{(r)}\right)\right|\mathbf{1}_{\{|V_2|=1,\ |V_6|=q-1\}}$$
$$= \left|E\left(\prod_{r=1}^{q}\psi_{l_r}^{(r)}\psi_{m_r}^{(r)}\right)\right|\mathbf{1}_{\{|V_2|=1,\ |V_6|=q-1\}}$$
$$= \left|\prod_{r=1}^{q-1}E\left(\psi_{l_r}^{(r)}\right)\right|\left|\prod_{r=1}^{q-1}E\left(\psi_{m_r}^{(r)}\right)\right|E\left[\left(\psi_{l_q}^{(q)}\right)^2\right]\,. \tag{4.42}$$

Observe that $(a_1, a_2, \ldots, a_{q-1})$ is a decreasing partition of $(s-1)$, since $a_q = 1$.

Hence if $s$ is even, then $a_r$ must be odd for some $1 \leq r \leq q-1$, and for that $r$, we will get $E\left(\psi_{l_r}^{(r)}\right) = 0$ . Since this makes the entire expression in (4.42) become 0 (thus making $S(V_1, \ldots, V_6, \nu) = 0$), we must not choose $|V_2| = 1$ for the bound of $E\left[\left(Y_{o,n,\delta}^{\lambda_q}\right)^2\right]$ in (4.40). Instead, we go with next lowest value, i.e. $|V_2| = 2$ to obtain,

$$E\left[\left(Y_{o,n,\delta}^{\lambda_q}\right)^2\right] \overset{o,n,\delta}{\ll} (n-o)\ \ l_{2,\sigma}(n-o). \tag{4.43}$$

Lemma 4.11 follows from (4.38,4.39,4.40,4.41) and (4.43). □

### 4.3.3 Rate of Convergence for Theorem 4.6

We now return to the proof of Theorem 4.6, where we shall bound the second moment of $S_n(q, \lambda_q)$ (defined in 4.16). In Lemma 4.11, taking $\psi_{l_r}^{(r)} = \xi_{l_r}^{a_r}$ for $1 \leq r \leq q$, and $\delta = 1$ (since $E\left[\left(\xi_{l_1}^{a_1}\right)^2\right] \overset{o,n}{\ll} 1$), we see that $Y_{o,n,\delta}^{\lambda_q}$ becomes $S_n(q, \lambda_q) - S_o(q, \lambda_q)$, and

$$E\left[\left(S_n(q,\lambda_q) - S_o(q,\lambda_q)\right)^2\right] \overset{o,n}{\ll} \begin{cases} n-o, & a_q \geq 2 \\ (n-o)\ l_{s,\sigma}(n-o), & a_1 = 1 \\ (n-o)\ l_{1,\sigma}(n-o), & a_q = 1,\ a_1 \geq 2. \end{cases} \tag{4.44}$$

Furthermore, if $s$ is even and $\xi_l$ is a symmetric random variable, then $\xi_l^{a_r}$ is symmetric when $a_r$ is odd, implying that $E\left(\xi_l^{a_r}\right) = 0$ for odd $a_r$. Hence, taking $\psi_{l_r}^{(r)} = \xi_{l_r}^{a_r}\ \forall\ 1 \leq r \leq q$ in Lemma 4.11, we see that $E\left(\psi_l^{(r)}\right) = 0$ when $a_r$ is odd, and that for $a_q = 1$ and $a_1 \geq 2$,

$$E\left[\left(S_n(q,\lambda_q) - S_o(q,\lambda_q)\right)^2\right] \overset{o,n,\delta}{\ll} (\delta\ (n-o)) \vee ((n-o)\ l_{2,\sigma}(n-o)). \quad (4.45)$$

The bounds in (4.44) and (4.45) are given in terms of a partition $\lambda_q$. We can check which partitions are possible for a given $s$, and then apply (4.44) and (4.45) to bound the second moment of $\sum_{k=1}^{n}(d_k - d)$. Recall that $s = a_1 + a_2 + \ldots + a_q$ and $a_1 \geq a_2 \geq + \ldots \geq a_q \geq 1$. When $s = 1$, none of the cases except $a_1 = 1$ are possible, and when $s = 2$, the third case i.e. $a_q = 1,\ a_1 \geq 2$ is not possible. Hence, we get from (4.44), that

$$E\left[\left(S_n(q,\lambda_q) - S_o(q,\lambda_q)\right)^2\right] \overset{o,n}{\ll} \begin{cases} (n-o)\ l_{2,\sigma}(n-o), & s = 2 \\ (n-o)\ l_{1,\sigma}(n-o), & s \neq 2 \end{cases}, \quad (4.46)$$

and from (4.45), that if $s$ is even and $\xi_l$ is a symmetric random variable, then

$$E\left[\left(S_n(q,\lambda_q) - S_o(q,\lambda_q)\right)^2\right] \overset{o,n}{\ll} (n-o)\ l_{2,\sigma}(n-o)\ . \quad (4.47)$$

Let $n_r = 2^r$, $n \in [n_r, n_{r+1})$ and $r \in \mathbb{N}_0$. Then, putting $n = n_r$ and $o = 0$ in (4.46), we get,

$$E\left[\left(S_{n_r}(q,\lambda_q)\right)^2\right] \overset{r}{\ll} \begin{cases} n_r\ l_{2,\sigma}(n_r), & s = 2 \\ n_r\ l_{1,\sigma}(n_r), & s \neq 2 \end{cases}. \quad (4.48)$$

First, consider $s \neq 2$. Then for $n_r \leq o < n < n_{r+1}$, it follows from (4.46) by

Theorem 4.5, with $Z_i = S_i(q, \lambda_q) - S_{i-1}(q, \lambda_q)$ and $f(n) = n\ l_{1,\sigma}(n)$, that

$$E\left[\max_{n_r \le o < n < n_{r+1}} \big(S_n(q,\lambda_q) - S_o(q,\lambda_q)\big)^2\right] \overset{r}{\ll} r^2 n_r\ l_{1,\sigma}(n_r)\ . \tag{4.49}$$

Combining (4.48) and (4.49), we have that

$$\sum_{r=0}^{\infty} E\left[\max_{n_r \le n < n_{r+1}} \left(\frac{S_n(q,\lambda_q)}{n^{\frac{1}{p}}}\right)^2\right] \ll \sum_{r=0}^{\infty} r^2 n_r^{1-\frac{2}{p}}\ l_{1,\sigma}(n_r) < \infty, \tag{4.50}$$

provided $(3 - 2\sigma) < \frac{2}{p}$, i.e. $p < \frac{2}{3-2\sigma}$. From (4.50), it follows by Fubini's Theorem and $n^{\text{th}}$ term divergence that for $p < \frac{2}{3-2\sigma}$,

$$\lim_{n\to\infty} \frac{S_n(q,\lambda_q)}{n^{\frac{1}{p}}} = 0 \quad a.s. \tag{4.51}$$

Now let $s = 2$. Then for $n_r \le o < n < n_{r+1}$, it follows from (4.46) by Theorem 4.5, with $Z_i = S_i(q, \lambda_q) - S_{i-1}(q, \lambda_q)$ and $f(n) = n\ l_{2,\sigma}(n)$, that

$$E\left[\max_{n_r \le o < n < n_{r+1}} \big(S_n(q,\lambda_q) - S_o(q,\lambda_q)\big)^2\right] \overset{r}{\ll} r^2 n_r\ l_{2,\sigma}(n_r)\ . \tag{4.52}$$

Combining (4.48) and (4.52), we have that

$$\sum_{r=0}^{\infty} E\left[\max_{n_r \le n < n_{r+1}} \left(\frac{S_n(q,\lambda_q)}{n^{\frac{1}{p}}}\right)^2\right] \ll \sum_{r=0}^{\infty} r^2 n_r^{1-\frac{2}{p}}\ l_{2,\sigma}(n_r) < \infty, \tag{4.53}$$

provided $(4 - 4\sigma) \vee 1 < \frac{2}{p}$, i.e. $p < \frac{1}{2-2\sigma}$ when $\sigma < \frac{3}{4}$, and $p < 2$ when $\sigma \ge \frac{3}{4}$. From (4.53), it follows by Fubini's Theorem and $n^{\text{th}}$ term divergence that for

$p < 2 \wedge \frac{1}{2-2\sigma}$,

$$\lim_{n\to\infty} \frac{S_n(q,\lambda_q)}{n^{\frac{1}{p}}} = 0 \quad a.s. \tag{4.54}$$

Finally, we consider the case where s is even, and $\xi_l$ is a symmetric random variable. Then, notice that our result in (4.47) is the same as that in (4.46) for $s = 2$. Thus, (4.52) and (4.53) holds for this case as well, and for $p < 2 \wedge \frac{1}{2-2\sigma}$, we get that

$$\lim_{n\to\infty} \frac{S_n(q,\lambda_q)}{n^{\frac{1}{p}}} = 0 \quad a.s. \tag{4.55}$$

Since $\sum_{k=1}^{n}(d_k - d)$ is the sum of $S_n(q, \lambda_q)$ over all $q \in \{1, \ldots, s\}$ and all partitions $\lambda_q$ (which are finite in number), we get from (4.51,4.54) and (4.55), that

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n}(d_k - d)}{n^{\frac{1}{p}}} = 0 \quad a.s.$$

for the values of $p$ as mentioned in (4.14) and (4.15). This completes the proof of Theorem 4.6. □

## 4.4 Heavy Tailed Case of Theorem 3.1

From (3.2) in Theorem 3.1, we find that heavy tails can only arise when $0 \le i \le \lfloor \frac{s-1}{2} \rfloor$, i.e. for products of at least $s - \lfloor \frac{s-1}{2} \rfloor = \lceil \frac{s+1}{2} \rceil$ terms. Also when $s = 1$, (3.1) along with Remark 3.4 ensures that it will have only long range dependence since $\alpha_0 > 2$, so we do not have to consider that case here. Since we will deal only with those terms exhibiting heavy tails in this section, we will assume that $s \geq 2$, fix $i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}$, and assume (due to Remark 3.4) that $1 < \alpha_i < 2$.

**Remark 4.12.** For a given partition $\lambda_q = \{a_1, a_2, \ldots, a_q\}$, heavy tails can only come up in the innovation involving the highest power, i.e. $\xi_l^{a_1}$. This is because for a term to possess heavy tails, it's variance must be infinite, hence $a_1 > \frac{s}{2}$. But that would force the rest of the $a_r$'s to be less than $\frac{s}{2}$, thus precluding heavy tails in terms involving $\xi_l^{a_r}$ for $r \in \{2, \ldots, q\}$. This shows that heavy tails concerning $\alpha_i$ will arise only in the sum

$$S_n^*(i) = \sum_{k=1}^{n} \sum_{\substack{l_1, l_2, \ldots, l_{i+1} \\ l_1 \notin \{l_2, \ldots, l_{i+1}\}}} \left( c_{k-l_1}^{s-i} \prod_{r=2}^{i+1} c_{k-l_r} \right) \left( \xi_{l_1}^{s-i} \prod_{r=2}^{i+1} \xi_{l_r} - E\left( \xi_{l_1}^{s-i} \prod_{r=2}^{i+1} \xi_{l_r} \right) \right). \quad (4.56)$$

**Remark 4.13.** Alternatively, for heavy tails involving $\alpha_i$, we could also consider the sum $S_n(q, \lambda_q)$ (from (4.16)) with $a_1 = s - i$, i.e.

$$S_n(q, \lambda_q) = \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left( c_{k-l_1}^{s-i} \prod_{r=2}^{q} c_{k-l_r}^{a_r} \right) \left( \xi_{l_1}^{s-i} \prod_{r=2}^{q} \xi_{l_r}^{a_r} - E\left( \xi_{l_1}^{s-i} \prod_{r=2}^{q} \xi_{l_r}^{a_r} \right) \right),$$

where $\lambda_q = (s - i, a_2, \ldots, a_q)$. In fact, note that $S_n^*$ (from (4.56)) is the sum of $S_n(q, \lambda_q)$ over all $q$, and all partitions $\lambda_q$ with $a_1 = s - i$. Both $S_n^*(i)$ and $S_n(q, \lambda_q)$ have distinct advantages to work with. While $S_n^*(i)$ has the advantage of having only one $\xi_l$ with power greater than one, $S_n(q, \lambda_q)$ has the advantage of having $l_1 \neq l_2 \neq \ldots \neq l_q$, hence Lemma 4.10 can be easily applied to it. Hence, we will mostly use $S_n(q, \lambda_q)$ to deal with the truncated terms, and $S_n^*(i)$

for the error terms.

### 4.4.1 Conversion to continuous random variables

Recall that in this section, $i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor \right\}$ is fixed. We will first replace $\xi_l^{s-i}$ with continuous random variables $\zeta_l$, which will ensure below that the truncation does not take place at a jump or at a point with positive probability. Let $\{U_l\}_{l \in \mathbb{Z}}$ be independent $[-1, 1]$-uniform random variables that are independent of $\{\xi_l\}_{l \in \mathbb{Z}}$. Then, we have that

$$S_n(q, \lambda_q) = A_n - B_n \ ,$$

where we define,

$$A_n(q, \lambda_q) = \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left( \prod_{r=1}^{q} c_{k-l_r}^{a_r} \right) \left( \left( \xi_{l_1}^{s-i} + U_{l_1} \right) \prod_{r=2}^{q} \xi_{l_r}^{a_r} - E\left( \left( \xi_{l_1}^{s-i} + U_{l_1} \right) \prod_{r=2}^{q} \xi_{l_r}^{a_r} \right) \right)$$

$$B_n(q, \lambda_q) = \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left( \prod_{r=1}^{q} c_{k-l_r}^{a_r} \right) \left( U_{l_1} \prod_{r=2}^{q} \xi_{l_r}^{a_r} - E\left( U_{l_1} \prod_{r=2}^{q} \xi_{l_r}^{a_r} \right) \right) \ .$$

In Lemma 4.11, taking $\psi_{l_r}^{(r)} = \xi_{l_r}^{a_r} \ \ \forall\ 2 \leq r \leq q$, $\psi_{l_1}^{(1)} = U_{l_1}$, and $\delta = 1$ (since $E\left[ (U_{l_1})^2 \right]$ is constant), we get that $Y_{o,n,\delta}^{\lambda_q} = B_n - B_o$. This gives us,

$$E\left[ \left( B_n(q, \lambda_q) - B_o(q, \lambda_q) \right)^2 \right] \overset{o,n}{\ll} \begin{cases} n - o, & a_q \geq 2, \\ (n-o)\ l_{s,\sigma}(n-o), & a_1 = 1, \\ (n-o)\ l_{1,\sigma}(n-o), & a_q = 1,\ a_1 \geq 2. \end{cases} \qquad (4.57)$$

This bound is the same as that in (4.44), which is expected, since we can see that heavy tails do not arise in $B_n(q, \lambda_q)$. Like in (4.46), we check which partitions are possible for a given $s$. When $s = 1$, none of the cases except $a_1 = 1$ are possible, and when $s = 2$, the third case i.e. $a_q = 1,\ a_1 \geq 2$ is not possible. Hence from (4.57), we get that

$$E\left[(B_n(q,\lambda_q) - B_o(q,\lambda_q))^2\right] \quad \overset{o,n}{\ll} \quad \begin{cases} (n-o)\ l_{2,\sigma}(n-o), & s = 2 \\ (n-o)\ l_{1,\sigma}(n-o), & s \neq 2 \end{cases}.$$

Proceeding along the lines of (4.48 - 4.55), with $B_n(q, \lambda_q) - B_o(q, \lambda_q)$ instead of $S_n(q, \lambda_q) - S_o(q, \lambda_q)$, we get that

$$\lim_{n\to\infty} \frac{B_n(q,\lambda_q)}{n^{\frac{1}{p}}} \ = \ 0 \qquad \text{a.s.}$$

for the values of $p$ as mentioned in the statement of Theorem 4.6. Defining $\zeta_l = \xi_l^{s-i} + U_l$, which is a function of $i$, we note that $\zeta_l$ is a continuous random variable since it is a convolution of two random variables, one of which is absolutely continuous. Also, note that $\zeta_l$ has the same tail probability bound as $\xi_l^{s-i}$, since

$$\begin{aligned} \sup_{t\geq 2} t^{\alpha_i} P(\zeta_1 > t) \quad &\leq \quad \sup_{t\geq 2} t^{\alpha_i} P(|\zeta_1| > t) \\ &\leq \quad \sup_{t\geq 2} t^{\alpha_i} P\left(\left|\xi_1^{s-i}\right| > t-1\right) \\ &\ll \quad \sup_{t\geq 1} \left(\frac{t+1}{t}\right)^{\alpha_i} t^{\alpha_i} P\left(\left|\xi_1^{s-i}\right| > t\right) < \infty. \quad (4.58) \end{aligned}$$

Thus, convergence of $S_n(q, \lambda_q)$ is equivalent to that of

$$A_n(q, \lambda_q) \;=\; \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left( \prod_{r=1}^{q} c_{k-l_r}^{a_r} \right) \left( \zeta_{l_1} \prod_{r=2}^{q} \xi_{l_r}^{a_r} - E\left( \zeta_{l_1} \prod_{r=2}^{q} \xi_{l_r}^{a_r} \right) \right).$$

Summing over all $q$, and partitions $\lambda_q$ where $a_1 = s-i$, we find that convergence of $S_n^*(i)$ (from (4.56)) is equivalent to that of,

$$T_n(i) \;=\; \sum_{k=1}^{n} \sum_{\substack{l_1, l_2, \ldots, l_{i+1} \\ l_1 \notin \{l_2, \ldots, l_{i+1}\}}} \left( c_{k-l_1}^{s-i} \prod_{r=2}^{i+1} c_{k-l_r} \right) \left( \zeta_{l_1} \prod_{r=2}^{i+1} \xi_{l_r} - E\left( \zeta_{l_1} \prod_{r=2}^{i+1} \xi_{l_r} \right) \right). \quad (4.59)$$

### 4.4.2 Truncation of $\zeta$ with highest power

We now break $\zeta$ into truncated and error terms. This partitioning will be done in such a way that the second moment of the truncated term is finite, hence can be managed by Theorem 4.6. The convergence of the error terms will be proven later on using Jensen's Inequality, Holder's Inequality, Doob's $L_p$ Maximal Inequality and Borel-Cantelli Lemma.

Let $\kappa > 0$. Using condition (4.58), and fixing $v_r^+ = n_r^{\frac{\kappa}{2-\alpha_i}}$ (where $n_r = 2^r$) for $r \in \mathbb{N}_0$, and letting $v_r^- = -v_r^+$, we get

$$\begin{cases} 2\int_0^{v_r^+} P(\zeta_1 > s) s \, ds \;\overset{r}{\ll}\; 2\int_0^{v_r^+} s^{-\alpha_i} s \, ds \;\overset{r}{\ll}\; n_r^{\kappa} \\ 2\int_{v_r^-}^{0} P(\zeta_1 < s) s \, ds \;\overset{r}{\ll}\; 2\int_{v_r^-}^{0} s^{-\alpha_i} s \, ds \;\overset{r}{\ll}\; n_r^{\kappa}, \end{cases} \qquad \forall\, r \in \mathbb{N}\,. \quad (4.60)$$

Next, we define i.i.d random variables $\{\overline{\zeta}_l^{(r)}\}_{l\in\mathbb{Z}}$ and $\{\tilde{\zeta}_l^{(r)}\}_{l\in\mathbb{Z}}$ for $r \in \mathbb{N}$ such that,

$$\begin{cases} \overline{\zeta}_l^{(r)} &= v_r^- \vee \zeta_l \wedge v_r^+ \\ \tilde{\zeta}_l^{(r)} &= \zeta_l - \overline{\zeta}_l^r \end{cases} . \tag{4.61}$$

We shall call $\overline{\zeta}_l^{(r)}$ the truncated terms, and $\tilde{\zeta}_l^{(r)}$ the error terms. Observe that $\overline{\zeta}_l^{(r)}$ and $\tilde{\zeta}_l^{(r)}$ are both functions of $r$. Breaking $\zeta_l^{(r)}$ into $\overline{\zeta}_l^{(r)}$ and $\tilde{\zeta}_l^{(r)}$ also helps us break up $A_n(q,\lambda_q)$ as $A_n(q,\lambda_q) = \overline{A}_n^{(r)}(q,\lambda_q) + \tilde{A}_n^{(r)}(q,\lambda_q)$, where

$$\begin{aligned} \overline{A}_n^{(r)}(q,\lambda_q) &= \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left(\prod_{r=1}^{q} c_{k-l_r}^{a_r}\right)\left(\overline{\zeta}_{l_1}^{(r)} \prod_{r=2}^{q} \xi_{l_r}^{a_r} - E\left(\overline{\zeta}_{l_1}^{(r)} \prod_{r=2}^{q} \xi_{l_r}^{a_r}\right)\right) \\ \tilde{A}_n^{(r)}(q,\lambda_q) &= \sum_{k=1}^{n} \sum_{l_1 \neq l_2 \neq \ldots \neq l_q} \left(\prod_{r=1}^{q} c_{k-l_r}^{a_r}\right)\left(\tilde{\zeta}_{l_1}^{(r)} \prod_{r=2}^{q} \xi_{l_r}^{a_r} - E\left(\tilde{\zeta}_{l_1}^{(r)} \prod_{r=2}^{q} \xi_{l_r}^{a_r}\right)\right), \end{aligned}$$

and $T_n(i)$ (from (4.59)) as $T_n(i) = \overline{T}_n^{(r)}(i) + \tilde{T}_n^{(r)}(i)$, where

$$\begin{aligned} \overline{T}_n^{(r)}(i) &= \sum_{k=1}^{n} \sum_{\substack{l_1,l_2,\ldots,l_{i+1} \\ l_1 \notin \{l_2,\ldots,l_{i+1}\}}} \left(c_{k-l_1}^{s-i} \prod_{r=2}^{i+1} c_{k-l_r}\right)\left(\overline{\zeta}_{l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{l_r} - E\left(\overline{\zeta}_{l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{l_r}\right)\right) \\ \tilde{T}_n^{(r)}(i) &= \sum_{k=1}^{n} \sum_{\substack{l_1,l_2,\ldots,l_{i+1} \\ l_1 \notin \{l_2,\ldots,l_{i+1}\}}} \left(c_{k-l_1}^{s-i} \prod_{r=2}^{i+1} c_{k-l_r}\right)\left(\tilde{\zeta}_{l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{l_r} - E\left(\tilde{\zeta}_{l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{l_r}\right)\right). \end{aligned}$$

### 4.4.3 Bounding second moment of truncated terms

Recall that $\zeta_l$, $\overline{\zeta}_l^{(r)}$, $\tilde{\zeta}_l^{(r)}$, $A_n(q,\lambda_q)$, $\overline{A}_n^{(r)}(q,\lambda_q)$, $\tilde{A}_n^{(r)}(q,\lambda_q)$, $T_n(i)$, $\overline{T}_n^{(r)}(i)$, and $\tilde{T}_n^{(r)}(i)$ are defined in terms of a fixed $i \in \left\{0,1,\ldots,\left\lfloor \frac{s-1}{2} \right\rfloor\right\}$. We now bound the second moments for the truncated terms, $\overline{\zeta}_l^{(r)}$.

Using (4.58,4.61), and the formula

$$E[g(X)] = \int_0^\infty g'(t)P(X > t)\ dt \quad - \quad \int_{-\infty}^0 g'(t)P(X < t)\ dt, \tag{4.62}$$

for a continuously differentiable function $g$ and a random variable $X$, we find that,

$$\begin{aligned} E[\overline{\zeta}_l^{(r)}] &= \int_0^{v_r^+} P(\zeta_l > t)\ dt \quad - \quad \int_{v_r^-}^0 P(\zeta_l < t)\ dt \\ &\leq \int_0^\infty P(|\zeta_l| > t)\ dt \\ &\leq E|\zeta_l| \overset{r}{\ll} 1. \end{aligned} \tag{4.63}$$

Also, by (4.60) we have,

$$\begin{aligned} E\left[\left|\overline{\zeta}_l^{(r)}\right|^2\right] &= E\left[|v_r^- \vee \zeta_l \wedge v_r^+|^2\right] \\ &= 2\int_0^{v_r^+} P(\zeta_l > s)s\ ds \quad - \quad 2\int_{v_r^-}^0 P(\zeta_l < s)s\ ds \\ &\overset{r}{\ll} n_r^\kappa, \qquad \forall\ r \in \mathbb{N}. \end{aligned} \tag{4.64}$$

We shall now use (4.63) and (4.64) to bound the second moment of $\overline{A}_n^{(r)}(q, \lambda_q)$, in terms of $n_r^\kappa$. Recall that $\{\overline{\zeta}_l^{(r)}\}$ are i.i.d., and $E\left[\left|\overline{\zeta}_l^{(r)}\right|\right] < \infty$. Hence, taking $\psi_{l_1}^{(1)} = \overline{\zeta}_{l_1}^{(r)}$, $\psi_{l_r}^{(r)} = \xi_{l_r}^{a_r}$ for $2 \leq r \leq q$, and $\delta = n_r^\kappa$ in Lemma 4.11, we see that $Y_{o,n,r}$ becomes $\overline{A}_n^{(r)}(q, \lambda_q) - \overline{A}_o^{(r)}(q, \lambda_q)$. Taking $o = 0$ in Lemma 4.11, we get that

$$E\left[\left(\overline{A}_n^{(r)}(q,\lambda_q)\right)^2\right] \overset{n,r}{\ll} \begin{cases} n_r^{\kappa} n, & a_q \geq 2 \\ n_r^{\kappa} n \; l_{s,\sigma}(n), & a_1 = 1 \\ n_r^{\kappa} n \; \vee \left(n \; l_{1,\sigma}(n)\right), & a_q = 1, \; a_1 \geq 2 \end{cases} .$$

Note that when $a_1 = 1$, heavy tails do not arise since second moment of $\overline{A}_n^{(r)}(q,\lambda_q)$ will exist, hence we can discard this case. When $s = 2$, the third case i.e. $a_q = 1, \; a_1 \geq 2$ is not possible. Hence, we have

$$E\left[\left(\overline{A}_n^{(r)}(q,\lambda_q)\right)^2\right] \overset{n,r}{\ll} \begin{cases} n_r^{\kappa} n, & s = 2 \\ n_r^{\kappa} n \; \vee \left(n \; l_{1,\sigma}(n)\right), & s \neq 2 \end{cases} . \tag{4.65}$$

Let $s \neq 2$. Then for $n_r \leq n < n_{r+1}$, it follows from (4.65) by Theorem 4.5, with $Z_i = \overline{A}_i^{(r)}(q,\lambda_q) - \overline{A}_{i-1}^{(r)}(q,\lambda_q)$ and $f(n) = n_r^{\kappa} n \; \vee \left(n \; l_{1,\sigma}(n)\right)$, that

$$E\left[\max_{n_r \leq n < n_{r+1}} \left(\overline{A}_n^{(r)}(q,\lambda_q)\right)^2\right] \overset{r}{\ll} r^2 \left[n_r^{1+\kappa} \vee \; \left(n_r \; l_{1,\sigma}(n_r)\right)\right] .$$

Summing (4.66) over all $q$ and over all partitions $\lambda_q$ where $a_1 = s - i$ (recall that $i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}$ is fixed), gives us that

$$E\left[\max_{n_r \leq n < n_{r+1}} \left(\overline{T}_n^{(r)}(i)\right)^2\right] \overset{r}{\ll} r^2 \left[n_r^{1+\kappa} \vee \; \left(n_r \; l_{1,\sigma}(n_r)\right)\right] . \tag{4.66}$$

Now, let $s = 2$. Then for $n_r \le n < n_{r+1}$, it follows from (4.65) by Theorem 4.5, with $Z_i = \overline{A}_i^{(r)}(q, \lambda_q) - \overline{A}_{i-1}^{(r)}(q, \lambda_q)$ and $f(n) = n_r^\kappa n$, that

$$E\left[\max_{n_r \le n < n_{r+1}} \left(\overline{A}_n^{(r)}(q, \lambda_q)\right)^2\right] \overset{r}{\ll} r^2 n_r^{1+\kappa}. \tag{4.67}$$

Summing (4.67) over all $q$ and over all partitions $\lambda_q$ where $a_1 = s - i$, gives us that

$$E\left[\max_{n_r \le n < n_{r+1}} \left(\overline{T}_n^{(r)}(i)\right)^2\right] \overset{r}{\ll} r^2 n_r^{1+\kappa} . \tag{4.68}$$

Finally, we consider the situation where s is even, and $\xi_l$ is symmetric. Clearly $\xi_l^{a_j}$ will be symmetric when $a_j$ is odd, implying that $E\left(\xi_l^{a_j}\right) = 0$ for odd $a_j$, $2 \le j \le q$. Also, since $a_1 = s - i$, we see that $\xi_l^{a_1}$ will be symmetric when $a_1$ is odd, implying that both $\zeta_l$ and $\overline{\zeta}_l^{(r)}$ will be symmetric. Hence, taking $\psi_{l_j}^{(j)} = \xi_{l_j}^{a_j}$ $\forall\, 2 \le j \le q$, $\psi_{l_1}^{(1)} = \overline{\zeta}_l^{(r)}$, $\delta = n_r^\kappa$, and $o = 0$ in Lemma 4.11, we see that $E\left(\psi_l^{(j)}\right) = 0$ when $a_j$ is odd, and that

$$E\left[\left(\overline{A}_n^{(r)}(q, \lambda_q)\right)^2\right] \overset{n,r}{\ll} n_r^\kappa n \vee \big(n\, l_{2,\sigma}(n)\big). \tag{4.69}$$

Then for $n_r \le n < n_{r+1}$, it follows from (4.69) by Theorem 4.5, with $Z_i = \overline{A}_i^{(r)}(q, \lambda_q) - \overline{A}_{i-1}^{(r)}(q, \lambda_q)$ and $f(n) = n_r^\kappa n \vee \big(n\, l_{2,\sigma}(n)\big)$, that

$$E\left[\max_{n_r \le n < n_{r+1}} \left(\overline{A}_n^{(r)}(q, \lambda_q)\right)^2\right] \overset{r}{\ll} r^2\left[n_r^{1+\kappa} \vee \big(n_r\, l_{2,\sigma}(n_r)\big)\right] .$$

Thus, summing (4.70) over all $q$ and all partitions $\lambda_q$ where $a_1 = s - i$, gives us

$$E\left[\max_{n_r \le n < n_{r+1}} \left(\overline{T}_n^{(r)}(i)\right)^2\right] \overset{r}{\ll} r^2\Big[n_r^{1+\kappa} \vee \left(n_r\ l_{2,\sigma}(n_r)\right)\Big] \,, \tag{4.70}$$

which can be seen as an improvement over (4.66), since the function $l_{2,\sigma} \le l_{1,\sigma}$.

### 4.4.4 Bounding $\tau$th moment of error terms, $\tau \in (1, \alpha_i)$

We now bound the second moments for the error terms.

Taking $1 < z < \alpha_i$, and using our tail probability bound in (4.58) along with (4.62), we have that

$$\begin{aligned}
E\left|\left(\tilde{\zeta}_1^{(r)}\right)^+\right|^z &= z\int_0^\infty s^{z-1}P\left(\zeta_1^{(r)} - (\zeta_1^{(r)} \wedge v_r^+) > s\right)\ ds \\
&= z\int_0^\infty s^{z-1}P\left(\zeta_1^{(r)} > v_r^+ + s\right)\ ds \\
&\overset{r}{\ll} \int_{v_r^+}^\infty (s - v_r^+)^{z-1}s^{-\alpha_i}\ ds \\
&\le (v_r^+)^{-\alpha_i}\int_{v_r^+}^{2v_r^+}(s - v_r^+)^{z-1}\ ds \ + \ \int_{2v_r^+}^\infty (s - v_r^+)^{z-\alpha_i-1}\ ds \\
&\overset{r}{\ll} (v_r^+)^{z-\alpha_i} \overset{r}{\ll} n_r^{\frac{\kappa(z-\alpha_i)}{2-\alpha_i}}.
\end{aligned}$$

By symmetry $E\left|\left(\tilde{\zeta}_1^{(r)}\right)^-\right|^z$ has the same bound, hence for $1 < z < \alpha_i$, we get that

$$\|\tilde{\zeta}_1^{(r)}\|_z \overset{r}{\ll} n_r^{\frac{\kappa(z-\alpha_i)}{z(2-\alpha_i)}}. \tag{4.71}$$

Now we shall find conditions for the convergence rates of $\tilde{T}_n^{(r)}(i)$. If we take,

$$X_n \;=\; \sum_{k=1}^{n} \sum_{\substack{l_1,l_2,\ldots,l_{i+1} \\ l_1\notin\{l_2,\ldots,l_{i+1}\}}} \left( c_{k-l_1}^{s-i} \prod_{r=2}^{i+1} c_{k-l_r} \right) \left( \tilde{\zeta}_{l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{l_r} \right)$$

in Lemma 4.3, we have for $\tau \in (1,2)$ that,

$$\begin{aligned}
& E^{\frac{1}{\tau}}\left[ \sup_{n_r\le n<n_{r+1}} \left| \tilde{T}_n^{(r)}(i) \right|^{\tau} \right] \\
&= E^{\frac{1}{\tau}}\left[ \sup_{n_r\le n<n_{r+1}} \left| \sum_{k=1}^{n} \sum_{\substack{l_1,l_2,\ldots,l_{i+1} \\ l_1\notin\{l_2,\ldots,l_{i+1}\}}} \left( c_{k-l_1}^{s-i} \prod_{r=2}^{i+1} c_{k-l_r} \right)\left( \tilde{\zeta}_{l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{l_r} - E\left( \tilde{\zeta}_{l_1} \prod_{r=2}^{i+1} \xi_{l_r} \right)\right) \right|^{\tau} \right] \\
&\overset{r}{\ll}\; E^{\frac{1}{\tau}}\left[ \sup_{n_r\le n<n_{r+1}} \left| \sum_{k=1}^{n} \sum_{\substack{l_1,l_2,\ldots,l_{i+1} \\ l_1\notin\{l_2,\ldots,l_{i+1}\}}} \left( c_{l_1}^{s-i} \prod_{r=2}^{i+1} c_{l_r} \right) \left( \tilde{\zeta}_{k-l_1}^{(r)} \prod_{r=2}^{i+1} \xi_{k-l_r} \right) \right|^{\tau} \right] \\
&\le\; E^{\frac{1}{\tau}}\left[ \sup_{n_r\le n<n_{r+1}} \left| \sum_{k=1}^{n} \sum_{l_1=-\infty}^{\infty} \left| c_{l_1}^{s-i} \tilde{\zeta}_{k-l_1}^{(r)} \right| \left| \sum_{\substack{l=-\infty \\ l\neq l_1}}^{\infty} c_l \xi_{k-l} \right|^{i} \right|^{\tau} \right]. \qquad (4.72)
\end{aligned}$$

Define,

$$\phi_{k,l_1} \;=\; \left| \sum_{\substack{l=-\infty \\ l\neq l_1}}^{\infty} c_l \xi_{k-l} \right|^{i}. \qquad (4.73)$$

Noting that $\sum_{m=-\infty}^{\infty} |c_m^{s-i}| < \infty$ because $s - i \geq 2$, and then using Jensen's inequality (since norms are convex), we have from (4.72), that

$$
\begin{aligned}
& E^{\frac{1}{\tau}}\Bigg[\sup_{n_r \leq n < n_{r+1}} \left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\Bigg] \\
\overset{r}{\ll}\ & E^{\frac{1}{\tau}}\left[\ \left|\sum_{l_1=-\infty}^{\infty} \left|c_{l_1}^{s-i}\right| \sup_{n_r \leq n < n_{r+1}} \left(\sum_{k=1}^{n} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right| |\phi_{k,l_1}|\right)\right|^{\tau}\ \right] \\
=\ & \sum_{m=-\infty}^{\infty} |c_m^{s-i}|\ \ E^{\frac{1}{\tau}}\left[\ \left|\sum_{l_1=-\infty}^{\infty} \frac{|c_{l_1}^{s-i}|}{\sum_m |c_m^{s-i}|} \sup_{n_r \leq n < n_{r+1}} \left(\sum_{k=1}^{n} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right| |\phi_{k,l_1}|\right)\right|^{\tau}\ \right] \\
\leq\ & \sum_{l_1=-\infty}^{\infty} |c_{l_1}^{s-i}|\ \ E^{\frac{1}{\tau}}\left[\sup_{n_r \leq n < n_{r+1}} \left|\sum_{k=1}^{n} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right| |\phi_{k,l_1}|\ \right|^{\tau}\ \right]. \qquad (4.74)
\end{aligned}
$$

First, we consider that $i \geq 1$. Then, by two applications of Holder's inequality with $p_1 = \frac{s}{s-\tau i}$ and $p_2 = \frac{s}{\tau i}$ (both of which are positive, and their reciprocals sum to one), we get from (4.74), that

$$
\begin{aligned}
& E^{\frac{1}{\tau}}\Bigg[\sup_{n_r \leq n < n_{r+1}} \left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\Bigg] \\
\overset{r}{\ll}\ & \sum_{l_1=-\infty}^{\infty} |c_{l_1}^{s-i}|\ E^{\frac{1}{\tau}}\left[\sup_{n_r \leq n < n_{r+1}} \left|\sum_{k=1}^{n} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right|^{\frac{s}{s-\tau i}}\right|^{\frac{\tau(s-\tau i)}{s}} \left|\sum_{j=1}^{n} |\phi_{j,l_1}|^{\frac{s}{\tau i}}\right|^{\frac{\tau^2 i}{s}}\right] \\
\overset{r}{\ll}\ & \sum_{l_1=-\infty}^{\infty} |c_{l_1}^{s-i}|\ E^{\frac{s-\tau i}{s\tau}}\left[\sup_{n_r \leq n < n_{r+1}} \left|\sum_{k=1}^{n} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right|^{\frac{s}{s-\tau i}}\right|^{\tau}\right] \\
& \qquad\qquad E^{\frac{i}{s}}\left[\sup_{n_r \leq n < n_{r+1}} \left|\sum_{j=1}^{n} |\phi_{j,l_1}|^{\frac{s}{\tau i}}\right|^{\tau}\right]. \qquad (4.75)
\end{aligned}
$$

Since $\frac{s}{s-\tau i}$ and $\frac{s}{\tau i}$ are positive, we find that both $\sum_{k=1}^{n} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right|^{\frac{s}{s-\tau i}}$ and $\sum_{j=1}^{n} |\phi_{j,l_1}|^{\frac{s}{\tau i}}$ are non-negative submartingales, due to Example 4 in page 475 of Shiryaev [33]. Hence, using Doob's $L_p$ maximal inequality (Theorem 2, page 493 of Shiryaev [33]) and then Jensen's inequality (since $\tau > 1$), we get from (4.75), that

$$\begin{aligned}
& E^{\frac{1}{\tau}}\Big[\sup_{n_r \le n < n_{r+1}} \Big|\tilde{T}_n^{(r)}(i)\Big|^{\tau}\Big] \\
\overset{r}{\ll} & \sum_{l_1=-\infty}^{\infty} |c_{l_1}^{s-i}|\; E^{\frac{s-\tau i}{s\tau}} \left[\left|\sum_{k=1}^{n_{r+1}-1} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right|^{\frac{s}{s-\tau i}}\right|^{\tau}\right] E^{\frac{i}{s}} \left[\left|\sum_{j=1}^{n_{r+1}-1} |\phi_{j,l_1}|^{\frac{s}{\tau i}}\right|^{\tau}\right] \\
\overset{r}{\ll} & \sum_{l_1=-\infty}^{\infty} |c_{l_1}^{s-i}|\; E^{\frac{s-\tau i}{s\tau}} \left[(n_{r+1}-1)^{\tau-1} \sum_{k=1}^{n_{r+1}-1} \left|\tilde{\zeta}_{k-l_1}^{(r)}\right|^{\frac{s\tau}{s-\tau i}}\right] \\
& \qquad E^{\frac{i}{s}} \left[(n_{r+1}-1)^{\tau-1} \sum_{j=1}^{n_{r+1}-1} |\phi_{j,l_1}|^{\frac{s}{i}}\right]. \qquad (4.76)
\end{aligned}$$

We now claim that $\|\phi_{1,l_1}\|_{\frac{s}{i}} < \infty$. From (4.73), using Triangle inequality and the fact that $\{\xi_l\}_{l\in\mathbb{Z}}$ are i.i.d., we see that

$$\begin{aligned}
\left[\|\phi_{1,l_1}\|_{\frac{s}{i}}\right]^{\frac{1}{i}} = \left\|\sum_{\substack{l=-\infty \\ l\neq l_1}}^{\infty} c_l \xi_{1-l}\right\|_s & \le \left\|\sum_{l=-\infty}^{\infty} c_l \xi_{1-l}\right\|_s + \|c_{l_1}\xi_{k-l_1}\|_s \\
& \le \left\|\sum_{l=-\infty}^{\infty} c_l \xi_{1-l}\right\|_s + c_{l_1}\|\xi_1\|_s \qquad (4.77)
\end{aligned}$$

Using Lemma 4.4, along with the facts that $c_{l_1} \ll 1$ and $E\Big[|\xi_1|^s\Big] < \infty$ (from (3.1)), we get from (4.77) that $\|\phi_{1,l_1}\|_{\frac{s}{i}} < \infty$. Since $s - i \ge 2$, $\{\tilde{\zeta}_l^{(r)}\}_{l\in\mathbb{Z}}$ are i.i.d., as are $\{\phi_{j,l_1}\}_{j\in\mathbb{N}}$, we get from (4.76), that

$$\begin{aligned}
& E^{\frac{1}{\tau}}\Bigg[\sup_{n_r\le n<n_{r+1}}\left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\Bigg] \\
\stackrel{r}{\ll} \;& \sum_{l_1=-\infty}^{\infty}|c_{l_1}^{s-i}|\; E^{\frac{s-\tau i}{s\tau}}\left[(n_{r+1}-1)^{\tau}\left|\tilde{\zeta}_1^{(r)}\right|^{\frac{s\tau}{s-\tau i}}\right] E^{\frac{i}{s}}\left[(n_{r+1}-1)^{\tau}\left|\phi_{1,l_1}\right|^{\frac{s}{i}}\right] \\
\stackrel{r}{\ll} \;& \sum_{l_1=-\infty}^{\infty}|c_{l_1}^{s-i}|\; n_r\left\|\tilde{\zeta}_1^{(r)}\right\|_{\frac{s\tau}{s-\tau i}}\;\|\phi_{1,l_1}\|_{\frac{s}{i}} \\
\stackrel{r}{\ll} \;& n_r\|\tilde{\zeta}_1^{(r)}\|_{\frac{s\tau}{s-\tau i}}\;. \qquad (4.78)
\end{aligned}$$

Recall that after (4.74), we had assumed that $i \geq 1$. Now, if we consider $i = 0$, we get that $|\phi_{k,l_1}| = 1$, and from (4.74), we get that

$$E^{\frac{1}{\tau}}\Bigg[\sup_{n_r\le n<n_{r+1}}\left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\Bigg] \stackrel{r}{\ll} \sum_{l_1=-\infty}^{\infty}|c_{l_1}^{s-i}|\;\; E^{\frac{1}{\tau}}\left[\sup_{n_r\le n<n_{r+1}}\left|\sum_{k=1}^{n}\left|\tilde{\zeta}_{k-l_1}^{(r)}\right|\right|^{\tau}\right].$$

We see that $\sum_{k=1}^{n}\left|\tilde{\zeta}_{k-l_1}^{(r)}\right|$ is a non-negative submartingale, due to Example 4 in page 475 of Shiryaev [33]. Thus, using Doob's $L_p$ maximal inequality (Theorem 2, page 493 of Shiryaev [33]), Jensen's inequality (since $\tau > 1$), and the fact that $\{\tilde{\zeta}_l^{(r)}\}_{l\in\mathbb{Z}}$ are i.i.d., we get that

$$\begin{aligned}
E^{\frac{1}{\tau}}\left[\sup_{n_r\le n<n_{r+1}}\left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\right] \;&\stackrel{r}{\ll}\; \sum_{l_1=-\infty}^{\infty}|c_{l_1}^{s-i}|\; E^{\frac{1}{\tau}}\left[\left|\sum_{k=1}^{n_{r+1}-1}\left|\tilde{\zeta}_{k-l_1}^{(r)}\right|\right|^{\tau}\right] \\
&\stackrel{r}{\ll}\; \sum_{l_1=-\infty}^{\infty}|c_{l_1}^{s-i}|\; E^{\frac{1}{\tau}}\left[(n_{r+1}-1)^{\tau-1}\sum_{k=1}^{n_{r+1}-1}\left|\tilde{\zeta}_{k-l_1}^{(r)}\right|^{\tau}\right] \\
&\stackrel{r}{\ll}\; \sum_{l_1=-\infty}^{\infty}|c_{l_1}^{s-i}|\; E^{\frac{1}{\tau}}\left[(n_{r+1}-1)^{\tau}\left|\tilde{\zeta}_1^{(r)}\right|^{\tau}\right] \\
&\stackrel{r}{\ll}\; n_r\|\tilde{\zeta}_1^{(r)}\|_{\tau}\;. \qquad (4.79)
\end{aligned}$$

Thus $\forall\, i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}$, we get from (4.78) and (4.79), that

$$E^{\frac{1}{\tau}}\left[\sup_{n_r \le n < n_{r+1}} \left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\right] \overset{r}{\ll} n_r \|\tilde{\zeta}_1^{(r)}\|_{\frac{s\tau}{s-\tau i}} \ . \tag{4.80}$$

Now, we choose $\tau > 1$ small enough so that $\alpha_i > \frac{s\tau}{s-\tau i}$, which is possible since $\alpha_i = \frac{s}{s-i}\alpha_0 > \frac{s}{s-i}$, and $\frac{s\tau}{s-\tau i}$ is continuous and increasing for $\tau \in (1, \alpha_i)$. Hence by (4.71) with $z = \frac{s\tau}{s-\tau i}$, and (4.80), we see that there exists $\mathcal{T}_i \in (1, \alpha_i)$ such that $\forall\, \tau \in (1, \mathcal{T}_i)$,

$$E\left[\sup_{n_r \le n < n_{r+1}} \left|\tilde{T}_n^{(r)}(i)\right|^{\tau}\right] \overset{r}{\ll} n_r^{\tau - \frac{\kappa(\alpha_i - \frac{s\tau}{s-\tau i})}{\frac{s}{s-\tau i}(2-\alpha_i)}}. \tag{4.81}$$

## 4.5 Final Rate of Convergence for Theorem 3.1

Finally, we shall use the Borel-Cantelli Lemma to combine the results of the last two sections and prove Theorem 3.1. Notice that in $\sum_{k=1}^{n}(d_k - d)$ (from Theorem 3.1), the light tailed terms are $S_n(q, \lambda_q)$ (from (4.16)) over all partitions where $a_1 \le \frac{s}{2}$, since their second moments are finite. The heavy tailed terms are $S_n^*(i)$ (from (4.56)) over $i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}$. We thus have

$$\sum_{k=1}^{n}(d_k - d) = \sum_{\substack{\lambda_q = (a_1, \ldots, a_q) \\ a_1 \le \frac{s}{2}}} S_n(q, \lambda_q) + \sum_{i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}} S_n^*(i) \tag{4.82}$$

First, we handle the light tailed terms. In Lemma 4.11, taking $\psi_{l_r}^{(r)} = \xi_{l_r}^{a_r}$ for $1 \le r \le q$, and $\delta = 1$, we see that $Y_{o,n,\delta}^{\lambda_q}$ becomes $S_n(q,\lambda_q) - S_o(q,\lambda_q)$, and

$$E\left[\left(S_n(q,\lambda_q) - S_o(q,\lambda_q)\right)^2\right] \overset{o,n}{\ll} \begin{cases} n-o, & a_q \ge 2 \\ (n-o)\ l_{s,\sigma}(n-o), & a_1 = 1 \\ (n-o)\ l_{1,\sigma}(n-o), & a_q = 1,\ a_1 \ge 2. \end{cases}.$$

Further, when $s$ is even and $\xi_1$ is symmetric, we get

$$E\left[\left(S_n(q,\lambda_q) - S_o(q,\lambda_q)\right)^2\right] \overset{o,n,\delta}{\ll} (\delta\ (n-o)) \vee ((n-o)\ l_{2,\sigma}(n-o)).$$

These are the same results as in (4.44) and (4.45). Thus, proceeding along the lines of (4.46 - 4.55), we get that

$$\lim_{n\to\infty} \frac{S_n(q,\lambda_q)}{n^{\frac{1}{p}}} \;=\; 0 \qquad\qquad \text{a.s.} \tag{4.83}$$

for the values of $p$ as mentioned in (4.14) and (4.15), in the statement of Theorem 4.6.

Now we deal with the heavy tailed terms. We fix $i \in \left\{0, 1, \ldots, \left\lfloor \frac{s-1}{2} \right\rfloor\right\}$, which fixes $S_n^*(i)$, and due to (4.59), consider $T_n(i)$ instead of $S_n^*(i)$.

First, we consider the case where $s > 2$. From (4.66,4.81), Markov's Inequality, and the fact that $l_{1,\sigma}(n_r) = n_r^{2-2\sigma}$ (since $\sigma < 1$), we get that, there exists

$\mathcal{T}_i$ such that $\forall\ 1 < \tau < \mathcal{T}_i$,

$$\begin{aligned} & P\left(\sup_{n_r \le n < n_{r+1}} |T_n(i)| > 2\epsilon n_r^{\frac{1}{p}}\right) \\ \le\ & \frac{1}{\epsilon^2 n_r^{\frac{2}{p}}} E\left[\sup_{n_r \le n < n_{r+1}} \left|\overline{T}_n^{(r)}(i)\right|^2\right] \ +\ \frac{1}{\epsilon^\tau n_r^{\frac{\tau}{p}}} E\left[\sup_{n_r \le n < n_{r+1}} \left|\tilde{T}_n^{(r)}(i)\right|^\tau\right] \\ \overset{r}{\ll}\ & r^2\left[\left(n_r^{1-\frac{2}{p}} l_{1,\sigma}(n_r)\right) \vee \left(n_r^{1+\kappa-\frac{2}{p}}\right)\right] \ +\ n_r^{\tau - \frac{\kappa(\alpha_i - \frac{s\tau}{s-\tau i})}{\frac{s}{s-\tau i}(2-\alpha_i)} - \frac{\tau}{p}} \\ \overset{r}{\ll}\ & r^2\left[\left(n_r^{3-2\sigma-\frac{2}{p}}\right) \vee \left(n_r^{1-\frac{\alpha_i}{p}}\right)\right] \ +\ n_r^{\tau - \frac{\alpha_i(s-\tau i)}{ps}}, \end{aligned} \tag{4.84}$$

by letting $\kappa = \frac{2-\alpha_i}{p}$. Note that $(3 - 2\sigma - \frac{2}{p}) \vee (1 - \frac{\alpha_i}{p}) < 0$ implies that $p < \alpha_i \wedge \frac{2}{3-2\sigma}$. Next, note that $\tau - \frac{(s-\tau i)\alpha_i}{ps} < 0$ if and only if $p < \alpha_i\left(\frac{s-\tau i}{s\tau}\right)$. But for any $p < \alpha_0 = \alpha_i\left(\frac{s-i}{s}\right)$, we select $\tau > 1$ small enough such that $p < \alpha_i\left(\frac{s-\tau i}{s\tau}\right)$. Hence, from (4.84), we get that $\forall\ p < \alpha_0 \wedge \frac{2}{3-2\sigma}$,

$$\sum_{r=1}^{\infty} P\left(\sup_{n_r \le n < n_{r+1}} |T_n(i)| > 2\epsilon n_r^{\frac{1}{p}}\right) < \infty\ . \tag{4.85}$$

When $s = 2$, from (4.68, 4.81) and Markov's Inequality, we get that there exists $\mathcal{T}_i$ such that $\forall\ 1 < \tau < \mathcal{T}_i$,

$$P\left(\sup_{n_r \le n < n_{r+1}} |T_n(i)| > 2\epsilon n_r^{\frac{1}{p}}\right) \ \overset{r}{\ll}\ r^2\left(n_r^{1-\frac{\alpha_i}{p}}\right) \ +\ n_r^{\tau - \frac{\alpha_i(s-\tau i)}{ps}}. \tag{4.86}$$

Again, $\forall\ p < \alpha_0 = \alpha_i\left(\frac{s-i}{s}\right)$, we can select $\tau > 1$ small enough such that $p < \alpha_i\left(\frac{s-\tau i}{s\tau}\right)$. Thus, we get $\forall\ p < \alpha_0$, that

$$\sum_{r=1}^{\infty} P\left(\sup_{n_r \le n < n_{r+1}} |T_n(i)| > 2\epsilon n_r^{\frac{1}{p}}\right) < \infty\ . \tag{4.87}$$

Lastly, when s is even, and $\xi_1$ is symmetric, from (4.70, 4.81) and Markov's Inequality, we get that there exists $\mathcal{T}_i$ such that $\forall\, 1 < \tau < \mathcal{T}_i$,

$$P\left(\sup_{n_r \le n < n_{r+1}} |T_n(i)| > 2\epsilon n_r^{\frac{1}{p}}\right) \overset{r}{\ll} r^2 \left[n_r^{1-\frac{\alpha_i}{p}} \vee n_r^{((4-4\sigma)\vee 1)-\frac{2}{p}}\right] + n_r^{\tau - \frac{\alpha_i(s-\tau i)}{ps}} .$$

Now, $\left(((4-4\sigma)\vee 1) - \frac{2}{p}\right) \vee (1 - \frac{\alpha_i}{p}) < 0$ implies that $p < 2 \wedge \alpha_i \wedge \frac{1}{2-2\sigma}$. Again, we note that $\tau - \frac{(s-\tau i)\alpha_i}{ps} < 0$ if and only if $p < \alpha_i\left(\frac{s-\tau i}{s\tau}\right)$, so for any $p < \alpha_0 = \alpha_i\left(\frac{s-i}{s}\right)$, we select $\tau > 1$ small enough such that $p < \alpha_i\left(\frac{s-\tau i}{s\tau}\right)$. Hence, we get $\forall\, p < 2 \wedge \alpha_0 \wedge \frac{1}{2-2\sigma}$, that

$$\sum_{r=1}^{\infty} P\left(\sup_{n_r \le n < n_{r+1}} |T_n(i)| > 2\epsilon n_r^{\frac{1}{p}}\right) < \infty . \tag{4.88}$$

Hence, for the values of $p$ in (4.85,4.87,4.88), from the Borel-Cantelli Lemma, we get that

$$\begin{aligned} &\lim_{n\to\infty} \frac{T_n(i)}{n^{\frac{1}{p}}} = 0 \qquad \text{a.s.} \\ \Rightarrow\ &\lim_{n\to\infty} \frac{S_n^*(i)}{n^{\frac{1}{p}}} = 0 \qquad \text{a.s. ,} \end{aligned} \tag{4.89}$$

due to (4.56,4.59). From (4.83, 4.89) and Remark 3.5, we get that

$$\lim_{n\to\infty} n^{-\frac{1}{p}} \sum_{k=1}^{n} (d_k - d) = 0 \qquad \text{a.s. ,}$$

for the values of $p$ as claimed in the statement of Theorem 3.1. This completes the proof of Theorem 3.1. □

# Bibliography

[1] Benveniste, A., Metivier, M. and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag Berlin Heidelberg.

[2] Davis, R. and Resnick, S. (1985). Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities. *Annals of Probability*, **13**, 179-195.

[3] Dette, H., Preuss, P. and Sen, K. (2017). Detecting long-range dependence in non-stationary time series. *Electronic Journal of Statistics*, **11**, 1600-1659.

[4] Eweda, E. and Macchi, O. (1984). Convergence of an adaptive linear estimation algorithm. *IEEE Transactions on Automatic Control*, **29**, 119-127.

[5] Feller, W. (1951). The Asymptotic Distribution of the Range of Sums of Independent Random Variables. *The Annals of Mathematical Statistics*, **22**, 427-432.

[6] Foss, S., Korshunov, D. and Zachary, S. (2013). *An Introduction to Heavy-Tailed and Subexponential Distributions.* Springer Series in Operations Research and Financial Engineering, Springer.

[7] Graves, G., Gramacy, R., Watkins, N. and Franzke, C. (2017). A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951–1980. *Entropy*, **19**, 437.

[8] Grossglauser, M. and Bolot, J.-C. (1999). On the relevance of long-range dependence in network traffic. *IEEE/ACM Transactions on Networking*, **7**, 629-640.

[9] Hosking, J.R.M. (1981). Fractional Differencing. *Biometrika*, **68**, 165-176.

[10] Hurst, H.E. (1956). The Problem of Long-Term Storage in Reservoirs. *Hydrological Sciences Journal*, **1**, 13-27.

[11] Hurst, H.E. (1956). Methods of using long-term storage in Reservoirs. *Proceedings of the Institution of Civil Engineers*, **5**, 519–543.

[12] Karagiannis, T., Molle, M. and Faloutsos, M. (2004). Long-range dependence ten years of Internet traffic modeling. *IEEE Internet Computing*, **8**, 57-64.

[13] Koscielny-Bunde, E., Kantelhardt, J. W., Braun, P., Bunde, A., and Havlin, S. (2006). Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. *Journal of Hydrology*, **322**, 120-137.

[14] Kouritzin, M.A. (1996). On the convergence of linear stochastic approximation procedures. *IEEE Transactions on Information Theory*. **42**, 1305-1309.

[15] Kouritzin, M.A. (1996). On the interrelation of almost sure invariance principles for certain stochastic adaptive algorithms and for partial sums of random variables. *Journal of Theoretical Probability*, **9**, 811-840.

[16] Kouritzin, M.A. (1995). Strong approximation for cross-covariances of linear variables with long-range dependence, *Stochastic Processes and their Applications*, **60,** 343-353.

[17] Kouritzin, M. A. and Sadeghi, S. (2016). Marcinkiewicz Law of Large Numbers for Outer products of Heavy-Tailed, Long-range Dependent Data. *Advances in Applied Probability*, **48**, 349-368.

[18] Kouritzin, M. A. and Sadeghi, S. (2015). Convergence rates and decoupling in linear stochastic approximation algorithms. *SIAM Journal on Control and Optimization*, **53**, 1484–1508.

[19] Loeve, M. (1977). *Probability Theory I*. Graduate Texts in Mathematics, Springer-Verlag New York.

[20] Louhichi, S. and Soulier, P. (2000). Marcinkiewicz–Zygmund strong laws for infinite variance time series. *Statistical Inference for Stochastic Processes*, **3**, 31-40.

[21] MANDELBROT, B. (1972). Statistical methodology for non-periodic cycles: from the covariance to R/S analysis. *Annals of Economic and Social Measurement*, **1**, 259-290.

[22] MANDELBROT, B. (1967). Some noises with 1/f spectrum, a bridge between direct current and white noises *IEEE Transactions on Information Theory*, **13**, 289–298.

[23] MANDELBROT, B. AND VAN NESS J. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, 422-437.

[24] MANDELBROT, B. B. AND WALLIS J. R. (1968). Noah, Joseph and operational hydrology. *Water Resources Research*, **4**, 909-918.

[25] MONTANARI, A., TAQQU, M.S. AND TEVEROVSKY, V. (1999). Estimating long-range dependence in the presence of periodicity: An empirical study. *Mathematical and Computer Modelling*, **29**, 217-228.

[26] MORAN, P.A.P. (1964). On the range of cumulative sums. *Annals of the Institute of Statistical Mathematics*, **16**, 109–112.

[27] PIPIRAS, V. AND TAQQU, M. (2017). *Long-Range Dependence and Self-Similarity.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

[28] POTTER, K.W. (1976). Evidence for nonstationarity as a physical explanation of the Hurst Phenomenon. *Water Resources Research*, **12**, 1047-1052.

[29] RIO, E. (1995). A Maximal Inequality and Dependent Marcinkiewicz-Zygmund Strong laws. *Annals of Probability*, **23**, 918-937.

[30] RODRIGO, M.M. AND RENATO, A.Q. (2017). Long-range dependence in the runoff time series of the most important Patagonian river draining to the Pacific Ocean. *New Zealand Journal of Marine and Freshwater Research*, **52**, 264-283.

[31] SAMORODNITSKY, G. (2016). *Stochastic Processes and Long Range Dependence.* Springer Series in Operations Research and Financial Engineering, Springer.

[32] SAMORODNITSKY, G. (2007). Long Range Dependence. *Foundations and Trends® in Stochastic Systems*, **1**, 163-257.

[33] Shiryaev, A. N. (1996). *Probabilty*. Graduate Texts in Mathematics, Springer, New York.

[34] Smith, V.K. (1975). A Simulation Analysis of the Power of Several Tests for Detecting Heavy-Tailed Distributions. *Journal of the American Statistical Association*, **70**, 662–665.

[35] Sornette, D. (2009). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press.

[36] Stout, W. F. (1974). *Almost Sure Convergence.* Academic Press, New York.

[37] Torre, K., Delignieres, D. and Lemoine, L. (2007). Detection of long-range dependence and estimation of fractal exponents through ARFIMA modelling. *British Journal of Mathematical and Statistical Psychology*, **60**, 85-106.

[38] Wallis, J.R. and O'Connell, P.E. (1973). Firm reservoir yield — How reliable are historic hydrological records? *Hydrological Sciences Bulletin*, **18**, 347-365.

[39] Wu, W. B. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic Processes and their Applications*, **115**, 939-958.

[40] Wu, W. B., Huang, Y. and Zheng, W. (2010). Covariances estimation for long-memory processes. *Advances in Applied Probability*, **42**, 137-157.