UNIVERSITY OF CALIFORNIA, SAN DIEGO

Statistical Inference for Parameters of Time Series Exhibiting the Noah and Joseph Effects

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Tucker S. McElroy

Committee in charge:

Professor Dimitris Politis, Chair
Professor Patrick Fitzsimmons
Professor Anthony Gamst
Professor Halbert White
Professor Ruth Williams

2001

The dissertation of Tucker S. McElroy is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2001

To the glory of God

"That in the dispensation of the
fulness of times he might gather
together in one all things in Christ,
both which are in heaven,
and which are on earth;
even in him"
— Ephesians 1:10

TABLE OF CONTENTS

Signature Page . . . iii

Dedication . . . iv

Table of Contents . . . v

Acknowledgements . . . vii

Abstract of the Dissertation . . . x

1 Introduction . . . 1
1.1 Noah and Joseph Effects . . . 1
1.1.1 Noah Effect - Heavy Tails . . . 1
1.1.2 Joseph Effect - Long Range Dependence . . . 2
1.2 Marked Point Processes . . . 3
1.3 Subsampling . . . 4

2 Linear Processes . . . 6
2.1 Heavy-Tailed Linear Models: Background . . . 6
2.2 Self-Normalized Sample Mean . . . 9
2.3 Self-Normalized Sample Maximum . . . 14
2.4 Self-Normalized Periodogram . . . 18
2.5 Subsampling Applications . . . 19
2.5.1 The Sample Mean . . . 20
2.5.2 The Symmetrized Situation . . . 22
2.5.3 The Sample Maximum and Periodogram . . . 25
2.6 Simulation Studies . . . 27

3 Stable Processes . . . 32
3.1 Stable Moving Average Processes . . . 32
3.2 Self-Normalized Sample Mean . . . 34
3.3 Subsampling Applications . . . 43
3.4 Technical Proofs . . . 45

4 Random Fields . . . 51
4.1 Heavy-Tailed Linear Random Fields . . . 51
4.2 Self-Normalized Sample Mean . . . 53
4.2.1 Partial Sums . . . 53
4.2.2 Sample Variance . . . 59
4.2.3 Joint Convergence . . . 63
4.3 Self-Normalized Periodogram . . . 66
4.4 Subsampling Applications . . . 72

4.5 Technical Proofs . . . 75
4.5.1 Proof of Claim 4.2.1 . . . 75
4.5.2 Proof of Claim 4.2.2 . . . 78

5 Marked Point Processes . . . 79
5.1 Observation Regions : Background . . . 79
5.2 Discrete Time Results . . . 80
5.3 Continuous Time Results . . . 90
5.4 Poisson Random Measures . . . 98
5.5 Subsampling Applications . . . 106

6 Conclusion . . . 111

A Stable Variables . . . 112
A.1 Basic Definitions . . . 112
A.2 Product Decomposition of α-stable random variables . . . 114
A.3 Series Representation for α-stable random variables . . . 115

B Point Process Techniques . . . 118

C Stable Integrals . . . 121
C.1 Stable Stochastic Processes . . . 121
C.2 Existence of Stable Integrals . . . 122
C.3 α-stable Random Measures . . . 123
C.4 Construction of Stable Integrals . . . 124

D Central Limit Theorem . . . 127
D.1 Introduction and background . . . 127
D.2 Central Limit Theorem . . . 128
D.3 Proof of the Theorem . . . 131
D.3.1 Step 0 : Notation and preliminaries . . . 131
D.3.2 Step 1 : Estimate of the "mortar" . . . 132
D.3.3 Step 2: The Mixing Condition . . . 138
D.3.4 Step 3: Estimate of the Big Block . . . 138
D.3.5 Step 4 : The Lyapounov Condition . . . 140

Bibliography . . . 142

ACKNOWLEDGEMENTS

Firstly I thank my Lord Jesus Christ for guiding and sustaining me through an arduous intellectual and spiritual journey. In all my research and discovery, I am but His mouthpiece to reveal His name which is stamped upon all of Creation. His servants in the Church – including Pastor James Lee, Reverend Jim Dennison, and the many friends I am blessed by – have continued to encourage me in my vocation and calling as a mathematician.

My thesis advisor Dimitris Politis has been extremely generous with his time and energy, providing suggestions and discussions throughout this process. I appreciate his patience in explaining and teaching mathematics to me, as well as his support of my efforts. He has guided me in the topics of my research, while being a model of diligence himself. It has been a pleasure to be his student for these past years.

I have also benefitted from discussions with my academic colleagues: Rob Ellis was always ready to listen to my investigations, and Arif Dowla prompted me to increase my understanding of time series. My office mates Ben Raphael, Christian Gromoll, Steve Bell, and Weining Kang provided profitable mathematical intercourse as well. Among the faculty, Patrick Fitzsimmons, Ruth Williams, Bill Helton, Len Haff, Ian Abramson, Yosef Rinott, Murray Rosenblatt, and Bruce Driver gave their support and interest.

Some of the research presented in this document has been published separately, with Dimitris Politis as co-author. That mathematical material which is presented here is due to my own efforts.

My family has endured my exodus to California in search of knowledge; their contact and endorsement of my efforts is greatly appreciated. My father, P. T. McElroy, is remembered and dearly missed.

Lastly, I am grateful to my betrothed Autumn for her love and prayers, which have been instrumental to my finishing this race. Truly I am blessed by her presence in my life.

VITA

August 26, 1976	Born, Boston, Massachusetts
1996	B. A., Columbia University
1996 – 2000	Teaching Assistant, Department of Mathematics, University of California San Diego
2001	Ph. D., University of California San Diego

PUBLICATIONS

McElroy, T. and Politis, D. (2000) *Inference for the Mean in the Presence of Serial Correlation and Heavy Tailed Distributions.* Accepted for publication in *Econometric Theory.*

McElroy, T., and Politis, D. (1999) *A Central Limit Theorem for an Array of Strong Mixing Random Fields. Festschrift for George Roussas.*

ABSTRACT OF THE DISSERTATION

Statistical Inference for Parameters of Time Series Exhibiting the Noah and Joseph Effects

by

Tucker S. McElroy
Doctor of Philosophy in Mathematics

University of California San Diego, 2001

Professor Dimitris Politis, Chair

This work is a study of various long-range dependent heavy-tailed time series models, which are useful for modeling data in the fields of econometrics, communications, insurance, and finance. We examine self-normalized statistics for model parameters, such as the mean, in order provide theoretical backing for statistical inference in these models. These models have in common (in addition to strict stationarity) long-range dependence (the *Joseph Effect*) and heavy-tailed marginal distributions (the *Noah Effect*). Most of the models have a linear dependence structure (such as an infinite order moving average) and are strong mixing, i.e. asymptotically, the random variables are independent. In later chapters, generalizations to random fields are considered. My objective has been to conduct inference for model parameters – e.g. form asymptotically correct confidence intervals. Thus, the bulk of the results are weak limit theorems for statistics such as the sample mean.

Two theoretical/practical techniques occur as recurring themes in this work: self-normalization and subsampling. All of the statistics considered have "unknown" rates of convergence; this problem is solved by dividing or "normalizing" it by some other

statistic which converges at that rate. A second difficulty is that the limit distributions generally are complicated functions of stable random variables and filter coefficients, and thus their quantiles are unknown; this is resolved by using subsampling methods to estimate the limit cumulative distribution function. Indeed, this work has the motive of providing a wealth of practical and useful probability models to which the already existing subsampling methods in [48] may be applied.

Chapter 1

Introduction

There are several important concepts used throughout this work; the most important are the "Noah and Joseph Effects", Marked Point Processes, and Subsampling Methods – each of which is discussed below. In addition the term "self-normalization" is ubiquitous here. It simply refers to the division of the desired statistic by some other statistical measure of scale, such as the sample variance. This operation circumvents some unpleasant practical issues, although it makes the corresponding theory more complicated. See the second section of Chapter Two for a demonstration of self-normalization.

1.1 Noah and Joseph Effects

Noah and Joseph are intriguing epithets for the phenomena of heavy tails and long range dependence, respectively. Both terms were coined by Mandelbrot when he was studying hydrology (see [42] and [41]), and he observed that both phenomena seemed to occur together; see [53] as well as [29] . All of the models that I consider in this work exhibit both heavy tails and some nontrivial dependence.

1.1.1 Noah Effect - Heavy Tails

A random variable is said to have **heavy tails** if its probability density function (pdf) has tails that decay at a polynomial rate (Mandelbrot coined the term "hyperbolic" random variables — viz. [41], p.204 — due to the asymptotic shape of the tail); in particular the tails are not exponential, and hence not all of the moments will be finite. An example is the Cauchy distribution, which has no mean. Many of the heavy tailed random variables which I will consider lack a second moment, and thus are used in models which exhibit (or are thought to exhibit) "infinite variance." This has been called the **Noah Effect** due to the infinite variance, which shows high variability, viz. "...all the fountains of the great deep [were] broken up, and the windows of heaven were opened. And the rain was upon the earth forty days and forty nights." (Genesis 7:11-12).

Examples of Heavy-Tailed Data Good examples can be found in the data from [69] and [22], which exhibit Ethernet traffic and S & P 500 index data respectively. In the arena of finance, much debate is continuing over infinite versus finite variance models. "It is a fact...that most financial data are heavy-tailed!" (viz. [22], p. 406) In some cases second moments may exist, but third or fourth moments do not. In teletrafficking, Resnick claims "...such phenomena as file lengths, cpu time to complete a job, call holding times, interarrival times between packets in a network and lengths of on/off cycles appear to be generated by distributions which have heavy tails" ([52], p.1805). For references, see [3], [68], [19], [20], [44], and [69]. Data sets displaying heavy tails can be found in other fields such as hydrology (viz. [12]), economics and finance (viz. [33]), and reliability and structural engineering (viz. [27]).

1.1.2 Joseph Effect - Long Range Dependence

The term **long range dependence** refers to the presence of a slowly decaying autocorrelation/autocovariance function; indeed, the decay may be so slow (e.g. a polynomial rate) that the covariances are not summable. We don't strictly require the word "auto" above, which is appropriate for stationary data - more general heteroskedastic data may also be thought of as having long range dependence. It is common to measure dependence via the strong mixing coefficients of [60] – also see [61]; roughly speaking, exponential decay corresponds to short-range dependence (e.g. Markov dependence), while polynomial decay corresponds to long-range dependence. This persistence over long gaps or lags will also be seen in the sample correlation and covariance functions. This has been called the **Joseph Effect** due to the persistency of phenomenon over time, viz. "Behold, there shall come seven years of great plenty throughout all the land of Egypt: And there shall arise after them seven years of famine..." (Genesis 41:29-30).

Examples of Long Range Dependence Chapter 1 of Jan Beran's book [2] contains five diverse examples of long range dependence – hydrology, video conferencing , Ethernet networks, governmental standardization, and climatology. In all of them we observe the characterizing behavior: the sample autocorrelations have a persistency over a large number of lags. This phenomenon appears in a number of other areas, e.g. economics – see [39], [66]; turbulence – see [40]; weather – see [37]; and communications – see [38]. In finance, "The existence of long term dependence in common stock price series is not surprising since stock prices are related directly or indirectly to climatological variables, such as rainfall, in which the existence of long term dependence is well established" ([26]); also see [25]. See [42] and [43] for examples in hydrology.

Additional mathematical examples will be provided in some of the following chapters. Appendix A provides some background on stable random variables and their domains of attraction.

1.2 Marked Point Processes

In time series it often occurs that data is collected at random times, or from random locations if we are considering a random field. The indices of the random variables themselves are then random variables, and the common model for this is a Point Process. Often the Point Process is Poisson, which has certain nice properties, but we will consider more general processes in this work. One probability mechanism governs the scattering of points, where the observations are taken from, and another mechanism gives the values of the random field at those locations. It is a common assumption (which we will make) that these two probability laws are independent.

According to [48] , irregularly spaced data is encountered “in many important cases, e.g. queueing theory, spatial statistics, mining and geostatistics, meteorology, etc...As a matter of fact, in case $d > 1$, irregularly spaced data seem to be the rule rather than the exception.” (Here d refers to the dimension of the random field.) Also see [58], [32], and [13].

In this work, we examine some of the models of earlier chapters under the added assumption of the Marked Point Process, and derive weak limit results for the statistics. As discussed below, subsampling methods are still valid under the assumption of random observation locations.

1.3 Subsampling

The Central Limit Theorem has been used to construct asymptotically valid confidence intervals for the mean in classical time series. This feat relies upon knowledge of the quantiles of the limit distribution (i.e. the Gaussian). In some scenarios, the weak limit is not normal, and the quantiles cannot be calculated. Nevertheless, practitioners still desire to make inference for the mean; subsampling provides an elegant solution to this difficulty.

Given a weak convergence result, and some other mild assumptions, one can form an estimate of the sampling distribution via “subsamples” of the data. With these estimated quantiles, approximate confidence intervals for the desired parameters may be constructed. The first subsampling theorems assumed independent data, but they have since been generalized to time series which satisfy a certain asymptotic independence condition. See the book *Subsampling* [48] for an extensive treatment of this subject.

These methods are also valid when the statistic is self-normalized, so long as the resulting quotient statistic has a weak limit. Many of the results in this work involve self-normalized statistics whose resulting limit distribution is completely unknown, and cannot be calculated due to its dependence on other unknown model parameters. But the limiting distribution can be approximated by the subsampling distribution estimator, so that inference is possible.

Chapter 2

Linear Processes

In this first part our attention will be restricted to linear time series. We will look at causal and non-causal ARMA (Auto-Regressive Moving Average) processes, which have a representation as an infinite order moving average. This is a fairly weak type of dependence (an AR(1) is strong mixing, with an exponential rate of decay), but nevertheless these models are wide spread in time series applications.

2.1 Heavy-Tailed Linear Models: Background

Consider the *linear* time series:

$$X_t := \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j} \quad \forall t \in \mathbb{N}, \tag{2.1}$$

where $\{Z_t\}$ is an *iid* (independent and identically distributed) sequence, and $\{\psi_j\}$ are filter coefficients which satisfy a summability condition.

Throughout this chapter, we require that the filter $\{\psi_j\} \in l_p$ for some $p \in [0, 1]$ (see Chap. 13 of [11]) in order that the sum converges almost surely. The input variables

$\{Z_t\}$ (for any $t \in \mathbb{Z}$) satisfy the following two properties for some $\alpha \in (0, 2)$:

$$\mathbb{P}[|Z_t| > x] = x^{-\alpha} L(x) \tag{2.2}$$

$$\frac{\mathbb{P}[Z_t > x]}{\mathbb{P}[|Z_t| > x]} \to p, \qquad \frac{\mathbb{P}[Z_t \leqslant -x]}{\mathbb{P}[|Z_t| > x]} \to q \tag{2.3}$$

as $x \to \infty$. Here p and q are between 0 and 1 and add up to 1. $L(x)$ is a "slowly varying" function, i.e. $L(ux)/L(x) \to 1$ for any fixed u. An example of this is the logarithm. Note that it easily follows that the right and left tails of Z_t behave like

$$\mathbb{P}[Z_t > x] \sim p x^{-\alpha} L(x), \quad \mathbb{P}[Z_t \leqslant -x] \sim q x^{-\alpha} L(x) \tag{2.4}$$

where "$\sim$" denotes that the ratio tends to unity as $x \to \infty$.

A random variable which satisfies the above conditions is said to be "heavy-tailed" or to possess "fat tails." Such random variables are in the domain of attraction of an α-stable law, i.e. if we take an *iid* sequence of such Z's, then there exist constants $a_n > 0$ and b_n such that

$$a_n^{-1} \left(\sum_{t=1}^{n} Z_t - b_n \right) \stackrel{\mathcal{L}}{\Longrightarrow} S \tag{2.5}$$

where S is an α-stable random variable. We then write $\{Z_t\} \in DOM(\alpha)$. It is common to see conditions (2.2) and (2.3) stated as

$$x^{\alpha}(1 - F(x)) \to p, \quad x^{\alpha} F(-x) \to q, \tag{2.6}$$

where Z is a common version of the Z_t's, which have cdf $F(\cdot)$. A good treatment of this material can be found in the book *Non-Gaussian Random Processes* [64] ; or see Appendix A.

An example of this is given by Z in the "normal" domain of attraction of an α-stable law, which means that we can take $a_n = n^{\frac{1}{\alpha}}$. If Z is itself stable, then the hypotheses are certainly satisfied. If in addition it is symmetric (Z is sαs), then X_t has the law of a sαs as well, but scaled by $(\sum_j |\psi_j|^{\alpha})^{\frac{1}{\alpha}}$. Note that no generality is lost if one places a constant $C > 0$ on the right hand side of the convergences in line (2.6). This constant is called the "dispersion" of Z (written $disp(Z)$). If Z has the $S_\alpha(\sigma, \beta, \mu)$ law, then we can take $p = \frac{\beta+1}{2}$ and $q = \frac{\beta-1}{2}$, and the dispersion is

$$C = \begin{cases} \sigma/\Gamma(1-\alpha)\cos\frac{\pi\alpha}{2} & \alpha \neq 1 \\ 2\sigma/\pi & \alpha = 1. \end{cases} \tag{2.7}$$

We will let $C = 1$ for simplicity; also see [11].

There are a few facts about the choice of a_n : firstly, the sequence should satisfy

$$n\mathbb{P}[|Z_1| > a_n x] \to x^{-\alpha} \tag{2.8}$$

as $n \to \infty$ for every positive x. (So if we take a_n that satisfies this, then we can prove the limit result for the domain of attraction.) It is easy to check that $a_n := \inf\{x : \mathbb{P}[|Z_1| > x] \leqslant n^{-1}\}$ satisfies this condition. Condition (2.8) will be very important in what follows. A change of variable argument, using (2.2), implies that $a_n = n^{\frac{1}{\alpha}}L(n)$ (not the same slowly varying function in (2.2)); thus the "normal" domain of attraction has $L \equiv 1$. Given this, a suitable choice for b_n is $\mathbb{E}[Z_1; |Z_1| \leqslant a_n]$. This definition is interesting – it suggests a "natural" truncation for Z_1.

Notice that since $\{\psi_j\} \in l_p$, they are also in l_α since $p < \alpha$, so $(\sum_j |\psi|^\alpha)^{\frac{1}{\alpha}} < \infty$. Here is some notation: Ψ will denote the whole sequence of $\{\psi_j\}$, and Ψ_p will denote its l_p norm. It is true that $\{X_t\}$ forms a strictly stationary sequence, since applying a shift operator to the law for the Z-series does not affect the distribution.

When α is not less than 2, the situation is quite different. From the Central Limit Theorem, a partial sum of any *iid* collection of $\mathbb{L}^2$ random variables has a Gaussian limit (this is the $\alpha = 2$ stable random variable). Thus, $DOM(2) \supset \mathbb{L}^2$ (and the containment is proper). By contrast, the other domains of attraction for $\alpha < 2$ are characterized purely by the tail behavior. Our general assumption on Z is that either it is a heavy-tailed random variable in the space $DOM(\alpha)$ for $1 < \alpha < 2$ – and thus it satisfies properties (2.2), (2.3), and (2.8) – or that Z is square integrable, and hence is in $DOM(2)$. Throughout, by an abuse of language, we will abbreviate this dichotomy by saying that $\alpha \in (1, 2]$.

2.2 Self-Normalized Sample Mean

In this section we are interested in estimating the mean of the linear time series $\{X_t\}$. We focus on the sample mean $\overline{X} := n^{-1} \sum_{i=1}^{n} X_i$ as an estimator for $\theta := \mathbb{E}X_t$, which is assumed to be finite. Our point of view is one of generality: the sample mean is a ubiquitous estimate of location; in particular, it is generally consistent for θ even if the regularity condition of finite variance breaks down. Nevertheless, statistical inference (confidence intervals and tests) for θ are based on the distribution of $\overline{X}$, which is crucially affected by dependence and/or heavy tails. The purpose of this section is to provide a way of consistently estimating the distribution of the (normalized) sample mean *without* knowledge (or explicit estimation) of either the dependence or the heavy-tailed index (this tail index measures the heaviness of the tails); we will achieve this using the subsampling methodology —see [47] or [48].

It turns out that the normalized sum of *iid* (independent and identically distributed) heavy-tailed random variables converges weakly to a non-normal limit (a stable law); thus it satisfies a non-central limit theorem. In order to develop confidence intervals for θ, we need the quantiles of this stable law, which, unfortunately, are generally unknown, because both the scale and the index of stability (the heavy-tailed index) will generally be unknown. The recourse is to use subsampling methodology to estimate the limit quantiles. A second practical problem is that the rate of convergence of the sum is generally unknown (it is not the common $\sqrt{n}$ which occurs in the Central Limit Theorem), which prevents us from forming the correct statistic. This is solved by self-normalization, i.e by dividing by some appropriate measure of scale, such as the square root of the sample variance. If this is done, the limit is no longer a stable random variable, but has a well-defined continuous cdf (cumulative distribution function), so that subsampling theory can still be applied.

If the sequence $\{X_t\}$ is *iid* (and maybe heavy tailed), then the limit distribution of the normalized sample mean is well-studied in [36]. [59] used the limit theory in [36] in order to show that subsampling works in this case; see Chapter 10 of [48] for more details. Nothing is yet known if the series $\{X_t\}$ is not independent; this is the purpose of this section. To elaborate, we will prove that $\sqrt{n}\frac{\overline{X}-\theta}{\hat{\sigma}}$ has a limit distribution with a well-defined continuous cdf; here $\hat{\sigma}$ is a statistic which measures the scale or size of the random variables. In [36], both $\frac{n}{a_n}(\overline{X}-\theta)$ and $\frac{\sqrt{n}}{a_n}\hat{\sigma}$ are shown to have nondegenerate limit distributions for an iid time series; they both depend upon an unknown rate a_n (that corresponds to the unknown tail index α), and thus cannot be used statistically. In that paper, the authors form the ratio of these two quantities, and prove that the quotient converges to a continuous limit distribution. Since the factor a_n cancels in the division, the unknown rate is eliminated and the resulting object $\sqrt{n}\frac{\overline{X}-\theta}{\hat{\sigma}}$ is statistically useful. We prove the corresponding limit theorems in a time series context, and thus generalize the work of [36] . Other literature on this topic includes: [16], [17], [50], [51], [14]. The first two papers are primarily concerned with the limit behavior of sample autocorrelations for this linear model, while [50], [51] consider point process techniques used to prove many of these results. [14] examines models with long range dependence which are not linear.

Since the limit distribution depends on the unknown tail index parameter α, one cannot calculate its quantiles, which are neccessary for the formation of confidence intervals. For the unnormalized statistic one could explicitly estimate the parameter α, e.g. by Hill's estimator – viz. [52] and [56], and the scale of the limiting stable distribution, for the formation of the desired confidence intervals; also cf. [58]. Alternatively, subsampling with estimated rate may be used, as in [7]. Our approach in this chapter involves self-normalization; this bypasses the issue of the unknown rate of convergence but brings in a new difficulty: the limiting distribution is no longer stable. Subsampling, nonetheless, can still be applied since a well-defined limiting distribution exists; hence, this difficulty is alleviated as well.

So we first assume that the data X_t is integrable. To this end, let $\alpha \in (1,2]$ (with the meaning described at the end of section 2.1), so that the mean of Z exists, and denote it by η. By the choice of α, we see that the filter coefficients are summable; denote the sum by ψ_∞. Then it follows that $\theta := \mathbb{E}X_t = \psi_\infty \eta$. In this section we investigate the limiting behavior of $S_n(\theta)$, which is defined to be

$$S_n(\theta) := \frac{\sum_{t=1}^{n}(X_t - \theta)}{\sqrt{\sum_{t=1}^{n} X_t^2}}. \tag{2.9}$$

The following theorem gives the main result of the section:

Theorem 2.2.1. *Given the assumptions of the model, i.e. that lines (2.1) , (2.2), (2.3), and (2.8) hold, and that* $\{\psi_j\} \in l_p$,

$$S_n(\theta) \stackrel{\mathcal{L}}{\Longrightarrow} \frac{\psi_\infty}{\Psi_2} \frac{S}{\sqrt{S_2}} =: \tilde{S} \tag{2.10}$$

where S is an α-stable random variable, and S_2 has a totally right skewed (thus positive) $\frac{\alpha}{2}$-stable law; in particular, the limit random variable $\tilde{S}$ has a continuous nondegenerate distribution.

Remark S. Resnick [50] has proved a similar result for *iid* data (which is the special case where $\Psi = \{\cdots 0, \psi_0 = 1, 0, \cdots\}$). So the above theorem is the natural generalization of S. Resnick's result to a linear time series with $\alpha \in (1,2)$.

Proof in the case that $1 < \alpha < 2$ the proof follows from a more general version of this theorem for random fields, which is discussed in the second section of Chapter Four. The $\mathbb{L}^2$ case is well-known – for example, see [11]. †

Define the following "centered version" of $S_n(\theta)$:

$$T_n(\theta) := \frac{\sum_{t=1}^{n}(X_t - \theta)}{\sqrt{\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^2}} = \sqrt{n}\frac{\overline{X} - \theta}{\hat{\sigma}_n}, \tag{2.11}$$

where $\hat{\sigma}_n := \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^2}$ is an estimate of scale. Then we obtain the following corollary that includes the regular $\alpha = 2$ case as a possibility:

Corollary 2.2.1. *Consider a linear time series with summable filter coefficients, and let $\alpha \in (1, 2]$. For $1 < \alpha < 2$, this means that conditions (2.2) , (2.3) , (2.8) hold, but if $\alpha = 2$ it means that Z is square integrable. Then $T_n(\theta) \stackrel{\mathcal{L}}{\Longrightarrow} T$, where the random variable T is either $\frac{\psi_\infty}{\Psi_2}$ times a standard normal (for $\alpha = 2$) or is $\tilde{S}$ (for $1 < \alpha < 2$).*

Proof The notations $A \stackrel{\mathbb{P}}{\sim} B$ will mean $A = o_P(1) + B$ (which is a reflexive and symmetric relation). We have the following equivalences in probability:

$$\frac{\sum_{t=1}^{n}(X_t - \theta)}{\sqrt{\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^2}} \stackrel{\mathbb{P}}{\sim} \frac{\sum_{t=1}^{n}(X_t - \theta)}{\sqrt{\sum_{t=1}^{n}\left(X_t - \theta\right)^2}} \stackrel{\mathbb{P}}{\sim} \frac{\sum_{t=1}^{n}(X_t - \theta)}{\sqrt{\sum_{t=1}^{n} X_t^2}}, \tag{2.12}$$

where the second equivalence is only true for $\alpha < 2$. This statement is really two claims: the first claim is the left equivalence and the second claim is the right equivalence.

First Claim Let $U_n := a_n^{-1}\sum_{t=1}^{n} Y_t = \frac{n}{a_n}\overline{Y}_n$ where $Y_t := X_t - \theta$. It follows from the proof of Theorem 2.2.1 that $U_n = O_P(1)$. Hence

$$a_n^{-2}\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2 = a_n^{-2}\sum_{t=1}^{n}\left((X_t - \theta) + (\theta - \overline{X}_n)\right)^2 = a_n^{-2}\sum_{t=1}^{n}(X_t - \theta)^2 + \frac{-1}{n}U_n^2 \tag{2.13}$$

Thus

$$a_n^{-2}\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2 = O_P\left(\frac{1}{n}\right) + a_n^{-2}\sum_{t=1}^{n}(X_t - \theta)^2, \tag{2.14}$$

which also holds true for the joint law:

$$\left(a_n^{-1}\sum_{t=1}^{n}(X_t - \theta), a_n^{-2}\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2\right) \tag{2.15}$$

$$= O_P\left(\frac{1}{n}\right) + \left(a_n^{-1}\sum_{t=1}^{n}(X_t - \theta), a_n^{-2}\sum_{t=1}^{n}(X_t - \theta)^2\right).$$

Applying the continuous function $f(x,y) := \frac{x}{\sqrt{y}}$ to (2.15) finishes the claim.

Second Claim Here we suppose that $1 < \alpha < 2$. Then

$$a_n^{-2} \sum_{t=1}^{n} (X_t - \theta)^2 = a_n^{-2} \sum_{t=1}^{n} X_t^2 - \frac{2\theta}{a_n^2} \sum_{t=1}^{n} X_t + \frac{n\theta^2}{a_n^2}; \tag{2.16}$$

the second term is $-\frac{2\theta n}{a_n^2}\overline{X}_n = o(1)$ almost surely, since the sample mean tends to θ almost surely by the law of large numbers (this follows also from Theorem 2.2.1) and $\frac{n}{a_n^2} = o(1)$; the third term is $n^{1-\frac{2}{\alpha}} L(n)^{-1}\theta^2 = o(1)$, where $L(\cdot)$ is slowly varying. Note that we need $\alpha < 2$ for these conclusions. Therefore the claim holds, by applying f again to the joint vector.

Now the case of Corollary 2.2.1 where $\alpha \in (1,2)$ is clear, because $T_n(\theta) \overset{\mathbb{P}}{\sim} S_n(\theta)$ from the claims; so we can apply Theorem 2.2.1 to get the desired result. Now suppose that $\alpha = 2$, i.e., $Z \in \mathbb{L}^2$. Suppose that Z has variance σ^2; then by Theorem 7.1.2 of [11], we know that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \theta) \overset{\mathcal{L}}{\Longrightarrow} \psi_\infty \sigma N \tag{2.17}$$

where N has a standard normal distribution. As for the denominator, we have

$$\frac{1}{n} \sum_{t=1}^{n} (X_t - \theta)^2 \xrightarrow{P} \Psi_2^2 \sigma^2 \tag{2.18}$$

by Proposition 7.3.5 of [11]. Hence, using Slutsky's theorem (note that the limit in (2.18) is nonzero! – see [21]),

$$\begin{aligned} T_n(\theta) &\overset{\mathbb{P}}{\sim} \frac{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \theta)}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (X_t - \theta)^2}} \\ &\overset{\mathcal{L}}{\Longrightarrow} \frac{\psi_\infty}{\Psi_2 \sigma} \sigma N = \frac{\psi_\infty}{\Psi_2} N. \end{aligned} \tag{2.19}$$

This proves the corollary. †

2.3 Self-Normalized Sample Maximum

Classical extreme value theory explores the asymptotic properties of the sample maximum. The cumulative distribution function of the sample maximum is a quantity of interest to the financial and insurance communities, due to the need to calculate "value at risk" (VaR), as well as to probabilists, because of the intriguing mathematics involved in the asymptotic behavior.

This section examines the limit behavior of the sample maximum for linear heavy-tailed time series. Heavy-tailed random variables have become increasingly popular as models for financial and insurance data, and the literature has expanded in the past few decades to accomodate the growing interest: see [16], [17], [14], [22], and [52]. This movement was induced by the discovery of high kurtosis in financial and insurance data (as well as in teletraffic data and many data coming from natural phenomena) which was not adequately explained by light-tailed distributions such as the normal. Linear time series have continued to be a popular model due to both the ease of analysis as well as the extent of applicability. The following two examples demonstrate the utility of the sample maximum (also see [6]):

Example 1: Insurance Consider a positive claim amount X_t to be made at time t (viz. Chapter 1 of [22], where the claim times come according to a Poisson Process). If an insurance company has an amount of wealth x at time n, they are interested in knowing what the probability that the largest claim amount (i.e. the sample maximum) exceeds x, since this would induce bankruptcy. If we let M_n denote the sample maximum, we wish to estimate $\mathbb{P}\left[M_n > x\right]$; seen another way, the insurance company may wish to find x such that the above quantity is smaller than a specified probability p, which amounts to calculating the pth quantile of M_n's cdf.

Example 2: Finance In finance we may interpret X_t as the value of some asset at time t, and we wish to know the probability that the sample maximum stays above some minimum quantity x (this is a "worst case scenario" analysis). In this case we must estimate the same object and calculate the pth quantile. This number is used to measure "value at risk" (VaR), a quantity required of all corporate treasurers by the Securities and Exchange Commission.

So define the sample maximum and sample standard deviation by

$$M_n := \max_{1 \leqslant t \leqslant n} X_t \tag{2.20}$$

and

$$\hat{\sigma}_n := \sqrt{\sum_{t=1}^{n} X_t^2} \tag{2.21}$$

respectively. In addition, define the following features of the filter (where $a \vee b$ denotes the maximum of a and b):

$$\psi_+ := \max_{j \in \mathbb{Z}} (\psi_j \vee 0), \quad \psi_- := \max_{j \in \mathbb{Z}} (-\psi_j \vee 0). \tag{2.22}$$

The main result below is a limit theorem for the self-normalized maximum.

Theorem 2.3.1. *Let $\alpha \in (0, 2)$, and let $\{X_t\}$ be a strong mixing linear time series which satisfies conditions (2.2), (2.3), (2.8). Also require that either $\psi_+^\alpha p > 0$ or $\psi_-^\alpha q > 0$. Then the following weak convergence holds:*

$$\frac{M_n}{\hat{\sigma}_n} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{Y}{\Psi_2 \sqrt{S_2}}, \tag{2.23}$$

where S_2 is an $\frac{\alpha}{2}$-stable totally right skewed random variable, and Y has the following distribution:

$$\mathbb{P}[Y \leqslant x] = \begin{cases} \exp\{-x^{-\alpha}(\psi_+^\alpha p + \psi_-^\alpha q)\} & x > 0 \\ 0 & x \leqslant 0 \end{cases} \tag{2.24}$$

Proof This proof relies heavily upon Point Process techniques; a summary of the requisite background material is provided in Appendix B. We begin with the following Poisson process convergence result (B.7) which we repeat here for convenience:

$$\sum_{k=1}^{\infty} \varepsilon_{\left(\frac{k}{n}, \frac{X_k}{a_n}\right)} \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{i \in \mathbb{Z}} \sum_{k=1}^{\infty} \varepsilon_{t_k, \psi_i j_k} \tag{2.25}$$

on $M_p((0,\infty)\times(\mathbb{R}\setminus\{0\}))$ as $n \to \infty$. Next we define the mappings T_1 and T_2^δ as follows:

$$T_1\left(\sum_{k=1}^{\infty} \varepsilon_{(u_k,v_k)}\right) := \bigvee_{u_k \leqslant 1} v_k \tag{2.26}$$

and

$$T_2^\delta\left(\sum_{k=1}^{\infty} \varepsilon_{(u_k,v_k)}\right) := \sum_{u_k \leqslant 1} v_k^2 1_{[v_k^2 > \delta]} \tag{2.27}$$

for each positive δ. Notice that $\bigvee$ denotes the maximum of that collection of random variables. These maps are continuous with respect to the limit point process – viz. [50] and [51]. It follows that the joint mapping (T_1, T_2^δ) is continuous, and by applying this to line (2.25) above, we obtain the following joint convergence:

$$\begin{aligned}\left(\bigvee_{k=1}^{n} a_n^{-1} X_k, \sum_{k=1}^{n} a_n^{-2} X_k^2 1_{[X_k^2 > a_n^2 \delta]}\right) &\stackrel{\mathcal{L}}{\Longrightarrow} \left(\bigvee_{t_k \leqslant 1}\left(\bigvee_{i \in \mathbb{Z}} \psi_i j_k\right), \sum_{t_k \leqslant 1} \sum_{i \in \mathbb{Z}} \psi_i^2 j_k^2 1_{[|\psi_i j_k|^2 > \delta]}\right) \\ &= \left(\bigvee_{t_k \leqslant 1} (\psi_+ j_k \vee (-\psi_-) j_k), \sum_{i \in \mathbb{Z}} \psi_i^2 \sum_{t_k \leqslant 1} j_k^2 1_{[\psi_i^2 j_k^2 > \delta]}\right)\end{aligned} \tag{2.28}$$

Next we wish to let δ tend to zero on both sides of this convergence. The validity of this follows the exact same argument as in Theorem 4.2 of [16] , where they take the limit as δ tends to zero of the same map T_2^δ. The first component of the weak convergence does not depend on δ, so we obtain the joint convergence

$$\left(a_n^{-1} M_n, a_n^{-2} \sum_{k=1}^{n} X_k^2\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(\bigvee_{t_k \leqslant 1} (\psi_+ j_k \vee (-\psi_-) j_k), \Psi_2^2 \sum_{t_k \leqslant 1} j_k^2\right). \tag{2.29}$$

Let the first random variable on the right hand be denoted by Y; it follows that Y has the stated cdf – see Theorem 3.1 of [16] . As for $\sum_{t_k \leqslant 1} j_k^2$, we know that the j_k are the points of a $PRM(\nu)$ on the space $\mathbb{R}\setminus\{0\}$ (via projecting the $PRM(\mu)$ onto the second coordinate). As discussed in [51] , the mapping

$$\sum_k \varepsilon_{j_k} \mapsto \sum_k \varepsilon_{j_k^2} \tag{2.30}$$

induces a corresponding transformation on the mean measures:

$$\nu(dx) \mapsto \frac{\alpha}{2} x^{-\frac{\alpha}{2}} 1_{(0,\infty)}(x)\, dx =: \nu_2(dx). \tag{2.31}$$

Thus we see that $\sum_{t_k \leqslant 1} j_k^2$ is a $PRM(\nu_2)$, and by the Ito representation is an $\frac{\alpha}{2}$ stable totally right skewed random variable, which we will denote by S_2. Finally, it is easy to apply the continuous function $f(x,y) := \frac{x}{\sqrt{y}}$ to obtain the weak convergence result stated in the theorem. †

We end this section with an application of this result to "Value at Risk." We interpret the random variables X_t as a claim amount, and we are interested in the maximum of the claim amounts over a certain time horizon. Since the claim amount is positive, the series $\{X_t\}$ must be non-negative for every t (almost surely); to model this, we must impose that all the filter coefficients are non-negative, and that the inputs $\{Z_t\}$ are totally right skewed. This latter situation can be characterized by decreeing that $p = 1$ and $q = 0$ in line (2.3), and the limit random variable Y defined in (2.24) has cdf

$$\mathbb{P}[Y \leqslant x] = \begin{cases} \exp\{-x^{-\alpha}(\psi_+^\alpha)\} & x > 0 \\ 0 \quad x \leqslant 0 \end{cases} \tag{2.32}$$

We are therefore interested in the cumulative distribution function (cdf) of M_n, which we will denote by $F_M(x) = \mathbb{P}[M_n \leqslant x]$; knowing F_M, we can calculate all other interesting parameters/quantities of M_n. F_M also has the following interpretation: for a given threshhold x, the number $F_M(x)$ represents the probability that no claim amount exceeds that threshhold.

Thus we can use (2.32) along with subsampling to estimate $\frac{M_n}{\hat{\sigma}_n}$. An interesting question for the future is to obtain better practical methods for estimating the distribution of M_n with the limit random variable W. Even though the practitioner wishes to estimate M_n, at this point the best one can do is to estimate $\hat{\sigma}^{-1}M_n$ via W. Now W is a complicated random variable, which depends through ψ_+ , ψ_-, and Ψ_2 on the unknown filter coefficients; therefore it is impractical to assume knowledge of the exact distribution of W. Subsampling methods have proved very effective at estimating such limit distributions in similar contexts, e.g. the sample mean problem of the previous section. Section five of this chapter describes how subsampling is implemented.

2.4 Self-Normalized Periodogram

In this section we present a result which describes the asymptotic behavior of the self-normalized periodogram. In classical time series the spectral density function gives information about periodicities in the time series. It is typically estimated by its empirical version, the periodogram

$$\left| n^{-\frac{1}{2}} \sum_{t=1}^{n} X_t e^{-i\omega t} \right|^2, \tag{2.33}$$

where $\omega \in (-\pi, \pi]$. But since the classical spectral density does not exist in the heavy-tailed case (since it is a function of the covariances, which are infinite for the model (2.1)), we must adjust the definition slightly to

$$\left| a_n^{-1} \sum_{t=1}^{n} X_t e^{-i\omega t} \right|^2. \tag{2.34}$$

Now we wish to study a "self-normalized" periodogram, with the same motivation as for the sample mean and sample maximum; thus we must remove the unknown rate a_n from our statistic. The appropriate normalization is again the sample variance (this time without the square root), and so we arrive at our definition of the self-normalized periodogram at the frequency $\omega \in (-\pi, \pi]$:

$$\tilde{I}_n(\omega) := \frac{\left|\sum_{t=1}^{n} X_t e^{-i\omega t}\right|^2}{\sum_{t=1}^{n} X_t^2}. \tag{2.35}$$

This object was studied in its unnormalized form in [34], and (2.35) was examined in [35]. The authors restrict to the case that Z is in the normal domain of attraction of an α-stable law, and thus $a_n = n^{\frac{1}{\alpha}}$. Our results actually use a different method of proof – namely we truncate the model to an m-dependent version – and are slightly more general because Z need not be in the *normal* domain of attraction. In addition, the results here are proved for the more general scenario of a random field (treated in Chapter Four).

So we state the following weak convergence theorem, where α is now allowed to range over $(0, 2]$. As described at the end of section 2.1, this means square integrability for $\alpha = 2$, and it means $Z \in DOM(\alpha)$ and the other heavy-tailed assumptions for $\alpha \in (0, 2)$. Since it could be the case that $\alpha \leqslant 1$, the mean might not exist.

Theorem 2.4.1. *Let $\omega \in (-\pi, \pi]$. Then the following joint weak convergence result holds as $n \to \infty$:*

$$\left(a_n^{-2} \sum_{t=1}^{n} X_t^2, a_n^{-1} \sum_{t=1}^{n} X_t \cos \omega t, a_n^{-1} \sum_{t=1}^{n} X_t \sin \omega t \right) \stackrel{\mathcal{L}}{\Longrightarrow} (S_2, U, V) \tag{2.36}$$

where S_2 is a totally right skewed $\frac{\alpha}{2}$-stable random variable, and U and V are both α-stable random variables. From this we also know that

$$\tilde{I}_n(\omega) \stackrel{\mathcal{L}}{\Longrightarrow} \frac{U^2 + V^2}{S_2}. \tag{2.37}$$

Remarks Note that we do not need to normalize X_t by its mean (if the mean exists, which happens in the $\alpha > 1$ case), which is explained in the proof. The proof of this theorem is given in the third section of Chapter Four, where the general random field version is addressed.

2.5 Subsampling Applications

This final section of the chapter describes how subsampling methods may be used to apply the results of sections two and three. The object of subsampling is to use the "subsampling distribution estimator" as an approximation of the limit distribution. For more details and background on these methods, see the book Subsampling [48] .

Strong mixing is a sufficient condition on the dependence structure which insures the validity of subsampling. The strong mixing assumption requires that $\alpha_X(k) \to 0$ as $k \to \infty$; here $\alpha_X(k) := sup_{A,B}|\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|$, where A and B are events in the σ-fields generated by $\{X_t, t \leq l\}$ and $\{X_t, t \geq l+k\}$, respectively, for any $l \geq 0$. General conditions for a linear series to be strong mixing are given in [70] ; they require that the ψ_j tend to zero fast enough (with j), and that Z has an absolutely continuous distribution. The strong mixing condition is easily seen to be satisfied if the series is an $MA(m)$ model for some $m \in \mathbb{N}$, i.e., when only a finite number of the filter coefficients ψ_j are nonzero. In addition, if the series has an $AR(1)$ representation, i.e. if

$$X_t = \psi X_{t-1} + Z_t \tag{2.38}$$

for some ψ bounded by one in absolute value, then the time series is strong mixing; see [46] .

2.5.1 The Sample Mean

Define the "subsampling distribution estimator" of $T_n(\theta)$ to be the following empirical distribution function (edf):

$$K_b(x) := \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1_{\{T_{b,i} \leq x\}} \tag{2.39}$$

where $T_{b,i}$ is essentially the statistic $T_b(\theta)$ evaluated on the subseries $\{X_i, \cdots, X_{b+i-1}\}$ (but with the unknown θ replaced by the estimate $\overline{X}_n$) ; in other words,

$$T_{b,i} := b^{\frac{1}{2}} \frac{\overline{X}_{b,i} - \overline{X}_n}{\hat{\sigma}_{b,i}}. \tag{2.40}$$

The precise definitions of $\overline{X}_{b,i}$ and $\hat{\sigma}_{b,i}$ are as follows:

$$\overline{X}_{b,i} := \frac{1}{b} \sum_{t=i}^{b+i-1} X_t, \tag{2.41}$$

$$\hat{\sigma}_{b,i} := \sqrt{\frac{1}{b} \sum_{t=i}^{b+i-1} (X_t - \overline{X}_{b,i})^2}. \tag{2.42}$$

Corollary 2.5.1. *Let $\alpha \in (1, 2]$ (with the interpretation from the end of section 2.1), and consider a strong mixing linear time series with summable filter coefficients. So if $1 < \alpha < 2$, then we assume that conditions (2.2), (2.3), (2.8) hold. Then the subsampling distribution estimator K_b is consistent as an estimator of the true sampling distribution of T_n, denoted by $J_n(x) = \mathbb{P}\{T_n \leq x\}$. In other words, if $b \to \infty$ as $n \to \infty$ but with $b/n \to 0$, we have*

$$\sup_x |K_b(x) - J_n(x)| \xrightarrow{P} 0 \tag{2.43}$$

and in addition

$$K_b^{-1}(t) \xrightarrow{P} J^{-1}(t) \tag{2.44}$$

for any $t \in (0, 1)$; here $G^{-1}(t)$ is the t-quantile of distribution G, i.e., $G^{-1}(t) := \inf\{x : G(x) \geq t\}$.

Proof This result follows immediately from our Corollary 2.2.1 and Theorem 12.3.1 of [48] for self-normalized statistics under strong mixing - one of our standing hypotheses on the time series. The rate of convergence τ_n of $T_n(\theta)$ is $\sqrt{n}$, so it is clear that $\frac{\tau_b}{\tau_n} \to 0$. Finally, we note that the limit law in Corollary 2.2.1 has no point masses, since the random variables S and S_2 have continuous cumulative distribution functions (cdfs). Thus the limit cdf $J(x)$ is continuous, and the stated sup-norm convergence holds. †

From this result, we may use $K_b(x)$ as a cdf from which to draw quantiles and develop confidence intervals with asymptotic veracity. This is done as follows:

$$\begin{aligned} 1 - t &= \mathbb{P}[J_n^{-1}(t/2) \leq T_n \leq J_n^{-1}(1 - t/2)] \approx \mathbb{P}[K_b^{-1}(t/2) \leq T_n \leq K_b^{-1}(1 - t/2)] \qquad (2.45) \\ &= \mathbb{P}[\overline{X} - K_b^{-1}(1 - t/2)\frac{\hat{\sigma}}{\sqrt{n}} \leq \theta \leq \overline{X} - K_b^{-1}(t/2)\frac{\hat{\sigma}}{\sqrt{n}}] \end{aligned}$$

where $1 - t$ is the confidence level. Thus the approximate equal-tailed confidence interval for θ is

$$\left[\overline{X} - K_b^{-1}(1 - t/2) \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \overline{X} - K_b^{-1}(t/2) \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right]. \tag{2.46}$$

The benefit of this procedure is that the rate a_n need neither be known nor estimated in order to form the above confidence interval. If the $\{Z_t\}$ sequence were in the normal domain of attraction, then $a_n = n^{\frac{1}{\alpha}}$. The problem with using this rate is that the exact value of α is unknown. One can either estimate α empirically, or use self-normalization. The Hill estimator may be used to first estimate α, but its implementation requires the choice of a bandwidth parameter, and is not robust under dependence (cf. [52]). However, in the case that the domain of attraction is not normal, there will also be a slowly varying function to estimate, which is a problem of a higher order of difficulty. Providentially, self-normalization avoids the need to estimate a_n; in essence, self-normalization implicitly performs the estimation.

2.5.2 The Symmetrized Situation

We now consider a special subcase wherein the statistic has a symmetric distribution, i.e. $T_n(\theta)$ is equal to $-T_n(\theta)$ in law. This will permit us to make use of a symmetrized version of the subsampling distribution estimator. In the $\alpha = 2$ case this will give a faster rate of convergence, while still being robust under smaller values of α.

Throughout this section, we assume that the inputs $\{Z_t\}$ are α-stable random variables symmetric about their mean η. This will be abbreviated by sαs. In modeling of heavy-tailed data, such an additional assumption is commonplace; sαs random variables are one of the main examples we know of heavy-tailed random variables, which we can simulate and study fairly easily. If $\alpha = 2$, then we are referring to a normal random variable (which is clearly symmetric). The following result specializes Theorem 2.2.1 to the symmetric case:

Theorem 2.5.1. *Let $\alpha \in (1, 2]$ with the interpretation from the end of section 2.1, and consider a linear time series with summable filter coefficients and sαs inputs. So if $1 < \alpha < 2$, then assume that conditions (2.2), (2.3), and (2.8) hold as well. Then $T_n(\theta) \stackrel{\mathcal{L}}{\Longrightarrow} T$ as in Corollary 2.2.1, and the limit T is symmetric about zero.*

Proof Henceforth "symmetric" will mean symmetric about zero, unless another center is specified. First note that the sum of independent symmetric random variables is also symmetric. Now in the $\alpha = 2$ case, the limit is normal, and hence is symmetric, and the theorem holds (actually, it is true without assuming that the Z_t's are symmetric about η). So now we restrict to the case where $1 < \alpha < 2$.

The first assertion - convergence - is true from Corollary 2.2.1, as our hypotheses are only stronger. From the proof of Theorem 2.2.1 we may write $T_n(\theta)$ in the following manner:

$$T_n(\theta) = o_P(1) + \frac{\psi_\infty a_n^{-1} \sum_{t=1}^n (Z_t - \eta)}{\Psi_2 \sqrt{a_n^{-2} \sum_{t=1}^n (Z_t - \eta)^2}} \tag{2.47}$$

The right hand term is symmetric for each n, since it is a sum of iid symmetric random variables (note that the denominator is unchanged by the mapping $Z_t \mapsto -Z_t$). Hence the limit will also be symmetric (i.e. the random variable S from Theorem 2.2.1 will be symmetric). This proves the theorem. †

We now discuss the regular case, in which $Z \in \mathbb{L}^2$, and perhaps even moments of order higher than two exist. Consider the following moment and mixing conditions:

$$\alpha_X(k) \leq d^{-1} e^{-dk} \quad for \quad some \quad d > 0 \tag{2.48}$$

and

$$E|X_0|^s < \infty \quad for \quad some \quad s \geq 5. \tag{2.49}$$

If (2.48) and (2.49) hold true, and if the Cramér-type regularity conditions (2.3), (2.5) and (2.6) of [24] also hold true, then the following Edgeworth expansion holds uniformly in x:

$$P(T_n \leq x) = \Phi(x/|\psi_\infty|) + n^{-1/2} p(x) + O(n^{-1}) \tag{2.50}$$

where Φ is the standard normal cdf, and the $p(x)$ is a well-defined smooth function. If in addition to the above, conditions $A2[2]$, $\alpha 3[2]$, and $\alpha 4[2]$ of Bertail [5] hold true, then we also have

$$K_b(x) = \Phi(x/|\psi_\infty|) + b^{-1/2}p(x) + O_P(b^{-1}). \tag{2.51}$$

To make use of the higher order terms in the regular case discussed above, we will employ a notion of "robust interpolation" that was first proposed in [48] in the context of iid data. For this purpose, define

$$K_b^{symm}(x) := \frac{K_b(x) + 1 - K_b(-x)}{2} \tag{2.52}$$

and

$$K_b^{rob}(x) := \sqrt{\frac{b}{n}} K_b(x) + \left(1 - \sqrt{\frac{b}{n}}\right) K_b^{symm}(x). \tag{2.53}$$

The next result shows the efficacy of K_b^{rob}.

Corollary 2.5.2. *Make the same assumptions as in Theorem (2.5.1), and assume that $b \to \infty$ and $\frac{b}{n} \to 0$ as $n \to \infty$. Then we have*

$$\sup_x |K_b^{rob}(x) - J_n(x)| = o_P(1) \tag{2.54}$$

as $n \to \infty$.

If it so happens that $\alpha = 2$ (i.e., the inputs Z_t are iid $\mathbb{L}_2$ random variables), and the two Edgeworth expansions (2.50) and (2.51) hold true, then in addition we have

$$\sup_x |K_b^{rob}(x) - J_n(x)| = O_P(1/b). \tag{2.55}$$

Proof This follows from our Theorem 2.5.1 and Theorem 10.3.1 of [48]. †

Recall that equation (2.50) implies that

$$\sup_x |\Phi(x/\psi_\infty) - J_n(x)| = O\left(\frac{1}{\sqrt{n}}\right), \tag{2.56}$$

so that the error in the normal approximation to $J_n(x)$ is of order $O(\frac{1}{\sqrt{n}})$. Now observe that, if we let $\sqrt{n} = o(b)$ (and $b = o(n)$ as always), equation (2.55) implies

$$\sup_x |K_b^{rob}(x) - J_n(x)| = o_P\left(\frac{1}{\sqrt{n}}\right), \tag{2.57}$$

showing that $K_b^{rob}(x)$ exemplifies a higher-order-accuracy as compared to the normal approximation to J_n (when they are both applicable) under the Corollary's assumptions.

However, $K_b^{rob}(x)$ *remains* consistent even when the normal approximation breaks down, e.g. in the presence of heavy tails; thus the title "robust" is justified for $K_b^{rob}(x)$. As another attractive feature of $K_b^{rob}(x)$ versus the normal approximation (when the latter is applicable), is that the quantity ψ_∞ must be explicitly estimated for the normal approximation to be used in practice; subsampling does not suffer from this difficulty as it implicitly (and automatically) provides an estimate of ψ_∞.

2.5.3 The Sample Maximum and Periodogram

We first discuss how to apply subsampling to the sample maximum problem; all the notation and results will carry over for the periodogram as well. So let T_n denote $\frac{M_n}{\sqrt{\hat{\sigma}}}$, and define the "subsampling distribution estimator" of T_n to be the following empirical distribution function (edf):

$$K_b(x) := \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1_{\{T_{b,i} \leqslant x\}} \tag{2.58}$$

where $T_{b,i}$ is essentially the statistic T_b evaluated on the subseries $\{X_i, \cdots, X_{b+i-1}\}$; in other words,

$$T_{b,i} := \frac{M_{b,i}}{\hat{\sigma}_{b,i}}. \tag{2.59}$$

The precise definitions of $M_{b,i}$ and $\hat{\sigma}_{b,i}$ are as follows:

$$M_{b,i} := \bigvee_{t=i}^{b+i-1} X_t, \tag{2.60}$$

$$\hat{\sigma}_{b,i} := \sqrt{\sum_{t=i}^{b+i-1} X_t^2}. \tag{2.61}$$

Now define the sampling distribution cdf to be $W_n(x)$:

$$W_n(x) := \mathbb{P}\left[T_n \leqslant x\right] \tag{2.62}$$

and let $W(x)$ be the cdf of the random variable $W = \frac{Y}{\Psi_2\sqrt{S_2}}$ from 2.3.1 :

$$W(x) := \mathbb{P}\left[W \leqslant x\right]. \tag{2.63}$$

Then the following corollary results from a standard application of the subsampling theory of [48] to our Theorem 2.3.1 :

Corollary 2.5.3. *Under the $MA(\infty)$ model with the strong mixing hypothesis, the subsampling distribution estimator K_b is consistent as an estimator of the true sampling distribution of T_n, denoted by $W_n(x)$. In other words, if $b \to \infty$ as $n \to \infty$ but with $b/n \to 0$, we have*

$$\sup_x |K_b(x) - W_n(x)| \xrightarrow{P} 0 \tag{2.64}$$

and in addition

$$K_b^{-1}(t) \xrightarrow{P} W^{-1}(t) \tag{2.65}$$

for any $t \in (0,1)$, where $G^{-1}(t)$ denotes the t-th quantile of a given cdf G, i.e. $G^{-1}(t) = \inf\{x : G(t) \leqslant x\}$. Finally, we can also write

$$W_n\left(c_{n,b}(1-t)\right)\mathbb{P}\left[T_n(\theta) \leqslant c_{n,b}(1-t)\right] \to 1-t \tag{2.66}$$

as $n \to \infty$, where $c_{n,b}(1-t) = \inf\{x : K_b(x) \geqslant 1-t$ for any $t \in (0,1)$.

Proof We need to check the hypotheses of Theorem 11.3.1 of [48]. In that theorem, the rate τ_n is just $\frac{a_n}{a_n} = 1$, and $\frac{a_b}{a_n} \to 0$ as $n \to \infty$. The limit random variable in our Theorem 2.3.1 has no point masses, and its cdf $W(\cdot)$ is certainly continuous everywhere, so Assumption 11.3.1 of Politis, Romano, and Wolf (1999) is satisfied. Thus line (2.64) holds, so that (2.65) and (2.66) follow from this. †

Now for the self-normalized periodogram, simply let $T_n = \tilde{I}_n(\omega)$ for $\omega \in (-\pi, \pi]$, and define the subsampling distribution estimator by (2.58). In this case we define $T_{b,i}$ by

$$T_{b,i} := \frac{\left|\sum_{t=i}^{b+i-1} X_t e^{\sqrt{-1}\omega t}\right|^2}{\sum_{t=i}^{b+i-1} X_t^2}. \tag{2.67}$$

Then, using the same notation as above (W_n and W defined by equation (2.62) and (2.63) respectively), Corollary 2.5.3 holds true verbatim. In the proof we use the rate $\tau_n = \frac{a_n^2}{a_n^2} = 1$, which is the only difference.

This approximation for the self-normalized periodogram suffers from the same difficulties as the sample max, as far as the problem of estimating the distribution of the un-normalized periodogram is concerned. In terms of applications, the periodogram must first be smoothed over several frequencies to be consistent for the spectral density (see [35]); the limit behavior of the difference of the smoothed periodogram and the spectral density is an interesting topic for future research.

2.6 Simulation Studies

Herein we present the results of simulating the linear time series model given in (2.1), and we investigate the behavior of the subsampling distribution estimator for the sample mean problem. We specialize to having F be the cdf for a sαs random variable; for simplicity we take η (and thus θ too) to be zero. An $AR(1)$ process and a $MA(11)$ were considered separately, with the parameter α ranging between 1.2, 1.5, and 1.8. Then $10,000$ time series were generated of length 100 and 1000, each for a different subsampling length. The length b of the subsampling block was varied between 1 and 25. Confidence intervals for the mean – see (2.46) – were developed from the computed subsampling distribution estimator at the 90%, 95%, and 99% confidence levels, and we recorded the fraction of times that the constructed interval contained the mean of zero. Having done this for each subsampling length, the two tables below give the accuracy for certain values of b.

Table 2.1: **AR (1) Model** Entries are empirical averages of the equal-tailed subsampling confidence intervals for the mean; the nominal (target) coverage level is given at the top of each column. This is the case of data generated by an $AR(1)$ model.

AR (1) $n = 100$ $\alpha = 1.2$				AR (1) $n = 1000$ $\alpha = 1.2$			
b	90 %	95 %	99 %	b	90 %	95 %	99 %
1	.9670	.9820	1.0000	1	.9453	.9456	.9519
2	.8679	.9294	.9905	4	.7425	.8214	.9138
4	.7269	.8355	.9133	8	.6624	.7429	.9135
8	.6896	.7317	.8005	12	.6423	.7105	.8777
12	.6490	.6920	.7424	16	.6273	.7004	.8635
16	.6114	.6480	.6939	24	.6193	.6891	.7962
AR (1) $n = 100$ $\alpha = 1.5$				AR (1) $n = 1000$ $\alpha = 1.5$			
b	90 %	95 %	99 %	b	90 %	95 %	99 %
1	.9912	.9935	.9999	1	.9508	.9633	.9995
2	.9686	.9820	.9994	4	.8982	.9371	.9569
4	.8735	.9359	.9772	8	.8144	.8766	.9820
8	.7829	.8325	.8905	12	.7941	.8568	.9710
12	.7378	.7841	.8395	16	.7639	.8334	.9493
16	.7106	.7526	.8066	24	.7514	.8121	.9086
AR (1) $n = 100$ $\alpha = 1.8$				AR (1) $n = 1000$ $\alpha = 1.8$			
b	90 %	95 %	99 %	b	90 %	95 %	99 %
1	.9925	.9927	1.0000	1	.9999	1.0000	1.0000
2	.9930	.9937	1.0000	4	.9510	.9788	.9980
4	.9544	.9841	.9953	8	.8969	.9436	.9940
8	.8709	.9130	.9541	12	.8656	.9243	.9839
12	.8251	.8668	.9179	16	.8540	.9071	.9750
16	.7949	.8393	.8916	24	.8395	.8949	.9568

Table 2.2: **MA (11) Model** Entries are empirical averages of the equal-tailed subsampling confidence intervals for the mean; the nominal (target) coverage level is given at the top of each column. This is the case of data generated by an $MA(11)$ model.

MA (11) $n = 100$ $\alpha = 1.2$				MA (11) $n = 1000$ $\alpha = 1.2$			
b	90 %	95 %	99 %	b	90 %	95 %	99 %
1	.9992	.9998	.9999	1	.9446	.9464	1.0000
2	.9973	.9995	.9998	4	.8903	.9381	.9693
4	.9748	.9956	.9973	8	.8239	.9032	.9988
8	.8981	.9648	.9746	12	.7707	.8579	.9888
12	.8483	.8944	.9100	16	.7285	.8179	.9697
16	.7768	.8183	.8307	24	.6906	.7748	.9321
MA (11) $n = 100$ $\alpha = 1.5$				MA (11) $n = 1000$ $\alpha = 1.5$			
b	90 %	95 %	99 %	b	90 %	95 %	99 %
1	.9995	.9999	1.0000	1	.9999	1.0000	1.0000
2	.9991	.9995	1.0000	4	.9423	.9439	.9999
4	.9860	.9982	.9998	8	.9251	.9401	.9998
8	.9284	.9740	.9891	12	.8984	.9280	.9991
12	.8919	.9267	.9492	16	.8650	.9127	.9949
16	.8286	.8645	.8899	24	.8255	.8833	.9827
MA (11) $n = 100$ $\alpha = 1.8$				MA (11) $n = 1000$ $\alpha = 1.8$			
b	90 %	95 %	99 %	b	90 %	95 %	99 %
1	.9999	1.0000	1.0000	1	.9831	.9949	1.0000
2	.9998	1.0000	1.0000	4	.9696	.9811	1.0000
4	.9942	.9988	.9999	8	.9606	.9710	1.0000
8	.9608	.9855	.9939	12	.9424	.9597	1.0000
12	.9186	.9452	.9639	16	.9240	.9521	.9992
16	.8765	.9035	.9226	24	.9041	.9430	.9959

The AR(1) process is simply

$$X_t = .5\, X_{t-1} + Z_t$$

for $t = 1, 2, \cdots, n$, and $X_0 := Z_0$; the sequence $\{Z_t\}$ are *iid* sαs random variables with zero mean. It is elementary that the above time series can also be written as

$$X_t = \sum_{j=0}^{\infty} (0.5)^j Z_{t-j}$$

which fits our model assumptions in this chapter.

The MA(11) process is

$$X_t = \sum_{j=0}^{10} \psi_j Z_{t+j}$$

for $t = 1, 2, \cdots, n$, and for coefficients $\{\psi_j\}$ equal to

$$0.03, 0.05, 0.07, 0.10, 0.15, 0.20, 0.15, 0.10, 0.07, 0.05, 0.03.$$

Both models are strong mixing – see the discussion in the beginning of section 2.5 and line (2.38).

These simulation studies show that the optimal block size b depends both on α and the degree of dependence. The $MA(11)$ model has less dependence than the $AR(1)$ model, but the optimal block lengths were larger for the former model. In general, lower values of α resulted in undercoverage, especially in the $AR(1)$ model; the optimal block size tended to be quite small (sometimes 1 or 2). For $\alpha = 1.8$, the coverage was optimized for larger values of b. Thus there appears to be some relationship between α and the optimal b. The $MA(11)$ model had coverages closer to the true percentage, and thus performed better than the $AR(1)$; thus the amount of dependence also seems to affect the optimal block size.

Our simulation results are encouraging but also demonstrate the need for a practical way of picking the subsampling size b , since the finite-sample performance of the confidence intervals depends on b. Some insights on the theory and practice of optimally picking b are available in the regular *iid* case; see [28] or [48]. Little is known in the case treated here where dependence, heavy tails, and self-normalization complicate the picture considerably. Future work will focus on this important topic.

Chapter 3

Stable Processes

3.1 Stable Moving Average Processes

As discussed in Chapter Two, the common statistical problem of estimating the mean in a model of heavy-tailed and generally dependent random variables is made intractable by the following difficulty: the rate of convergence of the partial sums depends directly upon α, an unknown model parameter which measures the heaviness of the probability density function's tails. In the past, the Hill Estimator was utilized to first estimate α, and thereby construct the correct rate of convergence for the partial sums. The weaknesses of this approach, namely poor robustness under dependence and difficulty with optimally choosing the number of order statistics to be used in Hill's estimator – see [52] , have instigated the search for a better method; the obscurity surrounding the limit distribution's quantiles (since the limit is in general not Gaussian) has further compounded these troubles. A solution is given by self-normalization and subsampling in the case of discrete moving average processes, as in section two of Chapter Two.

The goal of this chapter is to apply these same tools to the stable moving average model presented in [54] (also see [55]). In the first paper the authors study the sample autocorrelations, with the purpose of demonstrating that a random limit is generally obtained. Here we study the partial sums and the periodogram, obtain their weak limits, and present a self-normalized limit theorem to which subsampling methods may be applied. The rest of this section contains background material on α-stable stochastic integrals, and the theoretical results are found in section two. The third section discusses subsampling and its relevant applications.

We now present a definition of the stable integral via α-stable random measures. Thus we first describe what an α-stable random measure is, and discuss its properties, and then we exhibit the construction of the stable integral. The principle example that we are interested in is the moving average stable integral, which is obtained from integrands with a special form. An excellent source for this material is Chapter Three of *Stable non-Gaussian Processes* [64]; additional background material is included in Appendix C. Throughout the rest of this chapter, we assume that $\alpha \in (1,2)$, so that all α-stable random variables have finite mean. Note that the $\alpha = 2$ case is well-defined too, and one obtains a Gaussian random measure. Most of the same results are already known to hold true in this situation, and thus will not be considered here.

Given an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathbb{L}_0(\Omega)$ denote the space of all positive random variables on it. Let $(E, \mathcal{E}, m)$ be a measure space, and $\beta : E \longrightarrow [-1,1]$ is a measurable function. Let $\mathcal{E}_0 = \{A \in \mathcal{E} : m(A) < \infty\}$ be the sets of finite m-measure.

Now we define the desired measure: let $\mathbb{M}$ be an independently scattered σ-additive set function

$$\mathbb{M} : \mathcal{E}_0 \longrightarrow \mathbb{L}_0(\Omega) \tag{3.1}$$

such that for every $A \in \mathcal{E}_0$, we have

$$\mathbb{M}(A) \sim S_\alpha \left((m(A))^{\frac{1}{\alpha}}, \frac{\int_A \beta(x)\, m(dx)}{m(A)}, 0 \right), \tag{3.2}$$

i.e. an α-stable random variable with scale $(m(A))^{\frac{1}{\alpha}}$, skewness $\frac{\int_A \beta(x)\, m(dx)}{m(A)}$, and location 0. Independently scattered means that disjoint sets produce independent random variables, and σ-additive refers to its properties as a measure. Then $\mathbb{M}$ is called an **α-stable random measure** on $(E, \mathcal{E})$ with **control measure** m and **skewness intensity** β.

The following example will be considered: let $(\mathbb{R}, \mathcal{R}, dx)$ be a measure space, i.e. the Borel sets on the real line under Lebesgue measure. We impose that the skewness function be constant, so $\beta(x) = \beta$. This defines an α-stable random measure with Lebesgue control measure. Next we define a family of functions f indexed by time t as follows:

$$f_t(x) := f(t+x) \quad \forall x \tag{3.3}$$

Here $t \in \mathbb{N} \subset \mathbb{R}$. To make the integrals well-defined, it suffices to take $f \in \mathbb{L}^\alpha(\mathbb{R}, dx)$; but we will assume that f is continuous and bounded for almost every $x \in \mathbb{R}$ (with respect to Lebesgue measure). This of course implies that f is in every $\mathbb{L}^p$ space, but the additional assumption is necessary: the function

$$\sum_{t \in \mathbb{Z}} f(t+x) 1_{(0,1)}(x) \tag{3.4}$$

appearing in line (3.4) need not be integrable if f is merely integrable. Thus we have

$$X_t = \int_{\mathbb{R}} f_t(x) \mathbb{M}(dx) = \int_{\mathbb{R}} f(t+x) \mathbb{M}(dx). \tag{3.5}$$

3.2 Self-Normalized Sample Mean

For the rest of the chapter we assume that the time series has the form

$$Y_t := \int_{\mathbb{R}} f(t+x) \mathbb{M}(dx) + \theta, \tag{3.6}$$

so that clearly $Y_t = X_t + \theta$, where X_t is given by (3.5). Thus we have observations

$$Y_1, Y_2, \cdots, Y_n. \tag{3.7}$$

The first result concerns the partial sums; thus we are interested in the asymptotic behavior of

$$\overline{Y} - \theta, \tag{3.8}$$

which we will show converges at rate $n^{1-\frac{1}{\alpha}}$ (here $\overline{Y}$ denotes the sample mean of the observations Y_t). This expression, when multiplied by this rate, is equal to

$$S_n := n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} X_t. \tag{3.9}$$

Theorem 3.2.1. *We have the following convergence:*

$$S_n \stackrel{\mathcal{L}}{\Longrightarrow} S \tag{3.10}$$

where

$$S := \int_0^1 \sum_{t \in \mathbb{Z}} f(t+x)\, \mathbb{M}(dx). \tag{3.11}$$

$\mathbb{M}$ *is a strictly* α*-stable random measure with constant skewness as described in (3.2), defined on* $(\mathbb{R}, dx)$*.*

Example : $MA(\infty)$ model Let f have the following form for a summable sequence $\{\psi_j\}$, as in (2.1) :

$$f(u) := \sum_{j \in \mathbb{Z}} \psi_j 1_{(j,j+1]}(u), \tag{3.12}$$

which is bounded and continuous except on the integers, which has measure zero anyways. It is a simple exercise to show that

$$X_t = \sum_{j \in \mathbb{Z}} \psi_j \mathbb{M}((j-t, j-t+1]) \tag{3.13}$$

and by the disjointness of the sets $(j-t, j-t+1]$ for fixed t and $j \in \mathbb{Z}$, we see that the latter terms in (3.13) are *iid* α-stable random variables – call them Z_{t-j}. Then we have

$$X_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j}, \tag{3.14}$$

the usual $MA(\infty)$ model. In [16] and (2.10) of this work,, the partial sums are shown to converge to

$$\sum_{j \in \mathbb{Z}} \psi_j \cdot G \tag{3.15}$$

for an α-stable random variable G of appropriate scale. Plugging our choice of f into (3.11) yields

$$\sum_{j \in \mathbb{Z}} \psi_j \cdot \int_0^1 \mathbb{M}(dx), \tag{3.16}$$

which gives perfect agreement with the numerator of (2.10) if we let $G := \int_0^1 \mathbb{M}(dx)$.

Proof of Theorem 3.2.1 We consider the random sum

$$n^{1-\frac{1}{\alpha}}\overline{X}_n = n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} X_t = \int_{\mathbb{R}} n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} f(t+x)M(dx) =: S_n \tag{3.17}$$

which is an α-stable random variable with scale

$$\left(\int_{\mathbb{R}} \left| n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} f(t+x)\right|^{\alpha} dx\right)^{\frac{1}{\alpha}}. \tag{3.18}$$

Our first objective is to show that this quantity is finite, so that the random variable is well-defined. It will follow that the skewness is finite as well. The αth power of the scale is the following quantity:

$$\begin{aligned}
&\int_{\mathbb{R}} \left| n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} f(t+x)\right|^{\alpha} dx \qquad (3.19)\\
&= \frac{1}{n}\int_{\mathbb{R}} \left|\sum_{t=1}^{n} f(t+x)\right|^{\alpha} dx\\
&= \frac{1}{n}\int_{\mathbb{R}} |g_n(x)|\,|g_n(x)|^{\alpha-1}\,dx\\
&\leqslant \frac{1}{n}\|g_n\|_{\infty}\||g_n|^{\alpha-1}\|_1
\end{aligned}$$

where $g_n(x) := \sum_{t=1}^{n} f(t+x)$, and we've used the Hölder Inequality.

Claim We claim that $h(x) := \sum_{t\in\mathbb{Z}} |f(t+x)|$ has bounded uniform norm, i.e. $\|h\|_{\infty} \leqslant C < \infty$, and therefore $\|g_n\|_{\infty} \leqslant C < \infty$ as well. First note that h is periodic of length one, so we only have to take its essential supremum over the unit interval. Moreover, $\sup_n |g_n(x)| \leqslant h(x)$, so that the second assertion follows at once. Let $A := \{x \in [0,1] : h(x) = \infty\}$, and suppose for the sake of contradiction that the Lebesgue measure (λ) of A is positive. Then

$$\begin{aligned}
\int_{\mathbb{R}} |f(x)|\,dx &= \sum_{t\in\mathbb{Z}}\int_t^{t+1} |f(x)|\,dx \qquad (3.20)\\
&= \int_0^1 \sum_{t\in\mathbb{Z}} |f(t+x)|\,dx\\
&= \int_0^1 h(x)\,dx \geqslant \int_A h(x)\,dx\\
&= \infty\cdot\lambda(A) = \infty,
\end{aligned}$$

which contradicts the fact that $f \in \mathbb{L}_1$. So $\lambda(A) = 0$ afterall, and h has finite uniform norm.

Returning to the inequalities above, we have

$$\begin{aligned} \int_{\mathbb{R}} \left| n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} f(t+x) \right|^{\alpha} dx &\leqslant \frac{1}{n} C \int_{\mathbb{R}} |g_n(x)|^{\alpha-1}\, dx \\ &\leqslant \frac{1}{n} C \int_{\mathbb{R}} \sum_{t=1}^{n} |f(t+x)|^{\alpha-1}\, dx \\ &= C \int_{\mathbb{R}} |f(x)|^{\alpha-1}\, dx < \infty, \end{aligned} \tag{3.21}$$

which shows that the partial sums have finite scale for every n arbitrarily large, as desired.

We must check that the limit random variable S in (3.11) has finite scale; the αth power of its scale is defined to be

$$\begin{aligned} \int_0^1 \left| \sum_{t\in\mathbb{Z}} f(t+x) \right|^{\alpha} dx &= \int_0^1 |g(x)|^{\alpha}\, dx \\ &= \int_0^1 |g(x)|\, |g(x)|^{\alpha-1}\, dx \\ &= \|g1_{(0,1)}\|_{\infty} \int_0^1 |g(x)|^{\alpha-1}\, dx \\ &= \|h\|_{\infty} \int_0^1 \sum_{t\in\mathbb{Z}} |f(t+x)|^{\alpha-1}\, dx \\ &\leqslant C \int_{\mathbb{R}} |f(x)|^{\alpha-1}\, dx < \infty, \end{aligned} \tag{3.22}$$

where $g(x) := \sum_{t\in\mathbb{Z}} f(t+x)$. We need one technical lemma in order to proceed with the analysis:

Lemma 3.2.1. *Define $x^{<\alpha>}$ for any real number x to be $x^{<\alpha>} := |x|^{\alpha}\, sign(x)$ as in (C.7). Let $1 < \alpha < 2$ and consider any $a, b, c \in \mathbb{R}$. Then the following inequality is true:*

$$|(a+b)^{<\alpha>} - (a+c)^{<\alpha>}| \leqslant 2^{\alpha}(|b|^{\alpha} + |c|^{\alpha}) \tag{3.23}$$

This proof is technical, and is expounded in section four of this chapter. With this result, we can now state the following critical Lemma:

Lemma 3.2.2.

$$\frac{1}{n}\int_{\mathbb{R}}\left(\sum_{t=1}^{n} f(t+x)\right)^{<\alpha>} dx \to \int_0^1 \left(\sum_{t\in\mathbb{Z}} f(t+x)\right)^{<\alpha>} dx \tag{3.24}$$

$$\frac{1}{n}\int_{\mathbb{R}}\left|\sum_{t=1}^{n} f(t+x)\right|^{\alpha} dx \to \int_0^1 \left|\sum_{t\in\mathbb{Z}} f(t+x)\right|^{\alpha} dx$$

The proof of this Lemma is deferred until the final section of the chapter, due to its length; we now conclude the proof of Theorem 3.2.1. We see that the scale and skewness parameters of S_n, which are decribed in Theorem 3.2.1, converge to the scale and skewness parameters of S; since both objects are well-defined α-stable random variables, weak convergence follows (also see Proposition 3.5.1 of [64]). †

Next we examine the heavy-tailed periodogram for the series Y_t, which is defined as

$$N^{-\frac{2}{\alpha}}\left|\sum_{t=1}^{n} Y_t e^{-i\omega t}\right|^2, \tag{3.25}$$

where $\omega \in (0, 2\pi]$. The periodogram is commonly used to give a rough estimate of the spectral density function of the time series when the latter exists (which is not the case in our setting); here we shall put it to another use. The following theorem gives the weak limit behavior of the periodogram of the X_t series, and its corollary shows that this is the same as for the Y_t series for certain choices of ω.

Theorem 3.2.2. *Let $\omega \in (0, 2\pi]$. Then we have the following convergence result:*

$$n^{-\frac{2}{\alpha}}\left|\sum_{t=1}^{n} X_t e^{-i\omega t}\right|^2 \stackrel{\mathcal{L}}{\Longrightarrow} W(\omega)^2 + U(\omega)^2 \tag{3.26}$$

where

$$W(\omega) := \int_0^1 \sum_{t\in\mathbb{Z}} \cos\omega t f(x+t)\mathbb{M}(dx) \tag{3.27}$$

and

$$U(\omega) := \int_0^1 \sum_{t\in\mathbb{Z}} \sin\omega t f(x+t)\mathbb{M}(dx) \tag{3.28}$$

are α-stable random variables.

Remark As mentioned in the introduction of this chapter, the case that $\alpha = 2$ is classical, and is well-known (for example, see [11]). Thus this result gives some robustness under the possibility of $\alpha < 2$, as discussed in Chapter Two.

Proof First we notice that the periodogram can be written as

$$\left|\sum_{t=1}^{n} X_t e^{-i\omega t}\right|^2 = \left(\sum_{t=1}^{n} X_t \cdot \cos\omega t\right)^2 + \left(\sum_{t=1}^{n} X_t \cdot \sin\omega t\right)^2. \tag{3.29}$$

According to the notation introduced in Theorem 3.2.2, this is equal to

$$\left(\sum_{t=1}^{n} W_t\right)^2 + \left(\sum_{t=1}^{n} U_t\right)^2. \tag{3.30}$$

We first claim that

$$n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} W_t \stackrel{\mathcal{L}}{\Longrightarrow} W(\omega) \tag{3.31}$$

$$n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} U_t \stackrel{\mathcal{L}}{\Longrightarrow} U(\omega), \tag{3.32}$$

which follow in the same manner as the proof of Lemma 3.2.1, since

$$W_t = X_t \cdot \cos\omega t = \int_{\mathbb{R}} f(t+x)\cos\omega t\, \mathbb{M}(dx) \tag{3.33}$$

and

$$U_t = X_t \cdot \sin\omega t = \int_{\mathbb{R}} f(t+x)\sin\omega t\, \mathbb{M}(dx). \tag{3.34}$$

Now the new integrands $f(t+x)\cos\omega t$ and $f(t+x)\sin\omega t$ no longer have the transport form, but this will not affect the proof for the following reason: in the proof of Lemma 3.2.2, we always eventually estimate the absolute value of $f(t+x)$, and clearly this is the same as the absolute value of $f(t+x)\ \cos\omega t$ and $f(t+x)\ \sin\omega t$. Thus the calculations in the proof of Theorem 3.2.1 are still valid when these trigonometric terms are included.

Moreover, a joint convergence result holds

$$n^{-\frac{1}{\alpha}}\left(\sum_{t=1}^{n} W_t, \sum_{t=1}^{n} U_t\right) \overset{\mathcal{L}}{\Longrightarrow} (W(\omega), U(\omega)), \tag{3.35}$$

which the following calculation demonstrates:

$$\begin{aligned}
&\mathbb{E}\exp\left(i\eta n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} W_t + i\tau n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} U_t\right) \qquad (3.36)\\
&= \mathbb{E}\exp\left(i\eta\int_{\mathbb{R}} n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} f(t+x)\cos\omega t\,\mathbb{M}(dx) + i\tau\int_{\mathbb{R}} n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} f(t+x)\sin\omega t\,\mathbb{M}(dx)\right)\\
&= \mathbb{E}\exp\left(i\int_{\mathbb{R}} n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} f(t+x)\,[\eta\cos\omega t + \tau\sin\omega t]\,\mathbb{M}(dx)\right)\\
&\to \mathbb{E}\exp\left(i\int_0^1 \sum_{t\in\mathbb{Z}} f(t+x)\,[\eta\cos\omega t + \tau\sin\omega t]\,\mathbb{M}(dx)\right)\\
&= \mathbb{E}\exp\left(i\eta W(\omega) + i\tau U(\omega)\right),
\end{aligned}$$

for any real η, τ. Since the integrand $f(t+x)\,[\eta\cos\omega t + \tau\sin\omega t]$ is quite similar to $f(t+x)$, the convergence in the penultimate line of (2.36) holds for the reasons given in the paragraph above.

Now we apply the continuous mapping $(x, y) \mapsto x^2 + y^2$, and obtain

$$n^{-\frac{2}{\alpha}}\left(\left(\sum_{t=1}^{n} W_t\right)^2 + \left(\sum_{t=1}^{n} U_t\right)^2\right) \overset{\mathcal{L}}{\Longrightarrow} W(\omega)^2 + U(\omega)^2, \tag{3.37}$$

as desired. †

In the following result, we replace X_t in Theorem 3.2.1 by Y_t :

Corollary 3.2.1. *There exists a choice of $\omega \in (0, 2\pi]$ such that*

$$n^{-\frac{2}{\alpha}} I_n(\omega) := n^{-\frac{2}{\alpha}}\left|\sum_{t=1}^{n} Y_t e^{-i\omega t}\right|^2 \overset{\mathcal{L}}{\Longrightarrow} W(\omega)^2 + U(\omega)^2. \tag{3.38}$$

In fact, as long as ω is not an integer multiple of 2π, equation (3.38) holds true.

Proof The periodogram for the Y_t series is

$$\left|\sum_{t=1}^{n} Y_t e^{-i\omega t}\right|^2 \tag{3.39}$$

$$= \left(\sum_{t=1}^{n}(X_t+\theta)\cos\omega t\right)^2 + \left(\sum_{t=1}^{n}(X_t+\theta)\sin\omega t\right)^2$$

$$= \left(\sum_{t=1}^{n} X_t\cos\omega t\right)^2 + 2\theta\left(\sum_{t=1}^{n} X_t\cos\omega t\right)\left(\sum_{s=1}^{n}\cos\omega s\right) + \theta^2\left(\sum_{t=1}^{n}\cos\omega t\right)^2 \tag{3.40}$$

$$+ \left(\sum_{t=1}^{n} X_t\sin\omega t\right)^2 + 2\theta\left(\sum_{t=1}^{n} X_t\sin\omega t\right)\left(\sum_{s=1}^{n}\sin\omega s\right) + \theta^2\left(\sum_{t=1}^{n}\sin\omega t\right)^2. \tag{3.41}$$

It is easy to see that the sums of deterministic trigonometric terms are bounded as n increases. Let $z = e^{i\omega}$, and suppose that $z \neq 1$ (i.e. $\omega \notin 2\pi\mathbb{Z}$). So then

$$\left|\sum_{t=1}^{n} e^{i\omega t}\right| = \left|\sum_{t=1}^{n} z^t\right| \leqslant \frac{2}{|1-z|} \tag{3.42}$$

holds for every n. However, we now see that (3.42) bounds the trigonometric terms. Let $Re(x)$ and $Im(x)$ denote the real and imaginary parts of a complex number x. Then

$$\left|\sum_{t=1}^{n}\cos\omega t\right| = \left|Re(\sum_{t=1}^{n} z^t)\right| \leqslant \left|\sum_{t=1}^{n} z^t\right| \leqslant \frac{2}{|1-z|} \tag{3.43}$$

$$\left|\sum_{t=1}^{n}\sin\omega t\right| = \left|Im(\sum_{t=1}^{n} z^t)\right| \leqslant \left|\sum_{t=1}^{n} z^t\right| \leqslant \frac{2}{|1-z|}$$

for all n, so that the latter two terms in lines (3.40) and (3.41) are $o_P(n^{\frac{1}{\alpha}})$ and $o_P(n^{\frac{2}{\alpha}})$ respectively. Therefore,

$$n^{-\frac{2}{\alpha}}\left|\sum_{t=1}^{n} Y_t e^{-i\omega t}\right|^2 = o_P(1) + n^{-\frac{2}{\alpha}}\left|\sum_{t=1}^{n} X_t e^{-i\omega t}\right|^2 \tag{3.44}$$

which establishes the Corollary. †

Finally, we use the square root of the periodogram to normalize the partial sums, and thereby obtain a statistic that does not depend upon α :

Theorem 3.2.3. *Let $\omega \in (0, 2\pi)$, and define the random variables W_t and U_t as in equations (3.33) and (3.34):*

$$W_t := X_t \cdot \cos \omega t \tag{3.45}$$

$$U_t := X_t \cdot \sin \omega t$$

Then we have the following joint convergence:

$$\left(S_n, n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} W_t, n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} U_t \right) \stackrel{\mathcal{L}}{\Longrightarrow} (S, W(\omega), U(\omega)), \tag{3.46}$$

where the limit random variables are defined by (3.11), (3.31), (3.32). In addition,

$$\frac{n\left(\overline{Y} - \theta\right)}{\sqrt{I_n(\omega)}} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{S}{\sqrt{W(\omega)^2 + U(\omega)^2}}. \tag{3.47}$$

Example : $MA(\infty)$ model Continuing the example given after Theorem 3.2.1, we see that the limit in the right hand side of line (3.47) reduces to

$$\frac{\sum_{j\in\mathbb{Z}} \psi_j \mathbb{M}((0,1])}{\sqrt{\left[\left(\sum_{j\in\mathbb{Z}} \psi_j cos\omega j\right)^2 + \left(\sum_{j\in\mathbb{Z}} \psi_j sin\omega j\right)^2\right] (\mathbb{M}((0,1]))^2}}. \tag{3.48}$$

Note that this is random, and not degenerate (the denominator has the square root of an $\frac{\alpha}{2}$-stable random variable, whereas the top has an α-stable random variable, and these do not cancel!).

Proof of Theorem 3.2.3 This proof follows along the same lines as the Proof of Theorem 3.2.2 : it is easy to establish the joint limit result

$$\left(n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} X_t, n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} W_t, n^{-\frac{1}{\alpha}} \sum_{t=1}^{n} U_t \right) \stackrel{\mathcal{L}}{\Longrightarrow} (S, W(\omega), U(\omega)) \tag{3.49}$$

by employing the same techniques. Applying the continuous mapping $(x, y, z) \mapsto \frac{x}{\sqrt{y^2+z^2}}$ (note that the random variables in the second and third argument are nonzero with probability one, i.e. their laws do not contain a point mass at zero), we obtain the desired result:

$$\frac{\sum_{t=1}^{n} (Y_t - \theta)}{\sqrt{\left(\sum_{t=1}^{n} Y_t \cdot \cos \omega t\right)^2 + \left(\sum_{t=1}^{n} Y_t \cdot \sin \omega t\right)^2}} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{S}{\sqrt{W(\omega)^2 + U(\omega)^2}} \quad \dagger \tag{3.50}$$

Theorem 3.2.3 demonstrates the power of self-normalization, but its failings are evident in the enigmatic limit on the right hand side of the convergence. We will not trouble ourselves with the calculation of its distribution – an endeavor already statistically fruitless, since the random variable depends explicitly upon the function f, which is part of the model (and therefore hidden in the data). The next section explains how subsampling technology advantageously exploits this scenario.

3.3 Subsampling Applications

As in the previous section, we will fix a choice of $\omega \in (0, 2\pi)$ which satisfies Corollary 3.2.1. Let $T_n(\theta)$ denote the ratio in (3.47), i.e.,

$$T_n(\theta) := \frac{n\left(\left(\frac{1}{n}\sum_{t=1}^{n} Y_t\right) - \theta\right)}{\sqrt{\left(\sum_{t=1}^{n} Y_t \cdot \cos\omega t\right)^2 + \left(\sum_{t=1}^{n} Y_t \cdot \sin\omega t\right)^2}}. \tag{3.51}$$

We wish to construct an estimator of the distribution of $T_n(\theta)$, which we do via evaluating it upon subsamples of the original time series which have length b; more will be said about the choice of b later on. So we define the "subsampling distribution estimator" of $T_n(\theta)$ to be the following empirical distribution function (edf):

$$K_b(x) := \frac{1}{n-b+1}\sum_{i=1}^{n-b+1} 1_{\{T_{b,i} \leq x\}} \tag{3.52}$$

where $T_{b,i}$ is essentially T_b evaluated on the subseries $\{Y_i, \cdots, Y_{b+i-1}\}$; in other words,

$$T_{b,i} := b\frac{\overline{Y}_{b,i} - \overline{Y}_n}{\sqrt{I_{b,i}(\omega)}}. \tag{3.53}$$

The precise definitions of $\overline{Y}_{b,i}$ and $I_{b,i}(\lambda)$ are as follows:

$$\overline{Y}_{b,i} := \frac{1}{b}\sum_{t=i}^{b+i-1} Y_t, \tag{3.54}$$

$$I_{b,i}(\omega) := \left|\sum_{t=i}^{b+i-1} Y_t e^{-it\omega}\right|^2.$$

The object of subsampling is to use the subsampling distribution estimator as an approximation of the limit distribution in Theorem 3.2.3 . For more details and background on these methods, see [48] . The subsample size b must be chosen significantly smaller than the real sample size n, and in fact we require that $b = b(n)$ is some function of n such that $\frac{b}{n} \to 0$ as $n \to \infty$, but also that $b \to \infty$.

Strong mixing is a sufficient condition on the dependence structure of the time series to insure the validity of subsampling. The strong mixing assumption requires that $\alpha_X(k) \to 0$ as $k \to \infty$; here $\alpha_X(k) := sup_{A,B}|\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|$, where A and B are events in the σ-fields generated by $\{X_t, t \leq l\}$ and $\{X_t, t \geq l + k\}$, respectively, for any $l \geq 0$.

The strong mixing condition is easily seen to be satisfied if the series is m-dependent, i.e. X_t and X_{t+m} are independent for every t. For the following corollary, we assume that the time series X_t is strong mixing; necessary conditions for stable moving averages of the form (3.5) to be strong mixing are generally unknown. However, if f is compactly supported (which is stronger than our previous assumptions) almost everywhere, then m-dependence follows, where m is the diameter of the support region. This is follows from a result in [64], which says that two random variables X_t and X_s given by (3.5) are independent if and only if

$$f_t(x) \cdot f_s(x) = 0 \quad a.e.(dx). \tag{3.55}$$

But if $|t - s| > m$ this is clearly the case, so this implies m-dependence. The following corollary is a standard application of subsampling to the sample mean problem:

Corollary 3.3.1. *The subsampling distribution estimator K_b defined above is consistent as an estimator of the true sampling distribution of $T_n(\theta)$, denoted by $J_n(x) = \mathbb{P}\{T_n(\theta) \leq x\}$. In other words, if $b \to \infty$ as $n \to \infty$ but with $b/n \to 0$, we have*

$$\sup_x |K_b(x) - J_n(x)| \xrightarrow{P} 0 \tag{3.56}$$

and in addition

$$K_b^{-1}(t) \xrightarrow{P} J^{-1}(t) \tag{3.57}$$

for any $t \in (0,1)$; here $G^{-1}(t)$ is the t-quantile of distribution G, i.e., $G^{-1}(t) := \inf\{x : G(x) \geq t\}$.

Proof This result follows immediately from Theorem 3.2.3 and Theorem 12.2.2 of [48], using the rate of convergence $\tau_n = n$ and the fact that the limit distribution has no point masses. †

From this result, we may use $K_b(x)$ as a cdf (cumulative distribution function) from which to draw quantiles and develop confidence intervals with asymptotic veracity. This is done as follows:

$$1 - t = \mathbb{P}[J_n^{-1}(t/2) \leq T_n(\theta) \leq J_n^{-1}(1-t/2)] \approx \mathbb{P}[K_b^{-1}(t/2) \leq T_n(\theta) \leq K_b^{-1}(1-t/2)] \tag{3.58}$$

$$= \mathbb{P}[\overline{Y} - K_b^{-1}(1-t/2)\frac{I_n(\omega)}{n} \leq \theta \leq \overline{Y} - K_b^{-1}(t/2)\frac{I_n(\omega)}{n}]$$

where $1-t$ is the desired confidence level. Thus the approximate equal-tailed confidence interval for θ is

$$\left[\overline{Y} - K_b^{-1}(1-t/2) \cdot \frac{\sqrt{I_n(\omega)}}{n}, \overline{Y} - K_b^{-1}(t/2) \cdot \frac{\sqrt{I_n(\omega)}}{n}\right]. \tag{3.59}$$

As alluded to above, the advantage of this whole procedure is that no explicit knowledge of the value of α (other than that it lies between one and two) is required to form these approximate confidence intervals.

3.4 Technical Proofs

This section is devoted to the proofs of Lemma 3.2.1 and Lemma 3.2.2.

Proof of Lemma 3.2.1 This result is based on Lemma 3.1 of [54], which states that

$$|(a+b)^{<\beta>} - (a+c)^{<\beta>}| \leqslant 2(|b|^\beta + |c|^\beta) \tag{3.60}$$

for any real numbers a, b, c, and for an exponent $\beta \in (0,1)$. So as a first step, let $\beta = \frac{\alpha}{2}$, which will be a positive number less than one. Consider the case that $(a+b)(a+c) \geqslant 0$, and thus let $\gamma := sign(a+b) = sign(a+c)$. For any real numbers d and e, it is simple to establish the following by the triangle inequality:

$$|d-e|^\beta \geqslant |d|^\beta - |e|^\beta \tag{3.61}$$

and also

$$|e-d|^\beta \geqslant |e|^\beta - |d|^\beta. \tag{3.62}$$

Since the left hand sides of (3.61) and (3.62) are the same, and the right hand sides only differ by a sign, we see that

$$|\,|d|^\beta - |e|^\beta\,| \leqslant |d-e|^\beta. \tag{3.63}$$

Next, if we square out (3.63), we obtain

$$|d-e|^{2\beta} \geqslant |d|^{2\beta} - 2|d|^\beta|e|^\beta + |e|^{2\beta}. \tag{3.64}$$

Now if $|d| \leqslant |e|$, then (3.64) is bounded below by $|d|^{2\beta} - |e|^{2\beta}$. In the other case that $|d| \geqslant |e|$, we have

$$|d-e|^{2\beta} = |e-d|^{2\beta} \geqslant |e|^{2\beta} - 2|e|^\beta|d|^\beta + |d|^{2\beta} \geqslant |e|^{2\beta} - |d|^{2\beta}; \tag{3.65}$$

thus in all cases, we have (letting $\alpha = 2\beta$ now)

$$|\,|d|^\alpha - |e|^\alpha\,| \leqslant |d-e|^\alpha. \tag{3.66}$$

Using this with the numbers $d = a+b$ and $e = a+c$, and utilizing Jensen's inequality, we obtain

$$\begin{aligned}
|(a+b)^{<\alpha>} - (a+c)^{<\alpha>}| &= |\gamma|a+b|^\alpha - \gamma|a+c|^\alpha| \qquad (3.67)\\
&= ||a+b|^\alpha - |a+c|^\alpha| \\
&\leqslant |(a+b)-(a+c)|^\alpha \\
&= |b-c|^\alpha \\
&\leqslant (|b|+|c|)^\alpha \\
&\leqslant 2^{\alpha-1}(|b|^\alpha + |c|^\alpha).
\end{aligned}$$

Now we consider the case where $(a+b)(a+c) < 0$. It must be that either $a(a+b) \leqslant 0$ or $a(a+c) \leqslant 0$. In the first case we have $ab \leqslant 0$ and $|a| \leqslant |b|$, so

$$\begin{aligned}
|(a+b)^{<\alpha>} - (a+c)^{<\alpha>}| &= |a+b|^{\alpha} + |a+c|^{\alpha} \\
&\leqslant |b|^{\alpha} + 2^{\alpha-1}(|a|^{\alpha} + |c|^{\alpha}) \\
&\leqslant 2^{\alpha-1}(2|b|^{\alpha} + |c|^{\alpha}).
\end{aligned} \tag{3.68}$$

Similarly if $a(a+c) \leqslant 0$ we find that

$$|(a+b)^{<\alpha>} - (a+c)^{<\alpha>}| \leqslant 2^{\alpha-1}(|b|^{\alpha} + 2|c|^{\alpha}), \tag{3.69}$$

so that the overall bound for all the cases, using (3.67), is

$$\max\{2^{\alpha}(|b|^{\alpha} + |c|^{\alpha}), 2^{\alpha-1}(|b|^{\alpha} + |c|^{\alpha})\} = 2^{\alpha}(|b|^{\alpha} + |c|^{\alpha}), \tag{3.70}$$

as desired. †

Proof of Lemma 3.2.2 We will establish the first result, which is actually harder; the second result has a very similar proof. First, we know that the limit is finite due to the calculation of the scale above. As in [54] , we just consider the case of an even sample size, say it is $2n$. Define the following terms:

$$A_n := \frac{1}{2n} \int_{\mathbb{R}} \left(\sum_{t=1}^{2n} f(t+x) \right)^{<\alpha>} dx \tag{3.71}$$

$$B_n := \frac{1}{2n} \int_{-n}^{n} \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} dx \tag{3.72}$$

$$C_n := \int_{0}^{1} \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} dx \tag{3.73}$$

First we look at the limit behavior of the term C_n :

$$\left| \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} \right| \leqslant \left(\sum_{t\in\mathbb{Z}} |f(t+x)| \right)^{\alpha} \tag{3.74}$$

and

$$\int_0^1 \left(\sum_{t \in \mathbb{Z}} |f(t+x)| \right)^{\alpha} dx < \infty, \tag{3.75}$$

as shown above. Thus we may use the dominated convergence theorem to obtain

$$\lim_{n \to \infty} C_n = \int_0^1 \left(\sum_{t \in \mathbb{Z}} f(t+x) \right)^{<\alpha>} dx. \tag{3.76}$$

Next, we will examine the difference between A_n and B_n. Note that by change of variable, we may write

$$A_n = \frac{1}{2n} \int_{\mathbb{R}} \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} dx. \tag{3.77}$$

Below we introduce the notation $g_n(x) := \sum_{t=-n}^{n-1} f(t+x)$, which is bounded almost surely by C, just as in (3.20) above. So we have the following estimate:

$$\begin{aligned} |A_n - B_n| &= \frac{1}{2n} \left| \int_{|x|>n} \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} dx \right| \\ &\leqslant \frac{1}{2n} \int_{|x|>n} \left| \sum_{t=-n}^{n} f(t+x) \right|^{\alpha} dx \\ &= \frac{1}{2n} \int_{|x|>n} |g_n(x)|^{\alpha} \, dx \\ &= \frac{1}{2n} \int_{|x|>n} |g_n(x)| \, |g_n(x)|^{\alpha-1} \, dx \\ &\leqslant \frac{C}{2n} \int_{|x|>n} |g_n(x)|^{\alpha-1} \, dx \\ &\leqslant \frac{C}{2n} \int_{|x|>n} \sum_{t=-n}^{n} |f(t+x)|^{\alpha-1} \, dx, \end{aligned} \tag{3.78}$$

which tends to zero as $n \to \infty$ by equation (3.4) of [54] , using the value of $\beta = \alpha - 1 < 1$.

Finally, it will be shown that the difference of B_n and C_n also tends to zero, which will conclude the proof:

$$\begin{aligned}
&|B_n - C_n| \qquad (3.79)\\
&= \left| \frac{1}{2n} \sum_{j=-n}^{n-1} \int_j^{j+1} \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} dx - C_n \right| \\
&= \frac{1}{2n} \left| \sum_{j=-n}^{n-1} \int_0^1 \left(\sum_{t=-n}^{n-1} f(t+j+x) \right)^{<\alpha>} dx - 2nC_n \right| \\
&= \frac{1}{2n} \left| \sum_{j=-n}^{n-1} \int_0^1 \left(\sum_{t=-n+j}^{n+j-1} f(t+x) \right)^{<\alpha>} dx - \sum_{j=-n}^{n-1} \int_0^1 \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} dx \right| \\
&\leqslant \frac{1}{2n} \sum_{j=-n}^{n-1} \int_0^1 \left| \left(\sum_{t=-n+j}^{n+j-1} f(t+x) \right)^{<\alpha>} - \left(\sum_{t=-n}^{n-1} f(t+x) \right)^{<\alpha>} \right| dx.
\end{aligned}$$

We then apply Lemma 3.2.1, to obtain the upper bound

$$\begin{aligned}
&2^\alpha \frac{1}{2n} \sum_{j=1}^{n} \int_0^1 \left| \sum_{t=n}^{n+j-1} f(t+x) \right|^\alpha dx + 2^\alpha \frac{1}{2n} \sum_{j=1}^{n} \int_0^1 \left| \sum_{t=-n}^{-n+j-1} f(t+x) \right|^\alpha dx \qquad (3.80)\\
&+ 2^\alpha \frac{1}{2n} \sum_{j=-n}^{-1} \int_0^1 \left| \sum_{t=-n+j}^{-n-1} f(t+x) \right|^\alpha dx + 2^\alpha \frac{1}{2n} \sum_{j=1}^{n} \int_0^1 \left| \sum_{t=n+j}^{n-1} f(t+x) \right|^\alpha dx \\
&=: I_1 + I_2 + I_3 + I_4
\end{aligned}$$

The first term I_1 is easily handled:

$$\begin{aligned}
&2^\alpha \frac{1}{2n} \sum_{j=1}^{n} \int_0^1 \left| \sum_{t=n}^{n+j-1} f(t+x) \right|^\alpha dx \qquad (3.81)\\
&= 2^{\alpha-1} \frac{1}{n} \sum_{j=1}^{n} \int_0^1 |g_n(x)|\, |g_n(x)|^{\alpha-1}\, dx \\
&\leqslant 2^{\alpha-1} \frac{C}{n} \sum_{j=1}^{n} \int_0^1 \sum_{t=n}^{\infty} |f(t+x)|^{\alpha-1}\, dx \\
&\leqslant 2^{\alpha-1} C \int_n^\infty |f(x)|^{\alpha-1}\, dx,
\end{aligned}$$

for $g_n(x) = \sum_{t=n}^{n+j-1} f(t+x)$ (which is bounded almost everywhere by C).

For I_2 we have:

$$2^{\alpha}\frac{1}{2n}\sum_{j=1}^{n}\int_0^1\left|\sum_{t=-n}^{-n+j-1} f(t+x)\right|^{\alpha}dx \tag{3.82}$$
$$\leqslant 2^{\alpha-1}\frac{C}{n}\sum_{j=1}^{n}\int_0^1\sum_{t=-\infty}^{-n+j-1}|f(t+x)|^{\alpha-1}\,dx$$
$$= 2^{\alpha-1}\frac{C}{n}\sum_{j=1}^{n}\int_{-\infty}^{-n+j}|f(x)|^{\alpha-1}\,dx.$$

The term I_3 is similar to I_1 , and we use the same techniques:

$$2^{\alpha}\frac{1}{2n}\sum_{j=-n}^{-1}\int_0^1\left|\sum_{t=-n+j}^{-n-1} f(t+x)\right|^{\alpha}dx \tag{3.83}$$
$$\leqslant 2^{\alpha-1}\frac{C}{n}\sum_{j=-n}^{-1}\int_0^1\sum_{t=-\infty}^{-n-1}|f(t+x)|^{\alpha-1}\,dx$$
$$= 2^{\alpha-1}C\int_{-\infty}^{-n}|f(x)|^{\alpha-1}\,dx.$$

And lastly, we have an estimate of I_4 :

$$2^{\alpha}\frac{1}{2n}\sum_{j=-n}^{-1}\int_0^1\left|\sum_{t=n+j}^{n-1} f(t+x)\right|^{\alpha}dx \tag{3.84}$$
$$\leqslant 2^{\alpha-1}\frac{C}{n}\sum_{j=-n}^{-1}\int_0^1\sum_{t=n+j}^{\infty}|f(t+x)|^{\alpha-1}\,dx$$
$$= 2^{\alpha-1}\frac{C}{n}\sum_{j=-n}^{-1}\int_{n+j}^{\infty}|f(x)|^{\alpha-1}\,dx.$$

Placing these four estimates together, we see that the difference between B_n and C_n is bounded by

$$2^{\alpha-1}C\int_{|x|>n}|f(x)|^{\alpha-1}\,dx + 2^{\alpha-1}C\frac{1}{n}\sum_{j=0}^{n-1}\int_{|x|>j}|f(x)|^{\alpha-1}\,dx. \tag{3.85}$$

The limit of the first term is clearly zero, by the Dominated Convergence Theorem; the second term likewise becomes negligible, since it is the Cesaro sum for a vanishing sequence. See [54] for more details. †

Chapter 4

Random Fields

4.1 Heavy-Tailed Linear Random Fields

Consider a strictly stationary random field $\{X_t, t \in K\}$, for some finite observation region $K \subset \mathbb{Z}^d$. We are interested in estimating the mean of this random field under the assumption that the marginal distributions are heavy-tailed. In the next section we focus on the sample mean $|K|^{-1} \sum_{t \in K} X_t$ as an estimator for $\theta := \mathbb{E}X_t$, which is assumed to be finite. See the second section of Chapter Two for the motivation of this study; the only difference is that now the observation region can be a subset of $\mathbb{Z}^d$, the finite-dimensional integer lattice. In the third section we study the asymptotics of the periodogram, as well as a self-normalized periodogram. This generalizes the one-dimensional results of section four of Chapter Two. See [62] for a reference on random fields, and also [15], [1], [31].

Therefore the propositions here obtained are very similar to those of Chapter Two. The proofs are presented here, and by taking $d = 1$ we easily obtain Theorem 2.2.1 as a special case. The requisite background material is quite similar to that of section 2.1, but there are some additional nuances in the higher dimensional case, which are discussed below.

Let $\mathbb{Z}^d$ denote the integer lattice in d-dimensional Euclidean space, and let K be a subset of $\mathbb{Z}^d$ which is the "observation region" of the data, i.e. the locations at which the data is collected. We consider a random field $\{X(t), t \in \mathbb{Z}^d\}$, which has a linear dependence structure:

$$X(t) = \sum_{j \in \mathbb{Z}^d} \psi(j) Z(t-j) \tag{4.1}$$

This is a generalization of infinite order moving average time series to random fields. Throughout this chapter we will use the term "linear" to denote this infinite order moving average with *iid* residuals. We require the filter coefficients $\{\psi(j)\}$ to be in l_p for some $p \in [0,1]$ (see Chapter 13 of [11]) in order to ensure that the sum on the right hand side of (4.1) converges almost surely; the random variables $\{Z_t,\, t \in \mathbb{Z}^d\}$ are independent and identically distributed (hereafter abbreviated as *iid*; we use Z without an index to denote a common version which is equal in distribution), and satisfy the same two "heavy-tailed" properties discussed in Chapter Two, i.e. for some $\alpha \in (1,2)$

$$\mathbb{P}[|Z| > x] = x^{-\alpha} L(x) \tag{4.2}$$

$$\frac{\mathbb{P}[Z > x]}{\mathbb{P}[|Z| > x]} \to p, \qquad \frac{\mathbb{P}[Z \leqslant -x]}{\mathbb{P}[|Z| > x]} \to q \tag{4.3}$$

as $x \to \infty$. Here p and q are between 0 and 1 and add up to 1. $L(x)$ is a "slowly varying" function, i.e. $L(ux)/L(x) \to 1$ as $x \to \infty$ for any fixed u ; an example of a slowly varying function is the logarithmic function. Note that it easily follows that the right and left tails of Z behave like

$$\mathbb{P}[Z > x] \sim p x^{-\alpha} L(x), \qquad \mathbb{P}[Z \leqslant -x] \sim q x^{-\alpha} L(x) \tag{4.4}$$

where "$\sim$" denotes that the ratio tends to unity as $x \to \infty$.

The rest of the material on heavy tails carries over verbatim, as well as most of the notation. As before, Ψ will denote the whole lattice of $\{\psi(j), j \in \mathbb{Z}^d\}$, and Ψ_p will denote its l_p norm when treated as a sequence, i.e.

$$\Psi_p := \left(\sum_{j \in \mathbb{Z}^d} |\psi(j)|^p \right)^{\frac{1}{p}}. \tag{4.5}$$

It is true that $\{X_t\}$ forms a strictly stationary random field, since applying a shift operator to the law for the Z-series does not affect the distribution. Now when $\alpha > 1$, as in section 4.2 below, the mean does exist, and we shall call it $\eta := \mathbb{E}(Z_t)$. Thus $\mathbb{E}X_t = \psi_\infty \eta =: \theta$, where $\psi_\infty := \sum_{j \in \mathbb{Z}^d} \psi(j)$.

4.2 Self-Normalized Sample Mean

For notation, let n be the d-dimensional vector with components $n_1, n_2, \cdots, n_d$, and let $N = \Pi_{i=1}^d n_i$. Also, let 1 be the vector in $\mathbb{Z}^d$ with a one in each component. It will be clear from the context that when we are refering to vectors and when we are refering to scalars. By $o_P(1)$ we denote a random variable that tends to zero in probability as each of the $n_i \to \infty$ (so that all components grow, though not necessarily at the same rate). The observation region K mentioned in the previous section will be the d-dimensional cube $(0, n_1] \times (0, n_2] \times \cdots \times (0, n_d]$ intersected with the integer lattice $\mathbb{Z}^d$.

This section is broken down into the following subsections: first there is a treatment of the convergence of the partial sums, and then a discussion of the partial sums of squares (the sample variance statistic), and finally these results are combined into the desired joint limit theorem.

4.2.1 Partial Sums

Let $\sum_{t=1}^n = \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} \cdots \sum_{t_d=1}^{n_d}$, and let a_n be the rate which satisfies line (2.8) for the given random field $\{Z_t, t \in \mathbb{Z}^d\}$. The size of our observation region is N, so we will use a_N as the appropriate rate. We begin with the following basic lemma:

Lemma 4.2.1.

$$\frac{1}{a_N} \sum_{t=1}^n Z(t) = o_P(1) + \frac{1}{a_N} \sum_{t=1}^n Z(t-j) \tag{4.6}$$

for any $j \in \mathbb{Z}^d$.

Proof For any fixed vector $j = (j_1, j_2, \cdots, j_d)$, we examine the difference

$$\frac{1}{a_N} \sum_{t=1}^{n} Z(t) - \frac{1}{a_N} \sum_{t=1}^{n} Z(t-j) \tag{4.7}$$

so we can clearly assume without loss of generality that Z has mean zero. The above difference (4.7) is equal to

$$\frac{1}{a_N} \sum_{t \in K \Delta (K-j)} Z(t) \tag{4.8}$$

where Δ denotes the symmetric difference of two sets, and the set $K - j$ denotes K shifted by the vector $-j$. Upon examination of the set $K\Delta(K-j)$, we see that we can chop it up into (overlapping) blocks of various sizes: there are two blocks of size $j_1 \times n_2 \times \cdots \times n_d$, and two blocks of size $n_1 \times j_2 \times \cdots \times n_d$, and so forth. Thus in the first pair of blocks, there are $n' = N \cdot \frac{j_1}{n_1}$ terms present in the sum; these terms are *iid*, and converge to an α-stable law at rate $a_{n'}^{-1}$. Therefore, if we denote this first block by A, then we have

$$\frac{1}{a_{n'}} \sum_{t \in A} Z(t) = O_P(1). \tag{4.9}$$

Now

$$a_N = a_{n'} \cdot \frac{n_1^{\frac{1}{\alpha}} L(N)}{j_1^{\frac{1}{\alpha}} L(n')} \tag{4.10}$$

so

$$\frac{1}{a_N} \sum_{t \in A} Z(t) = \frac{1}{a_{n'}} \left(\frac{j_1^{\frac{1}{\alpha}} L(n')}{n_1^{\frac{1}{\alpha}} L(N)} \right) \sum_{t \in A} Z(t) = O_P\left(n_1^{-\frac{1}{\alpha}} \frac{L(n')}{L(N)} \right). \tag{4.11}$$

The term $\frac{L(n')}{L(N)}$ cannot tend to infinity (if it diverges at all) faster than $n_1^{\frac{1}{\alpha}}$, since L is slowly varying. Therefore the whole expression above tends to zero. Since this can be easily established for each of the $2d$ blocks, the Lemma is proved. †

Lemma 4.2.2. *Assume that the random variables $Z(t)$ have been centered to have mean zero. Then*

$$\frac{1}{a_N} \sum_{t=1}^{n} Z(t) \overset{\mathcal{L}}{\Longrightarrow} S \tag{4.12}$$

as $\min_i n_i \to \infty$*, where* S *is an* α*-stable law with scale* σ *(the same as that of* Z*), skewness* β*, and location* 0*.*

Proof The random variables $Z(t)$ are summed over the region K; so the left hand side is a sum of $N = |K|$ *iid* random variables. The result then follows from the fact that $Z \in DOM(\alpha)$. †

Proposition 4.2.1.

$$\frac{1}{a_N} \sum_{t=1}^{n} (X(t) - \theta) \stackrel{\mathcal{L}}{\Longrightarrow} \psi_\infty \cdot S \tag{4.13}$$

as $\min_i n_i \to \infty$*, where* $\psi_\infty = \sum_{j \in \mathbb{Z}^d} \psi(j)$*.*

Proof The proof of this proposition will be broken into several parts, due to the intricacy of the calculations. For notational convenience, we introduce the centered versions

$$Y(t) = X(t) - \theta \qquad W(t) = Z(t) - \eta. \tag{4.14}$$

Let B_m be the cube in $\mathbb{Z}^d$ of width $2m+1$ centered at the origin, so that the coordinates of each side run between $-m$ and m. We arrange the sequence $W(t-j)$ for $j \in B_m$ in one vector of length $d \times (2m+1)$, and we also construct the corresponding vector of $\psi(j)$'s for $j \in B_m$. The result of Lemma 4.2.1 holds true for this vector of W's, and it also holds true under the continuous mapping of inner product against the vector of filter coefficients; thus we obtain

$$\frac{1}{a_N} \sum_{t=1}^{n} \sum_{j \in B_m} \psi(j) W(t-j) = o_P(1) + \frac{1}{a_N} \sum_{t=1}^{n} \sum_{j \in B_m} \psi(j) W(t). \tag{4.15}$$

Let us denote the sum on the left hand side by

$$\frac{1}{a_N} \sum_{t=1}^{n} Y^{(m)}(t), \tag{4.16}$$

where

$$Y^{(m)}(t) = \sum_{j \in B_m} \psi(j) W(t-j). \tag{4.17}$$

Then it follows from Lemma 4.2.2 – since $W(t)$ has mean zero – that for fixed m

$$\frac{1}{a_N} \sum_{t=1}^{n} Y^{(m)}(t) \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{j \in B_m} \psi(j) \cdot S, \tag{4.18}$$

where S is an α-stable random variable (the same as that occurring in Lemma 4.2.2). We wish to now let $m \to \infty$ on both sides of this convergence; the right hand side converges almost surely to $\psi_\infty \cdot S$. For the left hand side we have the following Lemma:

Lemma 4.2.3. *Consider the difference*

$$\frac{1}{a_N} \sum_{t=1}^{n} Y(t) - \frac{1}{a_N} \sum_{t=1}^{n} Y^{(m)}(t); \tag{4.19}$$

the limit as $m \to \infty$ of the limit superior in probability as $\min_i n_i \to \infty$ of this expression is zero.

Proof The difference easily decomposes into three terms:

$$\frac{1}{a_N} \sum_{t=1}^{n} \sum_{j \in B_m^c} \psi(j) \left(W(t-j) 1_{\{|W(t-j)| \leqslant a_N\}} - b_N \right) \tag{4.20}$$

$$+ \frac{1}{a_N} \sum_{t=1}^{n} \sum_{j \in B_m^c} \psi(j) W(t-j) 1_{\{|W(t-j)| > a_N\}} \tag{4.21}$$

$$+ \frac{N b_N}{a_N} \sum_{j \in B_m^c} \psi(j), \tag{4.22}$$

where b_N was defined in (2.5). We divide each of these terms up into $2d$ terms, according to a division of B_m^c into (overlapping) chunks. Each piece is defined by fixing one index j_i to range between either $m+1$ and ∞ or $-(m+1)$ and $-\infty$; all other indices may take on any integer value. This produces $2d$ blocks, and the sum over each individual block will be shown to tend to zero in probability. The proof for each block is quite similar, so we prove only the first case:

$$D_1 := \{ j \in \mathbb{Z}^d : j_1 > m \}. \tag{4.23}$$

Thus, we must show that

$$\frac{1}{a_N}\sum_{t=1}^{n}\sum_{j\in D_1}\psi(j)\left(W(t-j)1_{\{|W(t-j)|\leqslant a_N\}}-b_N\right) \tag{4.24}$$

$$+\frac{1}{a_N}\sum_{t=1}^{n}\sum_{j\in D_1}\psi(j)W(t-j)1_{\{|W(t-j)|>a_N\}} \tag{4.25}$$

$$+\frac{Nb_N}{a_N}\sum_{j\in D_1}\psi(j) \tag{4.26}$$

has the desired limit behavior described in (4.19) – see [8] for details on this method of proof.

The Third Term (4.26) First note that

$$b_N := \mathbb{E}\left[W1_{\{|W|\leqslant a_N\}}\right] = \mathbb{E}[W] - \mathbb{E}\left[W1_{\{|W|>a_N\}}\right] = -\mathbb{E}\left[W1_{\{|W|>a_N\}}\right], \tag{4.27}$$

so that the absolute value of the third term is bounded by

$$\frac{N}{a_N}\sum_{j\in D_1}\psi(j)|b_N| \leqslant \frac{N}{a_N}\sum_{j\in D_1}|\psi(j)|\mathbb{E}\left[|W|1_{\{|W|>a_N\}}\right] \to \frac{\alpha}{\alpha-1}\sum_{j\in D_1}|\psi(j)| \tag{4.28}$$

by Karamata's Theorem – see [23] – where the limit is taken as $N\to\infty$ (which is implied by $\min_i n_i \to \infty$); thus the limit of this as $m\to\infty$ is zero, due to the summability of the filter coefficients.

The Second Term (4.25) If we write out the second term in full vector form, we consider the following probability, and use Markov's Inequality (see [21]) for any real

$\gamma > 0$, with the $\mathbb{L}^1$ norm:

$$\begin{aligned}
&\mathbb{P}\left[a_N^{-1}\left|\sum_{t\in K}\sum_{j\in D_1}\psi(j)W(t-j)1_{\{|W(t-j)|>a_N\}}\right|>\gamma\right] \\
&\leqslant \frac{1}{\gamma}\frac{1}{a_N}\mathbb{E}\left[\sum_{t\in K}\sum_{j\in D_1}|\psi(j)||W(t-j)|1_{\{|W(t-j)|>a_N\}}\right] \\
&= \frac{1}{\gamma}\frac{1}{a_N}\sum_{t\in K}\sum_{j\in D_1}|\psi(j)|\mathbb{E}\left[|W|1_{\{|W|>a_N\}}\right] \\
&= \frac{1}{\gamma}\frac{N}{a_N}\sum_{j\in D_1}|\psi(j)|\mathbb{E}\left[|W|1_{\{|W|>a_N\}}\right] \\
&\to \frac{1}{\gamma}\frac{\alpha}{\alpha-1}\sum_{j\in D_1}|\psi(j)|
\end{aligned} \tag{4.29}$$

as $N \to \infty$, due again to Karamata's Theorem. Recall that the set D_1 was defined in line (4.23). Finally, we let m go to ∞ and obtain zero, due to the summability of the filter coefficients.

The First Term (4.24) First we introduce the notation $\tilde{D}_1 := \{k \in \mathbb{Z}^d : k_1 > m - n_1\}$ and $C_1 := (0, n_2] \times \cdots \times (0, n_d] \bigcap \mathbb{Z}^{d-1}$, which is a subset of the hyperplane on axes 2 through d. The first term has the following form:

$$\begin{aligned}
&a_N^{-1}\sum_{t\in K}\sum_{j\in D_1}\psi(j)\left(W(t-j)1_{\{|W(t-j)|\leqslant a_N\}}-b_N\right) \\
&= a_N^{-1}\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}\left(W(-k)1_{\{|W(-k)|\leqslant a_N\}}-b_N\right)
\end{aligned} \tag{4.30}$$

where Ξ is defined as follows:

$$\Xi_{k_1,n_1}(s) := \begin{cases} \psi(m+1,s)+\cdots+\psi(n_1+k_1,s) & -n_1+m+1 \leqslant k_1 \leqslant m \\ \psi(k_1+1,s)+\cdots+\psi(k_1+1,s) & k_1 > m \end{cases} \tag{4.31}$$

for any $s \in \mathbb{Z}^{d-1}$. Now we apply the Chebyshev Inequality (see [21]) to the following probability:

$$\mathbb{P}\left[a_N^{-1}\left|\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}\left(W(-k)1_{\{|W(-k)|\leqslant a_N\}}-b_N\right)\right|>\gamma\right] \tag{4.32}$$

$$\leqslant \frac{1}{\gamma^2}\frac{1}{N}\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}^2\frac{N}{a_N^2}Var\left(W1_{\{|W|\leqslant a_N\}}\right).$$

In squaring out (4.30), we note that any "off-diagonal" terms are independent, and thus the expectation of those terms is zero (since b_N is the centering of the random variables). So this leaves only the "diagonal" terms in the squaring, which are written in the second line of (4.32). The last term has finite limit superior as $\min_i n_i \to \infty$, due again to Karamata's Theorem. As for the sum of coefficients, the following claim holds:

Claim 4.2.1.

$$\lim_{m\to\infty}\limsup_{\min_i n_i\to\infty}\frac{1}{N}\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}^2=0 \tag{4.33}$$

The proof of this is quite technical, and is deferred to the final section. Together, the three parts of the difference over block D_1 tend to zero, and the Lemma is established. †

The Proof of Proposition 4.2.2 now follows immediately from (4.18) and (4.19). †

4.2.2 Sample Variance

The proofs for the sample variance are extremely similar, so some of the more laborious details are omitted. As in the previous section, we have the following lemma:

Lemma 4.2.4.

$$\frac{1}{a_N^2}\sum_{t=1}^{n}Z^2(t)=o_P(1)+\frac{1}{a_N^2}\sum_{t=1}^{n}Z^2(t-j) \tag{4.34}$$

for any $j \in \mathbb{Z}^d$.

Proof The proof is almost identical to Lemma 4.2.1, once we realize that the *iid* random variables $Z^2(t)$ are in $DOM(\frac{\alpha}{2})$, and thus (using the same notation as in Lemma 4.2.1)

$$\frac{1}{a_{n'}^2} \sum_{t \in B} Z^2(t) = O_P(1). \tag{4.35}$$

Since

$$a_N = a_{n'} \cdot \frac{n_1^{\frac{1}{\alpha}} L(N)}{j_1^{\frac{1}{\alpha}} L(n')}, \tag{4.36}$$

it follows that

$$\frac{1}{a_N^2} \sum_{t \in B} Z^2(t) = \frac{1}{a_{n'}^2} \left(\frac{j_1^{\frac{2}{\alpha}} L^2(n')}{n_1^{\frac{2}{\alpha}} L^2(N)} \right) \sum_{t \in B} Z^2(t) = O_P \left(n_1^{-\frac{2}{\alpha}} \frac{L^2(n')}{L^2(N)} \right). \tag{4.37}$$

This estimate establishes the Lemma. †

Lemma 4.2.5.

$$\frac{1}{a_N^2} \sum_{t=1}^{n} Z^2(t) \stackrel{\mathcal{L}}{\Longrightarrow} S_2 \tag{4.38}$$

as $\min_i n_i \to \infty$, *where* S_2 *is an* $\frac{\alpha}{2}$*-stable law with scale* σ *(the same as that of* Z*), skewness* 1*, and location* 0*, i.e. it is a totally right skewed stable random variable.*

Remark Note that, unlike in the previous section, these random variables need not be centered by the mean. In fact, it is easy to establish, using the fact that $\alpha \in (1, 2)$, that the same result is true if we replace $Z(t)$ by $Z(t) - \eta$ (recall that $\eta := \mathbb{E}(Z)$).

Proof The proof is identical to the proof of Lemma 4.2.2, since $Z^2 \in DOM(\frac{\alpha}{2})$. †

Proposition 4.2.2.

$$\frac{1}{a_N^2} \sum_{t=1}^{\mathbf{n}} X^2(t) \stackrel{\mathcal{L}}{\Longrightarrow} \Psi_2^2 S_2 \tag{4.39}$$

as $\min_i n_i \to \infty$ *, where* $\Psi_2 = (\sum_{j \in \mathbb{Z}^d} \psi^2(j))^{\frac{1}{2}}$ *and* S_2 *is the* $\frac{\alpha}{2}$ *totally right skewed stable random variable from Lemma 4.2.5.*

Proof Because the random variable $X(t)$ is squared, this proof is a bit more complicated than that of Proposition 4.2.1. Thus, we first establish the following preliminary Lemma:

Lemma 4.2.6.

$$\frac{1}{a_N^2}\sum_{t=1}^{n} X^2(t) = o_P(1) + \frac{1}{a_N^2}\sum_{t=1}^{n}\sum_{j\in\mathbb{Z}^d}\psi^2(j)Z^2(t-j) \tag{4.40}$$

Proof of Lemma The difference between the right and left hand sides is

$$\frac{1}{a_N^2}\sum_{t=1}^{n}\sum_{i\neq j\in\mathbb{Z}^d}\psi(i)\psi(j)Z(t-i)Z(t-j) \tag{4.41}$$

which in the $\mathbb{L}^1$ norm is bounded by

$$\frac{1}{a_N^2}\sum_{t=1}^{n}\sum_{i\neq j\in\mathbb{Z}^d}|\psi(i)||\psi(j)|(\mathbb{E}|Z(t)|)^2 \leqslant \frac{N}{a_N^2}(\mathbb{E}|Z|)^2\cdot\left(\sum_{i\in\mathbb{Z}^d}|\psi(i)|\right)^2, \tag{4.42}$$

and this tends to zero as $N\to\infty$. This proves the Lemma. †

Now we return to the proof of Proposition 4.2.2 , which follows similar lines to that of Proposition 4.2.1 . By Lemma 4.2.6, it suffices to examine the convergence of

$$\frac{1}{a_N^2}\sum_{t=1}^{n}\sum_{j\in\mathbb{Z}^d}\psi^2(j)Z^2(t-j). \tag{4.43}$$

Again we consider this sum on the d-dimensional cube B_m, and by Lemma 4.2.4 we have

$$a_N^{-2}\sum_{t=1}^{n}\sum_{j\in B_m}\psi^2(j)Z^2(t-j) = o_P(1) + a_N^{-2}\sum_{t=1}^{n}\sum_{j\in B_m}\psi^2(j)Z^2(t) \tag{4.44}$$

so that

$$a_N^{-2}\sum_{t=1}^{n}\sum_{j\in B_m}\psi^2(j)Z^2(t-j) \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{j\in B_m}\psi^2(j)\cdot S_2 \tag{4.45}$$

by Lemma 4.2.5 . The idea is now to let m increase to infinity on both sides of this convergence. On the right side this is clearly valid, and almost sure convergence to

$\Psi_2^2 \cdot S_2$ is obtained. As for the left hand side, we must demonstate that the limit as $m \to \infty$, for any choice of $\gamma > 0$, of

$$\limsup_{\min_i n_i \to \infty} \mathbb{P}\left[\left| a_N^{-2} \sum_{t=1}^{n} \sum_{j \in B_m^c} \psi^2(j) \right| > \gamma \right] \tag{4.46}$$

is zero, just as in (4.19). We decompose this sum into two terms:

$$\frac{1}{a_N^2} \sum_{t=1}^{n} \sum_{j \in B_m^c} \psi^2(j) Z^2(t-j) 1_{\{|Z(t-j)| \leqslant a_N\}} + \frac{1}{a_N^2} \sum_{t=1}^{n} \sum_{j \in B_m^c} \psi^2(j) Z^2(t-j) 1_{\{|Z(t-j)| > a_N\}} \tag{4.47}$$

and each term is further divided into $2d$ overlapping blocks as in. Considering only the sum over the first block D_1 – see (4.23), we have

$$\frac{1}{a_N^2} \sum_{t=1}^{n} \sum_{j \in D_1} \psi^2(j) Z^2(t-j) 1_{\{|Z(t-j)| \leqslant a_N\}} + \frac{1}{a_N^2} \sum_{t=1}^{n} \sum_{j \in D_1} \psi^2(j) Z^2(t-j) 1_{\{|Z(t-j)| > a_N\}}. \tag{4.48}$$

The Second Term of (4.48) Choose any $\gamma > 0$, then by the use of Chebyshev's inequality with $\mathbb{E}|\cdot|^{\frac{1}{2}}$, we have

$$\begin{aligned} &\mathbb{P}\left[a_N^{-2} \left| \sum_{t=1}^{n} \sum_{j \in D_1} \psi^2(j) Z^2(t-j) 1_{\{|Z(t-j)| > a_N\}} \right| > \gamma \right] \qquad (4.49) \\ &\leqslant \frac{1}{\sqrt{\gamma}} a_N^{-1} \sum_{t=1}^{n} \sum_{j \in D_1} |\psi(j)| \mathbb{E}\left[|Z(t-j)| 1_{\{|Z(t-j)| > a_N\}} \right] \\ &\leqslant \frac{1}{\sqrt{\gamma}} a_N^{-1} \sum_{t=1}^{n} \sum_{j \in D_1} |\psi(j)| \mathbb{E}\left[|Z| 1_{\{|Z| > a_N\}} \right] \\ &\leqslant \frac{1}{\sqrt{\gamma}} \frac{N}{a_N} \sum_{j \in D_1} |\psi(j)| \mathbb{E}\left[|Z| 1_{\{|Z| > a_N\}} \right] \\ &\to \frac{1}{\sqrt{\gamma}} \sum_{j \in D_1} |\psi(j)| \frac{\alpha}{\alpha - 1} \end{aligned}$$

where the limit is as $\min_i n_i \to \infty$, and we have used Karamata's Theorem. The sum of the coefficients now tends to zero as $m \to \infty$, and thus the second term is accounted for.

First Term of (4.48) Now the first term can be rewritten as

$$a_N^{-2}\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Omega_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}Z^2(-k)1_{\{|Z(-k)|\leqslant a_N\}} \tag{4.50}$$

with Ω defined for any $s\in\mathbb{Z}^{d-1}$ by

$$\Omega_{k_1,n_1}(s):=\begin{cases}\psi^2(m+1,s)+\cdots+\psi^2(n_1+k_1,s) & -n_1+m+1\leqslant k_1\leqslant m\\ \psi^2(k_1+1,s)+\cdots+\psi^2(k_1+1,s) & k_1>m\end{cases} \tag{4.51}$$

We next apply the Markov Inequality to get the $\mathbb{L}^1$ norm of the previous quantity, for any $\gamma>0$:

$$\mathbb{P}\left[a_N^{-2}\left|\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Omega_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}Z^2(-k)1_{\{|Z(-k)|\leqslant a_N\}}\right|>\gamma\right] \tag{4.52}$$

$$\leqslant\frac{1}{\gamma}\frac{1}{a_N^2}\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}|\Omega_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)|\right\}\mathbb{E}\left[Z^2(-k)1_{\{|Z(-k)|\leqslant a_N\}}\right]$$

$$=\frac{1}{\gamma}\frac{N}{a_N^2}\mathbb{E}\left[Z^21_{\{|Z|\leqslant a_N\}}\right]\cdot\frac{1}{N}\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}|\Omega_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)|\right\}$$

The first term in the product has finite limit superior as $\min_i n_i\to\infty$. The following claim finishes the proof of the proposition:

Claim 4.2.2. *The sum of the filter coefficients inside the curly brackets in the final line of (4.52) above are bounded as* $\min_i n_i\to\infty$, *and the limit of this as* $m\to\infty$ *is zero.*

The proof of this claim is in the final section of this chapter. †

4.2.3 Joint Convergence

In this next part we demonstrate the joint convergence of the random variables previously studied, i.e., sample mean and sample variance. As a consequence, a limit theorem for the self-normalized quantity

$$\frac{\sum_{t=1}^n(X_t-\theta)}{\sqrt{\sum_{t=1}^n X_t^2}}$$

is obtained.

Proposition 4.2.3.

$$\left(\frac{1}{a_N}\sum_{t=1}^{n}(X_t-\theta),\frac{1}{a_N^2}\sum_{t=1}^{n}X_t^2\right)\overset{\mathcal{L}}{\Longrightarrow}\left(\psi_\infty S,\Psi_2^2 S_2\right) \tag{4.53}$$

and thus

$$\frac{\sum_{t=1}^{n}(X_t-\theta)}{\sqrt{\sum_{t=1}^{n}X_t^2}}\overset{\mathcal{L}}{\Longrightarrow}\frac{\psi_\infty S}{\Psi_2\sqrt{S_2}} \tag{4.54}$$

as $\min_i n_i \to \infty$.

Proof We begin by introducing some notation:

$$T_N^{(m)} := \frac{1}{a_N}\sum_{t=1}^{n}\left(X_t^{(m)}-\theta^{(m)}\right) \tag{4.55}$$

$$W_N^{(m)} := \frac{1}{a_N^2}\sum_{t=1}^{n}\left(X_t^{(m)}\right)^2 \tag{4.56}$$

where

$$X_t^{(m)} := \sum_{j\in B_m}\psi(j)Z(t-j) \tag{4.57}$$

and

$$\theta^{(m)} := \mathbb{E}\left[X_t^{(m)}\right] = \sum_{j\in B_m}\psi(j)\cdot\eta \tag{4.58}$$

with B_m defined in the proof of Proposition 4.2.1. From that proof we also know that

$$T_N^{(m)}\overset{\mathcal{L}}{\Longrightarrow}T^{(m)},\qquad W_N^{(m)}\overset{\mathcal{L}}{\Longrightarrow}W^{(m)} \tag{4.59}$$

with

$$T^{(m)} := \sum_{j\in B_m}\psi(j)S,\qquad W^{(m)} := \sum_{j\in B_m}\psi^2(j)S_2 \tag{4.60}$$

(this restates lines (4.18) and (4.45)). More precisely, we can write, from (4.15) and (4.44) ,

$$T_N^{(m)} = o_P(1)+\tilde{T}_N^{(m)},\qquad \tilde{T}_N^{(m)}\overset{\mathcal{L}}{\Longrightarrow}T^{(m)} \tag{4.61}$$

$$W_N^{(m)} = o_P(1)+\tilde{W}_N^{(m)},\qquad \tilde{W}_N^{(m)}\overset{\mathcal{L}}{\Longrightarrow}W^{(m)} \tag{4.62}$$

where

$$\tilde{T}_N^{(m)} := \sum_{j \in B_m} \psi(j) \frac{1}{a_N} \sum_{t=1}^{n} (Z(t) - \eta), \qquad \tilde{W}_N^{(m)} := \sum_{j \in B_m} \psi^2(j) \frac{1}{a_N^2} \sum_{t=1}^{n} Z^2(t) \tag{4.63}$$

We may concatenate these statements to produce the joint convergence

$$\left(T_N^{(m)}, W_N^{(m)}\right) = o_P(1) + \left(\tilde{T}_N^{(m)}, \tilde{W}_N^{(m)}\right) \tag{4.64}$$

$$\left(\tilde{T}_N^{(m)}, \tilde{W}_N^{(m)}\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(T^{(m)}, W^{(m)}\right) \tag{4.65}$$

This second line (4.65) holds true because it holds true for *iid* sequences (see [36] for the first demonstration of this in the case that the inputs Z are actually stable random variable; page 95 of [50] handles the case for *iid* inputs in DOM (α)) and therefore also for finite linear combinations of such. Thus putting (2.64) and (2.65) together, we find that

$$\left(T_N^{(m)}, W_N^{(m)}\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(T^{(m)}, W^{(m)}\right) \tag{4.66}$$

All that remains at this point is to take the limit in probability of (4.66) as m tends to ∞, as in the other Propositions. Now we also know that

$$\lim_{m \to \infty} T_N^{(m)} = T_N, \qquad \lim_{m \to \infty} W_N^{(m)} = W_N \tag{4.67}$$

(the limits are in probability) from Lemma 4.2.3 and (4.46) , where

$$T_N := \frac{1}{a_N} \sum_{t=1}^{\mathbf{n}} (X_t - \theta), \qquad W_N := \frac{1}{a_N^2} \sum_{t=1}^{\mathbf{n}} X_t^2 \tag{4.68}$$

On the right side of (4.66), we also know that

$$T^{(m)} \xrightarrow{a.s.} T := \psi_\infty S, \qquad W^{(m)} \xrightarrow{a.s.} W := \Psi_2^2 S_2. \tag{4.69}$$

Thus we apply the continuous function $f(x, y) = \frac{x}{\sqrt{y}}$ to the above convergence (4.66), and obtain

$$\frac{T_N^{(m)}}{\sqrt{W_N^{(m)}}} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{T^{(m)}}{\sqrt{W^{(m)}}} \tag{4.70}$$

We may safely let m tend to infinity here, due to the continuity of f, as the following calculation demonstrates:

$$\begin{aligned}\lim_{m\to\infty} \frac{T_N^{(m)}}{\sqrt{W_N^{(m)}}} &= \lim_{m\to\infty} f\left(T_N^{(m)}, W_N^{(m)}\right) \\ &= f\left(\lim_{m\to\infty} T_N^{(m)}, \lim_{m\to\infty} W_N^{(m)}\right) \\ &= f\left(T_N, W_N\right) = \frac{T_N}{\sqrt{W_N}}\end{aligned} \tag{4.71}$$

with a similar argument for the right hand side. This proves the proposition. †

4.3 Self-Normalized Periodogram

For a discussion of the relevance of periodogram estimation, see section four of Chapter Two or section two of Chapter Three. Here we emply the same notations as in the previous section of this chapter. One major change is that $\alpha \in (0, 2]$, so that we no longer impose that $\alpha > 1$. Of course $\alpha = 2$ is our short-hand for the fact that $Z \in \mathbb{L}^2$. The periodogram for a heavy-tailed random field is defined as follows:

$$\tilde{I}_n(\omega) := \left| a_N^{-1} \sum_{t=1}^{n} e^{it'\omega} \right|^2 \tag{4.72}$$

where $\omega = (\omega_1, \cdots, \omega_d)$ for frequencies $\omega_i \in (-\pi, \pi]$, and $t'\omega$ denotes the scalar product of the vectors t and ω. Now equation (4.72) depends upon the rate a_N, which cannot be calculated from the data; hence it is advantageous to use self-normalization, as in the case with the sample mean, to remove that rate. We will normalize by the sample variance, though we do not take the square root in this case:

$$\tilde{J}_n(\omega) := \frac{\tilde{I}_n(\omega)}{a_N^{-2}\hat{\sigma^2}_n} = \frac{\left|\sum_{t=1}^n e^{it'\omega}\right|^2}{\sum_{t=1}^n X_t^2}. \tag{4.73}$$

We begin by establishing the following joint limit result, which is the building block for the full theorem on the convergence of $\tilde{J}$.

Proposition 4.3.1. *Fix ω such that each component is a rational multiple of 2π. Let $\alpha \in (0,2]$, with the meaning as in section 2.1, but if $\alpha > 1$, assume that $\mathbb{E}Z = 0$. Then the following joint convergence result holds:*

$$\left(a_N^{-2}\sum_{t=1}^{n} Z_t^2, a_N^{-1}\sum_{t=1}^{n} Z_t \cos t'\omega, a_N^{-1}\sum_{t=1}^{n} Z_t \sin t'\omega\right) \overset{\mathcal{L}}{\Longrightarrow} \left(\tilde{S}, U, V\right), \tag{4.74}$$

where S_2 is a totally right skewed $\frac{\alpha}{2}$-stable random variable, and U and V are α-stable random variables.

Proof Fix ω, and let $U_t := Z_t \cos t'\omega$ and $V_t := Z_t \sin t'\omega$. Then choose any real numbers η_1, η_2, η_3, so that the characteristic function of the left hand side of (4.74) is

$$\begin{aligned} &\mathbb{E}\exp\left\{i\eta_1 a_N^{-1}\sum_{t=1}^{n} U_t + i\eta_2 a_N^{-1}\sum_{t=1}^{n} V_t + i\eta_3 a_N^{-2}\sum_{t=1}^{n} Z_t^2\right\} \qquad (4.75)\\ &= \mathbb{E}\exp\left\{i a_N^{-1}\sum_{t=1}^{n} Z_t\left(\eta_1 \cos t'\omega + \eta_2 \sin t'\omega\right) + i\eta_3 a_N^{-2}\sum_{t=1}^{n} Z_t^2\right\}. \end{aligned}$$

Now let $f(t) := \eta_1 \cos t'\omega + \eta_2 \sin t'\omega$, and notice that this function is periodic in each component t_i if we fix the other components, due to the choice of the ω_i's. This makes f into a periodic function on $\mathbb{Z}^d$ with a finite orbit, say of length L. Thus we can partition the observation region K into regions G_l^N, each of which consists of all points $t \in K$ such that $f(t)$ is constant. Thus we have f_l to be the value of $f(t)$ on the set G_l^N, and we know the size of G_l^N is $G = \frac{N}{L}$. Now we use this in (4.75) along with independence of the inputs to obtain

$$\mathbb{E}\exp\left\{ia_N^{-1}\sum_{l=1}^{L}\sum_{t\in G_l^N} Z_t f(t) + i\eta_3 a_N^{-2}\sum_{l=1}^{L}\sum_{t\in G_l^N} Z_t^2\right\} \tag{4.76}$$

$$= \Pi_{l=1}^{L}\mathbb{E}\exp\left\{ia_N^{-1}\sum_{t\in G_l^N} Z_t f_l + i\eta_3 a_N^{-2}\sum_{t\in G_l^N} Z_t^2\right\}$$

$$= \Pi_{l=1}^{L}\mathbb{E}\exp\left\{if_l\left(L^{\frac{1}{\alpha}}\frac{L(N)}{L(G)}\right)^{-1} a_G^{-1}\sum_{t\in G_l^N} Z_t + i\eta_3\left(L^{\frac{1}{\alpha}}\frac{L(N)}{L(G)}\right)^{-2} a_N^{-2}\sum_{t\in G_l^N} Z_t^2\right\}$$

$$\to \Pi_{l=1}^{L}\mathbb{E}\exp\left\{if_l L^{-\frac{1}{\alpha}} S_l + i\eta_3 L^{-\frac{2}{\alpha}}\tilde{S}_l\right\}$$

$$= \mathbb{E}\exp\left\{iL^{-\frac{1}{\alpha}}\sum_{l=1}^{L} f_l S_l + iL^{-\frac{2}{\alpha}}\eta_3\sum_{l=1}^{L}\tilde{S}_l\right\}$$

$$= \mathbb{E}\exp\left\{i\eta_1 L^{-\frac{1}{\alpha}}\sum_{l=1}^{L} c_l^1 S_l + i\eta_2 L^{-\frac{1}{\alpha}}\sum_{l=1}^{L} c_l^2 S_l + i\eta_3 L^{1-\frac{2}{\alpha}}\sum_{t=1}^{L}\tilde{S}_l\right\}.$$

The limit in the middle lines was taken as $\min_i n_i \to \infty$, which forced $N \to \infty$ and $G \to \infty$. Also we used the simple identity $a_N = L^{\frac{1}{\alpha}}\frac{L(N)}{L(G)}a_G$, and $\frac{L(N)}{L(G)} = \frac{L(LG)}{L(G)} \to 1$ as $G \to \infty$ since the function $L(\cdot)$ is slowly varying. Finally, c_l^1 and c_l^2 are the constants obtained when we decompose $\sum_{l=1}^{L} f_l = \eta_1 c_l^1 + \eta_2 c_l^1$. Now S_l is an α-stable random variable, and $\tilde{S}_l$ is a totally right skewed $\frac{\alpha}{2}$-stable random variable. The third equality is valid due to the joint convergence of the terms for $l = 1, 2, \cdots, L$.

One other issue must be addressed in the above calculation: we used the fact that $a_G^{-1}\sum_{t\in G_l^N} Z_t \overset{\mathcal{L}}{\Longrightarrow} S$. Now if $\alpha > 1$, we need to use the normalization b_G as described in (2.5). Asymptotically this is the same as subtracting the mean $\mathbb{E}Z$. So we can place the mean into the statement (4.74) in the case that $\alpha > 1$, or just say that Z has mean zero in the $\alpha > 1$ case. The argument in (3.44) will show that when we finally look at $\tilde{I}_n$, we won't have to worry about the mean at all, since the periodogram is asymptotically the same with or without the mean. As for the partial sums of the Z_t^2's, no normalization by b_N is necessary.

Thus from (4.76) we may conclude that

$$\left(a_N^{-2}\sum_{t=1}^{n} Z_t^2, a_N^{-1}\sum_{t=1}^{n} U_t, \sum_{t=1}^{n} V_t\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(L^{-\frac{2}{\alpha}}\sum_{l=1}^{L}\tilde{S}_l, L^{-\frac{1}{\alpha}}\sum_{l=1}^{L} c_l^1 S_l, L^{-\frac{1}{\alpha}}\sum_{l=1}^{L} c_l^2 S_l\right), \tag{4.77}$$

which is the right hand side of (4.74) when we make the following associations: let $\tilde{S} := L^{-\frac{2}{\alpha}}\sum_{l=1}^{L}\tilde{S}_l$, $U := L^{-\frac{1}{\alpha}}\sum_{l=1}^{L} c_l^1 S_l$, and $V := L^{-\frac{1}{\alpha}}\sum_{l=1}^{L} c_l^2 S_l$. †

We will now develop this result to investigate the joint asymptotic properties of $\tilde{I}_n(\omega)$ and the sample variance $\hat{\sigma^2}_n$. First observe that

$$\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j Z(t-j)\cos t'\omega \tag{4.78}$$

$$= \sum_{j\in B_m}\psi_j \cos j'\omega \sum_{s=1-j}^{n-j} Z(s)\cos s'\omega - \sum_{j\in B_m}\psi_j \sin j'\omega \sum_{s=1-j}^{n-j} Z(s)\sin s'\omega$$

$$= o_P(a_N) + \sum_{j\in B_m}\psi_j \cos j'\omega \sum_{s=1}^{n} Z(s)\cos s'\omega - \sum_{j\in B_m}\psi_j \sin j'\omega \sum_{s=1}^{n} Z(s)\sin s'\omega$$

by the law of cosines and application of Lemma 4.2.1. In a similar fashion, we obtain

$$\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j Z(t-j)\sin t'\omega \tag{4.79}$$

$$= o_P(a_N) + \sum_{j\in B_m}\psi_j \cos j'\omega \sum_{s=1}^{n} Z(s)\sin s'\omega + \sum_{j\in B_m}\psi_j \sin j'\omega \sum_{s=1}^{n} Z(s)\cos s'\omega$$

by using the law of sines. These statements (4.78) and (4.79), together with (4.34), produce the joint statement

$$\begin{aligned}
&\left(a_N^{-2}\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j^2 Z_{t-j}^2, a_N^{-1}\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j Z_{t-j}\cos t'\omega, a_N^{-1}\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j Z_{t-j}\sin t'\omega\right)\\
&= o_P(1) + \Bigg(\sum_{j\in B_m}\psi_j^2 a_N^{-2}\sum_{t=1}^{n}\psi_j^2 Z^2(t),\\
&\sum_{j\in B_m}\psi_j\cos j'\omega\sum_{s=1}^{n}Z(s)\cos s'\omega - \sum_{j\in B_m}\psi_j\sin j'\omega\sum_{s=1}^{n}Z(s)\sin s'\omega,\\
&\sum_{j\in B_m}\psi_j\cos j'\omega\sum_{s=1}^{n}Z(s)\sin s'\omega + \sum_{j\in B_m}\psi_j\sin j'\omega\sum_{s=1}^{n}Z(s)\cos s'\omega\Bigg)\\
&\stackrel{\mathcal{L}}{\Longrightarrow}\left(\sum_{j\in B_m}\psi_j^2\tilde{S}, \psi_c^m U - \psi_s^m V, \psi_c^m V + \psi_s^m U\right)
\end{aligned} \tag{4.80}$$

by using Proposition 4.3.1. The constants ψ_c^m and ψ_s^m are defined by the formulas

$$\begin{aligned}
\psi_c^m &:= \sum_{j\in B_m}\psi_j\cos j'\omega \qquad (4.81)\\
\psi_s^m &:= \sum_{j\in B_m}\psi_j\sin j'\omega.
\end{aligned}$$

Next apply the continuous mapping $(x,y,z)\mapsto(x,y^2+z^2)$ to the weak convergence in (4.80), and we obtain

$$\begin{aligned}
&\left(a_N^{-2}\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j^2 Z^2(t-j), \left|a_N^{-1}\sum_{t=1}^{n}\sum_{j\in B_m}\psi_j Z(t-j)e^{it'\omega}\right|^2\right) \qquad (4.82)\\
&\stackrel{\mathcal{L}}{\Longrightarrow}\left(\sum_{j\in B_m}\psi_j^2\tilde{S}, ((\psi_c^m)^2+(\psi_s^m)^2)(U^2+V^2)\right)
\end{aligned}$$

after some delightful algebra. So we are finally in the situation of Lemma 4.2.3 , so that we should take the limit as $m\to\infty$ in the convergence (4.82). The right hand side clearly converges almost surely to $(\Psi_2^2\tilde{S}, |\sum_{j\in\mathbb{Z}^d}\psi_j e^{ij'\omega}|^2(U^2+V^2))$. So if we can handle the left hand side of (4.82), we have proved the following theorem:

Theorem 4.3.1. *Let $\alpha \in (0, 2]$, with the interpretation from section 2.1, and consider a vector of frequencies ω such that each component is a rational multiple of 2π. Then the periodogram and sample variance converge jointly*

$$\left(\tilde{I}_n(\omega), \hat{\sigma^2}_n\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(\left|\sum_{j \in \mathbb{Z}^d} \psi_j e^{ij'\omega}\right|^2 (U^2 + V^2), \Psi_2^2 \tilde{S}\right) \tag{4.83}$$

as $min_i n_i \to \infty$, and the self-normalized periodogram therefore obeys

$$\tilde{J}_n(\omega) \stackrel{\mathcal{L}}{\Longrightarrow} \frac{|\sum_{j \in \mathbb{Z}^d} \psi_j e^{ij'\omega}|^2 (U^2 + V^2)}{\Psi_2^2 \tilde{S}}. \tag{4.84}$$

Proof The previous discussion leading up to (4.82) is the bulk of the proof. We must show that the periodogram for the truncated series is asymptotically the same as the periodogram; as for the sample variance, this was already established in (4.40) and (4.44). But by applying the same techniques used to prove (4.19), we can establish

$$a_N^{-1} \sum_{t=1}^{n} X_t \cos t'\omega = o_P(1) + a_N^{-1} \sum_{t=1}^{n} \sum_{j \in B_m} \psi_j Z(t-j) \cos t'\omega \tag{4.85}$$

$$a_N^{-1} \sum_{t=1}^{n} X_t \sin t'\omega = o_P(1) + a_N^{-1} \sum_{t=1}^{n} \sum_{j \in B_m} \psi_j Z(t-j) \sin t'\omega$$

with some minor adjustments (since $Z_t \cos t'\omega$ and $Z_t \sin t'\omega$ are not identically distributed; however, by partitioning them into orbits, as in the proof of Proposition 4.3.1, after much labor we get the same result). Put in a vector format, we have

$$\begin{aligned} &\left(a_N^{-1} \sum_{t=1}^{n} X_t \cos t'\omega, a_N^{-1} \sum_{t=1}^{n} X_t \sin t'\omega\right) \qquad (4.86) \\ &= o_P(1) + \left(a_N^{-1} \sum_{t=1}^{n} \sum_{j \in B_m} \psi_j Z(t-j) \cos t'\omega, a_N^{-1} \sum_{t=1}^{n} \sum_{j \in B_m} \psi_j Z(t-j) \sin t'\omega\right) \end{aligned}$$

where $o_P(1)$ here is a short hand for the limit superior of the probability of the difference over $min_i n_i \to \infty$ tending to zero as m increases to infinity. Now applying the continuous functional $(x, y) \mapsto x^2 + y^2$, which preserves the $o_P(1)$ relation, we have

$$\tilde{I}_n(\omega) = o_P(1) + \left|a_N^{-1} \sum_{t=1}^{n} \sum_{j \in B_m} \psi_j Z(t-j) e^{it'\omega}\right|^2. \tag{4.87}$$

All of this argument goes smoothly for $\alpha \leqslant 1$; in the case that $\alpha > 1$, we should replace Z by $Z - \mathbb{E}Z$ to make Proposition 4.3.1 work out correctly. However, it is easy to check that this makes no difference asymptotically to $\tilde{J}_n(\omega)$, because both its numerator and denominator grow at rate a_N^2 (also see equations (3.39) through (3.44)). Finally, the case that $\alpha = 2$ – i.e. Z is an $\mathbb{L}^2$ random variable – is already well-known, but we include it for completeness (see the remark below). †

Remark As mentioned in [35], the self-normalized periodogram has the nice property of being independent of the possibly unknown parameter α. Just as the self-normalized sample mean enjoys robustness under $\alpha \in (1, 2]$ as discussed in section two of Chapter Two, the self-normalized periodogram is robust for $\alpha \in (0, 2]$.

4.4 Subsampling Applications

The objective of the limit results in section two is to establish confidence intervals for the mean via the quantiles of the limiting distribution. Self-normalization by the sample variance was used to remove the unknown rate a_N of convergence – see (4.54) – so that the ratio of partial sums and sample variance could be formed by the practitioner. The second ingredient we need is a way of estimating the quantiles of the limit, which is the complicated random variable

$$\frac{\psi_\infty \cdot S}{\Psi_2 \sqrt{S_2}}; \tag{4.88}$$

this can be accomplished by subsampling. Section five of Chapter Two gives a treatment of this method applied to the one-dimensional time series case.

The concept of subsampling is developed in the book [48] . Subsets of the observation region K are chosen, for each set K, and the statistic is calculated over the random variables in that subest. This is done for all the subsets that can fit into K, and then an empirical distribution function is calculated from those values. The result is an estimate of the limit cdf, and its quantiles may be used as approximations.

Let us denote the ratio in (4.54) by

$$T_K(\theta) := \sqrt{N}\frac{\hat{\theta}_K - \theta}{\hat{\sigma}_K} = \frac{\sum_{t\in K}(X_t - \theta)}{\sqrt{\frac{1}{N}\sum_{t\in K}X_t^2}}, \tag{4.89}$$

where $\hat{\theta}_K := \frac{1}{N}\sum_{t\in K}X_t$ and $\hat{\sigma}_K := \sqrt{\frac{1}{N}\sum_{t\in K}X_t^2}$. From here we utilize the notation of Chapter 5 of [48] , so let $\mathbf{b}$ be a vector whose components give a subset of K (and $b = \Pi_{i=1}^d b_i$), and B will denote this set. The vector $\mathbf{q}$ gives the positions of the various subsampling blocks, and so $q = \Pi_{i=1}^d q_i$ gives the total number of those blocks. Next we define the "subsampling distribution estimator" of $T_K(\theta)$ to be the following empirical distribution function (edf):

$$L_{\mathbf{n},\mathbf{b}}(x) := \frac{1}{q}\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{q}} 1_{\{T_{\mathbf{b},\mathbf{i}}\leq x\}} \tag{4.90}$$

where $T_{\mathbf{b},\mathbf{i}}$ is essentially the sum $T_K(\theta)$ evaluated on the subseries $\{X_t\}$ with t in a scaled version of K with side lengths given by the vector $\mathbf{b}$ (but with the unknown θ replaced by the estimate $\hat{\theta}_K$) . Thus

$$T_{\mathbf{b},\mathbf{i}} := \sqrt{b}\,\frac{\hat{\theta}_{B+\mathbf{i}} - \hat{\theta}_K}{\hat{\sigma}_{B+\mathbf{i}}}. \tag{4.91}$$

Now we must briefly discuss mixing conditions – again see chapter 5 of [48] for details. Let $\hat{\alpha}_X(k;l_1)$ be the mixing coefficients introduced on page 122 of [48], i.e.

$$\hat{\alpha}_X(k;l_1) := \sup_{E_2=E_1+\mathbf{t}} \left|\mathbb{P}\left(A_1\bigcap A_2\right) - \mathbb{P}(A_1)\,\mathbb{P}(A_2)\right| \tag{4.92}$$

with $A_1 \in \mathcal{F}(E_1)$, $A_2 \in \mathcal{F}(E_2)$, $|E_1| \leqslant l_1$, and $\rho(E_1,E_2) \geqslant k$. Also E_1 is a subset of $\mathbb{Z}^d$, and $\mathbf{t} \in \mathbb{Z}^d$. This is actually a weaker condition than strong mixing – general conditions for a linear series (the $d = 1$ case) to be strong mixing are given in [70]; they require that the $\psi(j)$ tend to zero fast enough (with j), and that the Z_ts have an absolutely continuous distribution. The strong mixing condition is easily seen to be satisfied if the random field has a compactly supported filter function ψ. We make the following assumption on the mixing coefficients of the random field:

$$N^{-1}\sum_{k=1}^{\hat{n}} k^{d-1}\hat{\alpha}_X(k;b) \to 0 \tag{4.93}$$

where $\hat{n} := \max_i n_i$. Now we can state the desired corollary:

Corollary 4.4.1. *Let $J(\cdot)$ be the cdf of the limit random variable (4.88) , and choose the vector $\mathbf{b} = \mathbf{b}_K$ such that $b_i \to \infty$ and $b_i/n_i \to 0$ as $n_i \to \infty$, for $i = 1, 2, \cdots d$; also assume that the mixing condition (4.93) holds. Then*

$$L_{\mathbf{n},\mathbf{b}}(x) \xrightarrow{P} J(x) \tag{4.94}$$

for every continuity point x of $J(\cdot)$.

Remark It follows from this result that we can form the asymptotically correct confidence intervals for θ :

$$\left[\hat{\theta}_K - L^{-1}_{\mathbf{n},\mathbf{b}}(1-t/2) \cdot \frac{\hat{\sigma}_K}{\sqrt{N}}, \hat{\theta}_K - L^{-1}_{\mathbf{n},\mathbf{b}}(t/2) \cdot \frac{\hat{\sigma}_K}{\sqrt{N}}\right] \tag{4.95}$$

for a $1-t$ confidence level (here, $L^{-1}(\cdot)$ denotes the quantile function of the edf $L(\cdot)$) – see (2.44) .

Proof This follows immediately from (4.54) and Corollary 5.3.1 of [48] (notice that $\tau_u = \sqrt{u}$, so $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, as required). †

As for the periodogram, subsampling can also be used to approximate the limit distribution of $\tilde{J}_n(\omega)$, although as mentioned in section five of Chapter Two there is no application to the formation of confidence intervals (since there is no parameter). But Corollary 4.4.1 above will hold for the self-normalized periodogram if we just let $J(\cdot)$ be the cdf of the limit random variable of $\tilde{J}_n(\omega)$, which is given by (4.84). Also we let $L_{\mathbf{n},\mathbf{b}}(x)$ be the subsampling distribution estimator for the self-normalized periodogram. Thus equation (4.90) remains the same, and now $T_{\mathbf{b},\mathbf{i}}$ is defined by

$$T_{\mathbf{b},\mathbf{i}} := \frac{\tilde{I}_{B+\mathbf{i}}(\omega)}{\hat{\sigma^2}_{B+\mathbf{i}}}; \tag{4.96}$$

then the corollary still holds as stated.

4.5 Technical Proofs

This final section contains the proofs of three claims, previously deferred.

4.5.1 Proof of Claim 4.2.1

Since the function ψ is summable, we also assume that it satisfies certain monotonicity properties, i.e.

$$|\psi(j)| \geqslant |\psi(j+1)| \tag{4.97}$$

if $j \geqslant 0$ and if $j < 0$, then

$$|\psi(j)| \geqslant |\psi(j-1)|. \tag{4.98}$$

We consider

$$\begin{aligned}
&\sum_{k\in\tilde{D}_1}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}^2 \qquad (4.99)\\
&=\sum_{k_1=m+1-n_1}^{m}\sum_{k_2\in\mathbb{Z}}\cdots\sum_{k_d\in\mathbb{Z}}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}^2\\
&+\sum_{k_1=m+1}^{\infty}\sum_{k_2\in\mathbb{Z}}\cdots\sum_{k_d\in\mathbb{Z}}\left\{\sum_{t\in C_1}\Xi_{k_1,n_1}(k_2+t_2,\cdots,k_d+t_d)\right\}^2\\
&=: I_1+I_2
\end{aligned}$$

Let $\tilde{\mathbb{Z}}^{d-1}$ denote the $d-1$-dimensional integer lattice on axes 2 through d, i.e. $\tilde{\mathbb{Z}}^{d-1} := \{(k_2, \cdots, k_d) : k_2 \in \mathbb{Z}, \cdots, k_d \in \mathbb{Z}\}$. For the first term I_1, we have

$$\frac{1}{N} I_1 \leqslant \frac{1}{n_1 n_2 \cdots n_d} n_1 \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{t \in C_1} \sum_{l=1}^{n_1} \psi(m+l, k_2+t_2, \cdots, k_d+t_d) \right\}^2 \tag{4.100}$$
$$\leqslant \frac{1}{n_2 \cdots n_d} \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{t \in C_1} \sum_{l=1}^{n_1} |\psi(m+l, k_2+t_2, \cdots, k_d+t_d)| \right\}^2$$
$$\leqslant \frac{1}{n_2 \cdots n_d} \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{l>m} \sum_{t \in C_1} |\psi(l, k_2+t_2, \cdots, k_d+t_d)| \right\} \cdot \left\{ \sum_{s \in \tilde{\mathbb{Z}}^{d-1}} \sum_{j=1}^{n_1} |\psi(m+j, s)| \right\}$$
$$\leqslant \frac{1}{n_2 \cdots n_d} \left\{ \sum_{t \in C_1} \sum_{l>m} \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} |\psi(l, k_2+t_2, \cdots, k_d+t_d)| \right\} \cdot \left\{ \sum_{s \in C_1} \sum_{j>m} |\psi(j, s)| \right\}$$

The second term here clearly tends to zero as $m \to \infty$, regardless of n. For the first term, restrict to the case that $d = 2$ (the higher dimensional cases are extremely similar, but difficult to write down):

$$\frac{1}{n_2} \sum_{t_2=1}^{n_2} \sum_{l_1>m} \sum_{k_2=0}^{\infty} |\psi(l_1, k_2+t_2)| + \frac{1}{n_2} \sum_{t_2=1}^{n_2} \sum_{l_1>m} \sum_{k_2=-1}^{-\infty} |\psi(l_1, k_2+t_2)| \tag{4.101}$$
$$\leqslant \frac{1}{n_2} n_2 \sum_{l_1>m} \sum_{k_2=0}^{\infty} |\psi(l_1, k_2+1)| + \frac{1}{n_2} n_2 \sum_{l_1>m} \sum_{k_2=-1}^{-\infty} |\psi(l_1, k_2+n_2)|$$
$$\leqslant \sum_{l_1>m} \sum_{k_2 \in \mathbb{Z}} |\psi(l_1, k_2)| + \sum_{l_1>m} \sum_{k_2 \in \mathbb{Z}} |\psi(l_1, k_2)|$$
$$= 2 \sum_{l_1>m} \sum_{k_2 \in \mathbb{Z}} |\psi(l_1, k_2)|$$

The bound in the d-dimensional case will be

$$2(d-1) \sum_{l_1>m} \sum_{k_2 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} |\psi(l_1, k_2, \cdots, k_d)| \tag{4.102}$$

which tends to zero as $m \to \infty$, regardless of n. Thus it is true that $\frac{1}{N} I_1 \to 0$.

Next we consider the term I_2. We have the following bounds:

$$\begin{aligned}
&\frac{1}{n_1 \cdots n_d} I_2 \qquad (4.103)\\
&\leqslant \frac{1}{n_1 \cdots n_d} \sum_{k_1 > m} \left[\sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{t \in C_1} \sum_{l=1}^{n_1} |\psi(k_1 + l, k_2 + t_2, \cdots, k_d + t_d)| \right\} \right] \\
&\cdot \left\{ \sum_{s \in \tilde{\mathbb{Z}}^{d-1}} \sum_{j=1}^{n_1} |\psi(k_1 + j, s)| \right\} \\
&\leqslant \frac{1}{n_1 \cdots n_d} \sum_{k_1 > m} \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{t \in C_1} \sum_{l=1}^{n_1} |\psi(k_1 + l, k_2 + t_2, \cdots, k_d + t_d)| \right\} \\
&\cdot \left\{ \sum_{s \in \mathbb{Z}} |\psi(s)| \right\}
\end{aligned}$$

As for the first term of (4.103) in curved brackets, it is bounded as follows:

$$\begin{aligned}
&\sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{t \in C_1} \sum_{l=1}^{n_1} |\psi(k_1 + l, k_2 + t_2, \cdots, k_d + t_d)| \right\} \qquad (4.104)\\
&\leqslant n_1 \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} \left\{ \sum_{t \in C_1} |\psi(k_1 + 1, k_2 + t_2, \cdots, k_d + t_d)| \right\} \\
&\leqslant n_1 \sum_{k_2=0}^{\infty} \sum_{k_3 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} \sum_{t_2=1}^{n_2} \cdots \sum_{t_d=1}^{n_d} |\psi(k_1 + 1, k_2 + t_2, \cdots, k_d + t_d)| \\
&+ n_1 \sum_{k_2=-1}^{-\infty} \sum_{k_3 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} \sum_{t_2=1}^{n_2} \cdots \sum_{t_d=1}^{n_d} |\psi(k_1 + 1, k_2 + t_2, \cdots, k_d + t_d)| \\
&\leqslant n_1 \sum_{k_2=0}^{\infty} \sum_{k_3 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} n_2 \sum_{t_3=1}^{n_3} \cdots \sum_{t_d=1}^{n_d} |\psi(k_1 + 1, k_2 + 1, \cdots, k_d + t_d)| \\
&+ n_1 \sum_{k_2=-1}^{-\infty} \sum_{k_3 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} n_2 \sum_{t_3=1}^{n_3} \cdots \sum_{t_d=1}^{n_d} |\psi(k_1 + 1, k_2 + n_2, \cdots, k_d + t_d)| \\
&\leqslant 2 n_1 n_2 \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} \sum_{t_3=1}^{n_3} \cdots \sum_{t_d=1}^{n_d} |\psi(k_1 + 1, k_2, \cdots, k_d + t_d)|.
\end{aligned}$$

By repeatedly breaking each sum over the k indices into two parts, and simplifying, we obtain the overall bound of

$$2^{d-1} n_1 n_2 \cdots n_d \sum_{k_2 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} |\psi(k_1+1, k_2, \cdots, k_d)|. \tag{4.105}$$

Putting this together, we have

$$\begin{aligned} &\frac{1}{N} I_2 \qquad (4.106) \\ &\leqslant \frac{1}{n_1 \cdots n_d} \sum_{k_1 > m} \left[2^{d-1} n_1 n_2 \cdots n_d \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} |\psi(k_1+1, k_2, \cdots, k_d)| \cdot \left\{ \sum_{s \in \mathbb{Z}^{d-1}} |\psi(s)| \right\} \right] \\ &\leqslant 2^{d-1} \left\{ \sum_{s \in \mathbb{Z}^d} |\psi(s)| \right\} \left[\sum_{k_1 > m} \sum_{k \in \tilde{\mathbb{Z}}^{d-1}} |\psi(k_1+1, k_2, \cdots, k_d)| \right]; \end{aligned}$$

notice that the term in the square brackets tends to zero as $m \to \infty$, independent of the behavior of n, while the other parts are bounded. Thus $N^{-1} I_2 \to 0$ as desired, and the claim is proved. †

4.5.2 Proof of Claim 4.2.2

Here we prove that the following term is bounded as n increases to ∞, and that the limit as $m \to \infty$ of this is zero:

$$\frac{1}{N} \sum_{k \in \tilde{D}_1}^{\infty} \left\{ \sum_{t \in C_1} |\Omega_{k_1, n_1}(k_2 + t_2, \cdots, k_d + t_d)| \right\} \tag{4.107}$$

Once we assume the same monotonicity conditions on the filter coefficients as in the previous claim, the result immediately follows, since

$$\begin{aligned} &\left\{ \sum_{t \in C_1} \Omega_{k_1, n_1}(k_2 + t_2, \cdots, k_d + t_d) \right\} \qquad (4.108) \\ &\leqslant \left\{ \sum_{t \in C_1} \tilde{\Xi}_{k_1, n_1}(k_2 + t_2, \cdots, k_d + t_d) \right\}^2, \end{aligned}$$

where $\tilde{\Xi}$ is the same as Ξ except that every $\psi(j)$ is replaced by a $|\psi(j)|$. If we then make these substitutions in the previous claim, we are finished. †

Chapter 5

Marked Point Processes

5.1 Observation Regions : Background

When random field data is collected, it can happen that the positions of the data occur randomly. According to [48] , "...in many important cases, e.g., queueing theory, spatial statistics, mining, and geostatistics, meteorology, etc., the data correspond to observations of $X(t)$ at nonlattice, irregularly spaced points."

This situation is often modelled by a Poisson Point Process, i.e. one assumes that the observation points t are generated by a Poisson Point Process on the (compact) observation region K. We investigate such a structure in this section, and assume that the Poisson Point Process is independent of the law of the random field. It is necessary to define the stochastic process over the entire plane, not just the lattice, because the observation points from the point process can be located anywhere in the plane. The observation region K in which the points are constrained to lie is $(0, n_1] \times \cdots \times (0, n_d]$, so that if we intersect this set with the integer lattice, we get the observation region of Chapter Four. Also, we still have $|K| = N$.

The next section of this chapter generalizes the results from Chapter Three to a discrete random field situation, and the third section generalizes this to a continuous random field (i.e. a stochastic process). Next, the fourth section builds on the continuous random field model by placing the additional assumption that the observation locations are distributed according to an independent Poisson Random Measure. Thus our model is

$$X_t = \int_{\mathbb{R}^d} \psi(t+x)\, \mathbb{M}(dx) \tag{5.1}$$

for any $t \in K$ our observation region. Depending on which section of the chapter, K could be a subset of $\mathbb{Z}^d$ (the discrete context), a subset of $\mathbb{R}^d$ (the continuous case), or in the marked point process case K is in $\mathbb{R}^d$ but the t's are the points of a Poisson Random Measure falling into K. In each case we are interested in the asymptotics of sample mean, sample variance, "heavy-tailed" periodogram (cf. equation (2.35)), and joint limiting behavior of these. Thus we will define and investigate the stochastic integral

$$\int_K X(t)\, \mathbb{N}(dt). \tag{5.2}$$

In the final section of the chapter on subsampling, we discuss some possible statistical applications of these weak convergence results.

5.2 Discrete Time Results

The purpose of this first section is to generalize the results of Chapter Three to a higher-dimensional random field model, as given by equation (5.1) with $t \in K \subset \mathbb{Z}^d$. The key observation is that we can generalize Lemma 3.2.2 from $\mathbb{R}$ to $\mathbb{R}^d$, by using arguments very similar to those presented in Chapter Four. We assume throughout that ψ is an integrable function (on $\mathbb{R}^d$) which is bounded and continuous almost everywhere. Also, let B denote the unit cube in $\mathbb{R}^d$.

Theorem 5.2.1. *Let the partial sums, periodogram, and sample variance of the random* *field be denoted by*

$$S_n := N^{-\frac{1}{\alpha}} \sum_{t=1}^{n} X_t \tag{5.3}$$

$$\tilde{I}_n(\omega) := N^{-\frac{2}{\alpha}} \left| \sum_{t=1}^{n} e^{it'\omega} \right|^2 \tag{5.4}$$

$$\hat{\sigma^2}_n := N^{-\frac{2}{\alpha}} \sum_{t=1}^{n} X_t^2, \tag{5.5}$$

respectively. Here $\omega = (\omega_1, \cdots, \omega_d)$ *with each* $\omega_i \in (-\pi, \pi]$, *and* $t'\omega$ *denotes the scalar product of the two vectors. Then the following convergences hold:*

$$\left(S_n, \tilde{I}_n(\omega)\right) \overset{\mathcal{L}}{\Longrightarrow} \left(S, W(\omega)^2 + U(\omega)^2\right) \tag{5.6}$$

$$\hat{\sigma^2}_n \overset{\mathcal{L}}{\Longrightarrow} S_2 \tag{5.7}$$

and hence

$$\frac{S_n}{\sqrt{\tilde{I}_n(\omega)}} \overset{\mathcal{L}}{\Longrightarrow} \frac{S}{\sqrt{W(\omega)^2 + U(\omega)^2}}, \tag{5.8}$$

where $W(\omega)$ *and* $U(\omega)$ *are defined similarly to (3.27) and (3.28):*

$$W(\omega) := \int_B \sum_{t \in \mathbb{Z}^d} \cos t'\omega \, \psi(x+t) \, \mathbb{M}(dx) \tag{5.9}$$

$$U(\omega) := \int_B \sum_{t \in \mathbb{Z}^d} \sin t'\omega \, \psi(x+t) \, \mathbb{M}(dx).$$

Also, S *and* S_2 *are* α*-stable and positive* $\frac{\alpha}{2}$*-stable random variables respectively. They are defined as follows:*

$$S := \int_B \sum_{t \in \mathbb{Z}^d} \psi(x+t) \, \mathbb{M}(dx) \tag{5.10}$$

$$S_2 := \int_B \sum_{t \in \mathbb{Z}^d} \psi^2(x+t) \, \tilde{\mathbb{M}}(dx)$$

where $\tilde{\mathbb{M}}$ *is a positive* $\frac{\alpha}{2}$*-stable random measure, and can be thought of as the "square" of* $\mathbb{M}$.

Proof This theorem has the same proof as the results in Chapter Three, but we need to generalize Lemma 3.2.2 to a random field setting. The convergence of the sample variance follows from the Lemma 5.2.1 below, following along the lines of the sample variance result in [54].

Lemma 5.2.1. *Let* $\alpha \in (0,2)$. *Then*

$$\frac{1}{N}\int_{\mathbb{R}^d}\left(\sum_{t=1}^{n}\psi(t+x)\right)^{<\alpha>}\lambda(dx) \to \int_B\left(\sum_{t\in\mathbb{Z}^d}\psi(t+x)\right)^{<\alpha>}\lambda(dx) \tag{5.11}$$
$$\frac{1}{N}\int_{\mathbb{R}^d}\left|\sum_{t=1}^{n}\psi(t+x)\right|^{\alpha}\lambda(dx) \to \int_B\left|\sum_{t\in\mathbb{Z}^d}\psi(t+x)\right|^{\alpha}\lambda(dx)$$

as each $n_i \to \infty$.

Proof This proof follows the same lines as Lemma 3.2.2, and much of the same notation will be used. We will establish the top line (the skewness) of equation (5.11). Below we will restrict to the case that $\alpha > 1$, since the $\alpha \leqslant 1$ case is easier, and is obtained by generalizing Lemma 3.2 of [54]. So we start by dilating K to $2K$, so that each component n_i is an even integer – of course N increases to $2^d N$. Thus $2n$ denotes the vector $(2n_1, 2n_2, \cdots, 2n_d)$, which gives the outer bounds of $2K$, and let $\tilde{K}$ denote the set of real vectors x with $-n_i \leqslant x_i \leqslant n_i$. Thus we have skewness A_n , and B_n and C_n defined as follows:

$$A_n := \frac{1}{2^d N}\int_{\mathbb{R}^d}\left(\sum_{t=0}^{2n}\psi(t+x)\right)^{<\alpha>}\lambda(dx) \tag{5.12}$$

$$B_n := \frac{1}{2^d N}\int_{\tilde{K}}\left(\sum_{t=-n}^{n}\psi(t+x)\right)^{<\alpha>}\lambda(dx) \tag{5.13}$$

$$C_n := \int_B\left(\sum_{t=-n}^{n}\psi(t+x)\right)^{<\alpha>}\lambda(dx). \tag{5.14}$$

Now the bounds in (3.74) and (3.75) will hold, since $\psi \in \mathbb{L}^1(\mathbb{R}^d)$, so we find that

$$C_n \to \int_B\left(\sum_{t\in\mathbb{Z}^d}\psi(t+x)\right)^{<\alpha>}\lambda(dx). \tag{5.15}$$

Now we can write A_n in the equivalent form

$$A_n = \frac{1}{2^d N} \int_{\mathbb{R}^d} \left(\sum_{t=-n}^{n} \psi(t+x) \right)^{<\alpha>} \lambda(dx) \tag{5.16}$$

analagous to (3.77). Next we examine the difference between A_n and B_n:

$$\begin{aligned} |A_n - B_n| &= \frac{1}{2^d N} \left| \int_{\mathbb{R}^d \setminus \tilde{K}} \left(\sum_{t=-n}^{n} \psi(t+x) \right)^{<\alpha>} \lambda(dx) \right| \\ &\leqslant \frac{1}{2^d N} \int_{\mathbb{R}^d \setminus \tilde{K}} \left| \sum_{t=-n}^{n} \psi(t+x) \right|^{\alpha} \lambda(dx). \end{aligned} \tag{5.17}$$

Notice that the set $\mathbb{R}^d \setminus \tilde{K}$ is just the complement of a d-dimensional cube, and may thus be segregated into overlapping portions $D_1, D_2, \cdots, D_d$, where D_i is defined by

$$D_i := \{x \in \mathbb{R}^d : |x_i| \geqslant n_i\} \tag{5.18}$$

which is a very similar construction to (4.23). Now it is simple to bound (5.17) by d integrals, each of which covers one region D_i; we will prove that for $i = 1$ the integral tends to zero, and the other cases are exactly similar. Thus we must examine

$$\begin{aligned} &\frac{1}{2^d N} \int_{D_1} \left| \sum_{t=-n}^{n} \psi(t+x) \right|^{\alpha} \lambda(dx) \\ &= \frac{1}{2^d N} \int_{|x_1| > n_1} \left[\int_{\mathbb{R}^{d-1}} \left| \sum_{t=-n}^{n} \psi(t+x) \right|^{\alpha} dx_2 \cdots dx_d \right] dx_1, \end{aligned} \tag{5.19}$$

where we have rewritten the expression as an iterated integral (so we use the notation $\lambda(dx) = dx_1 dx_2 \cdots dx_d$). Now the part of (5.19) in the square brackets is a function of x_1 which depends on n, so let's name it $g_n(x_1)$. We need to arrive at a convenient bound for this function, so

$$
\begin{aligned}
g_n(x_1) &= \int_{\mathbb{R}^{d-1}} \left| \sum_{t=-n}^{n} \psi(t+x) \right|^{\alpha} dx_2 \cdots dx_d \\
&\leqslant \sup_{y \in \mathbb{R}^{d-1}} \left| \sum_{t_2=-n_2}^{n_2} \cdots \sum_{t_d=-n_d}^{n_d} \psi(t_1+x_1, t_2+y_2, \cdots, t_d+y_d) \right| \\
&\cdot \int_{\mathbb{R}^{d-1}} \left| \sum_{t=-n}^{n} \psi(t+x) \right|^{\alpha-1} dx_2 \cdots dx_d \\
&\leqslant \sup_{y \in \mathbb{R}^{d-1}} \left| \sum_{t_2=-n_2}^{n_2} \cdots \sum_{t_d=-n_d}^{n_d} \psi(t_1+x_1, t_2+y_2, \cdots, t_d+y_d) \right| \\
&\cdot \sum_{t=-n}^{n} \int_{\mathbb{R}^{d-1}} |\psi(t+x)|^{\alpha-1} dx_2 \cdots dx_d
\end{aligned}
\tag{5.20}
$$

does the job. Putting this into (5.19), we obtain

$$
\begin{aligned}
&\frac{1}{2^d N} \int_{D_1} \left| \sum_{t=-n}^{n} \psi(t+x) \right|^{\alpha} \lambda(dx) \\
&\leqslant \sup_{x \in \mathbb{R}^d} \left| \sum_{t=-n}^{n} \psi(t+x) \right| \frac{1}{2^d N} \sum_{t=-n}^{n} \int_{D_1} |\psi(t+x)|^{\alpha-1} \lambda(dx) \\
&\leqslant \sup_{x \in B} \sum_{t \in \mathbb{Z}^d} |\psi(t+x)| \\
&\cdot \frac{1}{2n_1} \sum_{t_1=-n_1}^{n_1} \int_{|x_1|>n_1} \left[\frac{n_1}{2^{d-1}N} \int_{\mathbb{R}^{d-1}} \sum_{t_2=-n_2}^{n_2} \cdots \sum_{t_d=-n_d}^{n_d} |\psi(t+x)|^{\alpha-1} dx_2 \cdots dx_d \right] dx_1 \\
&\leqslant C \frac{1}{2n_1} \sum_{t_1=0}^{2n_1} \int_{|x_1|>t_1} l(t_1+x_1) dx_1
\end{aligned}
\tag{5.21}
$$

where C is a finite constant bounding $\sup_{x \in B} \sum_{t \in \mathbb{Z}^d} |\psi(t+x)|$ – see the claim following (3.19) – and $l(t_1+x_1) := \frac{n_1}{2^{d-1}N} \int_{\mathbb{R}^{d-1}} \sum_{t_2=-n_2}^{n_2} \cdots \sum_{t_d=-n_d}^{n_d} |\psi(t+x)|^{\alpha-1} dx_2 \cdots dx_d$. Now if we can show that l is in $\mathbb{L}^1(\mathbb{R})$, then (5.21) will tend to zero as $n_1 \to \infty$ as in equation 3.4 of [54], since it is a Cesaro sum.

Now if we first simplify $h(t_1+x_1)$ via a change of variable, we obtain

$$
h(t_1+x_1) = \int_{\mathbb{R}^{d-1}} |\psi(t_1+x_1, u_2, \cdots, u_d)|^{\alpha-1} du_2 \cdots du_d \tag{5.22}
$$

from which it follows that

$$\int_{\mathbb{R}} h(x_1)dx_1 = \int_{\mathbb{R}}\int_{\mathbb{R}^{d-1}} |\psi(x_1,u_2,\cdots,u_d)|^{\alpha-1}\, du_2\cdots du_d\, dx_1 \qquad (5.23)$$
$$= \int_{\mathbb{R}^d} |\psi(x)|^{\alpha-1}\,\lambda(dx) < \infty,$$

as desired. Since arguments of this type can be applied to each block D_i, we see that $|A_n - B_n| \to 0$ as the various $n_i \to \infty$.

Next we must show that the difference $|B_n - C_n| \to 0$ as well. First we introduce the following notation: $E := \tilde{K}\bigcap \mathbb{Z}^d$, and $F_j := E\bigcap(E+j)$.

$$\left|\frac{1}{2^dN}\sum_{j\in E}\int_{B+j}\left(\sum_{t\in E}\psi(t+x)\right)^{<\alpha>}\lambda(dx) - C_n\right| \qquad (5.24)$$
$$= \frac{1}{2^dN}\left|\sum_{j\in E}\int_B\left(\sum_{t\in E}\psi(t+j+x)\right)^{<\alpha>}\lambda(dx) - \sum_{j\in E}\int_B\left(\sum_{t\in E}\psi(t+x)\right)^{<\alpha>}\lambda(dx)\right|$$
$$\leqslant \frac{1}{2^dN}\sum_{j\in E}\int_B\left|\left(\sum_{t\in E+j}\psi(t+x)\right)^{<\alpha>} - \left(\sum_{t\in E}\psi(t+x)\right)^{<\alpha>}\right|\lambda(dx)$$
$$\leqslant \frac{2^\alpha}{2^dN}\sum_{j\in E}\int_B\left|\sum_{t\in E+j\setminus F_j}\psi(t+x)\right|^{\alpha} + \left|\sum_{t\in E\setminus F_j}\psi(t+x)\right|^{\alpha}\lambda(dx)$$
$$\leqslant \frac{C2^\alpha}{2^dN}\sum_{j\in E}\int_B\left|\sum_{t\in E+j\setminus F_j}\psi(t+x)\right|^{\alpha-1} + \left|\sum_{t\in E\setminus F_j}\psi(t+x)\right|^{\alpha-1}\lambda(dx)$$
$$\leqslant \frac{C2^\alpha}{2^dN}\sum_{j\in E}\int_B\sum_{t\in E+j\setminus F_j}|\psi(t+x)|^{\alpha-1} + \sum_{t\in E\setminus F_j}|\psi(t+x)|^{\alpha-1}\lambda(dx)$$

by using the bound C stated after equation (5.21), and also by using Lemma 3.2.1. Now we want to divide the sets $E+j\setminus F_j$ and $E\setminus F_j$ in a strategic manner. So we divide each of these sets into d overlapping subsets G_i and H_i respectively in the following manner: the ith coordinate of elements of G_i obey the constraints for the ith coordinates of elements of the set $E+j\setminus F_j$, and all other coordinates of elements of G_i are only restricted to lie in $E+j$. Also H_i has the analagous definition for $E\setminus F_j$. Since the arguments for the various blocks are identical, we will focus on the case $i=1$. Then we may rewrite the last part of (5.24) as

$$
\begin{aligned}
&\frac{C2^{\alpha}}{2n_1}\sum_{j_1=-n_1}^{n_1}\Big[\int_B \sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d}\frac{n_1}{2^{d-1}N}\sum_{t\in G_1}|\psi(t+x)|^{\alpha-1}\,\lambda(dx) \\
&+\int_B \sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d}\frac{n_1}{2^{d-1}N}\sum_{t\in H_1}|\psi(t+x)|^{\alpha-1}\,\lambda(dx)\Big] \\
&\leqslant \frac{C2^{\alpha}}{2n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=n_1}^{n_1+j_1-1}\Big[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d} \\
&\sum_{t_2=-n_2+j_2}^{n_2+j_2}\cdots\sum_{t_d=-n_d}^{n_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_d\Big]dx_1 \\
&+\frac{C2^{\alpha}}{2n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=-n_1}^{-n_1+j_1-1}\Big[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d} \\
&\sum_{t_2=-n_2}^{n_2}\cdots\sum_{t_d=-n_d}^{n_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_d\Big]dx_1 \\
&+\frac{C2^{\alpha}}{2n_1}\sum_{j_1=-n_1}^{0}\int_0^1\sum_{t_1=-n_1+j_1}^{-n_1}\Big[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d} \\
&\sum_{t_2=-n_2+j_2}^{n_2+j_2}\cdots\sum_{t_d=-n_d+j_d}^{n_d+j_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_d\Big]dx_1 \\
&+\frac{C2^{\alpha}}{2n_1}\sum_{j_1=-n_1}^{0}\int_0^1\sum_{t_1=n_1+j_1}^{n_1}\Big[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d} \\
&\sum_{t_2=-n_2}^{n_2}\cdots\sum_{t_d=-n_d}^{n_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_d\Big]dx_1.
\end{aligned}
\tag{5.25}
$$

Let us label each of these terms as I_1, I_2, I_3, I_4 resepectively as in (3.80). Then for I_1 we have

$$
\begin{aligned}
I_1 &= \frac{C2^{\alpha}}{2n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=n_1}^{n_1+j_1-1}\Bigg[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d}\\
&\quad\sum_{t_2=-n_2+j_2}^{n_2+j_2}\cdots\sum_{t_d=-n_d}^{n_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_n\Bigg]dx_1\\
&\leqslant \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=n_1}^{n_1+j_1-1}\left[\int_0^1\cdots\int_0^1\sum_{t_2\in\mathbb{Z}}\cdots\sum_{t_d\in\mathbb{Z}}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_n\right]dx_1\\
&= \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=n_1}^{n_1+j_1-1}\left[\int_{\mathbb{R}^{d-1}}|\psi(t_1+x_1,u_2,\cdots,u_d)|^{\alpha-1}du_2\cdots du_d\right]dx_1\\
&\leqslant \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=n_1}^{\infty}k(t_1+x_1)\,dx_1\\
&= C2^{\alpha-1}\int_{n_1}^{\infty}k(x_1)\,dx_1
\end{aligned}
\tag{5.26}
$$

where $k(y) := \int_{\mathbb{R}^{d-1}}|\psi(y,u_2,\cdots,u_d)|^{\alpha-1}du_2\cdots du_d$ is an integrable function. We are now in the same framework as (3.81). Similar calculations will produce bounds for the other three terms:

$$
\begin{aligned}
I_2 &= \frac{C2^{\alpha}}{2n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=-n_1}^{-n_1+j_1-1}\Bigg[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d}\\
&\quad \sum_{t_2=-n_2}^{n_2}\cdots\sum_{t_d=-n_d}^{n_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_d\Bigg]dx_1\\
&\leqslant \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=-n_1}^{-n_1+j_1-1}\left[\int_0^1\cdots\int_0^1\sum_{t_2\in\mathbb{Z}}\cdots\sum_{t_d\in\mathbb{Z}}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_n\right]dx_1\\
&= \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=-n_1}^{-n_1+j_1-1}\left[\int_{\mathbb{R}^{d-1}}|\psi(t_1+x_1,u_2,\cdots,u_d)|^{\alpha-1}du_2\cdots du_d\right]dx_1\\
&\leqslant \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_0^1\sum_{t_1=-\infty}^{-n_1+j_1+1}k(t_1+x_1)\,dx_1\\
&= \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_{-\infty}^{-n_1+j_1}k(x_1)\,dx_1\\
&= \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=0}^{n_1}\int_{-\infty}^{-j_1}k(x_1)\,dx_1
\end{aligned}
\tag{5.27}
$$

and

$$
\begin{aligned}
I_3 &= \frac{C2^{\alpha}}{2n_1}\sum_{j_1=-n_1}^{0}\int_0^1\sum_{t_1=-n_1+j_1}^{-n_1}\Bigg[\int_0^1\cdots\int_0^1\frac{n_1}{2^{d-1}N}\sum_{j_2=-n_2}^{n_2}\cdots\sum_{j_d=-n_d}^{n_d}\\
&\quad \sum_{t_2=-n_2+j_2}^{n_2+j_2}\cdots\sum_{t_d=-n_d+j_d}^{n_d+j_d}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_d\Bigg]dx_1\\
&\leqslant \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=-n_1}^{0}\int_0^1\sum_{t_1=-n_1+j_1}^{-n_1}\left[\int_0^1\cdots\int_0^1\sum_{t_2\in\mathbb{Z}}\cdots\sum_{t_d\in\mathbb{Z}}|\psi(t+x)|^{\alpha-1}dx_2\cdots dx_n\right]dx_1\\
&= \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=-n_1}^{0}\int_0^1\sum_{t_1=-n_1+j_1}^{-n_1}\left[\int_{\mathbb{R}^{d-1}}|\psi(t_1+x_1,u_2,\cdots,u_d)|^{\alpha-1}du_2\cdots du_d\right]dx_1\\
&\leqslant \frac{C2^{\alpha-1}}{n_1}\sum_{j_1=-n_1}^{0}\int_0^1\sum_{t_1=-\infty}^{-n_1}k(t_1+x_1)\,dx_1\\
&=\leqslant C2^{\alpha-1}\int_{-\infty}^{-n_1}k(x_1)\,dx_1
\end{aligned}
\tag{5.28}
$$

and

$$
\begin{aligned}
I_4 = {} & \frac{C2^{\alpha}}{2n_1} \sum_{j_1=-n_1}^{0} \int_0^1 \sum_{t_1=n_1+j_1}^{n_1} \left[\int_0^1 \cdots \int_0^1 \frac{n_1}{2^{d-1}N} \sum_{j_2=-n_2}^{n_2} \cdots \sum_{j_d=-n_d}^{n_d} \right. \\
& \left. \sum_{t_2=-n_2}^{n_2} \cdots \sum_{t_d=-n_d}^{n_d} |\psi(t+x)|^{\alpha-1} dx_2 \cdots dx_d \right] dx_1 \\
\leqslant {} & \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=-n_1}^{0} \int_0^1 \sum_{t_1=n_1+j_1}^{n_1} \left[\int_0^1 \cdots \int_0^1 \sum_{t_2\in\mathbb{Z}} \cdots \sum_{t_d\in\mathbb{Z}} |\psi(t+x)|^{\alpha-1} dx_2 \cdots dx_n \right] dx_1 \\
= {} & \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=-n_1}^{0} \int_0^1 \sum_{t_1=n_1+j_1}^{n_1} \left[\int_{\mathbb{R}^{d-1}} |\psi(t_1+x_1, u_2, \cdots, u_d)|^{\alpha-1} du_2 \cdots du_d \right] dx_1 \\
\leqslant {} & \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=-n_1}^{0} \int_0^1 \sum_{t_1=n_1+j_1}^{\infty} k(t_1+x_1)\, dx_1 \\
= {} & \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=-n_1}^{0} \int_{n_1+j_1}^{\infty} k(x_1)\, dx_1 \\
= {} & \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=-n_1}^{0} \int_{-j_1}^{\infty} k(x_1)\, dx_1.
\end{aligned}
\tag{5.29}
$$

So putting these bounds for the four terms together, we see that (5.24) is bounded by

$$
\begin{aligned}
& C2^{\alpha-1} \int_{n_1}^{\infty} k(x_1)\, dx_1 + C2^{\alpha-1} \int_{-\infty}^{-n_1} k(x_1)\, dx_1 \\
& + \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=0}^{n_1} \int_{-\infty}^{-j_1} k(x_1)\, dx_1 + \frac{C2^{\alpha-1}}{n_1} \sum_{j_1=-n_1}^{0} \int_{-j_1}^{\infty} k(x_1)\, dx_1 \\
& = C2^{\alpha-1} \left[\int_{|x_1|>n_1} k(x_1)\, dx_1 + \frac{1}{n_1} \sum_{j_1=0}^{n_1} \int_{|x_1|>j_1} k(x_1)\, dx_1 \right]
\end{aligned}
\tag{5.30}
$$

which will tend to zero as $n_i \to \infty$ by the Dominated Convergence Theorem and the property of Cesaro Sums. Since this result can be produced for each i, this shows that $|B_n - C_n| \to 0$ as each $n_i \to \infty$. This completes the proof of the Lemma in the case that $\alpha > 1$. For $\alpha \leqslant 1$, similar techniques can be used, and the proof is actually easier (we can use the triangle inequality directly in many cases where we had to first factor and take a supremum). †

Remark All of the other corollaries and statements made in Section 3.2 can be extended to the random field case, by using techniques similar to those used in the proof above.

5.3 Continuous Time Results

In this section we consider some limit results for a continuous time stochastic process. The first part deals purely with the theoretical results, whereas the second part delivers statistical applications. We first wish to consider the model

$$X_t = \int_{\mathbb{R}^d} \psi(t+x)\,\mathbb{M}(dx), \quad t \in \mathbb{R}^d \tag{5.31}$$

defined on all of Euclidean space. The integrand function ψ must satisfy some regularity conditions: $\psi \in \mathbb{L}^1(\mathbb{R}^d, \lambda)$ where λ is Lebesgue measure on $\mathbb{R}^d$; in addition, ψ must be continuous and bounded almost everywhere. This is called a "moving average" model, because of the form of ψ; there is a simple connection to "MA (∞)" models encountered in classical time series, as the following example demonstrates.

The following theorem deals with the statistics problem of the asymptotic distribution of the partial sums. We consider a sampling region $K = (0, n_1] \times \cdots \times (0, n_d]$ from which the random field is observed. Let $N = \Pi_{i=1}^{n} n_i$ be the total volume of the sampling region; the appopriate rate of convergence will then be $N^{\frac{1}{\alpha}}$. We wish to sum X_t over *all* points in K, and so to that end we must calculate

$$\int_K X_t \lambda(dt). \tag{5.32}$$

Now a discrete sum of the random field is certainly well–defined by the linearity of the α-stable random integral, but it is not a priori clear that (5.32) makes any sense. Thus we make the following definition:

Definition 5.3.1. *Let $G_t := \int_{\mathbb{R}^d} \psi(t+x)\mathbb{M}(dx)$, for any $t \in \mathbb{R}^d$. When we write (5.32),* *we mean the limit in probability as $m \to \infty$ of the following:*

$$\sum_{t_i \in K_m} \left[G_{t_{i+1}} - G_{t_i}\right] \Delta t_i, \tag{5.33}$$

where K_m is a mesh of m points t_i in K, and the mesh gets progressively more fine as m is increased (this is the usual Riemann sums construction). Assume for simplicity that the spacings between mesh points are uniform, and thus let $\Delta t_i = \Delta$ denote that distance.

We must prove that this definition makes sense. By using the linearity of the stable integral, line (5.33) becomes

$$\int_{\mathbb{R}^d} \sum_{t_i \in K_m} \left(\psi(t_{i+1} + x) - \psi(t_i + x)\right) \Delta t_i \, \mathbb{M}(dx) = \int_{\mathbb{R}^d} F_m(x) \, \mathbb{M}(dx) \tag{5.34}$$

say. Now it follows that the limit in probability as $m \to \infty$ of (5.34) is $\int_{\mathbb{R}^d} F(x)\mathbb{M}(dx)$ for $F(x) := \int_K \psi(t + x)\, \lambda(dt)$ so long as

$$\int_{\mathbb{R}^d} \left|F_m(x) - F(x)\right|^\alpha \lambda(dx) \to 0 \tag{5.35}$$

as $n \to \infty$. The integrand in (5.35) is clearly bounded above by an $\mathbb{L}^1$ function: first we may break it into three parts by gaining the constant $3^{\alpha-1}$. The first term is then bounded above as follows:

$$\begin{aligned} \left|\sum_{t_i \in K_n} \psi(t_{i+1} + x)\Delta t_i\right|^\alpha &= \left|\sum_{t_i \in K_n} \psi(t_{i+1} + x)\Delta t_i\right| \cdot \left|\sum_{t_i \in K_n} \psi(t_{i+1} + x)\Delta t_i\right|^{\alpha-1} \\ &\leqslant \|\psi\|_\infty N \cdot \sum_{t_i \in K_m} |\psi(t_{i+1} + x)|^{\alpha-1} |\Delta t_i|^{\alpha-1}. \end{aligned} \tag{5.36}$$

Taking the integral over $\mathbb{R}^d$ of this gives a bound (after swapping order of integration and using shift invariance of Lebesgue measure) of

$$\|\psi\|_\infty N \cdot \int_{\mathbb{R}^d} |\psi(u)|^{\alpha-1}\, du \sum_{t_i \in K_m} |\Delta t_i|^{\alpha-1} < \infty. \tag{5.37}$$

Therefore the function in the last line of (5.36) is an $\mathbb{L}^1$ function which bounds the first term. Similar bounds for the other two terms may be determined, so that the Lebesgue

Dominated Convergence Theorem can be applied. Taking the limit as $m \to \infty$ inside the outer integral in (5.35) produces the desired result, since F_n converges pointwise to F.

Thus we have established that the expression (5.32) makes sense, and also that it is equal to

$$\int_{\mathbb{R}^d} \int_K \psi(t+x)\, \lambda(dt) \mathbb{M}(dx). \tag{5.38}$$

We may now state the following theorem:

Theorem 5.3.1. *The weak convergence of the partial sums is given below:*

$$N^{-\frac{1}{\alpha}} \int_K X_t \, \lambda(dt) \stackrel{\mathcal{L}}{\Longrightarrow} \left(\int_{\mathbb{R}^d} \psi(u)\, \lambda(du) \right) \cdot \mathbb{M}(B) \tag{5.39}$$

where B denotes the unit cube. It follows that the right hand side is an α-stable random variable with scale $\int_{\mathbb{R}^d} \psi(u)\, \lambda(du)$, mean zero, and skewness zero.

Proof First we partition the set K into cubes of unit size (so they are vector shifts of B, the unit cube). Utilizing (5.38), we have

$$\begin{aligned} N^{-\frac{1}{\alpha}} \int_K X_t \, \lambda(dt) &= N^{-\frac{1}{\alpha}} \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=0}^{n_d-1} \int_{j_1}^{j_1+1} \cdots \int_{j_d}^{j_d+1} X_t \, \lambda(dt) \qquad (5.40) \\ &= N^{-\frac{1}{\alpha}} \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=0}^{n_d-1} \int_{\mathbb{R}^d} \left[\int_{j_1}^{j_1+1} \cdots \int_{j_d}^{j_d+1} \psi(t+x) \lambda(dt) \right] \mathbb{M}(dx) \\ &= N^{-\frac{1}{\alpha}} \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=0}^{n_d-1} \int_{\mathbb{R}^d} \left[\int_0^1 \cdots \int_0^1 \psi(t+j+x) \lambda(dt) \right] \mathbb{M}(dx) \\ &= N^{-\frac{1}{\alpha}} \sum_{j=0}^{n-1} \int_{\mathbb{R}^d} \left[\int_B \psi(t+j+x) \lambda(dt) \right] \mathbb{M}(dx), \end{aligned}$$

where we have abbreviated $\sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=0}^{n_d-1}$ by $\sum_{j=0}^{n-1}$. Now in Chapter Three, we demonstrated that

$$\int_{\mathbb{R}} n^{-1} \left(\sum_{j=1}^{n} \phi(j+x) \right)^{<\alpha>} \lambda(dx) \to \int_0^1 \left(\sum_{j \in \mathbb{Z}} \phi(j+x) \right)^{<\alpha>} \lambda(dx) \tag{5.41}$$

as $n \to \infty$. This was the $d = 1$ case, for a discrete sum of the X_t's. In the previous section, our Lemma 5.2.1 extended this to a greater dimension d, so that we have the result:

$$\int_{\mathbb{R}^d} N^{-1} \left(\sum_{j=1}^{n} \phi(j+x) \right)^{<\alpha>} \lambda(dx) \to \int_B \left(\sum_{j \in \mathbb{Z}^d} \phi(j+x) \right)^{<\alpha>} \lambda(dx) \tag{5.42}$$

where j and x are now vectors. The significance of (5.42) was to establish convergence of the skewness parameters of the partial sum to the skewness parameter of the α-stable limit random variable. Similar results hold for the convergence of the scale (just replace $|\cdot|^{<\alpha>}$ by $(\cdot)^\alpha$ in (5.41) and (5.42)). These convergences were shown to be valid for any function ϕ which is continuous and bounded almost everywhere (with respect to the control measure λ), and in $\mathbb{L}^1(\mathbb{R}^d, \lambda)$.

So let us now consider the scale and skewness parameters of the α-stable random variable

$$\begin{aligned} & N^{-\frac{1}{\alpha}} \sum_{j=0}^{n-1} \int_{\mathbb{R}^d} \left[\int_B \psi(t+j+x) \lambda(dt) \right] \mathbb{M}(dx) \\ &= \int_{\mathbb{R}^d} \left[N^{-\frac{1}{\alpha}} \sum_{j=0}^{n-1} \int_B \psi(t+j+x) \lambda(dt) \right] \mathbb{M}(dx) \end{aligned} \tag{5.43}$$

for every vector n. First we should establish that (5.43) is well-defined, which is accomplished by showing that is has finite scale for each fixed value of n. The αth power of the scale is defined to be:

$$\int_{\mathbb{R}^d} N^{-1} \left| \sum_{j=0}^{n-1} \int_B \psi(t+j+x) \lambda(dt) \right|^\alpha \lambda(dx). \tag{5.44}$$

Now let $\phi(u) := \int_B \psi(u+t)\, \lambda(dt)$; it then follows from elementary arguments that ϕ is continuous and bounded almost everywhere, just like ψ, and is an $\mathbb{L}^1$ function as well. Using this substitution, (5.44) now becomes

$$\int_{\mathbb{R}^d} N^{-1} \left| \sum_{j=0}^{n-1} \phi(j+x) \right|^\alpha \lambda(dx). \tag{5.45}$$

In a similar vein, the αth power of the skewness works out to be

$$\int_{\mathbb{R}^d} N^{-1}\left(\sum_{j=0}^{n-1}\phi(j+x)\right)^{<\alpha>}\lambda(dx). \tag{5.46}$$

To prove the finiteness of (5.45) – and also of (5.46) – we use the following calculation:

$$\begin{aligned}
&\int_{\mathbb{R}^d} N^{-1}\left|\sum_{j=0}^{n-1}\phi(j+x)\right|^{\alpha}\lambda(dx) &(5.47)\\
&= N^{-1}\int_{\mathbb{R}^d}\left|\sum_{j=0}^{n-1}\phi(j+x)\right|\cdot\left|\sum_{j=0}^{n-1}\phi(j+x)\right|^{\alpha-1}\lambda(dx)\\
&\leqslant N\cdot\|\phi\|_\infty\cdot N^{-1}\int_{\mathbb{R}^d}\sum_{j=0}^{n-1}|\phi(j+x)|^{\alpha-1}\lambda(dx)\\
&= N\cdot\|\phi\|_\infty\cdot\|\phi\|_{\alpha-1}^{\alpha-1}<\infty,
\end{aligned}$$

where we have used the fact that ϕ is bounded and summable. This bound is not particullarly tight, but it establishes the finiteness of (5.45) and (5.46). Thus (5.43) describes a well-defined α-stable random variable. Turning now to the proposed limit random variable

$$\int_B\left(\int_{\mathbb{R}^d}\psi(t+x)\lambda(dt)\right)\mathbb{M}(dx) \tag{5.48}$$

– which actually does equal in distribution the limit random variable described in (5.39) – we calculate the αth power of its scale as follows:

$$\int_B\left|\int_{\mathbb{R}^d}\psi(t+x)\lambda(dt)\right|^{\alpha}\lambda(dx)=\left|\int_{\mathbb{R}^d}\psi(t)\lambda(dt)\right|^{\alpha}\cdot\lambda(B)\leqslant\|\psi\|_\alpha^\alpha<\infty. \tag{5.49}$$

Thus the limit random variable is well-defined, and we only need prove that the scale (5.45) and skewness (5.46) parameters of our partial sums (5.43) converge to the corresponding parameters of the limit random variable (5.48) (all the random variables have mean zero). For skewness (and the argument for scale is even easier) we utilize (5.42), which is valid, because our choice of ϕ is continuous and bounded almost everywhere, and integrable. Thus, substituting ϕ into (5.42), we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^d} N^{-1}\left(\sum_{j=0}^{n-1}\phi(j+x)\right)^{<\alpha>}\lambda(dx) && (5.50)\\
&= \int_{\mathbb{R}^d} N^{-1}\left(\sum_{j=1}^{n}\int_B \psi(t+j+x)\lambda(dt)\right)^{<\alpha>}\lambda(dx)\\
&\to \int_B \left(\sum_{j\in\mathbb{Z}^d}\int_B \psi(t+j+x)\lambda(dt)\right)^{<\alpha>}\lambda(dx)\\
&= \int_B \left(\int_{\mathbb{R}^d}\psi(t+x)\lambda(dt)\right)^{<\alpha>}\lambda(dx)
\end{aligned}
$$

which is the skewness of (5.48), as desired. Since all the parameters converge, the weak convergence is establilshed. †.

Now some simple extensions of this proof technique give us a similar limit theorem for the periodogram. Recall that the periodogram is used in classical time series to estimate the spectral density function. For non-square integrable time series or random fields, the N^{-1} rate is adjusted to $N^{-\frac{2}{\alpha}}$, and the limit theory is somewhat different in the *iid* case. Our periodogram is defined by

$$
\begin{aligned}
\tilde{I}_n(\omega) &:= N^{-\frac{2}{\alpha}}\left|\int_K X_t e^{it'\omega}\lambda(dt)\right|^2 && (5.51)\\
&= N^{-\frac{2}{\alpha}}\left\{\left(\int_K X_t\cos t'\omega\lambda(dt)\right)^2+\left(\int_K X_t\sin t'\omega\lambda(dt)\right)^2\right\},
\end{aligned}
$$

where $\omega = (\omega_1, \cdots, \omega_d)$ is a vector of frequencies, so $\omega_i \in (0, 2\pi]$. It is of interest to explore the joint limit behavior of (5.51) with the partial sums (5.32), in order to form a normalization as in Chapter Three. It suffices to look at the joint convergence of

$$
\left(N^{-\frac{1}{\alpha}}\int_K X_t\,\lambda(dt), N^{-\frac{1}{\alpha}}\int_K X_t\,\cos t'\omega\,\lambda(dt), N^{-\frac{1}{\alpha}}\int_K X_t\,\sin t'\omega\,\lambda(dt)\right), \qquad (5.52)
$$

and then apply the continuous function $(x, y, z) \mapsto (x, y^2 + z^2)$. Let's first examine the middle term of (5.52). Using the same techniques as discussed above, we see that we have the following equivalence just like (5.43):

$$N^{-\frac{1}{\alpha}} \int_K X_t \, \cos t'\omega \, \lambda(dt) = N^{-\frac{1}{\alpha}} \sum_{j=0}^{n-1} \int_{\mathbb{R}^d} \int_B \psi(j+t+x) \, \cos(t+j)'\omega \, \lambda(dt) \mathbb{M}(dx). \tag{5.53}$$

Now the αth power of the skewness (and the scale is quite similar) is

$$\begin{aligned} &\int_{\mathbb{R}^d} N^{-1} \Bigg(\sum_{j=0}^{n-1} \cos j'\omega \int_B \psi(t+j+x) \, \cos t'\omega \, \lambda(dt) \\ &- \sin j'\omega \int_B \psi(t+j+x) \, \sin t'\omega \, \lambda(dt) \Bigg)^{<\alpha>} \lambda(dx) \\ &= \int_{\mathbb{R}^d} N^{-1} \left(\sum_{j=0}^{n-1} a_j \, \phi^c(j+x) - b_j \, \phi^s(j+x) \right)^{<\alpha>} \lambda(dx) \end{aligned} \tag{5.54}$$

for $\phi^c(u) = \int_B \psi(t+u) \, \cos t'\omega \, \lambda(dt)$ and $\phi^s(u) = \int_B \psi(t+u) \, \sin t'\omega \, \lambda(dt)$; $a_j = \cos j'\omega$ and $b_j = \sin j'\omega$. Now in the proof of (5.42), we must establish that $g_n(x) := \sum_{j=0}^{n-1} a_j \phi^c(j+x) + b_j \phi^s(j+x)$ has bounded $\mathbb{L}^\infty$ norm and bounded $\mathbb{L}^{\alpha-1}$ norm. But it is easy to see this:

$$|g_n(x)| \leqslant \sum_{j \in \mathbb{Z}^d} |\phi^c(j+x)| + |\phi^s(j+x)| \tag{5.55}$$

and both ϕ^c and ϕ^s are bounded functions, since they are integrals of the bounded function ψ over a compact set against the Radon measures $\cos t'\omega \, \lambda(dt)$ and $\sin t'\omega \, \lambda(dt)$ respectively. Thus $g_n \in \mathbb{L}^\infty$, and by a similar calculation

$$|g_n(x)|^{\alpha-1} \leqslant \sum_{j=0}^{n-1} |\phi^c(j+x)|^{\alpha-1} + |\phi^s(j+x)|^{\alpha-1} \tag{5.56}$$

so that we have

$$\begin{aligned} &N^{-1} \int_{\mathbb{R}^d} |g_n(x)|^{\alpha-1} \lambda(dx) \\ &\leqslant N^{-1} \int_{\mathbb{R}^d} \sum_{j=0}^{n-1} |\phi^c(j+x)|^{\alpha-1} + |\phi^s(j+x)|^{\alpha-1} \lambda(dx) \\ &= 2 \int_{\mathbb{R}^d} |\phi(x)|^{\alpha-1} \lambda(dx) < \infty. \end{aligned} \tag{5.57}$$

So we can apply all the same techniques to (5.53) that we did to (5.32). Also, the third term in (5.52) can be treated to the same analysis. So each of these terms can be shown to converge individually (their limits will be described below). But they also converge jointly, which is easily shown as follows:

$$\begin{aligned} &\mathbb{E}\exp\left\{ i\left(\eta_1 \int_K X_t\,\lambda(dt) + \eta_2 \int_K X_t\,\cos t'\omega\,\lambda(dt) + \eta_3 \int_K X_t\,\sin t'\omega\,\lambda(dt)\right)\right\} \qquad (5.58) \\ &= \mathbb{E}\exp\left\{ i\left(\int_{\mathbb{R}^d}\int_K \psi(x+t)\left(1+\cos t'\omega + \sin t'\omega\right)\,\lambda(dt)\,\mathbb{M}(dx)\right)\right\} \end{aligned}$$

so that we only need to consider the integrand function $\psi(x+t)\,(1+\cos t'\omega+\sin t'\omega)$, which we can analyze as we did above. Now (5.58) converges to the characteristic function of

$$\int_B \int_{\mathbb{R}^d} \psi(x+t)(1+\cos t'\omega + \sin t'\omega)\,\lambda(dt)\,\mathbb{M}(dx). \qquad (5.59)$$

Summarizing all of this, we have the following theorem:

Theorem 5.3.2. *Let $\alpha \in (1,2)$. Then we have the following joint convergence:*

$$\begin{aligned} &\left(N^{-\frac{1}{\alpha}} \int_K X_t\,\lambda(dt),\, N^{-\frac{1}{\alpha}} \int_K X_t\,\cos t'\omega\,\lambda(dt),\, N^{-\frac{1}{\alpha}} \int_K X_t\,\sin t'\omega\,\lambda(dt)\right) \\ &\stackrel{\mathcal{L}}{\Longrightarrow} \left(\int_B\int_{\mathbb{R}^d} \psi(x+t)\,\lambda(dt)\,\mathbb{M}(dx),\right. \\ &\int_B\int_{\mathbb{R}^d} \psi(x+t)\cos t'\omega\,\lambda(dt)\,\mathbb{M}(dx), \\ &\left.\int_B\int_{\mathbb{R}^d} \psi(x+t)\sin t'\omega\,\lambda(dt)\,\mathbb{M}(dx)\right). \end{aligned} \qquad (5.60)$$

Thus it follows that

$$\begin{aligned} &\frac{\int_K X_t\,\lambda(dt)}{\tilde{I}_n(\omega)} \stackrel{\mathcal{L}}{\Longrightarrow} \\ &\frac{\int_B\int_{\mathbb{R}^d}\psi(x+t)\,\lambda(dt)\,\mathbb{M}(dx)}{\sqrt{\left(\int_B\int_{\mathbb{R}^d}\psi(x+t)\cos t'\omega\,\lambda(dt)\,\mathbb{M}(dx)\right)^2 + \left(\int_B\int_{\mathbb{R}^d}\psi(x+t)\sin t'\omega\,\lambda(dt)\,\mathbb{M}(dx)\right)^2}}. \end{aligned} \qquad (5.61)$$

Remark Thus the periodogram provides a useful "self-normalization" for the sample mean. The statistic on the left hand side of (5.61) can be calculated without foreknowledge of the model parameter α.

Finally, we can prove a similar result for the sample variance $\int_K X_t^2 \lambda(dt)$. First, from the paper [54], we have

$$N^{-\frac{2}{\alpha}} \int_K X_t^2 \, \lambda(dt) = o_P(1) + N^{-\frac{2}{\alpha}} \int_{\mathbb{R}^d} \int_K \psi^2(x+t) \, \lambda(dt) \tilde{\mathbb{M}}(dx) \tag{5.62}$$

where $\tilde{\mathbb{M}}$ is an $\frac{\alpha}{2}$-stable totally right skewed random measure. Now if we look at the $\frac{\alpha}{2}$th power of the skewness of this, we obtain

$$\int_{\mathbb{R}^d} N^{-1} \left(\sum_{j=0}^{n-1} \int_B \psi^2(t+j+x) \, \lambda(dt) \right)^{<\frac{\alpha}{2}>} \lambda(dx) = \int_{\mathbb{R}^d} N^{-1} \left(\sum_{j=0}^{n-1} \phi(j+x) \right)^{<\frac{\alpha}{2}>} \lambda(dx) \tag{5.63}$$

where $\phi(u) := \int_B \psi^2(u+t) \, \lambda(dt)$. Now this function ϕ will be continuous and bounded almost everywhere, and is in $\mathbb{L}^{\frac{\alpha}{2}}$ as well. Thus by applying (5.42) with α replaced by $\frac{\alpha}{2}$, we obtain the following result:

Theorem 5.3.3. *The sample variance obeys the following weak convergence:*

$$N^{-\frac{2}{\alpha}} \int_K X_t^2 \, \lambda(dt) \xrightarrow{\mathcal{L}} \int_{\mathbb{R}^d} \psi^2(u) \, \lambda(du) \mathbb{M}(B). \tag{5.64}$$

In the next section we will consider the additional assumption that the location of the data points are given by an independent Poisson Point Process.

5.4 Poisson Random Measures

We will now consider a more intricate situation, wherein the locations of the data are themselves random. It occurs many times in statistical problems that random field data is not distributed at lattice points, but instead is scattered all around the observation region. Generally speaking such a situation can be modeled by a Point Process on the region, which is simply a positive integer-valued random measure. If we wish to impose that the distribution of points does not depend on the location of the observation region, but only on its size and shape, then we say the random measure is spatially homogeneous (this is like a stationarity assumption). Also it is sensible that the distribution of points in one observation region should be completely independent of the distribution of points in another disjoint observation region – this is called the "independent scattering" property. It turns out that Poisson Random Measures (PRM) satisfy these properties, and will serve as a decent model.

Therefore let $\mathbb{N}$ denote a PRM with mean measure Λ (so $\mathbb{N}$ is sometimes denoted $PRM(\Lambda)$), i.e.

Definition 5.4.1. *We say that $\mathbb{N}$ is $PRM(\Lambda)$ on the measure space $\{\mathbb{R}^d, \mathcal{B}, \Lambda\}$ (where $\mathcal{B}$ are the Borel sets in $\mathbb{R}^d$) if and only if it is an independently scattered, σ-additive random measure which satisfies*

$$\mathbb{P}\left[\mathbb{N}(A) = k\right] = \exp\{-\Lambda(A)\}\frac{\Lambda(A)^k}{k!}, \quad k = 0, 1 \cdots \tag{5.65}$$

for every $A \in \mathcal{B}_0 := \{B \subset \mathcal{B} : \Lambda(A) < \infty\}$. In other words, $\mathbb{N}(A) \sim Pois(\Lambda(A))$. We call Λ the mean measure of $\mathbb{N}$.

Remarks More generally, we could just define the mean measure to be $\Lambda(\cdot) := \mathbb{E}\mathbb{N}(\cdot)$. Now if we impose that Λ be a translation invariant measure on $\mathbb{R}^d$, then spatial homogeneity follows at once from (5.38). Of course, Λ must then be Lebesgue measure (denoted by λ, as in the previous section) modulo some constant positive multiplicative factor, i.e. $\Lambda = r\lambda$ for some $r \in \mathbb{R}^+$.

Now we are interested in investigating the partial sums' limit behavior over the observation region K (we preserve the notation from last section), where the data locations are now determined by the random measure $\mathbb{N}$. Thus we wish to study

$$\mathbb{N}(K)^{-\frac{1}{\alpha}} \int_K X_t \, \mathbb{N}(dt) \tag{5.66}$$

as K expands to fill the positive orthant of $\mathbb{R}^d$, i.e. taking the limit as the vector n tends to infinity in each component. We have included the factor $\Lambda(K)^{-\frac{1}{\alpha}}$ since it is the appropriate rate of convergence (and can be calculated from the data, whereas $N^{-\frac{1}{\alpha}}$ cannot – see the later discussion on statistical applications). First, we must ensure that (5.66) makes sense.

It is a fact that any point measure m can be written as

$$m := \sum_i \varepsilon_{t_i} \tag{5.67}$$

where t_i are a collection of points in the underlying space (in our case this is $\mathbb{R}^d$, so the t_i would be real-valued d-vectors), and ε_x denotes the "point-mass" measure at x, i.e. $\varepsilon_x(A) := 1_A(x)$. So we see that $m(A)$ just counts up the number of points t_i that fall in the set A. Now our random measure $\mathbb{N}$ can be viewed as a mapping from the underlying probability space into a space of point measures on $\mathbb{R}^d$ – that space will be denoted by $M_p(\mathbb{R}^d)$ from now on. In other words, for each $\omega \in \Omega$ the underlying probability space, we have $\mathbb{N}(\omega)$ is some point measure m, which is completely determined by its points t_i. It follows that we can write

$$\mathbb{N}(\cdot) = \sum_i \varepsilon_{T_i}(\cdot) \tag{5.68}$$

where the T_i are now random elements of $\mathbb{R}^d$. Then we may interpret (5.66) as

$$\mathbb{N}(K)^{-\frac{1}{\alpha}} \int_K X_t \, \mathbb{N}(dt) = \mathbb{N}(K)^{-\frac{1}{\alpha}} \sum_{i:T_i \in K} X(T_i), \tag{5.69}$$

which is still a bit troubling – what is the meaning of $X(T_i)$ for a random T_i ? Generally speaking, α-stable integrals are only well-defined objects for deterministic integrands. The precise answer is to look at the characteristic function of (5.66), and define the summation conditionally on $\mathbb{N}$ being a particular point measure m:

$$\int_K X_t \, \mathbb{N}(dt) \Big| \{\mathbb{N} = m\} =^d \int_K X_t \, m(dt) = \sum_{i:t_i \in K} X(t_i). \tag{5.70}$$

Both (5.69) and (5.70) are useful; the former for the calculation of the statistic and visualization of the inference problem, and the latter for carrying out the probabilistic analysis.

Before stating the main theorems, we note that it follows from the Law of Large Numbers, independent scattering, spatial homogeneity, and shift invariance of Λ that

$\frac{\mathbb{N}(K)}{\Lambda(K)} \xrightarrow{a.s.} 1$ as each component of n tends to infinity. Thus we will simply replace $\mathbb{N}(K)$ by $\Lambda(K)$ in the statements of all asymptotic results below. But note that $\Lambda(K)^{-\frac{1}{\alpha}} = r^{-\frac{1}{\alpha}} N^{-\frac{1}{\alpha}}$, so up to a constant factor, in the limit, we can just use $N^{-\frac{1}{\alpha}}$ as the rate of convergence.

Theorem 5.4.1. *The following limit result holds true:*

$$N^{-\frac{1}{\alpha}} \int_K X_t \, \mathbb{N}(dt) \overset{\mathcal{L}}{\Longrightarrow} \int_{\mathbb{R}^d} \left(\int_B \psi(t+x) \, \mathbb{M}(dx) \right) \mathbb{N}(dt), \tag{5.71}$$

where the left and right hand sides of (5.71) are defined conditionally on the event $\{\mathbb{N} = m\}$. *We may also formulate the same result, by the comments in the previous paragraph, as*

$$\mathbb{N}(K)^{-\frac{1}{\alpha}} \int_K X_t \, \mathbb{N}(dt) \overset{\mathcal{L}}{\Longrightarrow} \int_{\mathbb{R}^d} \left(\int_B \psi(t+x) \, \mathbb{M}(dx) \right) \mathbb{N}(dt). \tag{5.72}$$

Proof It suffices to prove the convergence conditional on the event $\{\mathbb{N} = m\}$ for each point measure m, since that is the interpretation (via the comments after (5.69) and (5.70)) that we should give to (5.71). Let's first demonstrate that the right hand side of (5.71) is bounded in probability. So let $S := \int_{\mathbb{R}^d} \left(\int_B \psi(t+x) \, \mathbb{M}(dx) \right) \mathbb{N}(dt)$, and consider $\mathbb{E}[|S| \,\big|\, \mathbb{M}]$. The random measure $\mathbb{M}$ can be viewed as a random variable taking values in the space of signed measures on $\mathbb{R}^d$; taking one element μ of this space, we have

$$
\begin{aligned}
&\mathbb{E}\left[|S|\middle|\mathbb{M}=\mu\right] &(5.73)\\
&=\mathbb{E}\left[\left|\int_{\mathbb{R}^d}\int_B \psi(t+x)\mathbb{M}(dx)\mathbb{N}(dt)\right|\middle|\mathbb{M}=\mu\right]\\
&=\mathbb{E}\left[\left|\int_{\mathbb{R}^d}\int_B \psi(t+x)\mu(dx)\mathbb{N}(dt)\right|\right]\\
&\leqslant \mathbb{E}\left[\int_{\mathbb{R}^d}\left|\int_B \psi(t+x)\mu(dx)\right|\mathbb{N}(dt)\right]\\
&=\int_{\mathbb{R}^d}\left|\int_B \psi(t+x)\mu(dx)\right|\Lambda(dt)\\
&\leqslant \int_{\mathbb{R}^d}\int_B |\psi(t+x)|\,|\mu|(dx)\Lambda(dt)\\
&=\int_B\int_{\mathbb{R}^d}|\psi(t+x)|\,\Lambda(dt)|\mu|(dx)\\
&=\int_{\mathbb{R}^d}|\psi(u)|\,\Lambda(du)|\mu|(B)
\end{aligned}
$$

by using the triangle inequality for signed measures, Tonelli's Theorem, and shift invariance of the measure Λ. Thus in other words, we have $\mathbb{E}[|S|\big|\mathbb{M}] \leqslant \int_{\mathbb{R}^d}|\psi(u)|\,\Lambda(du)|\mathbb{M}|(B)$. If we now remove the conditioning on $\mathbb{M}$ by taking a second expectation, we have

$$
\mathbb{E}[|S|] \leqslant \int_{\mathbb{R}^d}|\psi(u)|\,\Lambda(du)\mathbb{E}[|\mathbb{M}|\,(B)] < \infty \tag{5.74}
$$

since $|\mathbb{M}|(B)$ is an $\mathbb{L}^1$ random variable. Thus S is bounded in probability, as desired.

We will now establish the convergence result (5.71). Let's fix a point measure m throughout the discussion, where m is given by (5.67). Then the situation is quite similar to the discrete case – see (5.42) – and the left hand side of (5.71) may be written

$$
\begin{aligned}
N^{-\frac{1}{\alpha}}\int_K X_t\, m(dt) &= N^{-\frac{1}{\alpha}}\int_{\mathbb{R}^d}\int_K \psi(x+t)\,m(dt)\,\mathbb{M}(dx) &(5.75)\\
&= N^{-\frac{1}{\alpha}}\int_{\mathbb{R}^d}\left(\sum_{j=0}^{n-1}\int_{B+j}\psi(t+x)\,m(dt)\right)\mathbb{M}(dx).
\end{aligned}
$$

So if we consider the αth power of the skewness, we obtain

$$\int_{\mathbb{R}^d} N^{-1}\left(\sum_{j=0}^{n-1}\int_{B+j}\psi(t+x)\,m(dt)\right)^{<\alpha>}\lambda(dx)=\int_{\mathbb{R}^d} N^{-1}\left(\sum_{j=0}^{n-1}\phi(j+x)\right)^{<\alpha>}\lambda(dx), \tag{5.76}$$

where $\phi(u) := \int_B \psi(t+u)\,m(dt)$. Note that since m is a point measure, it is shift invariant, so it follows that

$$\begin{aligned}\phi(j+x) &= \int_B \psi(t+j+x)\,m(dt) \\ &= \sum_{i:t_i\in B}\psi(t_i+j+x) \\ &= \sum_{i:t_i\in B+j}\psi(t_i+x) \\ &= \int_{B+j}\psi(t+x)\,m(dt)\end{aligned} \tag{5.77}$$

as desired. Now we can apply (5.42) if we show that our function ϕ in (5.77) is continuous and bounded almost everywhere, and is summable. In order the establish these properties, we must assume that $m(B)<\infty$ for each choice of m, or at least for almost every m with respect to $\mathbb{N}$ in the space $M_p(\mathbb{R}^d)$. In order to perform the following calculation, we must first view $\mathbb{N}$ as a mapping from the abstract probability space to $M_p(\mathbb{R}^d)$, so that we get the distribution Ξ as the push-forward measure of the random variable $\mathbb{N}$ (where its values are point measures!). If we let $E := \{m\in M_p(\mathbb{R}^d) : m(B)=\infty\}$ be the set of bisbehaving point measures; we wish to show that $\Xi(E)=0$. This is beneficial, because then

$$\begin{aligned}&\mathbb{E}\mathbb{E}\exp\{i\eta S\}\Big|\mathbb{N} = \int_{M_p(\mathbb{R}^d)}\mathbb{E}\left[\exp\{i\eta S\}\Big|\mathbb{N}=m\right]\Xi(dm) \\ &\int_E \mathbb{E}\left[\exp\{i\eta S\}\Big|\mathbb{N}=m\right]\Xi(dm)+\int_{M_p(\mathbb{R}^d)\setminus E}\mathbb{E}\left[\exp\{i\eta S\}\Big|\mathbb{N}=m\right]\Xi(dm) \\ &=0+\int_{M_p(\mathbb{R}^d)\setminus E}\mathbb{E}\left[\exp\{i\eta S\}\Big|\mathbb{N}=m\right]\Xi(dm)\end{aligned} \tag{5.78}$$

for any $\eta \in \mathbb{R}$ and any random variable S, so that we can restrict our attention without loss of generality to $M_p(\mathbb{R}^d) \setminus E$. But it is easy to see that

$$\Xi(E) = \mathbb{P}\{\omega : \mathbb{N}(\omega) \in E\} = \mathbb{P}[\mathbb{N}(B) = \infty] = 0 \tag{5.79}$$

since $B \in \mathcal{B}_0$ (recall that B is just the unit cube). So we have proved that $\mathbb{N}(B) < \infty$ with probability one.

So the boundedness and continuity of ϕ are now straight-forward to establish:

$$\|\phi\|_\infty \leqslant \int_B \|\psi\|_\infty \, m(dt) = \|\psi\|_\infty \cdot m(B) < \infty. \tag{5.80}$$

Also for any $m \in E$, the expression $\int_B \psi(t+u)\, m(dt)$ is really a finite sum of the almost everywhere continuous functions $\{\psi(t_i + \cdot) : t_i \in B\}$, and hence $\phi(\cdot)$ is also continuous almost everywhere. Lastly, integrability trivially follows from a calculation just like (5.78).

Hence we may apply (5.42) to the αth power of the skewness, and obtain the limit result

$$\begin{aligned} &\int_B \left(\sum_{j \in \mathbb{Z}^d} \phi(j + x) \right)^{<\alpha>} \lambda(dx) \qquad (5.81) \\ &= \int_B \left(\sum_{j \in \mathbb{Z}^d} \int_{B+j} \psi(t + x)\, m(dt) \right)^{<\alpha>} \lambda(dx) \\ &= \int_B \left(\int_{\mathbb{R}^d} \psi(t + x)\, m(dt) \right)^{<\alpha>} \lambda(dx). \end{aligned}$$

With a similar result for the skewness, we have established

$$\begin{aligned} &N^{-\frac{1}{\alpha}} \int_K X_t \,\mathbb{N}(dt) \bigg| \mathbb{N} = m \qquad (5.82) \\ &\stackrel{\mathcal{L}}{\Longrightarrow} \int_B \int_{\mathbb{R}^d} \psi(t + x)\, m(dt)\, \mathbb{M}(dx) \\ &= \int_{\mathbb{R}^d} \int_B \psi(t + x)\, \mathbb{M}(dx)\, m(dt). \end{aligned}$$

Now this is actually the desired result, because (5.71) is to be interpreted as true conditional on the set $\{\mathbb{N} = m\}$ for any $m \in M_p(\mathbb{R}^d)$. †

Similar results can be proved for the periodogram and the sample variance, by applying similar methods to the previous section. The periodogram in this case is defined by

$$\tilde{I}_n(\omega) = \mathbb{N}(K)^{-\frac{2}{\alpha}} \left| \int_K X_t \exp\{it'\omega\} \mathbb{N}(dt) \right|^2, \tag{5.83}$$

where $\omega = (\omega_1, \cdots, \omega_d)$ is a vector of frequencies, so $\omega_i \in (0, 2\pi]$. Here we need to prove the joint convergence, conditional on $\mathbb{N} = m$, of the vector

$$\left(N^{-\frac{1}{\alpha}} \int_K X_t \, \mathbb{N}(dt), N^{-\frac{1}{\alpha}} \int_K X_t \cos t'\omega \, \mathbb{N}(dt), N^{-\frac{1}{\alpha}} \int_K X_t \sin t'\omega \, \mathbb{N}(dt) \right). \tag{5.84}$$

The argument is just like the one used in the proof of Theorem 5.3.2, with $\phi^c(u) = \int_B \psi(t+u) \cos t'\omega \, m(dt)$ and $\phi^s(u) = \int_B \psi(t+u) \sin t'\omega \, m(dt)$. After establishing the appropriate continuity, boundedness, and summability properties, we have the following result:

Theorem 5.4.2. *Let $\alpha \in (1,2)$. Then we have the following joint convergence:*

$$\begin{aligned}
&\left(\mathbb{N}(K)^{-\frac{1}{\alpha}} \int_K X_t \, \mathbb{N}(dt), \mathbb{N}(K)^{-\frac{1}{\alpha}} \int_K X_t \cos t'\omega \, \mathbb{N}(dt), \mathbb{N}(K)^{-\frac{1}{\alpha}} \int_K X_t \sin t'\omega \, \mathbb{N}(dt) \right) \\
&\stackrel{\mathcal{L}}{\Longrightarrow} \left(\int_{\mathbb{R}^d} \left(\int_B \psi(t+x) \, \mathbb{M}(dx) \right) \mathbb{N}(dt), \right. \\
&\int_{\mathbb{R}^d} \left(\int_B \psi(t+x) \cos t'\omega \, \mathbb{M}(dx) \right) \mathbb{N}(dt), \\
&\left. \int_{\mathbb{R}^d} \left(\int_B \psi(t+x) \sin t'\omega \, \mathbb{M}(dx) \right) \mathbb{N}(dt) \right).
\end{aligned} \tag{5.85}$$

It also follows that we have a self-normalized convergence result, like (5.61).

Finally, with the choice $\phi(u) := \int_B \psi^2(u+t) \, m(dt)$, we obtain a sample variance result too:

Theorem 5.4.3. *The sample variance in the marked point process context has the following asymptotic behavior:*

$$\mathbb{N}(K)^{-\frac{2}{\alpha}} \int_K X_t^2 \, \mathbb{N}(dt) \stackrel{\mathcal{L}}{\Longrightarrow} \int_{\mathbb{R}^d} \int_B \psi^2(t+x) \mathbb{M}(dx) \mathbb{N}(dt). \tag{5.86}$$

5.5 Subsampling Applications

We now investigate subsampling when the observation points are given by a Poisson Point Process, as in Section 3. The discussion here follows Chapter 6 of [48] quite closely, and we will apply several of their theorems. First, we define the data to be of the form $Y_t = X_t + \theta$, so that the data has mean θ (when $\alpha > 1$). See section 3.2 for a similar setup. Next, let $\overline{\alpha}_X(k; l_1)$ be the mixing coefficients defined on page 141 of that book. Also make the following notation:

$$\hat{\theta}_K = \frac{1}{\mathbb{N}(K)} \int_K Y(t)\mathbb{N}(dt) \tag{5.87}$$

which estimates θ. By (5.71), we have the following asymptotic result:

$$\mathbb{N}(K)^{1-\frac{1}{\alpha}} \left(\hat{\theta}_K - \theta\right) \stackrel{\mathcal{L}}{\Longrightarrow} S \tag{5.88}$$

where S is an α-stable law. In fact, since $\mathbb{N}(K) \sim |K| = N$ as $\min_i n_i \to \infty$ almost surely, we may replace a_K by $a_{\mathbb{N}(K)}$ in the above statement. So we can define

$$\tau_u := u^{1-\frac{1}{\alpha}} = u^{\zeta} \tag{5.89}$$

and find that condition (6.6) of [48] is satisfied with $\zeta = 1 - \frac{1}{\alpha}$; also $\alpha \in (1, 2)$ implies that $\zeta \in (0, \frac{1}{2})$.

Thus it follows that

$$\tau_{\Lambda(K)} = \Lambda(K)^{1-\frac{1}{\alpha}} = r^{1-\frac{1}{\alpha}} N^{1-\frac{1}{\alpha}} \tag{5.90}$$

since Λ is $r > 0$ times Lebesgue measure. Therefore

$$\tau_{\Lambda(K)}\left(\hat{\theta}_K - \theta\right) = o_P(1) + \mathbb{N}(K)^{1-\frac{1}{\alpha}}\left(\hat{\theta}_K - \theta\right) \stackrel{\mathcal{L}}{\Longrightarrow} \cdot S. \tag{5.91}$$

Denote the diameter of K by $\delta(K)$; then $\delta(K) = \max_i n_i$, and it follows that $\delta(K) \to \infty$ as $\min_i n_i \to \infty$. Therefore Assumption 6.3.2 of [48] is satisfied. Finally, choose any sequence b_N such that $\frac{b_N}{N} \to 0$ as $N \to \infty$ but $b_N \to \infty$; choose $c := \frac{1}{b_N}$, so that $c = c(K) \in (0,1)$ tends to zero and

$$c \cdot \delta(K) = \frac{N}{b_N} \to \infty \tag{5.92}$$

as $N \to \infty$. We may now state a corollary that is essentially Theorem 6.3.1 of [48] applied to our context. Let $K_{1-c} := \{y \in K : B+y \subset K\}$ for $B := cK$, and $\hat{\theta}_{K,B,y}$ is the statistic $\hat{\theta}$ evaluated on the set $B+y$ for any $y \in K_{1-c}$; in this case it is

$$\hat{\theta}_{K,B,y} := \frac{1}{\mathbb{N}(B+y)} \int_{B+y} Y(t)\mathbb{N}(dt) \tag{5.93}$$

where we have restricted $\mathbb{N}$ to the region $B+y$. Finally, we need a condition on the mixing coefficients, just like (4.93) :

$$|K_{1-c}|^{-1} \int_0^{(1-c)\Delta(K)} u^{d-1}\overline{\alpha}_X \left(u; c^d|K|\right) \, du \to 0 \tag{5.94}$$

as $\min_i n_i \to \infty$ ($\Delta(K)$ is defined to be the infimum of the diameters of all l^∞ balls which contain K). Note that this condition is easily satisfied in the $d=1$ special case if the random field is strong mixing.

Corollary 5.5.1. *Make the above mixing assumption (5.94) , and let $J(x)$ be the cdf of the limit random variable S , and define the subsampling estimator to be*

$$L_{K,B}(x) = |K_{1-c}|^{-1} \int_{K_{1-c}} 1_{\{\tau_{\mathbb{N}(cK+y)}(\hat{\theta}_{K,B,y} - \hat{\theta}_K) \leqslant x\}} \, dy \tag{5.95}$$

Then $L_{K,B}(x) \xrightarrow{P} J(x)$ for every continuity point x of $J(x)$.

Proof This follows immediately from Theorem 6.3.1 of [48] , as all the hypotheses are satisfied.

This is a first result of some theoretical interest, but because the rate of convergence $\tau_{\Lambda(K)}$ depends on α, which are generally unknown in practice, this corollary has limited application. We wish to have a similar result for the self-normalized asymptotic result in Theorem 5.4.2, since the rate of convergence there does not depend on α. So we first present a theorem along the lines of Theorem 6.3.1 in [48] , which is valid for self-normalized statistics.

Consider now the statistic $\hat{\sigma}$ which is an estimate of scale (in our context, it will be the square root of the periodogram), and suppose we evaluate it on the observation region K. Then we may form the ratio

$$\tau_{\Lambda(K)} \frac{\left(\hat{\theta}_K - \theta\right)}{\hat{\sigma}_K} \tag{5.96}$$

where $\tau_u = u$ has been chosen as the appropriate rate of convergence, such that the ratio has a nontrivial weak limit. Now we adjust the definition of the subsampling distribution estimator accordingly:

$$L_{K,B}(x) := |K_{1-c}|^{-1} \int_{K_{1-c}} 1_{\left\{\tau_{\mathbb{N}(B+y)}\left(\frac{\hat{\theta}_{K,B,y} - \hat{\theta}_K}{\hat{\sigma}_{K,B,y}}\right) \leqslant x\right\}} dy, \tag{5.97}$$

with $\hat{\sigma}_{K,B,y}$ defined similarly to $\hat{\theta}_{K,B,y}$.

Theorem 5.5.1. *Assume that the ratio*

$$\tau_{\Lambda(K)} \frac{\hat{\theta}_K - \theta}{\hat{\sigma}_K} \tag{5.98}$$

converges weakly to a nondegenerate random variable. Let θ be real-valued and assume that τ_u has the form given by equation (6.6) . Let $\delta(K) \to \infty$, and let $c = c(K) \in (0,1)$ be such that $c \to 0$ but $c\delta(K) \to \infty$. Finally, assume the mixing condition (5.94) . Then the following conclusions hold:

i. $L_{K,B}(x) \xrightarrow{P} J(x)$ for every continuity point x of $J(x)$.

ii. If $J(\cdot)$ is continuous, then $\sup_x |L_{K,B}(x) - J(x)| \xrightarrow{P} 0$.

iii. Let

$$c_{K,B}(1-t) = \inf\{x : L_{K,B}(x) \geqslant 1-t\}. \tag{5.99}$$

If $J(x)$ is continuous at $x = \inf\{x : J(x) \geqslant 1-t\}$, then

$$\mathbb{P}\{\tau_{\Lambda(K)}\left(\hat{\theta}_K - \theta\right) \leqslant c_{K,B}(1-t)\} \to 1-t. \tag{5.100}$$

Thus the asymptotic coverage probability of the interval $[\hat{\theta}_K - \tau_{\mathbb{N}(K)}{}^{-1} c_{K,B}(1-t), \infty)$ is the nominal level $1-t$.

Proof This proof follows the same structure as Theorem 6.3.1 of [48] . Using the same arguments, we quickly see that $\tau_{\Lambda(B)}/\tau_{\Lambda(K)} \to 0$, and $\tau_{\Lambda(B)}/\tau_{\mathbb{N}(B+y)} \xrightarrow{a.s.} 1$. Let x be a continuity point of $J(x)$. We use an argument similar to the one used in the proof of Theorem 11.3.1, together with the fact that $\tau_{\mathbb{N}(A)}/\tau_{\Lambda(A)} \xrightarrow{a.s.} 1$, to show that we can replace

$$\{\tau_{\mathbb{N}(B+y)}\left(\frac{\hat{\theta}_{K,B,y}-\hat{\theta}_K}{\hat{\sigma}_{K,B,y}}\right) \leqslant x\} \tag{5.101}$$

by

$$\{\tau_{\Lambda(B+y)}\left(\frac{\hat{\theta}_{K,B,y}-\theta}{\hat{\sigma}_{K,B,y}}\right) \leqslant x\} \tag{5.102}$$

up to a penalty that is $o_P(1)$ (note that the only difference here is that $\hat{\theta}_K$ in (5.101) is replaced by θ in (5.102)). Therefore, it is sufficient to show that $U_K(x) \xrightarrow{P} J(x)$, where

$$U_K(x) := |K_{1-c}|^{-1}\int_{K_{1-c}} 1_{\{\tau_{\Lambda(B+y)}\left(\frac{\hat{\theta}_{K,B,y}-\theta}{\hat{\sigma}_{K,B,y}}\right)\leqslant x\}}\, dy. \tag{5.103}$$

From here onwards, the remainder of the proof is identical to the proof of Theorem 6.3.1 of [48]. †

We will immediately apply this to (5.96), with the periodogram as the estimator of scale:

$$\hat{\sigma}_K := \left|\int_K Y(t)e^{it'\omega}\,\mathbb{N}(dt)\right|. \tag{5.104}$$

Here ω is a vector of frequencies ω_i each of which is in $(0, 2\pi)$ (so they are not an integer multiple of 2π). Now the convergence result in Theorem 5.4.2 can be written as a self-normalized result:

$$\mathbb{N}(K)\frac{\hat{\theta}_K-\theta}{\hat{\sigma}_K} \xrightarrow{\mathcal{L}} \tag{5.105}$$

$$\frac{\int_{\mathbb{R}^d}\left(\int_B \psi(t+x)\,\mathbb{M}(dx)\right)\mathbb{N}(dt)}{\sqrt{\left(\int_{\mathbb{R}^d}\left(\int_B \psi(t+x)cost'\omega\,\mathbb{M}(dx)\right)\mathbb{N}(dt)\right)^2+\left(\int_{\mathbb{R}^d}\left(\int_B \psi(t+x)sint'\omega\,\mathbb{M}(dx)\right)\mathbb{N}(dt)\right)^2}}.$$

This uses the fact that

$$\hat{\sigma}_K = \left| \int_K Y(t) e^{it'\omega}\, \mathbb{N}(dt) \right| = o_P(N^{\frac{1}{\alpha}}) + \left| \int_K X(t) e^{it'\omega}\, \mathbb{N}(dt) \right| \tag{5.106}$$

so long as the components of ω are not integer multiples of 2π – see (3.44). Finally, notice that in (5.105) we have $\tau_u = u$. Using these facts, we can establish the following corollary:

Corollary 5.5.2. *Let $J(x)$ denote the cdf of the limit random variable in (5.105) , and make the mixing assumption (5.94) . Then, with the subsampling distribution estimator as defined above, we have $L_{K,B}(x) \xrightarrow{P} J(x)$ for every continuity point x of $J(x)$.*

Proof We have the following simple calculation:

$$\begin{aligned}
&\tau_{\Lambda(K)} \frac{\left(\hat{\theta}_K - \theta\right)}{\hat{\sigma}_K} \qquad (5.107)\\
&= \Lambda(K) \frac{\hat{\theta}_K - \theta}{\hat{\sigma}_K} \\
&= \Lambda(K) \frac{\mathbb{N}(K)^{-1} \int_K (Y(t) - \theta)\, \mathbb{N}(dt)}{\hat{\sigma}_K} \\
&= \frac{\Lambda(K)}{\mathbb{N}(K)} \cdot \frac{N^{-\frac{1}{\alpha}} \int_K X(t) \mathbb{N}(dt)}{N^{-\frac{1}{\alpha}} \left| \int_K Y(t) e^{it'\omega}\, \mathbb{N}(dt) \right|} \xrightarrow{\mathcal{L}}
\end{aligned}$$

$$\frac{\int_{\mathbb{R}^d} \left(\int_B \psi(t+x)\, \mathbb{M}(dx) \right) \mathbb{N}(dt)}{\sqrt{\left(\int_{\mathbb{R}^d} \left(\int_B \psi(t+x) cos t'\omega\, \mathbb{M}(dx) \right) \mathbb{N}(dt) \right)^2 + \left(\int_{\mathbb{R}^d} \left(\int_B \psi(t+x) sin t'\omega\, \mathbb{M}(dx) \right) \mathbb{N}(dt) \right)^2}}.$$

This establishes the convergence result neccessary for Theorem 6.3.1 of [48] , and all the other hypotheses are satisfied, just as in Corollary 5.5.1. This finishes the proof. †

Remark We may now apply this result to develop asymptotically correct confidence intervals for the mean (see the previous subsection). Since the subsampling distribution estimator involves Riemann integration, some numerical approximation must be made in its calculation – see section 6.4 of [48] for further details.

Chapter 6

Conclusion

We have studied several time series models that exhibit long-range dependence and heavy-tailed marginal distributions – in particular, infinite order moving averages, stable integral moving averages, and marked point processes – and have repeatedly used self-normalization together with subsampling to handle some otherwise intractable problems. Data possessing the Noah and Joseph Effects seem to be extremely prevalent in nature, and thus there is a great need for effective methods for handling the corresponding models. We hope the reader is convinced of the efficacy of self-normalization and subsampling toward this end, even in the more difficult context of random fields and marked point processes. Future investigations should explore heavy-tailed models with more complicated dependence structures.

Following are four appendices. The first gives background material on stable random variables, and thus is relevant to the whole work. The second appendix sets up the notation for Point Process techniques, which are used in the third section of chapter two. The third appendix treats α-stable processes and random measures. The last appendix describes a central limit theorem result for triangular arrays of non-stationary random fields.

Appendix A

Stable Variables

Herein I will present several equivalent definitions of α-stable random variables. They are real-valued random variables, and I will consider only nondegenerate ones. I follow the notation of *Non-Gaussian Stable Processes* [64] .

A.1 Basic Definitions

Definition A.1.1. *A random variable X is said to be* **stable** *iff for all constants A, $B > 0$, $\exists C > 0$ and D such that*

$$AX_1 + BX_2 =^d CX + D \tag{A.1}$$

where X_1 and X_2 are independent copies of X.

We say X is **strictly stable** iff $D = 0$. It is a fact that $C^\alpha = A^\alpha + B^\alpha$ for some parameter $\alpha \in (0, 2]$ (see Feller for a proof).

Definition A.1.2. *A random variable X is* **stable** *iff $\forall n \geqslant 2, \exists C_n > 0$ and D_n such that*

$$X_1 + X_2 + \cdots + X_n =^d C_n X + D_n \tag{A.2}$$

where $X_1, X_2, \cdots, X_n$ are independent copies of X. Note: we can take $C_n = n^{\frac{1}{\alpha}}$.

Definition A.1.3. *A random variable X is* **stable** *iff it has a "domain of attraction," i.e. $\exists Y_1, Y_2, \cdots$ and positive numbers d_n and real numbers a_n such that*

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{d_n} + a_n \Rightarrow X. \tag{A.3}$$

We can take $d_n = n^{\frac{1}{\alpha}} L(n)$ where L is a slowly varying function, i.e.

$$\frac{L(ux)}{L(x)} \longrightarrow 1 \tag{A.4}$$

as x tends to ∞. We say that X has a "normal domain of attraction" iff $d_n = n^{\frac{1}{\alpha}}$, i.e. we can do without the slowly varying function. We say that the Y_i's above belong to the domain of attraction of X, which is written $Y_i \in DOM(\alpha)$.

Definition A.1.4. *A random variable X is* **stable** *iff $\exists \alpha \in (0, 2]$, $\sigma \geqslant 0$, $\beta \in [-1, 1]$ and real μ such that*

$$\mathbb{E}e^{i\theta X} = \begin{cases} \exp\{-\sigma^\alpha |\theta|^\alpha (1 - i\beta(sign\theta) \tan \frac{\pi\alpha}{2}) + i\mu\theta\} & \alpha \neq 1 \\ \exp\{-\sigma^\alpha |\theta|^\alpha (1 + i\beta \frac{2}{\pi}(sign\theta) \log|\theta|) + i\mu\theta\} & \alpha = 1 \end{cases} \tag{A.5}$$

The parameters σ, β, and μ are unique.

Note a few facts about these parameters: $\alpha = 2$ corresponds to the normal, and in this case the β parameter is irrelevant. $sigma$ is the scale parameter, β is skewness, and μ is location, which is the mean if $\alpha > 1$. Thus we write

$$X \sim S_\alpha(\sigma, \beta, \mu). \tag{A.6}$$

We say X is symmetric α-stable iff $\beta = 0$ (shorthand: X is sαs). A 1-stable random variable is a Cauchy.

The tails of these random variables possess many interesting properties:

Lemma A.1.1. *Suppose $X \sim S_\alpha(\sigma, \beta, \mu)$ for $0 < \alpha < 2$. Then we have*

$$x^\alpha \mathbb{P}[X > x] \sim C_\alpha \frac{1+\beta}{2} \sigma^\alpha \tag{A.7}$$

$$x^\alpha \mathbb{P}[X < -x] \sim C_\alpha \frac{1-\beta}{2} \sigma^\alpha \tag{A.8}$$

where $C_\alpha := \left(\int_0^\infty x^{-\alpha}\sin x\,dx\right)^{-1}$, *and* $\sim$ *means that the ratio tends to unity as* x *tends to* ∞. *We will call* $C_\alpha\sigma^\alpha$ *the "dispersion" of* X.

Remark This result also holds for random variables in $DOM(\alpha)$. It is here that we see the "heavy-tails" rigorously presented; the decay of the left and right tails is roughly like $x^{-\alpha}$. This justifies Mandelbrot's designation as "hyperbolic" random variables. Note that, of course, the $\alpha = 2$ case is excluded, since Gaussian tails decay at an exponential rate.

A.2 Product Decomposition of α-stable random variables

First we consider the case of a symmetric α-stable random variable, i.e.

$$X \sim S_\alpha(\sigma, 0, 0)$$

and has characteristic function

$$\mathbb{E}\left[e^{itX}\right] = \exp\left(-\sigma^\alpha |t|^\alpha\right), \tag{A.9}$$

and we will denote this by saying X is $s\alpha s$. It has scale parameter σ; we say X is "standard" if $\sigma = 1$.

Proposition A.2.1. *Let* X *be* $s\alpha' s$ *where* $0 < \alpha' \leqslant 2$ *and* $0 < \alpha < \alpha'$. *Also let* A *be an* $\frac{\alpha}{\alpha'}$*- stable random variable totally skewed to the right with Laplace transform*

$$\mathbb{E}\exp\left(-\gamma A\right) = \exp\left(-\gamma^{\frac{\alpha}{\alpha'}}\right) \tag{A.10}$$

for $\gamma > 0$. *In other words,*

$$A \sim S_{\frac{\alpha}{\alpha'}}\left(\left(\cos\frac{\pi\alpha}{2\alpha'}\right)^{\frac{\alpha'}{\alpha}}, 1, 0\right). \tag{A.11}$$

Assume that X *and* A *are independent; then*

$$Z := A^{\frac{1}{\alpha'}} X \sim S_\alpha(\sigma, 0, 0). \tag{A.12}$$

Remark 1 This is easily proved by looking at characteristic functions and Laplace transforms.

Remark 2 If we let $\alpha' = 2$ in the above, we obtain the representation

$$Z = A^{\frac{1}{2}} X \sim s\alpha s \tag{A.13}$$

where X is Gaussian.

Remark 3 If X is skewed, e.g.

$$X \sim S_{\alpha'} \left(\sigma', \beta', 0\right) \tag{A.14}$$

and $\alpha' \neq 1$, then we obtain

$$Z = A^{\frac{1}{\alpha'}} X \sim S_{\alpha} \left(\sigma, \beta, 0\right) \tag{A.15}$$

if $\alpha \neq 1$, and if $\alpha = 1$ we find

$$Z = A^{\frac{1}{\alpha'}} X \sim S_{1} \left(\sigma, 0, \mu\right). \tag{A.16}$$

A.3 Series Representation for α-stable random variables

Let N_t be the number of "customer arrivals" in $[0, t]$. The process $\{N_t, t \geqslant 0\}$ is a Poisson process with rate λ. ($PP(\lambda)$) iff the interarrival times are exponentially distributed:

$$\tau_{i+1} - \tau_{t_i} \sim iid \exp(\lambda) \tag{A.17}$$

and $\mathbb{E}N_t = \lambda t$.

There are two key properties of this process to enunciate:

(1) Thinning: We can eliminate an arrival with probability $p \in [0, 1)$, independently of one another, to get a $PP(\lambda(1-p))$.

(2) Adding Poissons: Let N^1 and N^2 be independent Poissons with rates λ_1 and λ_2; then $N^1 + N^2$ is $PP(\lambda_1 + \lambda_2)$.

Proposition A.3.1. *Let $\{\tau_i\}$ be the arrival times of a $PP(1)$; let $\{R_i\}$ be iid random variables independent of the $\{\tau_i\}$ sequence. If*

$$\sum_{i=1}^{\infty} \tau_i^{-\frac{1}{\alpha}} R_i < \infty \tag{A.18}$$

almost surely, then the sum converges to a strictly α-stable random variable.

Remark The idea of the proof is to verify A.1.1. If we let X denote the convergent sum, and take two independent copies X^1 and X^2, and any constants A and B such that $A^\alpha + B^\alpha = 1$, then

$$AX^1 + BX^2 = \sum_{i=1}^{\infty} \left(A^{-\alpha}\tau_i^1\right)^{-\frac{1}{\alpha}} R_i^1 + \sum_{i=1}^{\infty} \left(B^{-\alpha}\tau_i^2\right)^{-\frac{1}{\alpha}} R_i^2. \tag{A.19}$$

Now $A^{-\alpha}\tau_i^1$ and $B^{-\alpha}\tau_i^2$ are the arrival times of Poisson processes of rates A^α and B^α respectively. Then superimpose these two independent Poisson processes, and the sum will be X.

We wish to generalize this a little more. Consider the following setup:

$\{\varepsilon_i\} \sim$ *iid* Rad $(\frac{1}{2})$

$\{W_i\} \sim$ *iid* elements of $\mathbb{L}_\alpha$

$\{\Gamma_i\}$ arrival times of a $PP(1)$,

where all the above sequences are independent. *Rad* refers to a Rademacher random variable which takes on values 1 and -1 with probability $\frac{1}{2}$ each. Then the following theorem holds:

Theorem A.3.1. *Suppose $0 < \alpha < 2$. Then*

$$\sum_{i=1}^{\infty} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i \tag{A.20}$$

converges almost surely to

$$X \sim S_\alpha\left((C_\alpha^{-1}\mathbb{E}|W_1|^\alpha)^{\frac{1}{\alpha}}, 0, 0\right) \tag{A.21}$$

where $C_\alpha \sigma^\alpha$ is the "dispersion.'

Corollary A.3.1. *Any sαs random variable for $0 < \alpha < 2$, i.e. it has law $S_\alpha(\sigma, 0, 0)$, has the series representation, i.e. it is equal in law to,*

$$\sigma \left(\frac{C_\alpha}{\mathbb{E}|W_1|^\alpha} \right)^{\frac{1}{\alpha}} \sum_{i=1}^{\infty} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i \tag{A.22}$$

Remark 1 This corollary is immediate from the theorem, and the proof of the theorem can be found in [64] . Nevertheless, the following heuristic gives us some intuition into why it is true: the waiting times are strictly increasing, so

$$\Gamma_1^{-\frac{1}{\alpha}} \geqslant \Gamma_2^{-\frac{1}{\alpha}} \geqslant \cdots \tag{A.23}$$

Thus the first term of the sum is the dominating term; let's investigate its tail behavior:

$$\begin{aligned} \mathbb{P}\left[|\varepsilon_1 \Gamma_1^{-\frac{1}{\alpha}} W_1| > x \right] &= \int_0^\infty \mathbb{P}\left[\Gamma_1 < w^\alpha x^{-\alpha} \right] F_{|W_1|}(dw) \\ &= \int_0^\infty \left(1 - \exp(-w^\alpha x^{-\alpha}) \right) F_{|W_1|}(dw) \sim \mathbb{E}|W_1|^\alpha x^{-\alpha} \end{aligned} \tag{A.24}$$

as $x \longrightarrow \infty$, by a Taylor series expansion. But this is the tail behavior of an α-stable random variable. The other terms give the neccessary corrections to make the sum stable.

Remark 2 A similar representation can be found in the nonsymmetric case, i.e. when the skewness β of X is nonzero.

Appendix B

Point Process Techniques

This is a brief introduction to point process techniques – see [51] for additional details.

Let E be the space $(0, \infty) \times \mathbb{R} \setminus \{0\}$ with the σ-field $\mathcal{E}$ of open sets, and define

$$\varepsilon_x(F) := 1_F(x) = \begin{cases} 1 & x \in F \\ 0 & x \notin F \end{cases} \tag{B.1}$$

for $x \in E$ and $F \in \mathcal{E}$.

A **point measure** m has the form $\sum_{i \in I} \varepsilon_{x_i}$, and is defined on all relatively compact subsets of E. The class of such measures is denoted by $M_p(E)$, and $\mathcal{M}_p(E)$ is the smallest σ-field such that the mapping for $F \in \mathcal{E}$

$$m \mapsto m(F), \tag{B.2}$$

i.e. the evaluation maps, are measurable.

A **Point Process** on E is a measurable map from a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into the measure space $(M_p(E), \mathcal{M}_p(E))$. Thus a point process is a random point measure.

Let $C_K^+(E)$ be the set of continuous functions from E to $\mathbb{R}^+$ that have compact support. A useful topology for $M_p(E)$ is the "vague" topology, under which $M_p(E)$ is a complete separable metric space. If $\mu_n \in M_p(E)$, then $\mu_n \longrightarrow \mu_0$ "vaguely" iff $\mu_n(f) \longrightarrow \mu_0(f) \quad \forall f \in C_K^+(E)$, in which case we write $\mu_n \xrightarrow{v} \mu_0$.

A Poisson process on $(E, \mathcal{E})$ with mean measure μ is a point process ξ satisfying for all $F \in \mathcal{E}$:

$$\mathbb{P}\left[\xi(F) = k\right] = \begin{cases} exp\{-\mu(A)\}\frac{(\mu(A))^k}{k!} & if \mu(A) < \infty \\ 0 \quad if \mu(A) = \infty \end{cases} \tag{B.3}$$

which has the independent scattering property, i.e. if $F_1, \cdots, F_n \in \mathcal{E}$ are all disjoint, then $\xi(F_1), \cdots, \xi(F_n)$ are independent random variables. We say that ξ is a Poisson Random Measure with mean measure μ, which is abbreviated $PRM(\mu)$.

Now suppose we have an *iid* sequence $\{Z_t\}$ which satisfies the tail behavior described by conditions (2.2), (2.3), (2.8). By defining the Levy measure ν on $\mathbb{R} \setminus \{0\}$ as follows

$$\nu(dx) := \alpha p x^{-\alpha-1} 1_{(0,\infty)}(x)dx + \alpha q(-x)^{-\alpha-1} 1_{(-\infty,0)}(x)dx, \tag{B.4}$$

we can formulate condition 2.8 of the previous section as

$$n\mathbb{P}\left[a_n^{-1} Z_1 \in \cdot\right] = nF(a_n \cdot) \xrightarrow{v} \nu. \tag{B.5}$$

We define the measure $\mu(dt, dx) := dt \times \nu(dx)$ on $(0, \infty) \times (\mathbb{R} \setminus \{0\})$. Then the following convergence holds true:

$$\sum_{k=1}^{\infty} \varepsilon_{\left(\frac{k}{n}, \frac{Z_k}{a_n}\right)} \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{k=1}^{\infty} \varepsilon_{(t_k, j_k)} \tag{B.6}$$

on $M_p\left((0,\infty) \times (\mathbb{R} \setminus \{0\})\right)$, where the limit is a $PRM(\mu)$. We use here the convention that a point that falls outside the state space does not contribute to the sum.

Now we cite a result from [16] for linear time series $\{X_t\}$ which satisfy our assumptions (2.2), (2.3), (2.8):

$$\sum_{k=1}^{\infty} \varepsilon_{\left(\frac{k}{n}, \frac{X_k}{a_n}\right)} \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{k=1}^{\infty} \sum_{i \in \mathbb{Z}} \varepsilon_{(t_k, \psi_i j_k)}, \tag{B.7}$$

where $\{(t_k, j_k)\}$ are, as above, the points of a $PRM(\mu)$ on the space $(0, \infty) \times (\mathbb{R} \setminus \{0\})$.

Appendix C

Stable Integrals

In this section several topics on α-stable random measures and integrals are presented, starting with basic definitions and moving to some representation results. This gives the construction of α-stable processes, e.g., a moving average of sαs r.v.'s. All this material is drawn from Chapters 1, 3 , and 7 of the book *Non-Gaussian Stable Processes* [64] .

C.1 Stable Stochastic Processes

A stochastic process $\{X_t\}_{t\in T}$ is **stable** iff all its finite dimensional distributions are stable. We define **strictly stable** and **symmetric stable** processes similarly.

Theorem C.1.1. *(a)* $\{X_t\}$ *is strictly stable iff all linear combinations*

$$\sum_{k=1}^{d} b_k X_{t_k} \quad b_1, b_2, \cdots, b_d \in \mathbb{R} \tag{C.1}$$

are strictly stable.

(b) Similarly, X *is symmetric stable iff all its linear combinations are*

(c) If $\alpha \geqslant 1$, *then* X *is* α*-stable iff all its linear combinations are.*

Example : α-stable standard Levy motion Consider the process with the following properties:

(1) $X(0) = 0$ almost surely

(2) X has independent increments

(3) $X_t - X_s \sim S_\alpha((t-s)^{\frac{1}{\alpha}}, \beta, 0) \quad 0 < \alpha \leqslant 2$.

Notice that when $\alpha = 2$ this is just Brownian Motion (here $\beta = 0$). When we stipulate that $\beta = 0$, we obtain **symmetric α stable Levy Motion** or sαs Levy Motion. It has the property of being $\frac{1}{\alpha}$ self-similar, i.e. $\{X_{ct}\}$ and $\{c^{\frac{1}{\alpha}}\{X_t\}$ have the same finite dimensional distributions.

C.2 Existence of Stable Integrals

We wish to define a stochastic process $\{I(f) : f \in F\}$ as integrals by specifying the finite dimensional distributions. Let $(E, \mathcal{E}, m)$ be a measure space and let

$$\beta : E \longrightarrow [-1, 1] \tag{C.2}$$

be a measurable mapping. Choose the space F to be

$$F = \begin{cases} \mathbb{L}_\alpha(E, \mathcal{E}, m) & \alpha \neq 1 \\ \mathcal{F}(m, \beta) & \alpha = 1 \end{cases} \tag{C.3}$$

where

$$\mathcal{F}(m, \beta) := \{f : f \in \mathbb{L}_1(E, \mathcal{E}, m) \quad and \quad \int_E |f(x)\beta(x) \log|f(x)||m(dx) < \infty\}. \tag{C.4}$$

Assume without loss of generality that m is σ-finite.

Given $f_1, \cdots, f_d \in F$, define $\mathbb{P}_{f_1,\cdots,f_d}$ in $\mathcal{R}^d$ by characteristic functions:

$$\phi_{f_1,\cdots,f_d}(\theta_1, \cdots, \theta_d) \tag{C.5}$$
$$= \exp\left\{ -\int_E |\sum_{j=1}^d \theta_j f_j(x)|^\alpha \left(1 - i\beta(x) sign\left(\sum_{j=1}^d \theta_j f_j(x)\right) \tan\frac{\pi\alpha}{2}\right) m(dx)\right\}$$

if $\alpha \neq 1$ and

$$\phi_{f_1,\cdots,f_d}(\theta_1, \cdots, \theta_d) \tag{C.6}$$
$$= \exp\left\{ -\int_E |\sum_{j=1}^d \theta_j f_j(x)| \left(1 + i\frac{2}{\pi}\beta(x) sign\left(\sum_{j=1}^d \theta_j f_j(x)\right) \log|\sum_{j=1}^d \theta_j f_j(x)|\right) m(dx)\right\}$$

if $\alpha = 1$.

These measures have the characteristic functions of jointly α-stable random variables on $\mathcal{R}^d$. They are consistent and hence, via Kolmogorov's Extension Theorem (see [21]), there exists a stochastic process $\{I(f) : f \in F\}$ whose finite-dimensional distributions have the characteristic functions given above. Then $I(f)$ is the **α-stable integral of f**; m is said to be the **control measure** and β is the **skewness intensity.**

We see the following interesting properties from this representation:
(1) For any $f_1, f_2, \cdots f_d \in F$, we have $I(f_1), I(f_2), \cdots, I(f_d)$ are jointly α-stable with the joint characteristic function given above; moreover they are jointly sαs iff $\beta = 0$.
(2) If $f \in F$, then $I(f) \sim S_\alpha(\sigma_f, \beta_f, \mu_f)$ where

$$\sigma_f := \left(\int_E |f(x)|^\alpha \, m(dx) \right)^{\frac{1}{\alpha}} \tag{C.7}$$

$$\beta_f := \frac{\int_E f(x)^{<\alpha>} \, \beta(x) \, m(dx)}{\int_E |f(x)|^\alpha \, m(dx)}$$

$$\mu_f = \begin{cases} 0 & \alpha \neq 1 \\ -\frac{2}{\pi} \int_E f(x) \, \beta(x) \, \log |f(x)| \, m(dx) & \alpha = 1 \end{cases}$$

We have used the notation $g^{<\alpha>} = |g|^\alpha sign(g)$.
(3) I is a linear functional.

C.3 α-stable Random Measures

Assume an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{L}_0(\Omega)$ denote the space of all random variables on it. Let $(E, \mathcal{E}, m)$ be a measure space, and $\beta : E \longrightarrow [-1, 1]$ is a measurable function. Let $\mathcal{E}_0 = \{A \in \mathcal{E} : m(A) < \infty\}$ be the sets of finite m-measure.

Now we define the desired measure: let $\mathbb{M}$ be an independently scattered σ-additive set function

$$\mathbb{M} : \mathcal{E}_0 \longrightarrow \mathbb{L}_0(\Omega) \tag{C.8}$$

such that for every $A \in \mathcal{E}_0$, we have

$$\mathbb{M}(A) \sim S_\alpha\left((m(A))^{\frac{1}{\alpha}}, \frac{\int_A \beta(x)\, m(dx)}{m(A)}, 0\right). \tag{C.9}$$

Independently scattered means that disjoint sets produce independent random variables, and σ-additive refers to its properties as a measure. Then $\mathbb{M}$ is called an **α-stable random measure** on $(E, \mathcal{E})$ with **control measure** m and **skewness intensity** β .

The first question one might ask is: does it exist? Define the stochastic process $\{\mathbb{M}(A), A \in \mathcal{E}_0\}$ via

$$\mathcal{M}(A) := I(1_A) \tag{C.10}$$

for $A \in \mathcal{E}_0$, where I denotes the stable integral defined in the previous section. Thus we just use indicator functions as a nice simple class of integrands in F to define a stochastic process (one way of viewing a random measure is as a stochastic process indexed by sets). To check that these functions are in F, observe that

$$\int (1_A)^\alpha\, dm = m(A) < \infty \tag{C.11}$$

due to the fact that $A \in \mathcal{E}_0$. We say that $\mathbb{M}$ is a sαs random measure iff $\beta = 0$.

Example Take $\mathbb{M}$ to be an α-stable random measure on $([0,\infty), \mathcal{B}(\mathbb{R}^+))$ with Lebesgue control measure and constant skewness $\beta(x) = \beta$. Define $X_t := \mathbb{M}([0,t])$ for $0 \leqslant t < \infty$, which is well-defined since $m([0,t]) < \infty \quad \forall t \geqslant 0$. Then we see that

$$X_t - X_s = \mathbb{M}((s,t]) \sim S_\alpha(|t-s|^{\frac{1}{\alpha}}, \beta, 0). \tag{C.12}$$

This process also has independent increments, which follows from the independent scattering properties of $\mathbb{M}$. Thus we see that X is a standard α-stable Levy motion.

C.4 Construction of Stable Integrals

We now wish to define $I(f) = \int_E f(x)\mathbb{M}(dx)$ for appropriate integrands f (e.g. for $f \in F$). Let $\mathbb{M}$ be an α-stable random measure on $(E, \mathcal{E})$ with control measure m and skewness intensity β. Let $\mathcal{E}_0$ denote the measurable sets with finite mass.

If $f(x) = \sum_{j=1}^{n} c_j 1_{A_j}(x)$ with A_j disjoint in $\mathcal{E}_0$ and c_j constants, then define

$$I(f) := \int_E f(x)\mathbb{M}(dx) = \sum_{j=1}^{n} c_j \mathbb{M}(A_j). \tag{C.13}$$

The above integral is a sum of independent random variables, due to independent scattering. In fact, $I(f) \sim S_\alpha(\sigma_f, \beta_f, \mu_f)$. We extend this definition to more general $f \in F$ via approximating by the above simple integrands (analagous to the Ito integral construction), and by showing convergence in probability (see [64] for details).

Example : Moving Averages Take f such that $\int_{-\infty}^{\infty} |f(x)|^\alpha \, dx < \infty$ for $0 < \alpha \leqslant 2$, and define

$$X_t := \int_{-\infty}^{\infty} f(t-x)\mathbb{M}(dx) \quad t \in \mathbb{R} \tag{C.14}$$

where $\mathbb{M}$ is a sαs random measure with Lebesgue control measure on $\mathbb{R}$. We call this a **sαs MA process**; it is strictly stationary.

Example : Ornstein-Uhlenbeck Process Take $f(x) = e^{-\lambda x} 1_{[0,\infty)}(x)$ in the above MA process. Then

$$X_t = \int_{-\infty}^{t} e^{-\lambda(t-x)}\mathbb{M}(dx) \tag{C.15}$$

for any $t \in \mathbb{R}$. Note that by the following random variable

$$X_t - e^{-\lambda(t-s)}X_s = \int_s^t e^{-\lambda(t-x)}\mathbb{M}(dx) \tag{C.16}$$

is independent of $\sigma\{X_u, u \leqslant s\}$. Thus the Ornstein-Uhlenbeck process is a Markov process.

Example : Sub-Gaussian Process Now consider the decomposition given in section 3.2, and consider $X_t = A^{\frac{1}{2}} G_t$ for $t \in T$, where G_t is a Gaussian process and $A \sim S_{\frac{\alpha}{2}}((\cos\frac{\pi\alpha}{4})^{\frac{2}{\alpha}}, 1, 0)$ independent of the Gaussians. Such a process is called a **sub-Gaussian process.** Now let $\mathbb{M}$ be a sαs random measure on the underlying probability space Ω, with control measure $\mathbb{P}$ (the probability measure); then

$$\left\{X_t\} =^d \{\frac{1}{d_\alpha\sqrt{2}}\int_\Omega G(t,\omega)\mathbb{M}(d\omega)\right\} \tag{C.17}$$

for $d_\alpha = (\mathbb{E}|Z|^\alpha)^{\frac{1}{\alpha}}$ and $Z \sim \mathcal{N}(0,1)$.

Appendix D

Central Limit Theorem

D.1 Introduction and background

Since the pioneering paper of Rosenblatt ([60]) there have been many results on the Central Limit Theorem (CLT) in the context of random variables satisfying a strong mixing condition; see e.g. [18] and the references therein.

With respect to a Central Limit Theorem for a triangular array of weakly dependent random variables there has been work by [65] who considers triangular arrays of stationary ϕ-mixing random variables taking values in general state spaces but with the ϕ-mixing coefficients being the same for each row of the array. [67] studied triangular arrays of random variables with ϕ-mixing coefficients being (possibly) different for each row of the array. In addition, [45] has a Central Limit Theorem under $\rho^\star$-mixing in a strictly stationary context. Recently, [57] and [49] derived a Central Limit Theorem for random variables with strong mixing (as opposed to ϕ-mixing or $\rho^\star$-mixing) coefficients being (possibly) different for each row of the array; however, their methods are very different and the respective sufficient conditions for the Central Limit Theorem to hold are not immediately comparable.

[9] was perhaps the first to prove a Central Limit Theorem in the context of a stationary random fields that satisfy a condition with strong mixing coefficients that depend on the sizes of the sets considered; see line (D.1) below. However, no result so far addresses the case of a triangular array of random fields. In this paper, we fill this gap and provide sufficient conditions for the Central Limit Theorem in the context of a triangular array of possibly nonstationary random fields satisfying a strong mixing condition that may be different for each row of the array. To do this, we employ the ubiquitous big block—small block method of [4]. Also see [45] for a similar result under a different type of mixing.

D.2 Central Limit Theorem

Fix $m \in \mathbb{N}$, and consider the real-valued random variables $X_{n,\mathbf{i}}$ defined on a common underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$; here $n \in \mathbb{N}$ and $\mathbf{i} \in \mathbb{N}^m$. For simplicity, we assume that the random variables $X_{n,\mathbf{i}}$ have mean zero. The random variables $\{X_{n,\mathbf{i}}$ for $n \in \mathbb{N}$ and $\mathbf{i} \in \mathbb{N}^m\}$ constitute a triangular array of random fields with kth 'row' consisting of the random field $\{X_{k,\mathbf{i}}, \mathbf{i} \in \mathbb{N}^m\}$. We equip $\mathbb{N}^m$ with the l_1 (taxicab) metric denoted by ρ (though any other metric l_p for $1 \leqslant p \leqslant \infty$ would work equally well).

In terms of notation: $\mathbf{i}$ denotes a $\mathbb{N}^m$ vector, so $\mathbf{i} = (i^1, i^2, \cdots, i^m)$. Also $\mathbf{i}^\sharp$ will be the maximum of the components, $\mathbf{i}^\flat$ will be the minimum of the components, and $|\mathbf{i}|$ will be the product of the components (though in other situations $|\cdot|$ will denote cardinality without possibility of confusion). All vectors will be in bold, with their components written as superscripts.

Define the strong mixing coefficient corresponding to the nth row of the array by

$$\alpha_n(j; l, k) := \sup_{\Theta, \Lambda \subset \mathbb{N}^m} \left\{ |\mathbb{P}(A\, B) - \mathbb{P}(A)\mathbb{P}(B)| : |\Theta| \leqslant l, |\Lambda| \leqslant k, \rho(\Theta, \Lambda) \geqslant j \right\}, \qquad \text{(D.1)}$$

where $A \in \sigma\{X_{n,\mathbf{i}} : \mathbf{i} \in \Theta\}$, $B \in \sigma\{X_{n,\mathbf{i}} : \mathbf{i} \in \Lambda\}$, and $\sigma\{X_{n,\mathbf{i}} : \mathbf{i} \in \Theta\}$ denotes the σ–algebra generated by the random variables $\{X_{n,\mathbf{i}} : \mathbf{i} \in \Theta\}$.

For $\mathbf{d}_n \in \mathbb{N}^m$, let the set $D_n = \{\mathbf{i} : 1 \leq i^k \leq d_n^k \text{ for } k = 1, \ldots, m\}$. We will work under an assumption that requires the set D_n to expand in all m directions at an almost uniform fashion. More precisely, assume that $\mathbf{d}_n^\flat \longrightarrow \infty$ in such a way that $D_n \subset Ball(c(|\mathbf{d}_n|)^{\frac{1}{m}})$, where $Ball(r)$ is the Euclidean ball of radius r in $\mathbb{R}^m$, and c is some positive constant. We will write $\mathbf{d}_n \Rightarrow \infty$ to denote this condition.

Theorem D.2.1. *Let $\{X_{n,\mathbf{i}}, n \in \mathbb{N}, \mathbf{i} \in \mathbb{N}^m\}$ be an array of mean zero random variables. Denote the mixing sequence corresponding to the nth row of the array by $\alpha_n(\cdot\,; l, k)$, and define*

$$S_{n,\mathbf{k},\mathbf{a}} := \sum_{\mathbf{i}=\mathbf{a}}^{\mathbf{a}+\mathbf{k}-\mathbf{1}} X_{n,\mathbf{i}} = \sum_{i_1=a_1}^{a_1+k_1-1} \sum_{i_2=a_2}^{a_2+k_2-1} \cdots \sum_{i_m=a_m}^{a_m+k_m-1} X_{n,(i_1,i_2,\cdots,i_m)}, \tag{D.2}$$

where $\mathbf{1} = (1, 1, \ldots, 1)$, and

$$\sigma_{n,\mathbf{k},\mathbf{a}}^2 := Var(|\mathbf{k}|^{-\frac{1}{2}} S_{n,\mathbf{k},\mathbf{a}}). \tag{D.3}$$

Assume the following conditions hold, for some $\delta > 0$:

$$\|X_{n,\mathbf{i}}\|_{2+2\delta} \leqslant \Delta \; for\, all\, n \; and\, \mathbf{i} \tag{D.4}$$

$$\sigma_{n,\mathbf{k},\mathbf{a}}^2 \longrightarrow \sigma^2 > 0 \; uniformly^1 \; in\, \mathbf{a} \tag{D.5}$$

$$C_n^m(\tau) := \sum\nolimits_{k=0}^{|\mathbf{d}_n|} (k+1)^{m-1} \, \alpha_n(k; 1, 1)^\tau = o(|\mathbf{r}_n|^{\frac{\delta}{2+\delta}}) \tag{D.6}$$

$$\alpha_n(\mathbf{l}_n^\flat; |\mathbf{d}_n|, |\mathbf{b}_n|) = o(|\mathbf{l}_n|/|\mathbf{d}_n|) \tag{D.7}$$

$$\alpha_n(\mathbf{b}_n^j; |\mathbf{d}_n|/|\mathbf{l}_n|, |\mathbf{b}_n|/|\mathbf{l}_n|) = o\left(\frac{1}{\mathbf{r}_n^j}\right) \quad \forall j \tag{D.8}$$

In the above, $\tau := \frac{\delta}{c(1+\delta)}$, where c is the smallest even integer that is greater or equal to $2 + \delta$; Δ is a finite constant independent of n and $\mathbf{i}$.

Note that conditions (D.6) , (D.7) , and (D.8) should hold for some choice of vector sequences $\mathbf{l}_n$*,* $\mathbf{b}_n$*, and* $\mathbf{d}_n$ *in* $\mathbb{N}^m$ *such that* $\mathbf{l}_n \Rightarrow \infty$*,* $\mathbf{b}_n \Rightarrow \infty$*, and* $\mathbf{d}_n \Rightarrow \infty$*, and which satisfy*

$$\frac{\mathbf{l}_n^i}{\mathbf{b}_n^i} \longrightarrow 0, \quad \frac{\mathbf{b}_n^i}{\mathbf{d}_n^i} \longrightarrow 0, \tag{D.9}$$

for each $i = 1, \ldots, m$*.*

Then:

$$|\mathbf{d}_n|^{-\frac{1}{2}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{d}_n} X_{n,\mathbf{i}} \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \sigma^2\right) \tag{D.10}$$

as $\mathbf{d}_n \Rightarrow \infty$*, where* $\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{d}_n} X_{n,\mathbf{i}} \equiv S_{n,\mathbf{d}_n,\mathbf{1}}$*.*

Remark Note that if we have stationarity in each row of the array, condition (D.5) can be replaced by condition (D.11) given below:

$$\sigma^2_{n,\mathbf{k}_n,\mathbf{1}} \to \sigma^2 \, asn \to \infty \tag{D.11}$$

for any sequence of vectors $\{\mathbf{k}_n\}$ such that $\mathbf{k}_n^\flat \to \infty$.

The hypotheses (D.6) , (D.7) , and (D.8) are quite general; below a result is formulated under a specific choice of all the block sizes.

Corollary D.2.1. *Assume the same general model as in Theorem D.2.1, and suppose that the index set under consideration is an m-dimensional cube with side lengths* $\mathbf{d}_n^i = n \quad \forall i$*. Let the block sizes be*

$$\mathbf{l}_n^i = \lfloor n^{\frac{1}{4}} \rfloor \qquad \mathbf{b}_n^i = \lfloor n^{\frac{3}{4}} \rfloor. \tag{D.12}$$

Assume (D.4) and (D.5), and the following new hypotheses:

$$C_n^m(\tau) = O(1) \tag{D.13}$$

$$\alpha_n(n^{\frac{1}{4}}; n^{\frac{3m}{4}}, n^m) = o(n^{-\frac{3m}{4}}) \qquad . \tag{D.14}$$

Then

$$n^{-\frac{m}{2}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} X_{n,\mathbf{i}} \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \sigma^2\right) \tag{D.15}$$

as $n \longrightarrow \infty$, *where* $\mathbf{n} = (n, n, \cdots, n)$.

Note that the corollary follows from the theorem, since (D.13) is clearly stronger than (D.6) . Also, (D.14) implies (D.7) and (D.8) :

$$\alpha_n(\mathbf{l}_n^\flat; |\mathbf{d}_n|, |\mathbf{b}_n|)\frac{|\mathbf{d}_n|}{|\mathbf{l}_n|} \sim \alpha_n(n^{\frac{1}{4}}; n^{\frac{3m}{4}}, n^m)n^{\frac{3m}{4}} \to 0 \tag{D.16}$$

and also

$$\begin{aligned} \alpha_n(\mathbf{b}_n^j; |\mathbf{d}_n|/|\mathbf{l}_n|, |\mathbf{b}_n|/|\mathbf{l}_n|)\mathbf{r}_n^j &\sim \alpha_n(n^{\frac{3}{4}}; n^{\frac{3m}{4}}, n^{\frac{m}{2}})n^{\frac{1}{4}} \\ &\leqslant \alpha_n(n^{\frac{1}{4}}; n^m, n^{\frac{3m}{4}})n^{\frac{3m}{4}} \longrightarrow 0. \end{aligned} \tag{D.17}$$

This inequality is due to the fact that strong mixing coefficients are monotonically decreasing in their first argument, and monotonically increasing in the second two arguments (which is easily seen from the definition). It also uses the fact that the α_n is symmetric in the latter two arguments.

D.3 Proof of the Theorem

The proof strategy for Theorem D.2.1 consists of several steps. The overall technique is to use a multidimensional generalization of the Bernstein big block — small block technique, combined with the use of the Lyapunov Central Limit Theorem.

D.3.1 Step 0 : Notation and preliminaries

We "chop" each of the m dimensions into big and small blocks (segments) whose dimensions are given by the vectors $\mathbf{b}_n^i$ and $\mathbf{l}_n^i$ respectively chosen to satisfy (D.7). Let $\mathbf{r}_n^i$ be the largest integer k such that $(k-1)(\mathbf{b}_n^i + \mathbf{l}_n^i) + \mathbf{b}_n < \mathbf{d}_n^i$. It follows that $\mathbf{r}_n^i(\mathbf{l}_n^i + \mathbf{b}_n^i) \sim \mathbf{d}_n^i$ and $\mathbf{r}_n^i\,\mathbf{b}_n^i \sim \mathbf{d}_n^i$ as well.

Next, let

$$U_{n,\mathbf{i}} := \sum_{\mathbf{j}\in B_n+\mathbf{i}^\diamondsuit} X_{n,\mathbf{j}}, \tag{D.18}$$

where $1 \leqslant \mathbf{i}^k \leqslant \mathbf{r}_n^k$ for $1 \leqslant k \leqslant m$, and $B_n := \left\{\mathbf{a} : 1 \leqslant \mathbf{a}^k \leqslant \mathbf{b}_n^k,\ 1 \leqslant k \leqslant m\right\}$, and $\mathbf{i}^\diamondsuit$ is a vector whose kth component is $(\mathbf{i}^k - 1)(\mathbf{b}_n^k + \mathbf{l}_n^k)$. So $\mathbf{i}$ ranges as an index over all the big blocks, where B_n is the "first" or bottom left corner block. Let V_n be the remaining random variables (that accounts for the "mortar" between the big blocks). We can imagine that V_n consists of a large number of little blocks of kth side length $\mathbf{l_n}^k$. Thus

$$S_{n,\mathbf{d_n},\mathbf{1}} = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r_n}} U_{n,\mathbf{i}} + V_n. \tag{D.19}$$

Since $\frac{\delta}{1+\delta} \geqslant \tau$ and the mixing coefficients are less than 1,

$$C_n^m(\tau) = \sum_{k=0}^{|\mathbf{d}_n|} (k+1)^{m-1}\, \alpha_n(k;1,1)^\tau \geqslant \sum_{k=0}^{|\mathbf{d}_n|} (k+1)^{m-1} \alpha_n(k;1,1)^{\frac{\delta}{1+\delta}} \geqslant C_n^m\left(\frac{\delta}{1+\delta}\right). \tag{D.20}$$

So by (D.6), we get

$$C_n^m\left(\frac{\delta}{1+\delta}\right) \cdot \frac{1}{|\mathbf{d}_n|} \leqslant C_n^m(\tau) \cdot \frac{1}{|\mathbf{r}_n|^{\frac{\delta}{2+\delta}}} \longrightarrow 0, \tag{D.21}$$

and thus

$$C_n^m\left(\frac{\delta}{1+\delta}\right) = o\left(|\mathbf{d}_n|\right). \tag{D.22}$$

D.3.2 Step 1 : Estimate of the "mortar"

We will chop the mortar into several chunks; some of these chunks will be asymptotically independent, thus ensuring that the mean square of the mortar is small. First, we introduce the following notations:

V_n is the normalized sum of all r.v.'s in the mortar. The jth base face of D_n denotes the set of all indices of the form

$$(a_1, a_2, \cdots, a_{j-1}, 1, a_{j+1}, \cdots, a_m). \tag{D.23}$$

For example if $m = 3$, these are the hyperplanes through $x = 1$, $y = 1$, and $z = 1$. Denote a strip of the jth base face by T_j. It is a sub-block with dimensions

$$\mathbf{d}_n^1 \times \cdots \times \mathbf{d}_n^{j-1} \times \mathbf{l}_n^j \times \mathbf{d}_n^{j+1} \times \cdots \times \mathbf{d}_n^m. \tag{D.24}$$

There are $\mathbf{r}_n^j$ of such T_j's within the jth base base, for each j. We won't bother to distinguish between the different T_j's of the jth base face; even though they have different distributions, we are only concerned with their second moments, which will be given the same bound due to our moment bounds and mixing assumptions.

Now imagine breaking each T_j into a multidimensional "checkerboard" which will consist of several distinct colors. We break it up in such a way that no block shares an edge or corner with a block of the same color, and the blocks will be of size $\mathbf{l}_n$, i.e. their dimension is $\mathbf{l}_n^1 \times \cdots \times \mathbf{l}_n^m$.

Claim The number of colors neccessary for this "checkerboard" is 2^{m-1}.

We prove this claim by induction on m; the case $m = 2$ is clear, since in this situation we have a flat strip which we alternately color with two different colors. So suppose the claim holds true for dimension m. We obtain the next highest dimensional strip by stacking up $\mathbf{r}_n^j$ m-dimensional strips in direction j (for an appropriate direction j). Without loss of generality say $j = m + 1$. The first "new stack," i.e. the second stack, must have all new colors, in order for the desired "checkerboard" separation to hold; for this we need 2^{m-1} new colors. The next stack after that (i.e. the third stack) will have the same pattern as the first 2^{m-1} colors, as the first and third layers will be separated by distance $\mathbf{l}_n^{m+1}$. Then the fourth layer can follow the coloring of the second layer, and so on. Thus we need a total of $2^{m-1} + 2^{m-1} = 2 \cdot 2^{m-1} = 2^m$ colors for the $(m+1)$-dimensional strip. This proves the claim.

Since the number of such little blocks in each T_j is

$$\frac{\mathbf{d}_n^1 \cdots \mathbf{d}_n^{j-1} \mathbf{d}_n^{j+1} \cdots \mathbf{d}_n^m}{\mathbf{l}_n^1 \cdots \mathbf{l}_n^{j-1} \mathbf{l}_n^{j+1} \cdots \mathbf{l}_n^m} = \frac{|\mathbf{d}_n| \mathbf{l}_n^j}{|\mathbf{l}_n| \mathbf{d}_n^j}, \tag{D.25}$$

the number of little blocks within each "color grouping" of T_j is 2^{1-m} times the above number. The normalized sum of blocks within a color grouping will be denoted by R_j, where we again suppress the fact that there are 2^{m-1} of these within each T_j. We will denote an arbitrary little block by L_n.

We will need the following lemma, which generalizes an instrumental result by Ibragimov ([30]):

Lemma D.3.1. *Let ϕ_Z be the characteristic function of the random variable Z.*

$$\left\|\phi_{\sum_{i=1}^{k} U_{n,i}+Y} - \phi_{\sum_{i=1}^{k} U'_{n,i}+Y'}\right\| \leqslant 16\, k\, \alpha_n(\lambda_n; k\, \beta_n, \gamma_n), \tag{D.26}$$

where Y and Y' have the same distribution, are measurable with respect to sigma algebras of size γ_n and β_n respectively, such that $\gamma_n \geqslant \beta_n$, and are independent of each other. Also Y' is independent of the $U'_{n,i}$'s, and all blocks are at least λ_n distance apart.

Proof of the Lemma Note that the $k = 1$ case is just Ibragimov's Lemma, where we can replace $\alpha_n(\lambda_n)$ by $\alpha_n(\lambda_n; \beta_n, \gamma_n)$ since we can make this replacement in the covariance inequalities for $U_{n,1}$ and Y. Proceeding by induction, we assume the statement is true for $j = k - 1$. The following notation will be used: let $W_{i_1,i_2,\cdots,i_t}$ be a random variable with the same distribution as $U_{n,i_1} + U_{n,i_2} + \cdots + U_{n,i_t}$. Thus we have

$$\begin{aligned}\left\|\phi_{\sum_{i=1}^{k} U_{n,i}+Y} - \phi_{\sum_{i=1}^{k} U'_{n,i}+Y'}\right\| \leqslant{}& \left\|\phi_{\sum_{i=1}^{k} U_{n,i}+Y} - \phi_{W_{1,2,\cdots,k}+Y'}\right\| \\ &+ \left\|\phi_{W_{1,2,\cdots,k}+Y'} - \phi_{\sum_{i=1}^{k} U'_{n,i}+Y'}\right\|.\end{aligned} \tag{D.27}$$

To the first term we apply Ibragimov's Lemma to $\sum_{i=1}^{k} U_{n,i}$ and Y, with their independent (but identically distributed versions) $W_{1,2,\cdots,k}$ and Y', and discover that the first part is bounded by $16\alpha_n(\lambda_n; k\,\beta_n, \gamma_n)$. For the second term, we use the independence to factor a term $\|\phi_{Y'}\|$ (which is bounded by 1) from both summands, leaving $\left\|\phi_{W_{1,2,\cdots,k}} - \phi_{\sum_{i=1}^{k} U'_{n,i}}\right\|$. By inductive hypothesis, since $W_{1,2,\cdots,k}$ and $\sum_{i=1}^{k-1} U_{n,i} + U_{n,k}$ are equal in distribution, the term is bounded by

$$16\,(k-1)\,\alpha_n(\lambda_n;\,(k-1)\,\beta_n; \beta_n) \leqslant 16\,(k-1)\,\alpha_n(\lambda_n; k\,\beta_n; \gamma_n). \tag{D.28}$$

Summing the two parts yields the desired result. $\dagger$

Now we will proceed to estimate V_n:

$$Var^{\frac{1}{2}}(V_n) \leqslant \sum_{j=1}^{m} Var^{\frac{1}{2}}\left(\sum_{k=1}^{\mathbf{r}_n^j} T_j\right) \tag{D.29}$$

where we suppress the fact that T_j depends on k for ease of notation. We claim that it suffices to show that

$$\sum_{j=1}^{m} (\mathbf{r}_n^j)^{\frac{1}{2}} Var^{\frac{1}{2}}(T_j) = o(1). \tag{D.30}$$

Applying Lemma D.3.1, we see that

$$\begin{aligned} \|\phi_{\sum_{k=1}^{\mathbf{r}_n^j} T_j} - \phi_{\sum_{k=1}^{\mathbf{r}_n^j} T_j'}\| &\leqslant 16\mathbf{r}_n^j \alpha_n \left(\mathbf{b}_n^j; \mathbf{r}_n^j \frac{\mathbf{l}_n^j|\mathbf{d}_n|}{\mathbf{d}_n^j|\mathbf{l}_n|}, \frac{\mathbf{l}_n^j|\mathbf{d}_n|}{\mathbf{d}_n^j|\mathbf{l}_n|}\right) \\ &\leqslant 16\mathbf{r}_n^j \alpha_n \left(\mathbf{b}_n^j; \frac{|\mathbf{d}_n|}{|\mathbf{l}_n|}, \frac{|\mathbf{d}_n|}{|\mathbf{l}_n|}\right) \longrightarrow 0. \end{aligned} \tag{D.31}$$

The limit at the end is due to (D.8). Here, T_j' is an independent identically distributed copy of T_j. Therefore the difference

$$\sum_{k=1}^{\mathbf{r}_n^j} T_j - \sum_{k=1}^{\mathbf{r}_n^j} T_j' \tag{D.32}$$

converges weakly to zero, and hence is $o_P(1)$ as well. Now the variance of the independent version is just

$$\sum_{k=1}^{\mathbf{r}_n^j} Var(T_j), \tag{D.33}$$

so if T_j tends to zero in mean square, it also does in probability, and then so will the non-independent version.

Next, we find that

$$Var^{\frac{1}{2}}(T_j) \leqslant 2^{m-1} Var^{\frac{1}{2}}(R_j). \tag{D.34}$$

We claim that all the little blocks within each R_j are asymptotically independent; this is again via Lemma (D.3.1) :

$$\|\phi_{\sum L_n} - \phi_{\sum L_n'}\| \leqslant 16 \left(2^{1-m} \frac{|\mathbf{d}_n| \mathbf{l}_n^j}{|\mathbf{l}_n| \mathbf{d}_n^j}\right) \alpha_n \left(\mathbf{l}_n^\flat; |\mathbf{d}_n|, |\mathbf{l}_n|\right) \longrightarrow 0, \tag{D.35}$$

where the sum is over all little blocks L_n within each color grouping R_j; as usual, L_n' denotes an *iid* (independent and indentically distributed) copy of L_n. The limit at the end is due to (D.7). This gives us asymptotic independence of different "colors" within the checkerboard.

Now we expand the little block L_n:

$$L_n^2 = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{l}_n} Y_{\mathbf{i}}{}^2 + \sum_{\mathbf{i} \neq \mathbf{j}} Y_{\mathbf{i}} Y_{\mathbf{j}}, \tag{D.36}$$

where $Y_{\mathbf{i}}$ is the random variable of the "first" little block with coordinates given by the vector $\mathbf{i}$. Here, $\mathbf{i} \neq \mathbf{j}$ means at least one coordinate of the two vectors differs. The estimates are the same for every little block in D_n, because the covariances will be bounded by the mixing coefficients, which are only a function of distance between variables, not their absolute location. We take the expectation of this sum to get variances and covariances. The first term becomes:

$$\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{l}_n} VarY_{\mathbf{i}} = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{l}_n} \|Y_{\mathbf{i}}\|_2^2 \leqslant \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{l}_n} \|Y_{\mathbf{i}}\|_{2+2\delta}^2 \leqslant \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{l}_n} \Delta^2 = |\mathbf{l}_n| \, \Delta^2. \tag{D.37}$$

For the second term, consider the covariance between random variables $Y_{\mathbf{i}}$ and $Y_{\mathbf{i}+\mathbf{k}}$; by holding all but the first coordinate fixed, we see that the number of covariances where the first coordinate of the variables differs by k^1 is $2\,(\mathbf{l}_n^1 - k^1)$. Repeating for each coordinate and using the fundamental mixing inequality (see [63]) gives a bound on the second term:

$$\sum_{\mathbf{k}>\mathbf{0}}^{\mathbf{l_n}} \prod_{j=1}^{m} 2 \cdot (\mathbf{l_n}{}^j - \mathbf{k}^j) \cdot 10 \cdot \alpha_n(\mathbf{k}^1 + \cdots + \mathbf{k}^m; 1, 1)^{\frac{\delta}{1+\delta}} \Delta^2, \tag{D.38}$$

where we have used $p = q = 2 + 2\delta$ so that $r = 1 - 2/p = \frac{\delta}{1+\delta}$, and (D.4) for moment bounds. Also, $\mathbf{j} > \mathbf{0}$ means, of course, that at least one component of $\mathbf{j}$ is positive. Simplifying the above expression gives:

$$2^m \cdot 10 \cdot |\mathbf{l}_n| \, \Delta^2 \sum_{\mathbf{k} > \mathbf{0}}^{\mathbf{l}_n} \alpha_n(\mathbf{k}^1 + \cdots + \mathbf{k}^m; 1, 1)^{\frac{\delta}{1+\delta}}. \tag{D.39}$$

The latter expression is simplified by counting the number of times $\alpha_n(k)$ occurs in the sum, for each fixed $k \in \mathbb{N}$. The points of the lattice corresponding to this $\alpha_n(k)$ form a "sphere" under the given metric. So the number of times $\alpha_n(k)$ appears in the sum is equal to the number of random variables of distance k from a fixed random variable. This is the number of lattice points on the sphere of radius k. It turns out that the number of points on this sphere is of the order of $(k+1)^{m-1}$. In other words, the number of solutions in $\mathbb{N}^m$ of $z_1 + z_2 + \cdots + z_m = k$ is bounded by $K\,(k+1)^{m-1}$, for some constant K only depending on m. Thus we obtain:

$$\sum_{\mathbf{k} > \mathbf{0}}^{\mathbf{l_n}} \alpha_n(\mathbf{k}^1 + \cdots + \mathbf{k}^m; 1, 1)^{\frac{\delta}{1+\delta}} \leqslant \sum_{k=1}^{\infty} K\,(k+1)^{m-1}\, \alpha_n(k; 1, 1)^{\frac{\delta}{1+\delta}}. \tag{D.40}$$

Putting it together, we get

$$\begin{aligned}
\mathbb{E}L_n^2 &\leqslant |\mathbf{l}_n| \cdot \Delta^2 + 2^m \cdot |\mathbf{l}_n| \cdot 10 \cdot \Delta^2 \cdot K \sum_{k=1}^{\infty} (k+1)^{m-1}\, \alpha_n(k; 1, 1)^{\frac{\delta}{1+\delta}} \qquad \text{(D.41)} \\
&\leqslant |\mathbf{l}_n| \cdot \Delta^2 \left\{ 1 + 2^m \cdot 10 \cdot K \sum_{k=1}^{\infty} (k+1)^{m-1}\, \alpha_n(k; 1, 1)^{\frac{\delta}{1+\delta}} \right\} \\
&\leqslant |\mathbf{l}_n| \cdot \Delta^2 \left\{ 2^m \cdot 10 \cdot K \sum_{k=0}^{\infty} (k+1)^{m-1}\, \alpha_n(k; 1, 1)^{\frac{\delta}{1+\delta}} \right\} \\
&= |\mathbf{l}_n| \cdot \Xi \cdot C_n^m \left(\frac{\delta}{1+\delta} \right),
\end{aligned}$$

where $\Xi := \Delta^2 \cdot 2^m \cdot 10 \cdot K$.

Thus we get the following estimate for the mortar:

$$Var^{\frac{1}{2}}(V_n') \leqslant \sum_{j=1}^{m} \left[(\mathbf{r}_n^j)^{\frac{1}{2}} 2^{m-1} \left(\frac{|\mathbf{d}_n|\mathbf{l}_n^j}{|\mathbf{l}_n|\mathbf{d}_n^j} \right)^{\frac{1}{2}} \frac{1}{|\mathbf{d}_n|} |\mathbf{l}_n|^{\frac{1}{2}} C_n^m (\frac{\delta}{1+\delta})^{\frac{1}{2}} \right] \tag{D.42}$$

$$= O\left(\frac{1}{|\mathbf{d}_n|^{\frac{1}{2}}} \sum_{j=1}^{m} \left(\frac{\mathbf{r}_n^j \mathbf{l}_n^j}{\mathbf{d}_n^j} \right)^{\frac{1}{2}} C_n^m (\frac{\delta}{1+\delta})^{\frac{1}{2}} \right)$$

$$= O\left(\frac{1}{|\mathbf{d}_n|^{\frac{1}{2}}} \left(\frac{\mathbf{r}_n^\sharp \mathbf{l}_n^\sharp}{\mathbf{d}_n^\sharp} \right)^{\frac{1}{2}} C_n^m (\frac{\delta}{1+\delta})^{\frac{1}{2}} \right) \longrightarrow 0.$$

Here, V_n' denotes V_n with the appropriate random variables replaced by *iid* versions, as described in the two claims above on the asymptotic situation. The limit is due to (D.22). Thus $V_n' = o_P(1)$, so $V_n = o_P(1)$ as well, as required.

D.3.3 Step 2: The Mixing Condition

Let ϕ and ζ be the characteristic functions of $|\mathbf{d}_n|^{-\frac{1}{2}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} U_{n,\mathbf{i}}$ and $|\mathbf{d}_n|^{-\frac{1}{2}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} U'_{n,\mathbf{i}}$. Then

$$\|\phi - \zeta\| \leqslant 16 \cdot |\mathbf{r}_n| \cdot \alpha_n(\mathbf{l_n}^\flat; |\mathbf{d}_n|, |\mathbf{b}_n|) \tag{D.43}$$

where $\|\cdot\|$ denotes the sup norm.

This follows immediately from Lemma D.3.1 (see the previous section) where Y is the last big block, so that $\beta_n = |\mathbf{b}_n| = \gamma_n$. Also $\lambda_n = \mathbf{l}_n^\flat$, and $k = |\mathbf{r}_n| - 1$; rounding upwards gives the desired bound. Now this upper bound goes to zero as n tends to infinity by (D.7), which establishes the asymptotic independence of the big blocks.

D.3.4 Step 3: Estimate of the Big Block

Much of the computations are the same as for the small block; here we use a trick found in [18]. So for the given δ, let c be the smallest even integer that is greater or equal to $2+\delta$; let $d := \frac{c}{2}$. Then

$$\mathbb{E}|U'_{n,\mathbf{i}}|^{2+\delta} = \mathbb{E}\left|\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} Z_{\mathbf{j}}\right|^{\frac{2+\delta}{c}c} \leqslant \mathbb{E}\left(\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} |Z_{\mathbf{j}}|^{\frac{2+\delta}{c}}\right)^{c} \tag{D.44}$$
$$= \mathbb{E}\left[\left(\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} Y_{\mathbf{j}}\right)^2\right]^d = \mathbb{E}\left[\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} Y_{\mathbf{j}}^2 + \sum_{\mathbf{i}\neq\mathbf{j}} Y_{\mathbf{i}}Y_{\mathbf{j}}\right]^d$$
$$= \left\|\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} Y_{\mathbf{j}}^2 + \sum_{\mathbf{i}\neq\mathbf{j}} Y_{\mathbf{i}}Y_{\mathbf{j}}\right\|_d^d$$
$$\leqslant \left[\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} \|Y_{\mathbf{j}}^2\|_d + \sum_{\mathbf{i}\neq\mathbf{j}} \|Y_{\mathbf{i}}Y_{\mathbf{j}}\|_d\right]^d,$$

where $Z_{\mathbf{j}}$ is the random variable of the "first" big block with coordinate vector $\mathbf{j}$, and $Y_{\mathbf{j}} := |Z_{\mathbf{j}}|^{\frac{2+\delta}{c}}$. The first inequality is merely the triangle inequality $|x+y|^{\epsilon} \leqslant |x|^{\epsilon} + |y|^{\epsilon}$ for $0 < \epsilon \leqslant 1$. The other inequality is the Holder Inequality. The first term is:

$$\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} \|Y_{\mathbf{j}}^2\|_d = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} (\mathbb{E}|Z_{\mathbf{j}}|^{2+\delta})^{\frac{1}{d}} = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} (\|Z_{\mathbf{j}}\|_{2+\delta}^{2+\delta})^{\frac{1}{d}} \tag{D.45}$$
$$\leqslant \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}_n} (\|Z_{\mathbf{j}}\|_{2+2\delta}^{2+\delta})^{\frac{1}{d}} \leqslant |\mathbf{b}_n|\, \Delta^{\frac{2+\delta}{d}},$$

and the second term is:

$$\sum_{\mathbf{i}\neq\mathbf{j}} \|Y_{\mathbf{i}}Y_{\mathbf{j}}\|_d = \sum_{\mathbf{i}\neq\mathbf{j}} (\mathbb{E}|W_{\mathbf{i}}W_{\mathbf{j}}|)^{\frac{1}{d}} \tag{D.46}$$
$$\leqslant \sum_{\mathbf{i}\neq\mathbf{j}} \left\{\Delta^{2+\delta} \cdot 10 \cdot \alpha_n(|\mathbf{i}-\mathbf{j}|;1,1)^{\frac{\delta}{2+2\delta}}\right\}^{\frac{1}{d}}$$
$$= \Delta^{\frac{2+\delta}{d}}\, 10^{\frac{1}{d}} \sum_{\mathbf{i}\neq\mathbf{j}} \alpha_n(|\mathbf{i}-\mathbf{j}|;1,1)^{\frac{\delta}{c(1+\delta)}},$$

where $W_{\mathbf{i}} := |Z_{\mathbf{i}}|^{\frac{2+\delta}{2}}$. Now we use the same type of counting argument on the sum of mixing coefficients as in Step 1, and obtain:

$$\sum_{\mathbf{i}\neq\mathbf{j}} \alpha_n(|\mathbf{i}-\mathbf{j}|;1,1)^{\tau} \leqslant |\mathbf{b}_n|\, 2^m \cdot K \sum_{k=1}^{\infty} (k+1)^{m-1} \alpha_n(k;1,1)^{\tau}, \tag{D.47}$$

where $\tau := \frac{\delta}{c\,(1+\delta)}$. Therefore we get the following bound:

$$\begin{aligned}
\mathbb{E}\left|\sum_{\mathbf{j}=1}^{\mathbf{b_n}} Z_{\mathbf{j}}\right|^{2+\delta} &\leqslant \left\{ |\mathbf{b}_n|\,\Delta^{\frac{2+\delta}{d}} + \Delta^{\frac{2+\delta}{d}} \cdot 10^{\frac{1}{d}} \sum_{\mathbf{i}\neq\mathbf{j}} \alpha_n(|\mathbf{i}-\mathbf{j}|;1,1)^{\tau} \right\}^{d} && \text{(D.48)}\\
&\leqslant \Delta^{2+\delta} \left\{ |\mathbf{b}_n| + 10^{\frac{1}{d}} \cdot |\mathbf{b}_n| \cdot 2^{m} \cdot K \sum_{k=1}^{\infty} (k+1)^{m-1} \alpha_n(k;1,1)^{\tau} \right\}^{d}\\
&\leqslant \Delta^{2+\delta} \left\{ 10^{\frac{1}{d}} \cdot |\mathbf{b}_n| \cdot 2^{m} \cdot K \sum_{k=0}^{\infty} (k+1)^{m-1} \alpha_n(k;1,1)^{\tau} \right\}^{d}\\
&= \Delta^{2+\delta} |\mathbf{b}_n|^{d} \cdot 10 \cdot K^{d} \cdot 2^{md}\, [C_n^m(\tau)]^{d},
\end{aligned}$$

where $C_n^m(\tau) := \sum_{k=0}^{\infty} (k+1)^{m-1} \alpha_n(k;1,1)^{\tau}$. Let $\Xi := 10\,K^{d}\,2^{md}$. Finally, we get:

$$\begin{aligned}
\mathbb{E}\left| |\mathbf{b}_n|^{-\frac{1}{2}} U'_{n,\mathbf{i}} \right|^{2+\delta} &= |\mathbf{b}_n|^{-\frac{2+\delta}{2}} \mathbb{E}|U'_{n,\mathbf{i}}|^{2+\delta} && \text{(D.49)}\\
&\leqslant |\mathbf{b}_n|^{-\frac{2+\delta}{2}} \Delta^{2+\delta} \cdot |\mathbf{b}_n|^{d} \cdot \Xi \cdot [C_n^m(\tau)]^{d}\\
&\leqslant \Delta^{2+\delta} \cdot \Xi \cdot [C_n^m(\tau)]^{d},
\end{aligned}$$

since $|\mathbf{b}_n|^{d} \leqslant |\mathbf{b}_n|^{\frac{2+\delta}{2}}$.

D.3.5 Step 4 : The Lyapounov Condition

We wish to show that $|\mathbf{d}_n|^{-\frac{1}{2}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} U'_{n,\mathbf{i}} \overset{\mathcal{L}}{\Longrightarrow} \sigma N$, where N denotes a standard normal random variable. To this end, we must verify that

$$\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}_n} \left\{ \left[\frac{1}{|\mathbf{d}_n|} Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} U'_{n,\mathbf{i}} \right) \right]^{-\frac{2+\delta}{2}} \mathbb{E}\left| |\mathbf{d}_n|^{-\frac{1}{2}} U'_{n,\mathbf{j}} \right|^{2+\delta} \right\} \longrightarrow 0. \qquad \text{(D.50)}$$

Then we may apply the Lyapounov Central Limit Theorem. It is valid to consider the independent versions $U'_{n,\mathbf{j}}$ of $U_{n,\mathbf{j}}$, since their characteristic functions are asymptotically the same by Step 1. First by (D.5) we obtain:

$$\begin{aligned}
\frac{1}{|\mathbf{r}_n| \cdot |\mathbf{b}_n|} Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} U'_{n,\mathbf{i}} \right) &= \frac{1}{|\mathbf{r}_n| \cdot |\mathbf{b}_n|} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} \mathbb{E}\big(U'_{n,\mathbf{i}}\big)^2 && \text{(D.51)}\\
&= \frac{1}{|\mathbf{r}_n|} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n} \mathbb{E}\Big(|\mathbf{b}_n|^{-\frac{1}{2}} U'_{n,\mathbf{i}} \Big)^2 \longrightarrow \sigma^2.
\end{aligned}$$

Now from Step 3 we have

$$\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}_n}\left\{\left[\frac{1}{|\mathbf{d}_n|}Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}\right)\right]^{-\frac{2+\delta}{2}}\mathbb{E}\left||\mathbf{d}_n|^{-\frac{1}{2}}U'_{n,\mathbf{j}}\right|^{2+\delta}\right\} \tag{D.52}$$
$$= \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}_n}\left\{\left[Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}\right)\right]^{-\frac{2+\delta}{2}}\mathbb{E}|U'_{n,\mathbf{j}}|^{2+\delta}\right\}$$
$$= |\mathbf{r}_n|^{-\frac{2+\delta}{2}}\left\{\left[\frac{1}{|\mathbf{r}_n|\cdot|\mathbf{b}_n|}Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}\right)\right]^{-\frac{2+\delta}{2}}\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}_n}\mathbb{E}\left||\mathbf{b}_n|^{-\frac{1}{2}}U'_{n,\mathbf{j}}\right|^{2+\delta}\right\}$$
$$\leqslant |\mathbf{r}_n|^{-\frac{2+\delta}{2}}\,O(1)\,|\mathbf{r}_n|\cdot\Delta^{2+\delta}\cdot\Xi\,[C_n^m(\tau)]^d$$
$$\leqslant |\mathbf{r}_n|^{-\frac{\delta}{2}}\,O(1)\,[C_n^m(\tau)]^{\frac{2+\delta}{2}}$$
$$= O(1)\left\{|\mathbf{r}_n|^{-\frac{\delta}{2+\delta}}\,C_n^m(\tau)\right\}^{\frac{2+\delta}{2}} \longrightarrow 0,$$

where the last step follows from (D.6). Thus, by Lyapounov's CLT,

$$\frac{|\mathbf{d}_n|^{-\frac{1}{2}}\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}}{\left[Var\left(|\mathbf{d}_n|^{-\frac{1}{2}}\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}\right)\right]^{\frac{1}{2}}} \stackrel{\mathcal{L}}{\Longrightarrow} N. \tag{D.53}$$

Now if we look at the variance we find that

$$\lim_{n\to\infty}\frac{1}{|\mathbf{d}_n|}Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}\right) = \lim_{n\to\infty}\frac{|\mathbf{r}_n|\cdot|\mathbf{b}_n|}{|\mathbf{d}_n|}\cdot\lim_{n\to\infty}\left\{\frac{1}{|\mathbf{r}_n|\cdot|\mathbf{b}_n|}Var\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}}\right)\right\} = 1\cdot\sigma^2. \tag{D.54}$$

Then by Slutsky's Theorem, we obtain

$$|\mathbf{d}_n|^{-\frac{1}{2}}\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}U'_{n,\mathbf{i}} \stackrel{\mathcal{L}}{\Longrightarrow} \sigma N. \tag{D.55}$$

Now we add up the big blocks and the mortar, which goes to zero in probability by Step 1. So again by Slutsky, we obtain:

$$|\mathbf{d}_n|^{-\frac{1}{2}}S_{n,\mathbf{d_n},\mathbf{1}} = |\mathbf{d}_n|^{-\frac{1}{2}}\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{r}_n}\left(U'_{n,\mathbf{i}}+V_n\right) \stackrel{\mathcal{L}}{\Longrightarrow} \sigma N. \tag{D.56}$$

This completes the proof. †

Remark One can also use an l_∞ norm and the same results hold, as the estimate on the cardinality of the sphere also holds for the sup norm.

Bibliography

[1] Arcones, M. and Giné, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annals of the Institute Henri Poincaré* **25**, 457–481.

[2] Beran, J. (1994) *Statistics for Long-Memory Processes.* Chapman and Hall.

[3] Beran, J., Sherman R., Taqqu, M., and Willinger W. (1995) Long-range dependence in Variable-bit rate video traffic. *IEEE Trans. Comm.* **43** 1566–1579.

[4] Bernstein, S. (1927). Sur l'extension du théorème du calcul des probabilités aux sommes de quantités dependantes. *Mathematische Annalen* **97**, 1–59.

[5] Bertail, P. (1997) Second order properties of an extrapolated bootstrap without replacement: the i.i.d. and strong mixing cases. *Bernoulli* **3**, 149–179.

[6] Bertail, P., Haefke, C., Politis, D., and White, H. (2000–01) A Subsampling Approach to Estimating the Distribution of Diverging Statistics with Applications to Assessing Financial Market Risks. UCSD Department of Economics Discussion Paper 2000–01.

[7] Bertail, P., Politis, D., and Romano, J. (1999) On Subsampling Estimators with Unknown Rate of Convergence. *Journal of the American Statistical Association* .

[8] Billingsley, P. (1995) *Probability and Measure.* John Wiley and Sons.

[9] Bolthausen, E. (1982). On the central limit theorem for stationary random fields, *Annals of Probability* **10**, 1047–1050.

[10] Bradley, R.C. (1992). On the spectral density and asymptotic normality of weakly dependent random fields, *Journal of Theoretical Probability* **5**, No. 2, 355–373.

[11] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods.* Springer.

[12] Castillo, E. (1988) Extreme Value Theory in Engineering. Academic Press, San Diego.

[13] Cressie, N. (1993) *Statistics for Spatial Data*, 2nd edition. John Wiley, New York.

[14] Davis, R. and Hsing, T. (1995) Point Process and Partial Sum Convergence for Weakly Dependent Random Variables with Infinite Variance. *Annals of Probability* **23**, No. 2, 879–917.

[15] Davis, R., and Mikosch, T. (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* **26**, No. 5, 2049–2080.

[16] Davis, R., and Resnick, S. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* **13**, No. 1, 179–195.

[17] Davis, R., and Resnick, S. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* **14**, No. 2, 533–558.

[18] Doukhan, P. (1994), *Mixing: Properties and Examples*, Lecture Notes in Statistics No. 85, Springer, New York.

[19] Duffy, D., McIntosh, A., Rosenstein, M., and Willinger, W. (1993) Analyzing telecommunications traffic data from working common channel signaling networks. In *Proceedings of the 25th Interface, San Diego.* Interface Foundation of North America.

[20] Duffy, D., McIntosh, A., Rosenstein, M., and Willinger, W. (1994) Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks. *IEEE Journal on Selected Areas in Communications* **12** 544–551.

[21] Durrett, R. (1996) *Probability: Theory and Examples.* Duxbury Press.

[22] Embrechts, P., Mikosch, T., and Kluppelberg, C. (1997) *Modeling Extrema Events for Insurance and Finance.* Springer-Verlag, Berlin.

[23] Feller, W. (1971) . *An Introduction to Probability Theory and Its Applications, Vol. 2.* 2nd edition John Wiley, New York.

[24] Götze, F. and Hipp, C. (1983) . Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64**, 211–239.

[25] Greene, Myron T., and Fielitz, Bruce D. (1977) Long-Term Dependence in Common Stock Returns. *Journal of Financial Economics* **4**, 339-349.

[26] Greene, Myron T., and Fielitz, Bruce D. (1979) The Effect of Long Term Dependence on Risk-Return Models of Common Stocks. *Operations Research* **25**, No. 5, 944–951.

[27] Grigoriu, M. (1995) *Applied Non-Gaussian Processes.* Prentice Hall, Englewood Cliffs, NJ.

[28] Hall, P., Horowitz J.L., and Jing, B.-Y. (1996) On blocking rules for the bootstrap with dependent data. *Biometrika* **50**, 561–574.

[29] Heath, D., Resnick, S., and Samorodnitsky, G. (1997) Heavy Tails and Long Range Dependence in On/Off Processes and Associated Fluid Models. website: http://www.orie.cornell.edu/trlist/

[30] Ibragimov, I. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications* **7**, 349–382.

[31] Ivanov, I., and Leonenko, N. (1989). *Statistical Analysis of Random Fields.* Kluwer Academic Publishers, The Netherlands.

[32] Karr, A.F. (1991) *Point Processes and their Statistical Inference*, 2nd edition. Dekker, New York.

[33] Koedijk, K., Schafgans, M., and De Vries, C. (1990) The tail index of exchange rate returns. *Journal of International Economics* **29** 93–108.

[34] Klüppelberg, C., and Mikosch, T. (1993) Spectral Estimates and Stable Processes. *Stochastic Processes and their Applications* **47** 323–344.

[35] Klüppelberg, C., and Mikosch, T. (1994) Some Limit Theory for the Self-Normalized Periodogram of Stable Processes. *Scandinavian Journal of Statistics* **21** 485–491.

[36] Logan, B.F., Mallows, C.L., Rice, S.O., and Shepp, L.A. (1973) . Limit Distributions of Self-Normalized Sums. *Annals of Probability* **1** , 788 – 809.

[37] Lovejoy, S. (1982) Area-Perimeter relation for rain and cloud areas. *Science* **216**, 185–187.

[38] Mandelbrot, Benoit B. (1965) Self-similar error clusters in communications systems and the concept of conditional systems and the concept of conditional stationarity. *I.E.E.E. Transcations on Communications Technology* COM-13, 71–90.

[39] Mandelbrot, Benoit B. (1969) Long-run linearity, locally Gaussian processes, H-spectra and infinite variances. *International Economic Rev* **10**, 82-113.

[40] Mandelbrot, Benoit B. (1974) Intermittent turbulence in self similar cascades; divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics* **62**, 331–358.

[41] Mandelbrot, Benoit B. (1983) *The Fractal Geometry of Nature.* W.H.Freeman and Company.

[42] Mandelbrot, Benoit B., and Wallis, J.R. (1968) Noah, Joseph, and Operational Hydrology. *Water Resources Res.* **4**, 909-917.

[43] Mandelbrot, Benoit B., and Wallis, J.R. (1969) Some Long-Run Properties of Geophysical Records. *Water Resources Res* **5**, 321-340.

[44] Meier-Hellstern, K., Wirth, P., Yan, Y., and Hoeflin, D. (1991) Traffic models for ISDN data users: office automation application. In *Teletraffic and Datatraffic in a Period of Change. Proceedings of the 13th ITC* (A. Jensen and V.B. Iversen, eds.) 167–192. North Holland, Amsterdam.

[45] Miller, Curtis. Three Theorems on ρ^*-Mixing Random Fields. *Journal of Theoretical Probability*, 1994.

[46] Pham, Tuan D. and Tran, Lanh T. (1985) Some Mixing Properties of Time Series Models. *Stochastic Processes and Their Applications* **19**, 297–303.

[47] Politis, D. and Romano, J. (1994) Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions. *Annals of Statistics* **22**, No. 4, 2031–2050.

[48] Politis, D., Romano, J., and Wolf, M. (1999) *Subsampling.* Springer, New York.

[49] Politis, D., Romano, J., and Wolf, M. (1997). Subsampling for heteroskedastic time series. *Journal of Econometrics*, Vol. 81, No. 2, 1997, pp. 281-317.

[50] Resnick, S. (1986) Point processes, regular variation, and weak convergence. *Advances in Applied Probability* **18**, 66–138.

[51] Resnick, S. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer–Verlag.

[52] Resnick, S. (1997) Special Invited Paper: Heavy Tail Modeling and Teletraffic Data. *Annals of Statistics* **25**, No. 5, 1805–1849.

[53] Resnick, S., and Samorodnitsky, G. (1997) Performance decay in a single-server exponential queuing model with long-range dependence. *website: http://www.orie.cornell.edu/trlist/*

[54] Resnick, S., Samorodnitsky, G., and Xue, F. (1999) How Misleading Can Sample ACF's of Stable MA's be? (Very!). *Annals of Applied Probability* **9** , No. 3, 797–817.

[55] Resnick, S., Samorodnitsky, G., and Xue, F. (2000) Growth Rates of Sample Covariances of Stationary Symmetric α-Stable Processes Associated with Null Recurrent Markov Chains. *Stochastic Processes and Their Applications* **85**, 321–339.

[56] Resnick, S., and Stairca, C. (1998) Tail index estimation for dependent data. *Annals of Applied Probability* **8**, No. 4, 1156–1183.

[57] Rio, E. (1997), About the Lindeberg method for strongly mixing sequences, *ESAIM: Probability and Statistics (electronic journal: http://www.emath.fr/ps/psEng.html)* , Vol.1, Dec. 1997, pp. 35-61.

[58] Ripley, B. (1981) *Spatial Statistics.* John Wiley, New York.

[59] Romano, J., and Wolf, M. (1998). Inference for the Mean in the Heavy-Tailed Case. *Metrika.*

[60] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences* **42**, 43–47.

[61] Rosenblatt, M. (1984). Asymptotic normality, strong mixing and spectral density estimates, *Annals of Probability* **12**, 1167–1180.

[62] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Boston.

[63] Roussas, G.G. and Ioannides, D. (1987). Moment Inequalities for Mixing Sequences of Random Variables, *Stochastic Analysis and its Applications* **5**, 61–120.

[64] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes.* Chapman and Hall.

[65] Samur, J. D. (1987), On the invariance principle for stationary ϕ-mixing triangular arrays with infinitely divisible limits, *Probab. Theory Related Fields*, 75, no.2, 245-259.

[66] Taqqu, M., and Levy, J. (1985) Using renewal processes to generate long-range dependence and high variability. *Dependence in Probability and Statistics.* Birkhauser.

[67] Utev, S.A. (1989), Central limit theorem for dependent random variables, *Prob. Theory and Math. Statistics, Vol. II (Vilnius, 1989)*, pp. 519-528

[68] Willinger, W., Taqqu, M., Leland, W., and Wilson, D. (1995) Self-similarity in high-speed packet traffic: analysis and modeling of ethernet traffic measurements. *Statist. Sci.* **10** 67–85.

[69] Willinger, W., Taqqu, M., Sherman, R., and Wilson, D. (1997) Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking* **5** 71–96.

[70] Withers, C.S. (1981) Conditions for Linear Processes to be Strong Mixing. *Probability Theory and Related Fields* **57**, 477–480