UNIVERSITE CATHOLIQUE DE LOUVAIN

Institute of Multidisciplinary Research for Quantitative Modelling and Analysis

Institute of Statistics, Biostatistics and Actuarial sciences

Heavy tailed functional time series

Ph.D. Thesis

Membres du jury:

Prof. Johan Segers *(UCL Belgium)*
Prof. Michel Denuit *(UCL Belgium)*
Prof. Jan Beirlant *(KUL AK Belgium)*
Prof. Henrik Hult *(KTH Sweden)*
Prof. Ingrid Van Keilegom *(UCL Belgium)*
Prof. Rainer Von Sachs *(UCL Belgium)*

Thèse présentée en vue de l'obtention du grade de Docteur en Sciences (orientation statistique) par :

Thomas Meinguet

Louvain-la-Neuve, 2010

Acknowledgment

I cordially thank Professors Johan Segers, Sidney Resnick, Henrik Hult, Ingrid Van Keilegom Michel Denuit and Jan Beirlant for the scientific discussions that we had.

For their financial support, I mainly thank the Université catholique de Louvain (BE), then Cornell University (NY, US), the Belgian Science Policy of the Belgian government (IAP research network, grant nr. P6/03), the Ministère de la Communauté française of Belgium for the Projet d'Actions de Recherche Concertées (contract 07/12/002) and the traveling grant BV09-04, the Fond National de la Recherche Scientifique (BE), the National Center for Atmospheric Research (US), the Institut de Statistique (UCL), the Institut de mathématiques pures et appliquées (UCL) and the School of Operations Research and Industrial Engineering (CU).

Contents

Chapter 1

Global introduction

1.1 Quick overview

The goal of this thesis is to treat the temporal tail dependence and the cross-sectional tail dependence of heavy tailed functional processes. The word "*tail*" refers to the extreme values that the process can have, the tail of a positive random variable originally indicating the right-hand part of its probability distribution, for large values. A *heavy tail* (or *fat tail*) can be considered by the reader to that degree as a polynomially decreasing distribution of large values whereas, for the well-known normal distribution, it is exponentially decreasing. For a normal distribution, a value equal to its expected value plus three times its standard deviation can already be considered as extreme. For heavy tailed models the extreme values taken into account can be much larger.

Spatio-temporal phenomena are for instance rain, temperature, pollution, rays coming from the sun. This incorporates everything that can be viewed as satellite pictures, with temporally dependent observations. More difficult phenomena to apprehend are for instance the sea level along a dike or the windspeed along the faces of a building etc. Physicists model spatial phenomena by functions. Adding a stochastic component brings us to consider random functions. A time series of such objects is called a *functional time series* or a *functional process*. The advantage of treating spatial phenomena as random functions, as opposed to random vectors, is not to lose information. When designing a dike, one is interested in the fact that the water level does not exceed the top of the dike anywhere, not only at some measurement points. Concerning pollution in the atmosphere, we are interest in the fact that in a city the concentration level does not exceed a dangerous threshold anywhere and not only at the measurement stations.

The general motivation to model the extremal behavior of spatio-temporal phenomena is simple to apprehend. The occurrence of weather related extreme events, such as precipitation, floods, storms, heat waves has an impact for instance on the capital of insurance companies. The companies should take into account these events in a healthy risk management. Further examples of spatio-temporal phenomena are the wave height at sea affecting the design of dikes, oil platforms or ships. Other studies could be image analysis with pattern recognition, analysis of medical data, applications in material science etc.

Thinking of the concentration of a polluting gas as a spatial phenomenon, for instance CO, NO or NO_2 levels, an idea of what could be a temporal model involving functions is

$$X_t(x) = \int K(x-y)X_{t-1}(y)dy + \varepsilon_t(x), \tag{1.1}$$

where K is a kernel representing what becomes at time $t+1$ a particle of pollution deposited at the origin at time t and ε_t is the new amount of pollution added at time t. In other terms, the pollution level today at location x is the average pollution in a neighborhood of x the day before plus the new contribution of the day. In this model there is a dependence in space and over time. If there is a high value somewhere, it is likely to have simultaneously a high value somewhere else. If there is a peak of pollution today, it can persists for some days and it may have been caused by a peak of pollution during the previous days. When the innovations ε_t are heavy tailed, Equation (1.1) is a particular instance of functional process for which the theory of this thesis is worked out.

The *temporal tail dependence* answers the question “If I have an extreme value today, what can I know about the probability to have an extreme value tomorrow”. Because of the temporal dependence extremes may appear in temporal clusters for instance. One of the goals is to study these clusters and to provide models for this dependence. The *cross-sectional tail dependence* answers the question “If I have an extreme value at some location, what can I know about the probability to have an extreme value somewhere else", the source of a particular extreme event possibly affecting different locations.

The main tools introduced in this thesis for a heavy tailed functional time series are its *tail process* and its *spectral process*. They are both unifying objects capturing all the aspects of the probabilistic distribution of extreme values over time *and* over space. They will be defined as the conditional distribution to have an extreme value yesterday, today, tomorrow etc at every location given that there is an extreme value somewhere today. The difference between the tail process and the spectral process is the way of rescaling. The tail process includes an intensity factor, while the spectral process only indicates the directions where the extremes probably lie today, in the future and in the past, given that there is an extreme value today. By construction, the knowledge of the tail process or spectral process enables researchers and practitioners to know the probabilistic distribution of the spatio-temporal dependence between extremes.

In this thesis, the development of the tail and spectral process for heavy tailed functional time series is followed by three theoretical applications.

The first application is a characterization in terms of the tail and spectral processes only of a variety of indices and objects describing the extremal behavior of the series: the extremal index, tail dependence coefficients, the extremogram and the point process of extremes.

Another important limit object to which this theory applies is the point process of extremes. It counts the number of extremes in a given spatial area. The temporal dependence causes extremes to appear in clusters and changes the nature of the limit process from Poisson in the independent case to compound Poisson in the stationary case.

The second application is the computation of an explicit expression of the tail and spectral processes for heavy tailed *linear* functional time series. Equation (1.1) is a model of first order autoregressive functional time series.

The third and last application is the introduction and the study of a model for the spatio-temporal dependence called *maxima of moving maxima of continuous functions*, in short CM3. Some theoretical properties of this process are derived: the expression of its spectral process and the conditions under which CM3 processes satisfy the finite-cluster property and the strong mixing property. The last part of this study consists of an estimation methodology for CM3 processes, with extensive Monte Carlo simulations.

The contributions of this thesis will be described in detail in the begining of each chapter. To run over them here in a few words, the tail and spectral processes of a time series of random functions (infinite dimensional objects) are adaptations of these concepts that already existed for random vectors (finite dimensional objects). This adaptation has required different mathematical proofs. It should be mentioned that the structure developed here to tackle the problem in infinite dimensional spaces sheds new light even on the finite dimensional case. Afterward, the study of linear functional processes with heavy tailed innovations is new, with an expression of their spectral processes and the related steps to obtain these results. Finally, although CM3 processes generalize their finite dimensional equivalent, the properties and estimation methods derived here are new even in the finite dimensional case.

1.2 Pareto tails, regular variation

The construction leading to the concept of heavy tails is the following. Let Y be a random variable such that $P(Y > y) = 1/y^{\alpha}$, $y > 1$. Such a distribution is called Pareto(α). The parameter $\alpha > 0$ controls the density of high values: the closer to 0 it is, the likelier Y is to have large values or the heavier is its tail. If α belongs to $(0, 2]$ then the variance of Y does not exist. If furthermore α lies in $(0, 1]$ then the expected value of Y does not exist either. The concepts of first and second moments can thus not be assumed to exist with models that have heavy tails.

Pareto random variables are aimed at modeling the tail of a distribution. It is rare to be able to directly model a phenomenon with a Pareto random variable. To have an "*heavy tail*" only the tail of the model should be Pareto-distributed. Concretely, the general idea illustrated on a sample $x_1, \dots, x_n$ of n ordered values is the following. We look at the k largest values $x_{n-k+1}, \dots, x_n$, which form the upper part of the sample. We divide them by the threshold x_{n-k} to get a normalized sample $y_1 := x_{n-k+1}/x_{n-k}, \dots, y_k := x_n/x_{n-k}$ located above the value 1. It is this sample that we assume to be distributed as a Pareto(α) random variable.

This reasoning leads us to consider a wider class than the Pareto(α) distributions called the *regularly varying* distributions. The threshold

x_{n-k} above is replaced in theoretical settings by a value x and the problem to split the sample in two at a thresholding value is replaced by a limit as $x \to \infty$. Given a positive random variable X we say that the distribution of X is regularly varying if there exists $\alpha > 0$, called the index of regular variation, such that

$$\mathcal{L}\left(\frac{X}{x} \;\middle|\; X > x\right) \xrightarrow{d} \text{Pareto}(\alpha), \qquad x \to \infty, \tag{1.2}$$

where $\mathcal{L}(X|A)$ indicates the conditional law of X given the event A. This convergence in distribution can be rewritten here, for every $y > 1$,

$$P\left(\frac{X}{x} > y \;\middle|\; X > x\right) \to y^{-\alpha}, \qquad x \to \infty. \tag{1.3}$$

The function $f(x) := P(X > x)$ can thus still be $x^{-\alpha}$, $x > 1$, but may also be another function (starting from 1 and decreasing towards 0) that has the same kind of behavior at infinity as $x^{-\alpha}$; (1.3) is equivalent to

$$\lim_{x\to\infty} \frac{f(xy)}{f(x)} = y^{-\alpha}, \qquad y > 0.$$

This family of course includes the Pareto random variables but also, for instance, the Student, Fréchet and Cauchy random variables. For all those distributions, the properties of existence of expected value and variance – that only depend on the tail of the distribution – stay the same as in the Pareto case with respect to α.

1.3 Motivation chapter by chapter

Chapter 2 mainly has two goals. The first one is to extend the theory of regular variation to more general objects than random variables, such as random functions or random sequences for instance for particular applications. The general ambient space will be a Banach space $(\mathbb{B}, \|\cdot\|)$, that is, the set $\mathbb{B}$ equipped with the norm $\|\cdot\|$ is a complete vector space. Banach spaces provide a unique structure to host all these infinite dimensional objects as well as the finite dimensional ones. The second goal of Chapter 2 is to take into account in the modelization of extremes the temporal dependence that might exist by, introducing the concept of *joint regular variation*.

One more assumption that will be required is the separability of the Banach space $\mathbb{B}$, mainly for technical reasons. For instance, the Banach space of the continuous functions defined on a compact domain of $\mathbb{R}^n$ with the sup-norm is separable but, when the functions are defined on the whole space $\mathbb{R}^n$ and are assumed bounded, it is not.

To build a model for a heavy tailed random element X in a separable Banach space $\mathbb{B}$, (1.2) can be replaced by

$$\mathcal{L}\left(\frac{X}{x}\;\middle|\; \|X\| > x\right) \xrightarrow{d} \mathcal{L}(Y), \qquad x \to \infty, \tag{1.4}$$

where Y is a random element in $\mathbb{B}$ with $\|Y\| \sim \text{Pareto}(\alpha)$. This is equivalent to require that

$$\mathcal{L}\left(\frac{X}{\|X\|}\;\middle|\; \|X\| > x\right) \xrightarrow{d} \lambda, \qquad x \to \infty, \tag{1.5}$$

where λ is a probability measure on the unit sphere, called the *spectral measure* of X, indicating the propensity to observe extreme values in the different directions.

Thanks to this construction all available information about cross-sectional tail dependence is contained in the distribution of Y or in the measure λ. For instance, for a regularly varying random continuous function X, if there is an extreme value $X(a)$ at point a what can I know about the probability distribution of the value $X(b)$ of the field at point b? Let $\|X\|$ be the sup-norm of X. For $y > 0$ the following computation follows from (1.4) only and reflects a basic argument of this thesis. One readily checks that $P(Y(a) = y) = 0$ and $P(Y(b) = 1) = 0$ thanks to homogeneity property of Proposition 2.2.2. By definition of the sup-norm, $X(b) > x$ implies $\|X\| > x$, so that

$$\begin{aligned}
&\lim_{x\to\infty} P\left(X(a)/x > y \mid X(b) > x\right) \\
&= \lim_{x\to\infty} \frac{P\left(X(a)/x > y, X(b)/x > 1\right)}{P\left(X(b)/x > 1\right)} \\
&= \lim_{x\to\infty} \frac{P\left((X(a)/x, X(b)/x) \in (y,+\infty) \times (1,+\infty) \mid \|X\| > x\right)}{P\left(X(b)/x > 1 \mid \|X\| > x\right)} \\
&= \frac{P\left(Y(a) > y, Y(b) > 1\right)}{P\left(Y(b) > 1\right)} \\
&= P\left(Y(a) > y \mid Y(b) > 1\right).
\end{aligned}$$

To model the temporal dependence of a series $(X_t)_{t\in\mathbb{Z}}$, the first assumption made here is the stationarity of the series: the law of the joint vector $(X_{t-h},\dots,X_t)$ is the same for any ending time position t and any length h in the past. The concept of *joint* regular variation (over time) is then characterized by the existence of the tail and spectral processes of the original series. To take into account the temporal dependence in (1.4), the existence of Y with $\|Y\| \sim \text{Pareto}(\alpha)$ is replaced by the existence of a process $(Y_t)_{t\in\mathbb{Z}}$ with $\|Y_0\| \sim \text{Pareto}(\alpha)$ such that

$$\mathcal{L}\left(\frac{(X_t)_{t\in\mathbb{Z}}}{x} \;\middle|\; \|X_0\| \geqslant x \right) \xrightarrow{d} \mathcal{L}\left((Y_t)_{t\in\mathbb{Z}}\right), \qquad x\to\infty, \tag{1.6}$$

where the convergence in distribution holds in the proper product space. The central result of this chapter is that this limit process $(Y_t)_{t\in\mathbb{Z}}$, called *tail process*, necessarily admits a polar decomposition in two independent components: on the one hand, a real-valued radial component $\|Y_0\|$ whose distribution is Pareto(α); on the other hand, a sequence-valued component $(\Theta_t)_{t\in\mathbb{Z}}$ defined by $\Theta_t = Y_t/\|Y_0\|$, called the *spectral process*. The spectral process can alternatively be obtained through the limit

$$\mathcal{L}\left(\frac{(X_t)_{t\in\mathbb{Z}}}{\|X_0\|} \;\middle|\; \|X_0\| \geqslant x \right) \xrightarrow{d} \mathcal{L}\left((\Theta_t)_{t\in\mathbb{Z}}\right), \qquad x\to\infty, \tag{1.7}$$

with the same rescaling as in (1.5).

As an example, an important characteristic describing the tail dependence in a process is its *extremal index θ*. Under additional conditions, it admits the expression

$$\begin{aligned} \theta &= \lim_{t\to\infty}\lim_{x\to\infty} P(\max_{i=1,\dots,t}\|X_i\| \leqslant x \mid \|X_0\| > x) \\ &= P(\sup_{i\geqslant 1}\|Y_i\| \leqslant 1) \\ &= E[\sup_{i\geqslant 0}\|\Theta_i\|^\alpha - \sup_{i\geqslant 1}\|\Theta_i\|^\alpha], \end{aligned}$$

the last equalities coming from the properties of the spectral process.

To study the extremal dependence over time, a convenient plot playing the same role as the traditional autocorrelation function is the *extremogram*

$$\rho(h) := \lim_{x\to\infty} \frac{P\left(\|X_0\| > x, \|X_h\| > x\right)}{P(\|X_0\| > x)} = P(\|Y_0\| > 1, \|Y_h\| > 1),$$

whose expression in terms of the tail process is derived from (1.6).

Chapter 3 is devoted to the study of the limit distribution of the point process

$$N_n = \sum_{i=1}^{n} \delta_{X_i/a_n}$$

where $(X_t)_{t\in\mathbb{Z}}$ is a heavy tailed functional time series and $a_n \to \infty$ is a normalizing sequence such that $nP(\|X_0\| > a_n) \to 1$. It is an accurate tool to obtain pieces of information about extremes. If $n < +\infty$ the distribution of N_n is the full distribution of the number of values in a selected spatial area. Passing to the limit as $n \to \infty$ only leaves the values that the process can have "above its normal level", called here extreme values. If the X_i are independent and identically distributed, N_n converges in distribution to a Poisson process whose intensity measure is a radial, homogeneous extension of the distribution of Y_0 to the whole space. In the stationary case, temporal dependence causes extremes to appear in clusters. These clusters change the nature of the limit of N_n from Poisson to compound Poisson, the extremal index playing a major role in its distribution.

The theory of joint regular variation is applied to two particular models in this thesis. Chapter 4 deals with functional linear processes and Chapter 5 with maxima of moving maxima of continuous functions, that can be thought of as "max-linear" processes.

Particular linear processes are the well-known autoregressive moving average ARMA(p,q) time series

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where ε_t is the innovation of the process at time t (which is not assumed to be normally distributed here). For instance, if X_t is the level of a polluting gas at day t in the atmosphere, the quantities X_t and ε_t can be real numbers for the total pollution, vectors of dimension k for the pollution level at k measurement stations or functions to model the pollution level everywhere on a given area. By developing the expression of X_{t-i} in the formula, every ARMA(p,q) process can be seen as a moving average with infinitely many dependence coefficients:

$$X_t = \varepsilon_t + \sum_{i=1}^{\infty} \gamma_i \varepsilon_{t-i}.$$

In sets of vectors or functions the multiplication $\gamma_i \varepsilon_{t-i}$ can be replaced by another linear operation $T_i(\varepsilon_{t-i})$. These are the reasons to study *linear processes* which consist of the family of processes admitting the representation

$$X_t = \sum_{i=0}^{\infty} T_i(\varepsilon_{t-i}).$$

For instance the model (1.1) for the pollution over a city admits such a representation with $T_1(x) = K * x$, the convolution between the kernel K and the field x, and $T_i = T_1^i$ if $i \neq 1$.

In standard functional data analysis the innovations ε_t of linear processes are assumed to follow a Gaussian distribution. Here the normality of the innovations is replaced by a regularly varying law satisfying (1.4). The main result of Chapter 4 is an explicit expression of the spectral process of a linear process given the spectral measure of the innovations and the linear maps T_i. As a consequence, it yields for example an expression of the extremal index.

In Chapter 5 the processes of interest are maxima of moving maxima of continuous functions, in short CM3. If $a_i^{(j)}$ ($i \in \mathbb{Z}_+$, $j \in \mathbb{Z}$) are strictly positive, real, continuous functions defined on a compact $K \subset \mathbb{R}^d$ such that

$$\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} a_i^{(j)}(x) = 1,$$

a CM3 process is

$$X_t(x) = \sup_{j\in\mathbb{Z}} \sup_{i\geqslant 0} a_i^{(j)}(x) Z_{t-i}^{(j)}$$

where the innovations $Z_i^{(j)}$ are independent and identically distributed unit-Fréchet random variables. Replacing the double supremum by a sum would give the structure of a linear process with the multiplication as operator; this explains why these processes can be qualified as *max-linear*.

The fact that, given real numbers $\xi_i > 0$ such that $\xi_1 + \xi_2 + \ldots = 1$, the distribution of $\max(\xi_1 Z_1, \xi_2 Z_2, \ldots)$ stays unit-Fréchet implies that X_t still has unit-Fréchet margins. But the transformation from $(Z_t)_{t\in\mathbb{Z}}$ to $(X_t)_{t\in\mathbb{Z}}$ induces a dependence structure over time and in space. Extremes appear in temporal clusters and, at time t, a large value for X_t at location x causes large values at other locations. CM3 processes appear after a standardization of the margins of the process of interest into unit-Fréchet.

As it will be explained, CM3 processes are able to model a wide range of spatio-temporal dependence structures between extreme values.

The first part of Chapter 5 is a study of the general properties of CM3 processes, among which an expression of its spectral process is given in function of the parameter functions. It has a particularly simple form in terms of a discrete random variable. The second objective is to fit CM3 processes to samples with measurement errors. For that purpose several empirical methods to estimate the parameter functions are compared and suggested.

Chapter 2

Regularly varying time series in Banach spaces

Based on Meinguet and Segers (2010).

Abstract. When a spatial process is recorded over time and the observation at a given time instant is viewed as a point in a function space, the result is a time series taking values in a Banach space. To study the spatio-temporal extremal dynamics of such a time series, the latter is assumed to be jointly regularly varying. This assumption is shown to be equivalent to convergence in distribution of the rescaled time series conditionally on the event that at a given moment in time it is far away from the origin. Depending on the way of rescaling the limit is called the tail process or the spectral process. These processes provide convenient starting points to study, for instance, joint survival functions, tail dependence coefficients, extremograms, and extremal indices.

2.1 Introduction

A powerful way to model spatio-temporal phenomena is by means of time series of functional observations. For risk management purposes, the interest is often in the extremes of such processes; examples from the literature include sea levels along dikes in the Netherlands (de Haan and Lin 2001), windspeeds along the faces of a building (Davis and Mikosch 2008), and precipitation in the state of Colorado (Cooley, Nychka, and

Naveau 2007). In such cases, second moments cannot be assumed to exist, violating the basic assumption in standard functional data analysis based on the sequence of autocovariance operators (Bosq 2000, Pumo 1998).

While originally defined for univariate functions and random variables, the concept of regular variation has by now been defined and studied in quite abstract settings, including the one of stochastic processes (Giné, Hahn, and Vatan 1990, Hult and Lindskog 2005; 2006). As for random variables, regular variation provides the mathematical backbone for a coherent theory of extreme values. By considering our functional observations as points in a suitable function space, we are led to consider regularly varying time series taking values in separable Banach spaces.

The natural goal is to study the cross-sectional and the temporal dependence between extremes of time series of functional observations. In Section 2.3 the tail and spectral processes are developed. They are unifying objects containing all available information about this dependence structure. Precise examples will be given in Chapter 4 with heavy tailed linear time series and in Chapter 5 with maxima of moving maxima of continuous functions.

As in the finite-dimensional case in Basrak and Segers (2008), the joint regular variation is shown to be equivalent to the existence of the limit in distribution of the rescaled process

$$(X_t/x)_{t\in\mathbb{Z}} \text{ conditionally on } \|X_0\| > x \text{ as } x \to \infty$$

in the proper product space. This is Theorem 2.3.2. The limit in distribution, denoted $(Y_t)_{t\in\mathbb{Z}}$, is the *tail process*. According to Theorem 2.3.3, it admits a familiar-looking decomposition into independent radial and angular components. The radial component is fully determined by the index of regular variation of $\|X_0\|$, while the angular component, called the *spectral process*, effectively captures all aspects of extremal dependence: extremal indices, limits in distribution of the cluster process and the point processes of extremes etc. The spectral process $(\Theta_t)_{t\in\mathbb{Z}}$ is given by the limit in distribution

$$(X_t/\|X_0\|)_{t\in\mathbb{Z}} \text{ conditionally on } \|X_0\| > x \text{ as } x \to \infty.$$

Moreover the tail and spectral processes are uniquely determined by their future, the *forward tail process* $(Y_t)_{t\geqslant 0}$ and the *forward spectral process*

$(\Theta_t)_{t\geqslant 0}$. Theorem 2.3.6 is the *time change formula*, a property that describes how the distribution of the members of $(\Theta_t)_{t\in\mathbb{Z}}$ changes over time. Theorem 2.3.4 gives the distribution of the spectral process after a linear transformation between two Banach spaces.

Section 2.4 deals with the relationship between the tail and spectral processes and the features describing the extremal behavior of $(X_t)_{t\in\mathbb{Z}}$. The exponent and spectral measures, the coefficient of tail dependence, the extremal index of the univariate time series $(\|X_t\|)_{t\in\mathbb{Z}}$ and the extremogram are given in terms of $(Y_t)_{t\in\mathbb{Z}}$ or $(\Theta_t)_{t\in\mathbb{Z}}$ only. The advantage in comparison with their usual expressions is that a limit is skipped each time here, making the computations of these quantities easier. Example 2.3.5 illustrates this idea for GARCH(1,1) processes.

2.2 Regular variation

To begin with, we give a quick explanation about why we need completeness and separability assumptions in a theory of regular variation. They are indispensable for a couple of technical matters: they make Prohorov's theorem a biimplication, so that tightness is equivalent to weak convergence (Billingsley 1999); they make complete and separable the sets of measures to express vague convergence (Resnick 2008) or M_0-convergence (Hult and Lindskog 2006); they imply Kolmogorov's Extension theorem that guarantees that a suitably “consistent” collection of finite-dimensional distributions defines a stochastic process (Pollard 2002); and they allow to replace weak convergence by pointwise convergence of the Laplace transforms (Resnick 2007). Completeness and separability are always assumed to deal with random objects admitting regularly varying distributions. All the results cited here will be key steps in the proofs of the main assertions of this chapter.

On the contrary a classical assumption dropped here is the local compactness of the state space. Local compactness is a technical way to obtain limits, available if and only if the results of the experiment lie in a finite dimensional space. Since the goal of this thesis precisely is to consider function spaces and sequence spaces we cannot impose this hypothesis here. Some results previously obtained in extreme value theory without this assumption lie for instance in de Haan and Lin (2001) and de Haan and Ferreira (2006) for continuous functions. Here we consider a separable Banach space as ambient space. The results will be true for random functions on a compact subset of $\mathbb{R}^n$ with the metric of uniform convergence, for random sequences with a suitable norm and of course in every finite dimensional normed vector space.

The first building block for a theory of regularly varying random elements in separable Banach spaces is the property of regular variation of deterministic functions (Bingham, Goldie, and Teugels 1987). A measurable function $f : [a, \infty) \to \mathbb{R}_+$ is *regularly varying* if, for all $y > 0$, $f(xy)/f(x) \to g(y) > 0$ as $x \to \infty$. In this case $g(y) = y^\alpha$ for some $\alpha \in \mathbb{R}$ and we write $f \in RV_\alpha$. The extension to a positive random variable X is done by requiring the function $P(X > \cdot\,)$ to be in $RV_{-\alpha}$ for some $\alpha > 0$.

The fundamental relationship between regular variation and extreme value theory is that a random variable X lies in the domain of attraction of a Fréchet(α) distribution if and only if $P(X > \cdot\,)$ is regularly varying with index $-\alpha$. According to the definition of Fisher and Tippett (1928) and Gnedenko (1943), there are only three types of possible extreme values distributions: Fréchet, Gumbel and Weibull. Regularly varying random variables form thus the entire domain of attraction of the Fréchet extreme value distribution. No other distribution can be qualified as "heavy tailed". This is the reason why we only focus here on the regularly varying random objects, as in Davis and Hsing (1995) and in Davis and Mikosch (2008) for instance.

Regular variation of probability measures on Euclidean space is usually defined in terms of vague convergence of a sequence of Radon measures, defined on Euclidean space compactified at infinity and punctured at the origin (Resnick 2007). Since $\mathbb{B}$ is not assumed to be locally compact, such an approach does not work here. A possible way out is to replace vague convergence by $\hat{w}$-convergence described in Daley and Vere-Jones (1988). This, however, requires changing the metric on $\mathbb{B}$ and completing it at infinity (de Haan and Ferreira 2006, Davis and Mikosch 2008). These somewhat artificial steps are not needed in the equivalent approach of Hult and Lindskog (2006), which we follow here.

To define regular variation in a separable Banach space $\mathbb{B}$ we follow the approach of Hult and Lindskog (2006). For this purpose, set

$$\begin{aligned}
\mathcal{C}(\mathbb{B}) &:= \{f : \mathbb{B} \to \mathbb{R} \mid f \text{ is continuous}\}\,, \\
\mathcal{C}_b(\mathbb{B}) &:= \{f \in \mathcal{C} \mid f \text{ is bounded}\}\,, \\
\mathcal{C}_{bd}(\mathbb{B}) &:= \{f \in \mathcal{C}_b \mid f \text{ vanishes on an open disk containing } 0\}
\end{aligned}$$

and, if X is a topological space and $\mathcal{B}_X$ its Borel σ-field, i.e. the σ-field generated by the open sets of X,

$$\begin{aligned} M_b(X) &:= \{\mu \text{ is a measure on } \mathcal{B}_X \mid \mu(X) < +\infty\}\,, \\ M_0(\mathbb{B}) &:= \{\mu \in M_b(\mathbb{B} \setminus \{0\}) \mid \forall r > 0,\ \mu([\|x\| > r]) < +\infty\}\,. \end{aligned}$$

Recall that $\mu_n \to \mu$ with μ_n and μ in $M_b(\mathbb{B})$ if and only if, for every $f \in \mathcal{C}_b$, $\int f d\mu_n \to \int f d\mu$ (weak convergence, see for instance Billingsley (1999)). Similarly, $\mu_n \to \mu$ for μ_n and μ in $M_0(\mathbb{B})$ holds if, for every $f \in \mathcal{C}_{bd}$, $\int f d\mu_n \to \int f d\mu$. The topology on M_0 induced by this convergence is metrizable and makes M_0 a complete and separable metric space. There are versions in M_0 of the Portmanteau Lemma, Prohorov's theorem and the Continuous mapping theorem. If $\mathbb{B}$ is locally compact, for the multivariate theory, M_0-convergence reduces to vague convergence (Hult and Lindskog 2006).

We can now give the definition of a regularly varying element in $\mathbb{B}$.

Definition 2.2.1. *A random element X in $\mathbb{B}$ is* regularly varying *with index $\alpha > 0$ if there exists $\mu \in M_0(\mathbb{B})$ and $V(x) \in RV_{-\alpha}$ with $\mu \neq 0$ and*

$$\frac{P\,(X \in x\,\cdot\,)}{V(x)} \xrightarrow{M_0} \mu(\,\cdot\,) \qquad \text{as } x \to \infty. \tag{2.1}$$

The measure μ is called the exponent measure *of X.*

The name "exponent measure" for the limit measure μ in Banach spaces is set up to continue the tradition of its analogue in the multivariate case and the functional case (de Haan and Ferreira 2006, Section 9.3).

To derive first properties of the exponent measure, let $\operatorname{cl} A$ indicates the closure of $A \subset \mathbb{B}$, $\operatorname{int} A$ its interior and $\partial A = \operatorname{cl} A \setminus \operatorname{int} A$ its boundary.

Proposition 2.2.2. *The measure μ in* (2.1) *has the homogeneity property*

$$\mu(yA) = y^{-\alpha}\mu(A), \tag{2.2}$$

for every $y > 0$ and $A \in \mathcal{B}_{\mathbb{B}}$ such that $0 \notin \operatorname{cl} A$. Furthermore, for every $r > 0$, we have

$$\mu(\partial B(0, r)) = 0. \tag{2.3}$$

In particular μ has no atoms.

Proof. See page 39. □

A consequence of Definition 2.2.1 and Proposition 2.2.2 is that μ is unique up to a multiplicative constant depending on the choice of $V(x)$.

Definition 2.2.1 is equivalent to require that X has a spectral measure in following sense. Write $\mathbb{S} = \{x \in \mathbb{B} : \|x\| = 1\}$.

Proposition 2.2.3. *A random element X in $\mathbb{B}$ is regularly varying with index $\alpha > 0$ if and only if there exists a probability measure λ on $\mathbb{S}$ such that, for every $y > 0$,*

$$\frac{P\left(\|X\| > xy, X/\|X\| \in \cdot\,\right)}{P(\|X\| > x)} \xrightarrow{w} \frac{1}{y^\alpha}\lambda(\,\cdot\,) \qquad \text{as } x \to \infty. \tag{2.4}$$

Proof. See page 39. □

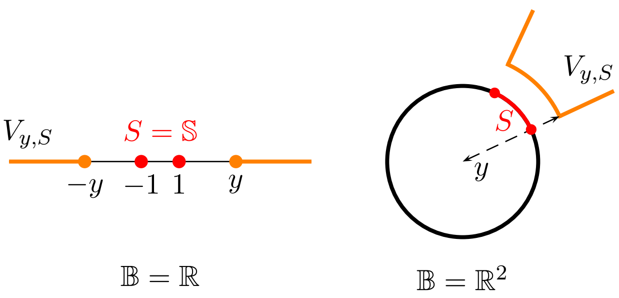


Figure 2.1: The set $V_{y,S} := \{x \in \mathbb{B} : \|x\| > y,\ x/\|x\| \in S\}$ in $\mathbb{B} = \mathbb{R}$ and $\mathbb{B} = \mathbb{R}^2$.

The measure λ is called the *spectral* (or *angular*) *measure* of X. Initially the existence of a spectral measure was the way to describe regular variation for random variables. The regular variation is spread in a balanced way on the two tails. A random variable $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is regularly varying with index $\alpha > 0$ if there exist two numbers $p \geqslant 0$ and $q \geqslant 0$ such that $p + q = 1$ and, for every $y > 0$,

$$\frac{P(|X| > xy, X > 0)}{P(|X| > x)} \to \frac{p}{y^\alpha} \quad \text{as } x \to \infty$$

and

$$\frac{P(|X| > xy, X < 0)}{P(|X| > x)} \to \frac{q}{y^\alpha} \quad \text{as } x \to \infty.$$

This is equivalent to the existence of a probability measure λ on $\{-1,+1\} = \partial B(0,1) \subset \mathbb{R}$ such that, for every $y > 0$,

$$\frac{P(|X| > xy, X/|X| \in \cdot)}{P(|X| > x)} \xrightarrow{w} \frac{1}{y^\alpha}\lambda(\cdot) \quad \text{as } x \to \infty.$$

The construction in Banach spaces generalizes this idea of a measure of the unit sphere.

It is obvious from the proof of Proposition 2.2.3 (page 22) that, calling

$$c := \mu(\{x \in \mathbb{B} : \|x\| > 1\}),$$

the spectral measure and the exponent measure are connected through

$$\lambda(\,\cdot\,) = \frac{\mu(\{x \in \mathbb{B} : \|x\| > 1,\, x/\|x\| \in \,\cdot\,\})}{c} \qquad \text{and} \qquad \mu(V_{y,S}) = \frac{c}{y^\alpha}\lambda(S),$$

where, given $S \in \mathbb{S}$ and $y > 0$, $V_{y,S} := \{x \in \mathbb{B} : \|x\| > y,\, x/\|x\| \in S\}$. In other words, as the sets $V_{y,S}$ form a π-system generating $\mathcal{B}_{\mathbb{B}_0}$, if $h_{\#}\mu$ indicates the pushforward of the measure μ by the map h, that is $h_{\#}\mu(A) = \mu(h^{-1}(A))$, we have

$$T_{\#}\mu = c(\nu \otimes \lambda) \tag{2.5}$$

with the polar decomposition

$$T : \mathbb{B}_0 \to (0,+\infty) \times \mathbb{S} : x \mapsto (\|x\|, x/\|x\|)$$

and $d\nu(r) = d(-r^{-\alpha})$ (i.e. $\nu((r,\infty)) = r^{-\alpha}$) for $r > 0$. Thanks to the identity $\int (f \circ h)\, d\mu = \int f\, d(h_{\#}\mu)$ applied with $h = T^{-1}$, (2.5) becomes

$$\int f d\mu = c \int_{\mathbb{S}} \int_0^\infty f(r\theta) d(-r^{-\alpha}) d\lambda(\theta) \tag{2.6}$$

for every μ-integrable $f : \mathbb{B}_0 \to \mathbb{R}$.

Remark 2.2.4. Proposition 2.2.3 implies that $V(x) = P(\|X\| > x)$ is regularly varying with index $-\alpha$ hence this choice suits for (2.1). An immediate result is that in this case $c = 1$ in (2.6).

There also is a common characterization of regular variation in terms of a normalizing sequence $(a_n)_{n\in\mathbb{N}}$ that holds here.

Proposition 2.2.5. *A random element X in a separable Banach space is regularly varying with index $\alpha > 0$ if there exists a probability measure λ on $\mathbb{S}$ and a sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n \to \infty$ and, for every $u > 0$,*

$$nP\left(\|X\| > ya_n, X/\|X\| \in \cdot\right) \xrightarrow{w} \frac{1}{y^\alpha}\lambda(\cdot) \quad \text{if } n \to \infty. \tag{2.7}$$

Proof. See page 43. □

If one of these definitions is fulfilled, it follows that the sequence $(a_n)_{n\in\mathbb{N}}$, the function $f(x) := P(\|X\| > x)$ and, for $S \in \mathbb{S}$ such that $\lambda(S) \neq 0$, the function $g(x) := P\left(\|X\| > x, X/\|X\| \in S\right)$ are regularly varying with index $-\alpha$.

For a characterization of regular variation with a normalizing sequence in terms of the convergence in M_0, the index of regular variation needs to be specified for instance as follows.

Proposition 2.2.6. *A random element X in a separable Banach space is regularly varying with index $\alpha > 0$ if there exists a measure $\mu \in M_0$ and a regularly varying sequence $(a_n)_{n\in\mathbb{N}}$ with index $1/\alpha$ such that*

$$nP\left(\frac{X}{a_n} \in \cdot\right) \xrightarrow{M_0} \mu(\cdot) \quad \text{as } n \to \infty. \tag{2.8}$$

Proof. See page 44. □

Note that if α is not specified is Definition 2.2.6 but if (2.8) holds for some sequence $(a_n)_{n\in\mathbb{N}}$ of positive numbers, it implies that $(a_n)_{n\in\mathbb{N}}$ is regularly varying.

2.3 Joint regular variation

Joint regular variation is a model for heavy tailed temporal dependent phenomena under the assumption of stationarity. This model essentially characterized by the tail process and the spectral process of the series. It was originally introduced for multivariate time series in Basrak and Segers (2008). The advantage to carry over this theory to separable Banach spaces is to possess a tool to study the cross-sectional tail dependence and the temporal tail dependence in series of functional observations. Functional data analysis has been recently developed involving the autocovariance operators as main tool, see for instance Allam and Mourid (2002), Antoniadisa and Sapatinas (2003), Bosq (1996), Mas (2002; 2007), Mas and Menneteau (2003), Mas and Pumo (2007), Pumo (1998). For heavy tailed phenomena the first and second moments cannot be assumed to exist.

In this section we first explain how the tail process is obtained, with the property that is future is uniquely determined by its past, and conversely. Then we give examples in simple cases, for an independent and a fully dependent time series. Next we derive the *time change formula*, a relation describing how the distribution of the tail and the spectral process changes over time. The time change formula is a property of the tail and spectral processes yielded by the stationarity of the initial process. The last result of this chapter is an expression what becomes the spectral process of a series after a linear transformation to another Banach space. It will be useful in Chapter 4 to apply this theory to functional linear time series.

A time series $(X_t)_{t\in\mathbb{Z}}$ with values in a Banach space $\mathbb{B}$ is said here to be *stationary* if it is stationary in the strong sense, that is if

$$\mathcal{L}(X_t,\dots,X_{t+h}) = \mathcal{L}(X_0,\dots,X_h)$$

for any time position t and any length h.

To treat the joint regular variation of a stationary time series $(X_t)_{t\in\mathbb{Z}}$ with values in $\mathbb{B}$, we need to operate with the k-fold Cartesian product $\mathbb{B}^k$ as a Banach space itself. For making $\mathbb{B}^k$ itself a separable Banach space, we endow $\mathbb{B}^k$ with the max-norm

$$\|(x_1,\dots,x_k)\| = \max(\|x_1\|,\dots,\|x_k\|). \tag{2.9}$$

Definition 2.3.1. *A stationary process $(X_t)_{t\in\mathbb{Z}}$ in a separable Banach space $\mathbb{B}$ is* jointly regularly varying *with index $\alpha > 0$ if, for every $k \geqslant 1$, the joint random element $(X_1,\dots,X_k)$ with values in $\mathbb{B}^k$ endowed with the max-norm is regularly varying with index $\alpha > 0$.*

If $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying, there is thus a sequence of exponent measures μ_k characterized by

$$\frac{P\left((X_1,\dots,X_k)\in x\,\cdot\right)}{V_k(x)} \xrightarrow{M_0} \mu_k(\,\cdot\,) \qquad \text{as } x\to\infty, \tag{2.10}$$

where $V_k \in RV_{-\alpha}$. As in the previous section, we may require that $V_k(x) = U_k(x)$ where $U_k(x) := P(\|(X_1, \dots, X_k)\| > x)$ to obtain $c_k := \mu_k(\{x \in \mathbb{B}^k : \|x\| > 1\}) = 1$ for every k. The choice $V_k(x) = W(x)$ where $W(x) := P(\|X_1\| > x)$ so that these functions do not depend on k is also possible. Indeed, on the one hand, by (2.9),

$$\begin{aligned} W(x) &= P(\|X_1\| > x) \\ &\leqslant P(\|X_1\| > x \text{ or } \dots \text{ or } \|X_k\| > x) \\ &= U_k(x) \\ &\leqslant P(\|X_1\| > x) + \dots + P(\|X_k\| > x) \\ &= kW(x), \end{aligned}$$

which implies $1 \leqslant \dfrac{U_k(x)}{W(x)} \leqslant k$. On the other hand, by (2.1),

$$\lim_{x\to\infty} \frac{W(x)}{U_k(x)} = \mu_k([\|x_1\| > 1]).$$

So, with $V_k(x) = W(x)$ in (2.10), we have $1/k \leqslant c_k \leqslant 1$.

To that degree the joint regular variation of a stationary process is characterized by a sequence of exponent or spectral measures, each of them in a different vector space. A natural goal is thus to obtain a more tractable condition. The *spectral process* answer this question as a unique object containing the whole spatio-temporal distribution of extremes of the original process.

The first step towards the spectral process is the equivalence between the joint regular variation and the existence of a tail process or, alternatively, a forward tail process.

Theorem 2.3.2 (Tail process)**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary process in a separable Banach space $\mathbb{B}$. The following are equivalent:*

(i) $(X_t)_{t\in\mathbb{Z}}$ *is jointly regularly varying with index $\alpha > 0$.*

(ii) There exists a process $(Y_n)_{n\in\mathbb{N}}$ in $\mathbb{B}$, called forward tail process, *with $P(\|Y_0\| > y) = y^{-\alpha}$ for $y \geqslant 1$ and, as $x \to \infty$,*

$$\mathcal{L}\left((X_n/x)_{n\in\mathbb{N}} \;\middle|\; \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left((Y_n)_{n\in\mathbb{N}}\right).$$

(iii) There exists a process $(Y_t)_{t\in\mathbb{Z}}$ in $\mathbb{B}$, called tail process, *with $P(\|Y_0\| > y) = y^{-\alpha}$ for $y \geqslant 1$ and, as $x \to \infty$,*

$$\mathcal{L}\left((X_t/x)_{t\in\mathbb{Z}} \;\middle|\; \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left((Y_t)_{t\in\mathbb{Z}}\right).$$

Proof. See page 44. □

The convergences in distribution hold in the proper product spaces $\mathbb{B}^{\mathbb{N}}$ and $\mathbb{B}^{\mathbb{Z}}$ respectively which is the convergence in distribution of all finite stretches made with components of the original random sequence.

The tail process admits a decomposition into two components. A radial part Y materializes the index of regular variation as a Pareto(α) variable. The resulting angular component $(\Theta_t)_{t\in\mathbb{Z}}$ is the *spectral process* and then captures all the others features of the joint regular variation. Furthermore the random elements Y and $(\Theta_t)_{t\in\mathbb{Z}}$ are independent of each other.

Theorem 2.3.3 (Spectral process). *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary process in a separable Banach space $\mathbb{B}$. Assume that $P(\|X_0\| > \cdot) \in RV_{-\alpha}$ for some $\alpha > 0$. The following are equivalent:*

(i) $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying with index $\alpha > 0$.

(ii) There exists a process $(\Theta_n)_{n\in\mathbb{N}}$ in $\mathbb{B}$, called forward spectral process, *such that, as $x \to \infty$,*

$$\mathcal{L}\left((X_n/\|X_0\|)_{n\in\mathbb{N}} \;\middle|\; \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left((\Theta_n)_{n\in\mathbb{N}}\right).$$

(iii) There exists a process $(\Theta_t)_{t\in\mathbb{Z}}$ in $\mathbb{B}$, called spectral process,*such that, as $x \to \infty$,*

$$\mathcal{L}\left((X_t/\|X_0\|)_{t\in\mathbb{Z}} \;\middle|\; \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left((\Theta_t)_{t\in\mathbb{Z}}\right).$$

In this case, the tail process $(Y_t)_{t\in\mathbb{Z}}$ of $(X_t)_{t\in\mathbb{Z}}$ is given by $(Y_t)_{t\in\mathbb{Z}} \overset{d}{=} (Y\Theta_t)_{t\in\mathbb{Z}}$, where Y is independent of $(\Theta_t)_{t\in\mathbb{Z}}$ and is Pareto(α)*-distributed.*

Proof. See page 57. □

A couple of remarks are helpful to better understand this spectral process. First observe that the law of Θ_0 is equal to the spectral measure λ. Indeed, the weak convergence of (iii) in Theorem 2.3.3 compounded with the projection $\pi_0((x_t)_{t\in\mathbb{Z}}) = x_0$ yields (2.4) with $y = 1$.

The second fact to be mentioned is that an independent and identically distributed time series $(X_t)_{t\in\mathbb{Z}}$ whose members are regularly varying with index $\alpha > 0$ and spectral measure λ is jointly regularly varying with tail and spectral processes

$$Y_t \overset{d}{=} Y\Theta_t \qquad \text{where} \qquad \mathcal{L}(\Theta_t) = \begin{cases} \lambda & \text{if } t = 0, \\ 0 & \text{if } t \neq 0, \end{cases}$$

and $Y \sim \text{Pareto}(\alpha)$. This follows from a direct computation.

In the full-dependence case, when all members of $(X_t)_{t\in\mathbb{Z}}$ are equal to a regularly varying element with index $\alpha > 0$, the series is jointly regularly varying, with tail and spectral processes

$$Y_t \overset{d}{=} Y\Theta_t \qquad \text{with} \qquad \Theta_t = \Theta_0$$

where $Y \sim \text{Pareto}(\alpha)$ and $\mathcal{L}(\Theta_0) = \lambda$. The spectral process of any time series lies in between these two "extreme" cases.

The next proposition describes the spectral process after a linear transformation of the type $\xi_t = T(X_t)$ between two separable Banach spaces. This proposition is designed to be helpful in Chapter 4 about linear time series.

Proposition 2.3.4 (Linear transformations)**.** *Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be two separable Banach spaces and let $T : \mathbb{B}_1 \to \mathbb{B}_2$ be linear and continuous. Assume that $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying in $\mathbb{B}_1$ with spectral process $(\Theta_t^X)_{t\in\mathbb{Z}}$. Define the time series $(\xi_t)_{t\in\mathbb{Z}}$ in $\mathbb{B}_2$ by $\xi_t = T(X_t)$. Then $(\xi_t)_{t\in\mathbb{Z}}$ is jointly regularly varying in $\mathbb{B}_2$ with spectral process $(\Theta_t^\xi)_{t\in\mathbb{Z}}$, whose law is determined by*

$$\begin{aligned} &E[f(\Theta_{-s}^\xi, \dots, \Theta_t^\xi)] \\ &\qquad = \frac{1}{E[\|T(\Theta_0)\|^\alpha]} E\left[f\left(\frac{T(\Theta_{-s}^X)}{\|T(\Theta_0^X)\|}, \dots, \frac{T(\Theta_t^X)}{\|T(\Theta_0^X)\|} \right) \|T(\Theta_0^X)\|^\alpha \right] \end{aligned}$$

for every $f : \mathbb{B}_2^{s+t+1} \to \mathbb{R}$ continuous and bounded.

Proof. See page 58. □

The same result also holds for the forward spectral process. Moreover, setting $s = t = 0$, this proposition provides a way to compute the spectral measure of a random element after linear transformation.

An example that formally enlightens many concepts of this chapter is the GARCH(1,1) process. Specific examples of jointly regularly varying time series in infinite dimensional Banach spaces are the object of the next chapters.

Example 2.3.5. A GARCH(1,1) process $(\xi_t)_{t\in\mathbb{Z}}$ is a real valued process solution of the equations

$$\begin{cases} \xi_t = \sigma_t Z_t, \\ \sigma_t^2 = \alpha_0 + \alpha_1 \xi_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{cases}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid $N(0,1)$ innovations. These can be rewritten

$$\underbrace{\begin{pmatrix} \xi_t^2 \\ \sigma_t^2 \end{pmatrix}}_{X_t} = \underbrace{\begin{pmatrix} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha_1 & \beta_1 \end{pmatrix}}_{A_t} \underbrace{\begin{pmatrix} \xi_{t-1}^2 \\ \sigma_{t-1}^2 \end{pmatrix}}_{X_{t-1}} + \underbrace{\begin{pmatrix} \alpha_0 Z_t^2 \\ \alpha_0 \end{pmatrix}}_{B_t}$$

so that they appear as the (multivariate) stochastic recurrence equation

$$X_t = A_t X_{t-1} + B_t. \tag{2.11}$$

Under certain conditions (2.11) possesses a regularly varying solution $(X_t)_{t\in\mathbb{Z}}$, see Basrak et al. (2002), Kesten (1973), Kallenberg (1986). Assume that, as $x \to \infty$,

$$\mathcal{L}(X_0/x \mid \|X_0\| \geqslant x) \xrightarrow{d} Y_0.$$

Then, conditionally on $\|X_0\| \geqslant x$, as $x \to \infty$,

$$\begin{aligned} X_1/x = A_1(X_0/x) + B_1/x &\xrightarrow{d} A_1 Y_0 = Y_1, \\ X_2/x = A_2(X_1/x) + B_2/x &\xrightarrow{d} A_2 Y_1 = Y_2, \\ \dots \\ X_t/x = A_t(X_{t-1}/x) + B_t/x &\xrightarrow{d} A_t Y_{t-1} = Y_t. \end{aligned}$$

Hence the tail process $(Y_t)_{t\in\mathbb{Z}}$ of $(X_t)_{t\in\mathbb{Z}}$ has the property

$$Y_t = A_t \dots A_2 A_1 Y_0$$

and the spectral process $(\Theta_t)_{t\in\mathbb{Z}}$ of $(X_t)_{t\in\mathbb{Z}}$ satisfies

$$\Theta_t = A_t \dots A_2 A_1 \Theta_0.$$

The spectral process of $(\xi_t^2)_{t\in\mathbb{Z}}$ can be recovered using Proposition 2.3.4, the projection on the first component being a linear map. Further details and simulations can be found in Basrak and Segers (2008).

As stated in Theorem 2.3.3, a first property of the joint regular variation and the spectral process is that they are uniquely determined by the forward spectral process only. The probabilistic future of the series is determined by its past and conversely. The *time change formula* precises how the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ changes over time.

Theorem 2.3.6 (Time change formula)**.** *Let $s, t \in \{0, 1, 2, \ldots\}$ and let $f : \mathbb{B}^{t+s+1} \to \mathbb{R}$ integrable with respect to the law of $(\Theta_{-s}, \ldots, \Theta_t)$ such that $f(0, \ldots) = 0$. Then*

$$E\left[f(\Theta_{-s}, \ldots, \Theta_t)\right] = E\left[f\left(\frac{\Theta_0}{\|\Theta_s\|}, \ldots, \frac{\Theta_{t+s}}{\|\Theta_s\|}\right)\|\Theta_s\|^\alpha 1\{\Theta_s \neq 0\}\right].$$

Proof. See page 59. □

A simple case occurs when f only depends on its first component and verifies $f(0) = 0$; then

$$E\left[f(\Theta_{-s})\right] = E\left[f\left(\frac{\Theta_0}{\|\Theta_s\|}\right)\|\Theta_s\|^\alpha 1\{\Theta_s \neq 0\}\right].$$

This gives an expression of the distribution of Θ_{-s} in function of the law of Θ_0 and Θ_s still valid for $s \in \mathbb{Z}$. In particular taking $f(x_{-s}) = \|x_{-s}\|^\alpha$ yields

$$E\left[\|\Theta_{-s}\|^\alpha\right] = P(\Theta_s \neq 0), \ \ s \in \mathbb{Z},$$

a useful relation.

Note that the time change formula obtained in Basrak and Segers (2008) in the multivariate case holds for f continuous and the proof here implies this formula for f integrable as well.

2.4 Tail dependence features

Numerous quantities and objects describing the asymptotic tail behavior of a jointly regularly time series $(X_t)_{t\in\mathbb{Z}}$ admit an expression involving the tail process $(Y_t)_{t\in\mathbb{Z}}$ or the spectral process $(\Theta_t)_{t\in\mathbb{Z}}$ only. We will limit the discussion here to the exponent and spectral measures, the coefficients of tail dependence, the extremal indices and the extremograms.

Most of those quantities or objects are originally defined by means of a double limit. The knowledge of the tail process and the spectral process reduces the number of limits in their expressions to one. This is particularly efficient in the case where the tail process or the spectral process can be easily computed, as in Example 2.3.5.

Table 2.1 contains the tail quantities considered here and their expression if the series is jointly regularly varying. The notation $\mu \llcorner A$ indicates the restriction of μ to A.

Object	Reformulation
Exponent measure μ of X_i $\mu \llcorner [\|x\| > 1]$	$\mathcal{L}(Y_0)$
Spectral measure λ of X_i λ	$\mathcal{L}(\Theta_0)$
Exponent measure $\mu_{s,t}$ of $(X_{-s}, \ldots, X_t)$ $\mu_{s,t} \llcorner [\|x_0\| > 1]$	$\mathcal{L}(Y_{-s}, \ldots, Y_t)$
Coefficient of tail dependence $\lambda_t = \lim_{x\to\infty} P(\|X_t\| > x \mid \|X_0\| > x)$	$P(\|Y_t\| > 1)$ $= E[\min(\|\Theta_t\|^\alpha, 1)]$
Candidate extremal index $\theta^* =$ $\lim_{t\to\infty} \lim_{x\to\infty} P(\max_{i=1,\ldots,t} \|X_i\| \leqslant x \mid \|X_0\| > x)$	$P(\sup_{i\geqslant 1} \|Y_i\| \leqslant 1)$ $= E[\sup_{i\geqslant 0} \|\Theta_i\|^\alpha - \sup_{i\geqslant 1} \|\Theta_i\|^\alpha]$
Extremogram $\rho_{A,B}(h) = \lim_{n\to\infty} nP\left(\frac{X_0}{a_n} \in A, \frac{X_h}{a_n} \in B\right)$	$P(Y_0 \in A, Y_h \in B)$

Table 2.1: Expressions of the main objects describing the tail behavior of a stationary series involving the tail and spectral processes only

The coefficients of tail dependence λ_t give the probability to have an extreme value at time t given that there is an extreme value today. The extremal index θ was introduced in Leadbetter (1983). The candidate extremal index θ^* in Table 2.1 is equal to θ if conditions (C) and $(\mathcal{A}(a_n))$ of Section 3.2 hold as explained in Example 3.2.3. Under condition (C), $1/\theta^*$ is the expected size of clusters of extremes (Section 3.6).

The extremogram was introduced in Davis and Mikosch (2009) as a correlogram for extreme events through

$$\rho_{A,B}(h) = \lim_{n\to\infty} nP\left(\frac{X_0}{a_n} \in A, \frac{X_h}{a_n} \in B\right),$$

for integer h and for regions A, B at least one of which stays away from the origin and where a_n is a positive sequence satisfying $nP(\|X_0\| >$

$a_n) \to 1$ as $n \to \infty$. The result in Table 2.1 is easy to see if A and B are continuity sets of the distributions of Y_0 and Y_h respectively and if at least one of them is included in $\{x \in \mathbb{B} : \|x\| > 1\}$.

Example 2.4.1. Figure 2.2 shows[1] two empirical extremograms from Davis and Mikosch (2009) for a simulated GARCH(1,1) process with parameters $\alpha_0 = 1.0e^{-6}$, $\alpha_1 = 0.92$, $\beta_1 = 0.07$, sample size $n = 100,000$ and $A = B = (1, \infty)$. The empirical extremogram admits the expression

$$\hat{\rho}_{A,B}(h) = \frac{\sum_{t=1}^{n-h} 1\{a_m^{-1} X_t \in A, a_m^{-1} X_{t+h} \in B\}}{\sum_{t=1}^{n} 1\{a_m^{-1} X_t \in A\}}$$

where a_m is a large empirical quantile of the sample. Here a_m was taken to be equal to $x_{n-k+1,n}$ for two different values of k.

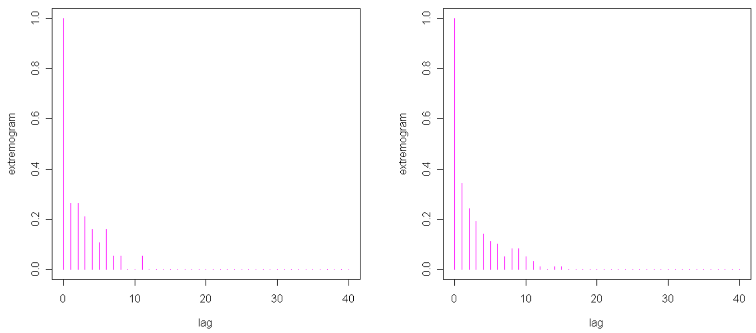


Figure 2.2: Extremogram of the GARCH(1,1) of Example 2.4.1 with $k = 20$ on the left and $k = 100$ on the right

According to its authors, the extremograms can be used for instance to detect the behavior of a GARCH(1,1) or an AR(1) process in a dataset.

2.5 Technical tools

2.5.1 Convergence of measures in M_0

The reference for the results not proved here is Hult and Lindskog (2006).

[1] Plots made using Ivor Cribben's software.

Let (X, d) is a complete separable metric space, $\mathcal{B}_X$ is its Borel σ-field and x_0 is a point of X that may be called the *origin*.

Definition 2.5.1. *For $x_0 \in X$, we write $X_0 = X \setminus \{x_0\}$. The topology on X_0 is the topology of X restricted to $X \setminus \{x_0\}$ and the Borel σ-field of X_0 is $\mathcal{B}_{X_0} = \{A \in \mathcal{B}_X \mid A \subset X_0\}$.*

Definition 2.5.2. *Let*

$$\begin{aligned}
\mathcal{C}(X) &:= \{f : X \to \mathbb{R} \mid f \text{ is continuous}\}, \\
\mathcal{C}_b(X) &:= \{f \in \mathcal{C}(X) \mid f \text{ is bounded}\}, \\
\mathcal{C}_{bd}(X) &:= \{f \in \mathcal{C}(X) \mid f \text{ is bounded and } (\exists r > 0) B(x_0, r)) \subset f^{-1}(\{0\})\}
\end{aligned}$$

where b states for "bounded" and d for "disk". The reference to X may be omitted if the context is clear.

Definition 2.5.3. *Given an origin x_0, we call M_0, $M_0(X)$ or $M_0(X, x_0)$ depending on the context, the set of all measures μ on $\mathcal{B}_{X_0}$ such that $\mu(X_0 \setminus B(x_0, r)) < +\infty$ for every $r > 0$.*

Definition 2.5.4. *Let $(\mu_n)_{n\in\mathbb{N}} \subset M_0$ and $\mu \in M_0$. We say that (μ_n) converges in M_0 to μ and we write*

$$\mu_n \xrightarrow{M_0} \mu$$

if $\int f d\mu_n \to \int f d\mu$ for every $f \in \mathcal{C}_{bd}$.

Definition 2.5.5. *For $\mu \in M_0$, $\mu^{(r)}$ is defined as the restriction of μ to $X_0 \setminus B(x_0, r)$.*

Theorem 2.5.6. *Let $(\mu_n)_{n\in\mathbb{N}} \subset M_0$ and $\mu_n \in M_0$. Then*

1. *If $\mu_n \xrightarrow{M_0} \mu$, then $\mu_n^{(r)} \xrightarrow{w} \mu^{(r)}$ in $X_0 \setminus B(x_0, r)$ for every $r > 0$ except possibly for an at most countable number of them.*
2. *If there exists $(r_k)_{k\in\mathbb{N}}$ with $r_k \downarrow 0$ such that $\mu_n^{(r_k)} \xrightarrow{w} \mu^{(r_k)}$ for every k in $X_0 \setminus B(x_0, r_k)$, then $\mu_n \xrightarrow{M_0} \mu$.*

Proposition 2.5.7. *Let $r > 0$. Consider $\mu, \nu \in M_0$. The map*

$$d_{M_0}(\mu, \nu) = \int_0^\infty e^{-r} d_r(\mu^{(r)}, \nu^{(r)})[1 + d_r(\mu^{(r)}, \nu^{(r)})]^{-1} dr,$$

where $\mu^{(r)}$ and $\nu^{(r)}$ are respectively the restrictions of μ and ν to $X_0 \setminus B(x_0, r)$ (for which, by definition, the total mass is finite) and d_r is the Prohorov's distance between the restrictions, is a distance on M_0.

Proposition 2.5.8. *The distance d_{M_0} metrizes the topology of M_0-convergence, i.e.*

$$\mu_n \xrightarrow{M_0} \mu \quad \textit{if and only if} \quad \lim_{n\to\infty} d_{M_0}(\mu_n, \mu) = 0.$$

Theorem 2.5.9. *The metric space (M_0, d_{M_0}) is complete and separable.*

This follows from the fact that (X, d) is complete and separable.

Theorem 2.5.10 (Portmanteau)**.** *Let $(\mu_n)_{n\in\mathbb{N}} \subset M_0$ and $\mu_n \in M_0$. The following condition are equivalent:*

1. $\mu_n \xrightarrow{M_0} \mu$,
2. $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for every* $A \in \mathcal{B}_X$ *with* $\mu(\partial A) = 0$ *and* $x_0 \notin \operatorname{cl} A$,
3. $\limsup_{n\to\infty} \mu_n(X_n \in C) \leqslant \mu(X \in C)$ *for every closed* $C \in \mathcal{B}_X$ *such that* $x_0 \notin C$ *and* $\liminf_{n\to\infty} \mu_n(X_n \in O) \geqslant \mu(X \in O)$ *for every open* $O \in \mathcal{B}_X$ *such that* $x_0 \notin \overline{O}$.

From these theorems we derive the characterizations of the relatively compact parts of M_0.

Theorem 2.5.11. *A subset $M \subset M_0$ is relatively compact if and only if there exists a sequence $(r_k)_{k\in\mathbb{N}}$ with $r_k \downarrow 0$ such that for every $k \in \mathbb{N}$ the set $M^{(r_k)} = \left\{\mu^{(r_k)} | \mu \in M\right\}$ is relatively compact for the topology of weak convergence in $X \setminus B(x_0, r_k)$.*

Prohorov's theorem yields the following equivalent statement.

Theorem 2.5.12. *A subset $M \subset M_0$ is relatively compact if and only if there exists a sequence $(r_k)_{k\in\mathbb{N}}$ with $r_k \downarrow 0$ such that for every $i \in \mathbb{N}$*

$$\sup_{\mu\in M} \mu(X \setminus B(x_0, r_i)) < +\infty$$

and, for every $\eta > 0$ there exists a compact $K_i \subset X \setminus B(x_0, r_i)$ such that

$$\sup_{\mu\in M} \mu(X \setminus (B(x_0, r_i) \cup K_i)) \leqslant \eta.$$

As a note, Theorem 2.5.12 applied to $X = \mathbb{R}^n$ becomes the following.

Theorem 2.5.13. *A subset $M \subset M_0(\mathbb{R}^n, x_0)$ is relatively compact if and only if for every $r > 0$ the following conditions*

a) $\sup_{\mu \in M} \mu(\mathbb{R}^n \setminus B(x_0, r)) < +\infty$,

b) $\lim_{R \to +\infty} \sup_{\mu \in M} \mu(\mathbb{R}^n \setminus B(x_0, R)) = 0$,

are fulfilled.

And Applying Theorem 2.5.12 to sets of continuous functions, Arzéla-Ascoli theorem gives the following characterization.

Theorem 2.5.14. *Let $\mathcal{C} = \mathcal{C}\left([0,1]^d, \mathbb{R}^q\right)$. A subset $M \subset M_0(\mathcal{C}, f_0)$ is relatively compact if and only if for every $r > 0$ and every $\varepsilon > 0$ the following conditions are fulfilled*

a) $\sup_{\mu \in M} \mu(\mathcal{C} \setminus B(f_0, r)) < +\infty$,

b) $\lim_{R \to +\infty} \sup_{\mu \in M} \mu(\{f \in \mathcal{C} \mid \|f(0)\|_\infty > R\}) = 0$,

c) $\lim_{\delta \to 0} \sup_{\mu \in M} \mu(\{f \in \mathcal{C} \mid \omega_f(\delta) \geqslant \varepsilon\}) = 0$,

where ω_f is the modulus of continuity of f.

Recall that a family $\Pi \subset X$ is a π-system if it is stable for finite intersections. The famous theorem about π-systems holds in M_0.

Theorem 2.5.15. *Let $\mathcal{A}$ be a π-system that generates $\mathcal{B}_{X_0}$. Then if two measures μ and $\nu \in M_0$ are σ-finite on $\mathcal{A}$ and equal on $\mathcal{A}$, they are equal on $\mathcal{B}_{X_0}$.*

The following proposition is useful to verify the hypotheses of Theorem 2.5.15.

Proposition 2.5.16. *Let $(X, \mathcal{O})$ be a separable topological space and $\mathcal{A}$ be a collection of Borel sets of X such that, for every $x \in X$ and every $r > 0$ there exists $A \in \mathcal{A}$ such that $x \in \operatorname{int} A$ and $A \subset B(x, r)$. Then $\sigma(\mathcal{A}) = \mathcal{B}_X$.*

Proof. $\sigma(\mathcal{A}) \subset \mathcal{B}_X$. The members of A are Borel sets.

$\sigma(\mathcal{A}) \supseteq \mathcal{B}_X$. Let $G \in \mathcal{O}$ be an open set. We will show that G is a union of elements of $\mathcal{A}$. Put $\mathcal{A}_0 = \{\operatorname{int} A \mid A \in \mathcal{A} \text{ and } A \subset G\}$. By assumption, for every $x \in G$ and every open set $V \subset G$ with $x \in V$, there exists $A_0 \in \mathcal{A}_0$ such that $x \in A_0 \subset V$. The family $\mathcal{A}_0$ is thus an open basis of the topological space X restricted to G. Since X is separable and G is open, X restricted to G is also separable. Consequently, as G is open in its restriction, (i) there exists a countable collection $(A_n)_{n\in\mathbb{N}}$ with $A_n \subset G$ for every $n \in \mathbb{N}$ and (ii) such that $G = \bigcup_{n\in\mathbb{N}} \operatorname{int} A_n$. Then $\bigcup_{n\in\mathbb{N}} A_n \subset G$ according to (i) and $G \subset \bigcup_{n\in\mathbb{N}} A_n$ according to (ii). □

There are useful characterizations of relative compactness in M_0 and of tightness of a sequence of random measures in M_0 after a reduction in $\mathbb{R}$.

Proposition 2.5.17. *A subset $M \subset M_0$ is relatively compact if and only if for every $f \in \mathcal{C}_{bd}^+$*

$$\sup_{\mu\in M} \mu(f) < +\infty.$$

Proof. ($\Rightarrow$) Let $f \in \mathcal{C}_{bd}^+$ and let $T_f : M_0 \to \mathbb{R}_+ : \mu \mapsto \mu(f)$. This application is continuous because by definition $\mu_n \xrightarrow{M_0} \mu$ implies $\mu_n(f) \to \mu(f)$. Thus by continuity and by assumption we have

$$\sup_{\mu\in M} \mu(f) = \sup_{\mu\in\overline{M}} T_f(\mu) < +\infty.$$

($\Leftarrow$) For $f \in \mathcal{C}_{bd}^+$, define $I_f = [0, \sup_{\mu\in\overline{M}} \mu(f)]$. Then

$$I := \prod_{f\in\mathcal{C}_{bd}^+} I_f \subset \mathbb{R}^{\mathcal{C}_{bd}^+}$$

is compact by assumption and by Tychonoff's theorem. One identifies $\mu \in M_0$ to its values $\{\mu(f) \mid f \in \mathcal{C}_{bd}^+\}$. This defines an application $T : \overline{M} \to I$. The topologies on $\overline{M}$ and $T(\overline{M})$ coincide since on the one hand $\mu_n \xrightarrow{M_0} \mu$ if and only if $\mu_n(f) \to \mu(f)$ and on the other hand the topology on $T(\overline{M})$ is given by componentwise convergence. Since $\overline{M}$ is closed and T is continuous, $T(\overline{M}) \subset I$ is closed and thus compact. □

Proposition 2.5.18. *A sequence $(\mu_n)_{n\in\mathbb{N}}$ of random measures in M_0 is tight if and only if for every $f \in \mathcal{C}_{bd}^+$ the sequence of random variables $(\mu_n(f))_{n\in\mathbb{N}}$ is tight in $\mathbb{R}$.*

Proof. ($\Rightarrow$) Let $f \in \mathcal{C}_{bd}^+$. We have to show that each subsequence $(\mu_{n(j)}(f))_{j\in\mathbb{N}}$ of $(\mu_n(f))_{n\in\mathbb{N}}$ contains a further subsequence that converges weakly. This is the case because $(\mu_n)_{n\in\mathbb{N}}$ is tight. As the map $M_0 \to \mathbb{R} : \mu \mapsto \mu(f)$ is continuous, it follows from the Continuous mapping theorem (Resnick 2008, p 152) that $(\mu_{n(j(k))}(f))_{k\in\mathbb{N}}$ possesses a weak limit in $\mathbb{R}$ as $k \to \infty$.

($\Leftarrow$) Let $\varepsilon > 0$. Let $g_i \in \mathcal{C}_{bd}^+$ be such that $g_i \uparrow 1_{\mathbb{B}_0}$. Since $(\mu_n(g_i))_{n\in\mathbb{N}}$ is tight, by assumption for every $i \geqslant 0$ there exists $c_i > 0$ such that, for every $n \in \mathbb{N}$,

$$P(\mu_n(g_i) > c_i) \leqslant \frac{\varepsilon}{2^{i+1}}.$$

The set

$$M = \bigcap_{i\geqslant 0} \{\mu \in M_0 \mid \mu(g_i) \leqslant c_i\}$$

is relatively compact in M_0 by Proposition 2.5.17. Indeed for each $f \in \mathcal{C}_{bd}^+$ there exists i_0 and K_0 such that $f \leqslant K_0 g_{i_0}$, so that

$$\sup_{\mu\in M} \mu(f) \leqslant K_0 \sup_{\mu\in M} \mu(g_{i_0}) \leqslant K_0 c_{i_0}.$$

Therefore $\overline{M}$ is compact and, for every $n \in \mathbb{N}$,

$$\begin{array}{l|l} P(\mu_n \notin \overline{M}) & \leqslant \sum_{i\geqslant 0} P(\mu_n(g_i) > c_i) \\ \leqslant P(\mu_n \notin M) & \leqslant \sum_{i\geqslant 0} \frac{\varepsilon}{2^{i+1}} \\ = P(\bigcup_{i\geqslant 0} \{\mu_n(g_i) > c_i\}) & = \varepsilon \end{array}$$

which is the definition of tightness. $\square$

In the sequel, do not confuse the relative compactness in M_0 with weak relative compactness for random measures in M_0. The same holds for the Continuous mapping theorem for convergence in M_0 (Hult and Lindskog 2006, Theorem 2.5) and the Continuous mapping theorem for weak convergence in metric spaces (Resnick 2008, p 152).

2.5.2 Lemmas

These two lemmas are useful to prove the Tail and Spectral theorems.

Lemma 2.5.19. *Let X and Y be two independent random variables such that X has a continuous law. Then conditionally on $Y \neq 0$ the random variable XY has a continuous law.*

Proof. Let $g : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Let $F(y)$ denote the cumulative distribution function of Y conditionally on $Y \neq 0$. Then, setting $z = xy$, by Fubini's theorem,

$$\begin{aligned}
&E\left[g(XY) \mid Y \neq 0\right] \\
&= \int_{\mathbb{R}} E\left[g(xY) \mid Y \neq 0\right] f_X(x)dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}\setminus 0} g(xy)dF(y) f_X(x)dx \\
&= \int_{\mathbb{R}} g(z) \left(\int_{\mathbb{R}\setminus 0} f_X(z/y)1/y \; dF(y) \right) dz,
\end{aligned}$$

which means that XY conditionally on $Y \neq 0$ has a density. □

For the second, recall that $\mu \llcorner E$ indicates the restriction of μ to E.

Lemma 2.5.20. *Let $(\mu_t)_{t \geqslant t_0}$ and μ be measures in $M_0(\mathbb{B})$ such that $\mu_t \xrightarrow{M_0} \mu$. Let $E \in \mathcal{B}_{\mathbb{B}}$ be such that $\mu(\partial E) = 0$ and $0 \notin \overline{E}$. Then, for every $f : E \to \mathbb{R}$ continuous and bounded, we have $\int f d(\mu_t \llcorner E) \to \int f d(\mu \llcorner E)$.*

Proof. By the Portmanteau Lemma, the conclusion is equivalent to

$$\mu_t \llcorner E \xrightarrow{d} \mu \llcorner E,$$

which is in turn equivalent to

$$(\forall A \in \mathcal{B}_E : \mu \llcorner E(\partial A) = 0)\, \mu_t \llcorner E(A) \to \mu \llcorner E(A).$$

So take $A \in \mathcal{B}_E$ be such that $\mu \llcorner E(\partial A) = 0$, i.e. $\mu(E \cap \partial A) = 0$. Since $\overline{E \cap A} \subset \overline{E} \not\ni 0$, the convergence holds if $\mu(\partial(E \cap A)) = 0$. However

$$\begin{aligned}
0 &\leqslant \mu(\partial(E \cap A)) \leqslant \mu(\partial E \cup \partial A) \leqslant \mu(\partial E) + \mu(\partial A) = \mu(\partial A) \\
&= \mu(\partial A \cap E) + \mu(\partial A \setminus E) = \mu(\partial A \setminus E) \leqslant \mu(\partial E) = 0,
\end{aligned}$$

which proves the lemma. The last inequality comes from the fact that, since $A \subset E$, we have $\partial A \subset \overline{E}$ hence $\partial A \setminus E \subset \partial E$. □

2.6 Proofs for Section 2.2

Proof of Proposition 2.2.2 (page 21). (2.1) implies (2.2) and (2.3). Let

$$\mu_x = \frac{1}{V(x)} P(X \in x\cdot).$$

Since $\mu_x \xrightarrow{M_0} \mu$, on the one hand, for every $f \in \mathcal{C}_{bd}$, $\int f d\mu_x \to \int f d\mu$. On the other hand, for every $y > 0$, $f(\frac{\cdot}{y}) \in \mathcal{C}_{bd}$. If $f \equiv 0$ on $B(0,r)$, then $f(\frac{\cdot}{y}) \equiv 0$ on $B(0,yr)$ and $f(\frac{\cdot}{y})$ is also continuous. As a consequence, for every $y > 0$,

$$\begin{aligned}
&\int f\left(\frac{\cdot}{y}\right) d\mu_x \\
&= \frac{1}{V(x)} E\left[f\left(\frac{X}{xy}\right)\right] \\
&= \frac{V(xy)}{V(x)} \frac{1}{V(xy)} E\left[f\left(\frac{X}{xy}\right)\right] \\
&= \frac{V(xy)}{V(x)} \frac{1}{V(xy)} \int f d\mu_{xy} \\
&\to \frac{1}{y^\alpha} \int f d\mu, \quad x \to \infty.
\end{aligned}$$

Thus, setting $\mu_1(f) := \int f(\frac{\cdot}{y}) d\mu$ and $\mu_2(f) := y^{-\alpha} \int f d\mu$, we have $\mu_1 = \mu_2$ which means that, for every $y > 0$ and every $A \in \mathcal{B}_{\mathbb{B}}$ such that $0 \notin \operatorname{cl} A$, $\mu(yA) = y^{-\alpha}\mu(A)$.

It follows that, for every $r > 0$, we have $\mu(\partial B(0,r)) = 0$, otherwise the function $y \mapsto \mu(\mathbb{B} \setminus B(0,y)) = y^{-\alpha}\mu(\mathbb{B} \setminus B(0,1))$ would be discontinuous at $y = r$ (and $\mu(\mathbb{B} \setminus B(0,1)) \neq 0$ because if not (2.2) would imply $\mu = 0$). This implies that μ has no atom. □

Proof of Proposition 2.2.3 (page 22). (2.1) implies (2.4). For $y > 0$ and $S \subset \mathbb{S}$, set

$$V_{y,S} = \{x \in \mathbb{B} \mid x/\|x\| \in S \text{ and } \|x\| > y\}.$$

For $S \in \mathcal{B}_{\mathbb{S}}$ define

$$\lambda(S) = \frac{\mu(V_{1,S})}{\mu(V_{1,\mathbb{S}})}.$$

It is a probability measure on $\mathbb{S}$ since

$$\lambda(\mathbb{S}) = \frac{\mu(V_{1,\mathbb{S}})}{\mu(V_{1,\mathbb{S}})} = \frac{\mu(V_{1,\mathbb{S}})}{\mu(V_{1,\mathbb{S}})} = 1.$$

Let $y>0$ and let $S \in \mathcal{B}_{\mathbb{S}}$ such that $\lambda(\partial S)=0$. We have $\mu(\partial V_{y,S})=0$. Indeed, for $A \times B$ in a general Cartesian product $X \times Y$, we have

$$\begin{aligned}
\partial(A \times B) &= \operatorname{cl}(A \times B) \setminus \operatorname{int}(A \times B) \\
&= (\operatorname{cl} A \times \operatorname{cl} B) \setminus (\operatorname{int} A \times \operatorname{int} B) \\
&= (\operatorname{cl} A \setminus \operatorname{int} A) \times \operatorname{cl} B \cup \operatorname{cl} A \times (\operatorname{cl} B \setminus \operatorname{int} B) \\
&= \partial A \times \operatorname{cl} B \cup \operatorname{cl} A \times \partial B,
\end{aligned}$$

which implies

$$\begin{aligned}
\partial V_{y,S} &\cong \partial((y,\infty) \times S) \\
&= \partial(y,\infty) \times \operatorname{cl} S \cup \operatorname{cl}(y,\infty) \times \partial S \\
&= \{y\} \times \operatorname{cl} S \cup [y,\infty) \times \partial S \\
&\subset \{y\} \times \mathbb{S} \ \cup \ (y,\infty) \times \partial S \\
&\cong \partial B(0,y) \cup V_{y,\partial S}.
\end{aligned}$$

As a consequence, according to (2.2) and (2.3),

$$\mu(\partial V_{y,S}) \leqslant \mu(\partial B(0,y)) + \mu(V_{y,\partial S}) = 0 + y^{-\alpha}\lambda(\partial S) = 0.$$

This implies

$$\begin{array}{l|l}
\lim\limits_{x\to\infty} \dfrac{P\left(\|X\|>xy, X/\|X\| \in S\right)}{P(\|X\|>x)} & = \dfrac{1}{\mu(V_{1,\mathbb{S}})}\mu(V_{y,S}) \\
= \lim\limits_{x\to\infty} \dfrac{P\left(X \in V_{xy,S}\right)}{P(X \in xV_{1,\mathbb{S}})} & = \dfrac{1}{y^\alpha}\dfrac{1}{\mu(V_{1,\mathbb{S}})}\mu(V_{1,S}) \\
= \lim\limits_{x\to\infty} \dfrac{P\left(X \in xV_{y,S}\right)}{P(X \in xV_{1,\mathbb{S}})} & = \dfrac{1}{y^\alpha}\lambda(S) \\
= \lim\limits_{x\to\infty} \dfrac{V(x)}{P(X \in xV_{1,\mathbb{S}})}\dfrac{P\left(X \in xV_{y,S}\right)}{V(x)} &
\end{array}$$

which yields the weak convergence.

(2.4) implies (2.1). For $y>0$ and $S \subset \mathbb{S}$, set

$$V_{y,S} = \{x \in \mathbb{B} \mid x/\|x\| \in S \text{ and } \|x\|>y\}.$$

There exists an unique measure $\mu \in M_0$ such that

$$\mu(V_{y,S}) = \frac{1}{y^\alpha}\lambda(S).$$

To show this, the idea developed below is to work in the Cartesian product $(0,\infty)\times\mathbb{S}$ and to use the homeomorphism

$$R:(0,\infty)\times\mathbb{S}\to\mathbb{B}_0:(r,z)\mapsto rz$$

to obtain the required homogeneity property. As

$$\{V_{y,S}\mid y>0 \text{ and } S\in\mathcal{B}_{\mathbb{S}}\}$$

is a π-system generating $\mathcal{B}_{\mathbb{B}_0}$, this measure will be unique.

Let $V(x)=P(X\in xV_{1,\mathbb{S}})=P(\|X\|>x)$ which is a regularly varying function with index $-\alpha$.

For $x\geqslant 1$ define

$$\mu_x=\frac{1}{V(x)}P\left(X\in x\cdot\right)$$

which is a measure on $\mathcal{B}_{\mathbb{B}_0}$. We will show that $\mu_x\to\mu$ in $M_0(\mathbb{B})$ as $x\to\infty$. To do this, we use the homeomorphism

$$T=(T_1,T_2):\mathbb{B}_0\to(0,+\infty)\times\mathbb{S}:z\mapsto\left(\|z\|,\frac{z}{\|z\|}\right)$$

$(T=R^{-1})$. For $y>0$ fixed, set

$$\nu_x^1=T_{1\#}\left(\mu_x\llcorner(\mathbb{B}\setminus B[0,y])\right)\qquad\text{and}\qquad\nu_x^2=T_{2\#}\left(\mu_x\llcorner(\mathbb{B}\setminus B[0,y])\right).$$

The family $(\nu_x^1)_{x\geqslant 1}$ is relatively compact in $M_b((y,+\infty))$. Indeed, for every $r>y$,

$$\begin{array}{l|l}
\lim_{x\to\infty}\nu_x^1((r,+\infty)) & =r^{-\alpha}\\
\lim_{x\to\infty}\dfrac{P(X\in x(\mathbb{B}\setminus B(0,r)))}{V(x)} & =\mu(\mathbb{B}\setminus B(0,r)).\\
\lim_{x\to\infty}\dfrac{P(\|X\|>xr)}{P(\|X\|>x)} &
\end{array}$$

The family $(\nu_x^1)_{x\geqslant 1}$ has thus a weak limit in $M_b((y,+\infty))$ which is, according to Borel-Stieltjes theorem, $\int_{\cdot}\alpha y^{-\alpha-1}d\mathcal{L}^1(y)$. Consequently, according to following Prohorov's theorem, for every $\varepsilon_1>0$, there exists a compact $K_1\subset(y,+\infty)$ such that $\sup_{x\geqslant 1}\nu_x^1(K_1^c)\leqslant\varepsilon_1$.

The family $(\nu_x^2)_{x\geqslant 1}$ is by assumption relatively compact in $M_b(\mathbb{S})$ for the topology of weak convergence, since it converges. According to

Prohorov's theorem, for every $\varepsilon_2 > 0$, there exists a compact $K_2 \subset \mathbb{S}$ such that $\sup_{x\geqslant 1} \nu_x^2(K_2^c) \leqslant \varepsilon_2$.

Set $K = T^{-1}(K_1 \times K_2)$, which depends on ε_1 and ε_2. According to Tychonoff's theorem, $K_1 \times K_2$ is compact and, as T is an homeomorphism, K is compact.

Let $\varepsilon > 0$ and choose $\varepsilon_1 = \varepsilon_2 = \dfrac{\varepsilon}{2}$. Then

$$
\begin{aligned}
&\sup_{x\geqslant 1} (\mu_x \llcorner (\mathbb{B} \setminus B[0,y]))\,(K^c)\\
&= \sup_{x\geqslant 1} \mu_x(\{z \in \mathbb{B} \ :\ \|z\| > y,\ \|z\| \notin K_1 \text{ or } z/\|z\| \notin K_2\})\\
&\leqslant \sup_{x\geqslant 1}[\ \mu_x(\{z \in \mathbb{B} \ :\ \|z\| > y,\ \|z\| \notin K_1\})\\
&\qquad\qquad + \mu_x(\{z \in \mathbb{B} \ :\ \|z\| > y,\ z/\|z\| \notin K_2\})\ ]\\
&\leqslant \sup_{x\geqslant 1} \mu_x(\{z \in \mathbb{B} \ :\ \|z\| > y,\ \|z\| \notin K_1\})\\
&\qquad\qquad + \sup_{x\geqslant 1} \mu_x(\{z \in \mathbb{B} \ :\ \|z\| > y,\ z/\|z\| \notin K_2\})\\
&= \sup_{x\geqslant 1} \nu_x^1\{K_1^c) + \sup_{x\geqslant 1} \nu_x^2(K_2^c)\\
&\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}
$$

which proves thanks to Prohorov's theorem that $(\mu_x \llcorner (\mathbb{B} \setminus B[0,y]))_{x\geqslant 1}$ is relatively compact in $\mathbb{B} \setminus B[0,y]$. As $y > 0$ was arbitrary, $(\mu_x)_{x\geqslant 1}$ is relatively compact in $M_0(\mathbb{B})$ (Hult and Lindskog 2006). Thus the family $(\mu_x)_{x\geqslant 1}$ possesses an accumulation point in M_0 when $x \to \infty$.

To show that every accumulation point of $(\mu_x)_{x\geqslant 1}$ is equal to μ, consider $\mathcal{V} = \{V_{y,S} \mid y > 0 \text{ and } S \in \mathcal{B}_{\mathbb{S}}\}$. It is a π-system generating $\mathcal{B}_{\mathbb{B}_0}$. Put $\mathcal{V}_0 = \{V_{y,S} \in \mathcal{V} \mid \lambda(\partial S) = 0\}$. The set $\mathcal{V}_0$ is still a π-system. Let indeed V_{y_1,S_1} and $V_{u_2,S_2} \in \mathcal{V}_0$. We have $V_{y_1,S_1} \cap V_{y_2,S_2} = V_{y_3,S_3}$ with $y_3 = \max(y_1, y_2)$ and $S_3 = S_1 \cap S_2$ and since

$$
\begin{aligned}
&\partial S_3 = \partial(S_1 \cap S_2) = \mathrm{cl}(S_1 \cap S_2) \setminus \mathrm{int}(S_1 \cap S_2)\\
&\subset (\mathrm{cl}\, S_1 \cap \mathrm{cl}\, S_2) \setminus \mathrm{int}(S_1 \cap S_2)\\
&= (\mathrm{cl}\, S_1 \cap \mathrm{cl}\, S_2) \setminus (\mathrm{int}\, S_1 \cap \mathrm{int}\, S_2)\\
&= (\mathrm{cl}\, S_1 \cap \mathrm{cl}\, S_2) \cap (\mathrm{int}\, S_1 \cap \mathrm{int}\, S_2)^c\\
&= (\mathrm{cl}\, S_1 \cap \mathrm{cl}\, S_2) \cap ((\mathrm{int}\, S_1)^c \cup (\mathrm{int}\, S_2)^c)\\
&= (\mathrm{cl}\, S_1 \cap \mathrm{cl}\, S_2 \cap (\mathrm{int}\, S_1)^c) \cup (\mathrm{cl}\, S_1 \cap \mathrm{cl}\, S_2 \cap (\mathrm{int}\, S_2)^c)\\
&\subset (\mathrm{cl}\, S_1 \cap (\mathrm{int}\, S_1)^c) \cup (\mathrm{cl}\, S_2 \cap (\mathrm{int}\, S_2)^c)\\
&= (\mathrm{cl}\, S_1 \setminus \mathrm{int}\, S_1) \cup (\mathrm{cl}\, S_2 \setminus \mathrm{int}\, S_2)\\
&= \partial S_1 \cup \partial S_2,
\end{aligned}
$$

this implies $\lambda(S_3) = 0$. Furthermore $\sigma(\mathcal{V}_0) = \mathcal{B}_{\mathbb{B}_0}$. To show it, by virtue of Proposition 2.5.16 it suffices to prove that $\sigma(\{S \in \mathcal{B}_{\mathbb{S}} \mid \lambda(\partial S) = 0\}) = \mathcal{B}_{\mathbb{S}}$. This is the case if for every $z \in \mathbb{S}$ and every $r > 0$, we can find $S \in \mathcal{B}_{\mathbb{S}}$ with $\lambda(\partial S) = 0$, $z \in \operatorname{int} S$ and $S \subset B(z,r)$. So let $z \in \mathbb{S}$ and $r > 0$. There exists a quantity at most countable $(r_i)_{i\in\mathbb{N}}$ such that $\lambda(\partial B(z,r_i) \cap \mathbb{S}) \neq 0$, otherwise we would have $\lambda(\mathbb{S}) = +\infty$ (and $\lambda(\mathbb{S}) = 1$). Thus the condition is satisfied with $S = B(z,\rho) \cap \mathbb{S}$ for every $0 < \rho < r$ such that $\rho \notin (r_i)_{i\in\mathbb{N}}$. Finally, for every $V_{y,S} \in \mathcal{V}_0$, we have

$$\mu_x(V_{y,S}) = \frac{P\left(\|X\| > xy, X/\|X\| \in S\right)}{P(\|X\| > x)} \to \frac{1}{y^\alpha}\lambda(S) = \mu(V_{y,S}), \quad x \to \infty.$$

Theorem 2.5.15 implies that all accumulation point M_0 for μ_x as $x \to \infty$ must be equal to μ. The measure μ is thus the limit of μ_x in M_0 as $x \to \infty$. □

Proof of Proposition 2.2.5 (page 24). (2.4) implies (2.7). Let $f(y) = P(\|X\| > y)$ which is nonincreasing and right continuous. Following the Portmanteau lemma, applying (2.4) to $\mathbb{S}$ which is such that $\lambda(\partial\mathbb{S}) = \lambda(\emptyset) = 0$ (inside $\mathbb{S}$, $\operatorname{cl}\mathbb{S} = \operatorname{int}\mathbb{S} = \mathbb{S}$), we obtain $\lim_{t\to\infty} \frac{f(ty)}{f(y)} = y^{-\alpha}$, that is $f \in RV_{-\alpha}$. Set $a_n = \inf\left\{x \in \mathbb{R} \mid f(x) \leqslant \frac{1}{n}\right\} = (\frac{1}{f})^{\leftarrow}(n)$. Since $\frac{1}{f} \in RV_\alpha$ with $\alpha > 0$ and as $\frac{1}{f}$ is nonincreasing, we have $(\frac{1}{f})^{\leftarrow} \in RV_{\frac{1}{\alpha}}$, hence $a_n \to \infty$ as $n \to \infty$ (de Haan and Ferreira 2006, Proposition B.1.9.1, $\lim_{t\to\infty}(\frac{1}{f})^{\leftarrow}(t) = +\infty$).

To see that $nP(\|X\| > a_n) \to 1$ as $n \to \infty$, fix $\varepsilon > 0$. Then, setting $U = \frac{1}{f}$, the inequalities

$$a_n(1-\varepsilon) < a_n = U^{\leftarrow}(n) \quad \text{and} \quad U^{\leftarrow}(n) = a_n < a_n(1+\varepsilon)$$

imply respectively

$$U[a_n(1-\varepsilon)] < n \quad \text{and} \quad n \leqslant U[a_n(1+\varepsilon)]$$

hence, inverting these inequalities and multiplying them by $U(a_n) = \frac{1}{f}(a_n)$, we have

$$\frac{U(a_n)}{U[a_n(1+\varepsilon)]} \leqslant \frac{1}{nf(a_n)} \quad \text{and} \quad \frac{1}{nf(a_n)} \leqslant \frac{U(a_n)}{U[a_n(1-\varepsilon)]}.$$

Consequently, since $U \in RV_\alpha$ and $a_n \to \infty$ as $n \to \infty$, we find

$$(1+\varepsilon)^{-\alpha} \leqslant \liminf_{n\to\infty} \frac{1}{nf(a_n)} \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{nf(a_n)} \leqslant (1-\varepsilon)^{-\alpha}.$$

Since $\varepsilon > 0$ was arbitrary, this implies that $nf(a_n) = nP(\|X\| > a_n) \to 1$ as $n \to \infty$.

(2.7) implies (2.4). Let S be a Borel set of $\mathbb{S}$ such that $\lambda(\partial S) = 0$. By assumption, $a_n \to \infty$ as $n \to \infty$. Hence for $t > 0$ large enough, there exists $n \in \mathbb{N}$ such that $a_n \leqslant t \leqslant a_{n+1}$. Then,

$$\begin{aligned} &\frac{P\left(\|X\| > ya_{n+1}, X/\|X\| \in S\right)}{P(\|X\| > a_n)} \\ &\leqslant \frac{P\left(\|X\| > yt, X/\|X\| \in S\right)}{P(\|X\| > t)} \\ &\leqslant \frac{P\left(\|X\| > ya_n, X/\|X\| \in S\right)}{P(\|X\| > a_{n+1})} \end{aligned}$$

Letting t tend to infinity in these expressions (which implies that n tends to infinity), we obtain on both sides the value

$$\frac{y^{-\alpha}\lambda(S)}{1^{-\alpha}\lambda(\mathbb{S})} = y^{-\alpha}\lambda(S).$$

The conclusion follows from the sandwich theorem. □

Proof of Proposition 2.2.6 (page 24). Similar to (2.4) ⇔ (2.7) above. □

2.7 Proofs for Section 2.3

The proof of Theorem 2.3.2 (page 26) is divided into a succession of steps, so that each of them separably provides enlightening results:

- (ii) implies (i) is Proposition 2.7.1 → Proposition 2.7.9,
- (i) implies (iii) is Proposition 2.7.10,
- (iii) implies (ii) is trivial.

Let us remember that

- (i) is the joint regular variation,
- (ii) is the existence of a forward tail process,
- (iii) is the existence of a tail process.

Note that the existence of a tail process implies existence of a forward tail process but the converse is nontrivial.

Then we have the proofs of Theorem 2.3.3, the equivalence between tail and spectral processes, the proof of Theorem 2.3.6, the time change formula, and the proof of Proposition 2.3.4 which yields the spectral process after a linear transformation.

Recall that is all this section $\mathbb{B}$ is a separable Banach space.

Proposition 2.7.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1. Then there exists $\alpha > 0$ such that $\|X_0\|$ is regularly varying with index α and $\|Y_0\| \sim \mathrm{Pareto}(\alpha)$.*

Proof. The definition of forward tail process compounded with the continuous function $g((x_t)_{t\geqslant 0}) = x_0$ yields the weak convergence

$$\mathcal{L}\left(\frac{X_0}{x} \;\middle|\; \|X_0\| > x\right) \xrightarrow{d} \mathcal{L}(Y_0), \qquad \text{as } x \to \infty.$$

On the one hand, since $\mathcal{L}(\|Y_0\|)$ possesses at most countably many atoms, it follows from the Portmanteau Lemma that, for almost every $y \geqslant 1$,

$$P(\|X_0\| > xy \mid \|X_0\| > x) \to P(\|Y_0\| > y)$$

as $x \to \infty$. On the other hand, since $y \geqslant 1$,

$$P(\|X_0\| > xy \mid \|X_0\| > x) = \frac{P(\|X_0\| > xy)}{P(\|X_0\| > x)}.$$

As $P(\|X_0\| > xy)/P(\|X_0\| > x)$ possesses a limit almost everywhere as $x \to \infty$, it must be $y^{-\alpha}$ for some $\alpha \in \mathbb{R}$ and every $y \geqslant 1$ (Bingham, Goldie, and Teugels 1987, Theorem 1.4.1). So $P(\|Y_0\| > y) = y^{-\alpha}$ for $y \geqslant 1$. According to the fact that $\mathcal{L}(\|Y_0\|)$ is a probability distribution, we have $\alpha \geqslant 0$ and $P(\|Y_0\| > y) = 1$ for $y < 1$. Since $\|Y_0\|$ is nondegenerate in 1, we obtain $\alpha > 0$. ☐

Now introduce, for $t \geqslant 0$, the random element

$$\Theta_t = \frac{Y_t}{\|Y_0\|},$$

so that $(\Theta_t)_{t\geqslant 0}$ is, by definition, the *forward spectral process* of $(X_t)_{t\in\mathbb{Z}}$.

Proposition 2.7.2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1. The random variable $\|Y_0\|$ is independent of the forward spectral process $(\Theta_t)_{t\geqslant 0}$.*

Proof. By definition, $\|Y_0\|$ and $(\Theta_t)_{t\geqslant 0}$ are independent if and only if the σ-fields $\mathcal{F} := \sigma(\|Y_0\|) = \left\{\|Y_0\|^{-1}(B) \mid B \in \mathcal{B}_{\mathbb{R}}\right\}$ and $\mathcal{G} := \sigma(\Theta_0, \Theta_1, \dots)$ $:= \sigma\left(\left\{\Theta_t^{-1}(B) \mid t \geqslant 0,\, B \in \mathcal{B}_{\mathbb{B}}\right\}\right)$ are independent:

$$(\forall A \in \mathcal{F})(\forall B \in \mathcal{G}) : P(A \cap B) = P(A)P(B).$$

The expression of $\mathcal{G}$ can be reformulated using a π-system: $\mathcal{G} = \sigma(\Pi)$ with

$$\Pi = \left\{ \bigcap_{i=1}^{k} \Theta_{t_i}^{-1}(B_i) \;\middle|\; k \geqslant 1,\, t_i \geqslant 0,\, B_i \in \mathcal{B}_{\mathbb{B}} \right\}.$$

Henceforth $\mathcal{F}$ and $\mathcal{G}$ are independent if and only if, according to Billingsley (1995), Theorem 4.2,

$$(\forall A \in \mathcal{F})(\forall B \in \Pi) : P(A \cap B) = P(A)P(B).$$

So $\|Y_0\|$ and $(\Theta_t)_{t\geqslant 0}$ are independent if and only if, for each $t \geqslant 0$, $\|Y_0\|$ and the finite stretch $(\Theta_0, \dots, \Theta_t)$ are independent.

Let us show that the joint law of $(\|Y_0\|, \Theta_0, \dots, \Theta_t)$ is equal to the product of the one of $\|Y_0\|$ and $(\Theta_0, \dots, \Theta_t)$. Lemma 1.4.2 of van der Vaart and Wellner (1996) yields that the joint law $(\|Y_0\|, \Theta_0, \dots, \Theta_t)$ is determined by its values on the functions of type fg where $f : \mathbb{R}_+ \to \mathbb{R}_+$ and $g : \mathbb{B}^{t+1} \to \mathbb{R}_+$ are continuous and bounded. Consequently $\|Y_0\|$ and $(\Theta_t)_{t\geqslant 0}$ are independent if and only if, for every f and g as above,

$$E\left[f(\|Y_0\|)g(\Theta_0, \dots, \Theta_t)\right] = E\left[f(\|Y_0\|)\right] E\left[g(\Theta_0, \dots, \Theta_t)\right].$$

Using Lebesgue's dominated convergence theorem and the fact that $\|Y_0\| \geqslant 1$, the functions f can be replaced by the family of functions $1\{\cdot > y\}$ where $y \geqslant 1$. Consequently $\|Y_0\|$ and $(\Theta_t)_{t\geqslant 0}$ are independent if and only if, for each $y \geqslant 1$ and every $g : \mathbb{B}^{t+1} \to \mathbb{R}_+$ continuous and bounded,

$$\begin{aligned} E\left[1\{\|Y_0\| > y\} g(\Theta_0, \ldots, \Theta_t)\right] &= E\left[1\{\|Y_0\| > y\}\right] E\left[g(\Theta_0, \ldots, \Theta_t)\right] \\ &= y^{-\alpha} E\left[g(\Theta_0, \ldots, \Theta_t)\right] \end{aligned}$$

by Proposition 2.7.1.

Applying the Continuous mapping theorem to the definition of forward tail process with the map

$$\zeta(x_0, \ldots x_t) = \left(\|x_0\|, \frac{x_0}{\|x_0\|}, \ldots, \frac{x_t}{\|x_0\|}\right),$$

we obtain, as $x \to \infty$,

$$\mathcal{L}\left(\left(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left(\|Y_0\|, \Theta_0, \ldots, \Theta_t\right),$$

Fix $y \geqslant 1$. Let $g : \mathbb{B}^{t+1} \to \mathbb{R}_+$ be continuous and bounded as above and set

$$f : \mathbb{R} \times \mathbb{B}^{t+1} \to \mathbb{R}_+ : (u, x_0, \ldots x_t) \mapsto 1\{u > y\} g(x_0, \ldots x_t).$$

We can apply the Portmanteau Lemma with f (adapt Lemma 2.5.20 to weak convergence), whence

$$\begin{aligned} & E\left[1\left\{\frac{\|X_0\|}{x} > y\right\} g\left(\frac{X_0}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| \geqslant x\right] \\ & \to E\left[1\{\|Y_0\| > y\} g\left(\Theta_0, \ldots, \Theta_t\right)\right] \end{aligned}$$

as $x \to \infty$. Furthermore, we also can compute the same limit as follows

$$\begin{aligned} & E\left[1\left\{\frac{\|X_0\|}{x} > y\right\} g\left(\frac{X_0}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| \geqslant x\right] \\ & = E\left[g\left(\frac{X_0}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| \geqslant xy\right] \frac{P(\|X_0\| \geqslant xy)}{P(\|X_0\| \geqslant x)} \\ & \to y^{-\alpha} E\left[g(\Theta_0, \ldots, \Theta_t)\right] \end{aligned}$$

as $x \to \infty$. Thus the conclusion follows from the uniqueness of the limit. $\square$

Next we have a relation which, given a large value for $\|X_0\|$, allows us to know some piece of information about the past.

Proposition 2.7.3 (Knowledge about the past). *Given a stationary time series $(X_t)_{t\in\mathbb{Z}}$ in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1, the following formula holds: for every $t \geqslant 0$,*

$$E\left[\|\Theta_t\|^\alpha\right] = \lim_{r\to 0}\lim_{x\to\infty} P\left(\|X_{-t}\| > rx \mid \|X_0\| > x\right).$$

Proof. Let $r > 0$. Note that by the Continuous mapping theorem applied to the definition of forward tail process with the projection on the t^{th} component

$$\mathcal{L}\left(\frac{X_t}{rx}\,\middle|\,\|X_0\| > rx\right) \xrightarrow{d} \mathcal{L}\left(\|Y_t\|\right)$$

as $x \to \infty$. According to the facts that $\|Y_t\| = \|Y_0\|\|\Theta_t\|$, $\|Y_0\|$ and $\|\Theta_t\|$ are independent random variables (Proposition 2.7.2), $\|Y_0\|$ is continuous and by Lemma 2.5.19, the law of $\|Y_t\|$ is continuous except maybe where $\Theta_t = 0$. Therefore $\mathcal{L}(\|Y_t\|)$ does not have any atom. Thus $P(\|Y_t\| = \frac{1}{r}) = 0$ for every $r > 0$. Reversing the time, by stationarity, and applying the Portmanteau Lemma, we get

$$\begin{aligned}
&P\left(\|X_{-t}\| > rx \mid \|X_0\| > x\right)\\
&= P\left(\|X_0\| > x \mid \|X_{-t}\| > rx\right)\frac{P(\|X_{-t}\| > rx)}{P(\|X_t\| > x)}\\
&= P\left(\frac{\|X_t\|}{xr} > \frac{1}{r}\,\middle|\,\|X_0\| > x\right)\frac{P(\|X_0\| > rx)}{P(\|X_0\| > x)}\\
&\to P\left(\|Y_t\| > \frac{1}{r}\right) r^{-\alpha}
\end{aligned}$$

as $x \to \infty$. Moreover, setting $x = ry$ and $v = x^{-\alpha}$,

$$\begin{aligned}
&P\left(\|Y_t\| > \frac{1}{r}\right) r^{-\alpha}\\
&= P\left(\|Y_0\|\|\Theta_t\| > \frac{1}{r}\right) r^{-\alpha}\\
&= \int_1^\infty P\left(y\|\Theta_t\| > \frac{1}{r}\right) d(-y^{-\alpha}) r^{-\alpha}
\end{aligned}
\qquad\left|\qquad
\begin{aligned}
&= \int_r^\infty P\left(\|\Theta_t\|^\alpha > x^{-\alpha}\right) d(-x^{-\alpha})\\
&= \int_0^{r^{-\alpha}} P\left(\|\Theta_t\|^\alpha > v\right) dv\\
&\to \int_0^\infty P\left(\|\Theta_t\|^\alpha > v\right) dv
\end{aligned}\right.$$

as $r \to 0$ which is equal to $E\left[\|\Theta_t\|^\alpha\right]$. □

In order to technically prepare Proposition 2.7.5, let us define the set function ν_t, $t \geqslant 0$, from $\mathcal{B}_{\mathbb{B}}$ to $\mathbb{R}$ by

$$\nu_t(A) = 1_A(0)\left(1 - E\left[\|\Theta_t\|^\alpha\right]\right) + \int_0^\infty P\left(v\Theta_0 \in A, \|\Theta_t\| > \frac{1}{v}\right) d(-v^{-\alpha}).$$

Let us first remark that ν_t is a probability measure.

Proposition 2.7.4. *If $(X_t)_{t\in\mathbb{Z}}$ is a stationary time series in $\mathbb{B}$ and admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1, then the map ν_t is a probability measure on $\mathbb{B}$. If $f : \mathbb{B} \to \mathbb{R}$ is ν_t-integrable, the expression of its integral is*

$$\int f d\nu_t = f(0)\left(1 - E\left[\|\Theta_t\|^\alpha\right]\right) + \int_0^\infty E\left[f(v\Theta_0)\, 1\{\|\Theta_t\| > \frac{1}{v}\}\right] d(-v^{-\alpha}).$$

Proof. Let $A \in \mathcal{B}_{\mathbb{B}}$. To see that $\nu_t(A) \geqslant 0$, it suffices to show that $1 - E\left[\|\Theta_t\|^\alpha\right]$ and $\int_0^\infty P\left(v\Theta_0 \in A, \|\Theta_t\| > \frac{1}{v}\right) d(-v^{-\alpha})$ are both nonnegative. The first term is nonnegative because $E\left[\|\Theta_t\|^\alpha\right]$ is a limit of probabilities (Proposition 2.7.3) and the second term is nonnegative because it is an integral of a probability.

We have $\nu_t(\emptyset) = 0$, indeed

$$\nu_t(\emptyset) = \int_0^\infty P\left(v\Theta_0 \in \emptyset, \|\Theta_t\| > \frac{1}{v}\right) d(-v^{-\alpha}) = \int_0^\infty P(\emptyset)\, d(-v^{-\alpha}) = 0.$$

Finally assume that $A_1, A_2, A_3, \ldots$ is a countable sequence of pairwise disjoint sets in $\mathcal{B}_{\mathbb{B}}$. Then $\nu_t\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \nu_t(A_i)$ since developing both sides yields

$$\begin{aligned} &1_{\bigcup_{i=1}^\infty A_i}(0)\left(1 - E\left[\|\Theta_t\|^\alpha\right]\right) + \int_0^\infty P\left(v\Theta_0 \in \bigcup_{i=1}^\infty A_i, \|\Theta_t\| > \frac{1}{v}\right) d(-v^{-\alpha}) \\ &= \sum_{i=1}^\infty \left(1_{A_i}(0)\left(1 - E\left[\|\Theta_t\|^\alpha\right]\right) + \int_0^\infty P\left(v\Theta_0 \in A_i, \|\Theta_t\| > \frac{1}{v}\right) d(-v^{-\alpha})\right). \end{aligned}$$

Consequently ν_t is a measure.

We have $\nu_t(\mathbb{B}) = 1$ since

$$E\left[\|\Theta_t\|^\alpha\right] = \int_0^\infty P\left(\|\Theta_t\| > \frac{1}{v}\right) d(-v^{-\alpha})$$

according to the proof of Proposition 2.7.3.

About the formula of $\int f d\nu_t$, remark that $1_A(0) = \delta_0(A)$, which proves the first half. For the integral part, remark that it is valid for functions of type $f = 1_A$, which makes it valid for every ν_t-integrable function tanks to the same rules as above and Lebesgue's dominated convergence theorem. $\square$

The probability measure ν_t helps to express the distribution of a shift in the past.

Proposition 2.7.5. *If $(X_t)_{t\in\mathbb{Z}}$ is a stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1, then the measure ν_t is the following weak limit*

$$\mathcal{L}\left(\frac{X_{-t}}{x}\,\middle|\, \|X_0\| > x\right) \xrightarrow{d} \nu_t$$

as $x \to \infty$.

Proof. Let $f : \mathbb{B} \to \mathbb{R}$ be Lipschitz continuous and bounded. According to the Portmanteau Lemma (van der Vaart 1998), we have to show that

$$E\left[f\left(\frac{X_{-t}}{x}\right)\,\middle|\, \|X_0\| > x\right] \to \int f d\nu_t =$$
$$f(0)\left(1 - E\left[\|\Theta_t\|^\alpha\right]\right) + \int_0^\infty E\left[f\left(v\Theta_0\right) 1\{\|\Theta_t\| > \frac{1}{v}\}\right] d(-v^{-\alpha})$$

as $x \to \infty$. Given $r > 0$, let us decompose the expected value of the left hand-side of the limit according to $\{\|X_{-t}\| \leqslant xr\}$ and $\{\|X_{-t}\| > xr\}$ before letting r tend towards 0.

Firstly, since $L = \operatorname{lip} f < +\infty$, we have $|f(\xi)| \leqslant f(0) + L\|\xi\|$. Then, using the notation $a \in A \subset B$ in $\mathbb{R}$,

$$\begin{aligned}
&E\left[f\left(\frac{X_{-t}}{x}\right) 1\{\|X_{-t}\| \leqslant xr\}\,\middle|\, \|X_0\| > x\right]\\
&\in E\left[f\left(0\right) 1\{\|X_{-t}\| \leqslant xr\} \mid \|X_0\| > x\right]\\
&\qquad\qquad \pm LE\left[\frac{\|X_{-t}\|}{x} 1\{\|X_{-t}\| \leqslant xr\}\,\middle|\, \|X_0\| > x\right]\\
&\subset f\left(0\right) E\left[1\{\|X_{-t}\| \leqslant xr\} \mid \|X_0\| > x\right] \pm LE\left[\frac{xr}{x}\,\middle|\, \|X_0\| > x\right]\\
&= f\left(0\right)\left(1 - P\left(\|X_{-t}\| > xr \mid \|X_0\| > x\right)\right) \pm Lr\\
&\to f(0)\left(1 - E\left[\|\Theta_t\|^\alpha\right]\right)
\end{aligned}$$

after $x \to \infty$ and $r \to 0$ by Proposition 2.7.3, which yields the first part of the claim.

Secondly,

$$\begin{aligned}
&E\left[f\left(\frac{X_{-t}}{x}\right) 1\{\|X_{-t}\| > xr\} \;\middle|\; \|X_0\| > x \right] \\
&= E\left[f\left(\frac{X_{-t}}{x}\right) 1\{\|X_0\| > x\} \;\middle|\; \|X_{-t}\| > xr \right] \frac{P(\|X_{-t}\| > xr)}{P(\|X_0\| > x)} \\
&= E\left[f\left(\frac{rX_0}{rx}\right) 1\left\{ \frac{\|X_t\|}{rx} > \frac{1}{r} \right\} \;\middle|\; \|X_0\| > xr \right] \frac{P(\|X_0\| > xr)}{P(\|X_0\| > x)} \\
&\to E\left[f(rY_0) 1\left\{ \|Y_t\| > \frac{1}{r} \right\} \right] r^{-\alpha}
\end{aligned}$$

as $x \to \infty$ according to the definition of the forward tail process. Finally

$$\begin{aligned}
&E\left[f(rY_0) 1\left\{ Y_t > \frac{1}{r} \right\} \right] r^{-\alpha} \\
&= \int_1^\infty E\left[f(ry\Theta_0) 1\left\{ y\|\Theta_t\| > \frac{1}{r} \right\} \right] d(-y^{-\alpha}) r^{-\alpha} \\
&= \int_r^\infty E\left[f(v\Theta_0) 1\left\{ \|\Theta_t\| > \frac{1}{v} \right\} \right] d(-v^{-\alpha}) \\
&\to \int_0^\infty E\left[f(v\Theta_0) 1\left\{ \|\Theta_t\| > \frac{1}{v} \right\} \right] d(-v^{-\alpha})
\end{aligned}$$

as $r \to 0$, which is the second part of the claim. $\square$

In order to prove Proposition 2.7.7 we detach a result about the tightness of the joint conditional vector below.

Proposition 2.7.6. *If $(X_t)_{t\in\mathbb{Z}}$ is stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1, then, for all $s, t \geqslant 0$, the sequence of probability measures*

$$\left(\left(\frac{X_{-s}}{x}, \dots, \frac{X_t}{x} \right) \;\middle|\; \|X_0\| > x \right)_{x \geqslant x_0}$$

is tight.

Proof. It follows from Proposition 2.7.5 and Prohorov's theorem that for every $r \in \mathbb{Z}$ the sequence

$$\left(\frac{X_r}{x} \,\middle|\, \|X_0\| > x \right)_{x \geqslant x_0}$$

is tight, i.e. for every $r \in \mathbb{Z}$ and for every $\varepsilon > 0$ there exists a compact $K_{r,\varepsilon} \subset \mathbb{B}$ such that for every $x \geqslant x_0$

$$P\left(\frac{X_r}{x} \notin K_{r,\varepsilon} \,\middle|\, \|X_0\| > x \right) \leqslant \varepsilon.$$

Let $\varepsilon > 0$ and apply this property with $\varepsilon/(s+t+1)$ for $r = -s$ to $r = t$. Set $K_\varepsilon = K_{-s,\varepsilon} \times \ldots \times K_{t,\varepsilon}$. The set K_ε obtained is compact by Tychonoff's Theorem and satisfies the property that, for every $x \geqslant x_0$,

$$P\left(\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x} \right) \notin K_\varepsilon \,\middle|\, \|X_0\| > x \right) \leqslant \varepsilon.$$

which is the statement. □

Proposition 2.7.7. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t \geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at* 1. *Let $s, t \geqslant 0$. Then, as $x \to \infty$,*

$$\mathcal{L}\left(\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x} \right) \,\middle|\, \|X_0\| > x \right) \xrightarrow{d} \nu_{s,t}$$

for some probability measure $\nu_{s,t}$ which we define to be $\mathcal{L}(Y_{-s}, \ldots, Y_t)$.

Proof. By Prohorov's theorem, Proposition 2.7.6 ensures the probabilities of the sequence to have an accumulation point as $x \to \infty$. So it suffices to see that every accumulation point as $x \to \infty$ is the same measure, that we will called $\nu_{s,t}$. By the Portmanteau Lemma, it suffices to show that for every $f : \mathbb{B}^{t+s+1} \to \mathbb{R}$ Lipschitz continuous and bounded, we have

$$\lim_{x \to \infty} E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x} \right) \,\middle|\, \|X_0\| > x \right] = \int f d\nu_{s,t}.$$

Let us proceed by induction on s. If $s = 0$, by the definition of forward tail process we have

$$\lim_{x \to \infty} E\left[f\left(\frac{X_0}{x}, \ldots, \frac{X_t}{x} \right) \,\middle|\, \|X_0\| > x \right] = E\left[f\left(Y_0, \ldots, Y_t \right) \right]$$

so that $\nu_{0,t} = \mathcal{L}$ is the law of $(Y_0, \ldots, Y_t)$ defined by assumption.

Suppose now that the statement is true for $s-1$ and let us prove it for s. As in Proposition 2.7.5 decompose the expected value according to $\{\|X_{-s}\| \leqslant xr\}$ and $\{\|X_{-s}\| > xr\}$ before letting r tend to 0. For the first event we have

$$\begin{aligned}
&E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|X_{-s}\| \leqslant xr\} \;\middle|\; \|X_0\| > x\right] \\
&= E\left[f\left(0, \frac{X_{-s+1}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|X_{-s}\| \leqslant xr\} \;\middle|\; \|X_0\| > x\right] \\
&\quad + E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|X_{-s}\| \leqslant xr\} \;\middle|\; \|X_0\| > x\right] \\
&\quad - E\left[f\left(0, \frac{X_{-s+1}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|X_{-s}\| \leqslant xr\} \;\middle|\; \|X_0\| > x\right].
\end{aligned}$$

Since $\operatorname{lip} f = L < +\infty$, the absolute value of the last difference is less than rL thus tends to 0 as r tends 0 (independently from x). Therefore we just have to consider the first term in the sequel. Writing $1\{\|X_{-s}\| \leqslant xr\} = 1 - 1\{\|X_{-s}\| > xr\}$ this one becomes

$$E\left[f\left(0, \ldots, \frac{X_t}{x}\right) \middle| \|X_0\| > x\right] - E\left[f\left(0, \ldots, \frac{X_t}{x}\right) 1_{\{\|X_{-s}\| > xr\}} \middle| \|X_0\| > x\right].$$

By the recurrence hypothesis, as $x \to \infty$, the left hand-side tends to

$$\int f(0, x_{-s+1}, \ldots, x_t) d\nu_{s,t} := \int \overline{f}(x_{-s+1}, \ldots, x_t) d\nu_{s-1,t}$$

where $\overline{f}(x_{-s+1}, \ldots, x_t) = f(0, x_{-s+1}, \ldots, x_t)$. The right hand-side can be easily regrouped with the decomposition according to the other event, $\{\|X_{-s}\| > xr\}$, to provide the value

$$E\left[g\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|X_{-s}\| > xr\} \;\middle|\; \|X_0\| > x\right]$$

where

$$g(x_{-s}, \ldots, x_t) = f(x_{-s}, \ldots, x_t) - f(0, x_{-s+1}, \ldots, x_t).$$

Note that if $x_{-s} = 0$ then $g \equiv 0$. Proceeding as in Proposition 2.7.5 we get

$$\begin{aligned}
&E\left[g\left(\frac{X_{-s}}{x},\dots,\frac{X_t}{x}\right)1\{\|X_{-s}\| > xr\} \;\middle|\; \|X_0\| > x\right] \\
&= E\left[g\left(\frac{rX_0}{rx},\dots,\frac{rX_{s+t}}{rx}\right)1\left\{\frac{\|X_s\|}{rx} > \frac{1}{r}\right\} \;\middle|\; \|X_0\| > xr\right]\frac{P(\|X_0\| > xr)}{P(\|X_0\| > x)} \\
&\to E\left[g\left(rY_0,\dots,rY_{s+t}\right)1\left\{\|Y_s\| > \frac{1}{r}\right\}\right]r^{-\alpha}
\end{aligned}$$

as $x \to \infty$. Now letting r tends towards 0, this expression becomes

$$\begin{aligned}
&\int_1^\infty E\left[g\left(ry\Theta_0,\dots,ry\Theta_{s+t}\right)1\left\{y\|\Theta_s\| > \frac{1}{r}\right\}\right]d(-y^{-\alpha})r^{-\alpha} \\
&= \int_r^\infty E\left[g\left(v\Theta_0,\dots,v\Theta_{s+t}\right)1\left\{\|\Theta_s\| > \frac{1}{v}\right\}\right]d(-v^{-\alpha}) \\
&\to \int_0^\infty E\left[g\left(v\Theta_0,\dots,v\Theta_{s+t}\right)1\left\{\|\Theta_s\| > \frac{1}{v}\right\}\right]d(-v^{-\alpha}),
\end{aligned}$$

which defines $\nu_{s,t}$ on the whole space $\mathbb{B}^{s+t+1}$. □

Now that a tail process $(Y_t)_{t\in\mathbb{Z}}$ is defined, proceed as for the forward tail process and set

$$\Theta_t = \frac{Y_t}{\|Y_0\|},\ t \geqslant \mathbb{Z},$$

which the *spectral process* of the series $(X_t)_{t\in\mathbb{Z}}$. We have again a result of independence.

Proposition 2.7.8. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is non-degenerate at* 1. *Then the random variable $\|Y_0\|$ is independent of the spectral process $(\Theta_t)_{t\in\mathbb{Z}}$.*

Proof. The proof is the exactly the same as the one of Proposition 2.7.2 but with the Continuous mapping theorem applied to the definition of tail process with the map

$$\zeta(x_{-s},\dots x_t) = \left(\|x_0\|,\frac{x_{-s}}{\|x_0\|},\dots,\frac{x_t}{\|x_0\|}\right),$$

nothing else changes. □

The existence of exponent measures establishes the fact that the series $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying.

Proposition 2.7.9 (Exponent measure). *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series in $\mathbb{B}$ that admits a forward tail process $(Y_t)_{t\geqslant 0}$ for which the law of $\|Y_0\|$ is nondegenerate at 1. Let $s,t \geqslant 0$. There exists $\mu_{s,t} \in M_0(\mathbb{B}^{s+t+1})$ such that, as $x \to \infty$,*

$$\frac{1}{P(\|X_0\| > x)} P\left(\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) \in \cdot\right) \xrightarrow{M_0} \mu_{s,t},$$

where, for $f \in \mathcal{C}_{bd}(\mathbb{B}^{s+t+1})$, the value $\mu_{s,t}(f)$ admits the expression

$$\int f d\mu_{s,t} \sum_{i=-s}^{t} \int_0^\infty E[f(0,\ldots,0,v\Theta_0,\ldots,v\Theta_{t-i}) \\ 1\{\Theta_{-s-i} = \ldots = \Theta_{-1} = 0\}]d(-v^{-\alpha}).$$

Proof. Set

$$\mu_{s,t;x}(\cdot) = \frac{1}{P(\|X_0\| > x)} P\left(\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) \in \cdot\right).$$

Let $f \in \mathcal{C}_{bd}(\mathbb{B}^{s+t+1})$ vanishing on a ball of radius $r > 0$. Then, using the convention $\bigvee \emptyset = -\infty$, the value of $\int f d\mu_{s,t;x}$ is equal to

$$\begin{aligned}
&\frac{1}{P(\|X_0\| > x)} E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right)\right] \\
&= \frac{1}{P(\|X_0\| > x)} E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\bigvee_{i=-s}^{t} \|X_i\| > rx\}\right] \\
&= \frac{1}{P(\|X_0\| > x)} E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) \sum_{i=-s}^{t} 1\{\bigvee_{j=-s}^{i-1} \|X_j\| \leqslant rx < \|X_i\|\}\right]
\end{aligned}$$

$$= \sum_{i=-s}^{t} E\left[f\left(\frac{X_{-s}}{x}, \dots, \frac{X_t}{x}\right) 1\{\bigvee_{j=-s}^{i-1} \|X_j\| \leqslant rx\} \middle| \|X_i\| > rx \right]$$
$$\cdot \frac{P(\|X_i\| > rx)}{P(\|X_0\| > x)}$$
$$= \sum_{i=-s}^{t} E\left[f\left(\frac{X_{-s-i}}{x}, \dots\right) 1\{\bigvee_{j=-s}^{i-1} \|X_{j-i}\| \leqslant rx\} \middle| \|X_0\| > rx \right]$$
$$\cdot \frac{P(\|X_0\| > rx)}{P(\|X_0\| > x)}$$

and, as $x \to \infty$, by Proposition 2.7.7, $\int f d\mu_{s,t;x}$ tends to

$$\sum_{i=-s}^{t} E\left[f\left(rY_{-s-i}, \dots, rY_{t-i}\right) 1\{\bigvee_{j=-s}^{i-1} \|Y_{j-i}\| \leqslant 1\} \right] r^{-\alpha}$$
$$= \sum_{i=-s}^{t} \int_1^{\infty} E\left[f(ry\Theta_{-s-i}, \dots, ry\Theta_{t-i}) 1\{\bigvee_{j=-s}^{i-1} \|y\Theta_{j-i}\| \leqslant 1\} \right] d(-y^{-\alpha}) r^{-\alpha}$$
$$= \sum_{i=-s}^{t} \int_r^{\infty} E\left[f\left(v\Theta_{-s-i}, \dots, v\Theta_{t-i}\right) 1\{\bigvee_{j=-s}^{i-1} \|v\Theta_{j-i}\| \leqslant r\} \right] d(-v^{-\alpha}),$$

which defines $\mu_{s,t}$. By dominated convergence, we can pass to the limit as $r \to 0$, obtaining the announced expression for $\mu_{s,t}$. □

That proves (ii) implies (i). The direction (i) implies (iii) is the next proposition. To obtain (ii) implies (iii) simply take $s = 0$.

Proposition 2.7.10. *If the series $(X_t)_{t \in \mathbb{Z}}$ is regularly varying with index $\alpha > 0$, then $(X_t)_{t \in \mathbb{Z}}$ has a tail process with $\|Y_0\|$ Pareto(α)-distributed.*

Proof. Let $s, t \geqslant 0$. By definition, the random object $X = (X_{-s}, \dots, X_t)$ in $\mathbb{B}^{t+s+1}$ is regularly varying with index $\alpha > 0$, so that there exists $\mu_{s,t} \in M_0(\mathbb{B}^{t+s+1})$ such that

$$\frac{1}{V(x)} P\left(X \in x\cdot\right) \xrightarrow{M_0} \mu_{s,t}(\cdot) \quad \text{as } x \to \infty,$$

where $V(x) = P(\|X_0\| > x) \in RV_{-\alpha}$. Set $\nu_{s,t}$ the restriction of $\mu_{s,t}$ to $\{(x_{-s}, \ldots, x_t) \in \mathbb{B}^{t+s+1} \mid \|x_0\| > 1\}$. By construction $\nu_{s,t}$ is a probability measure.

Let $(Y_{-s}, \ldots Y_t)$ be a random function having $\nu_{s,t}$ as law and let $f \in \mathcal{C}_b(\mathbb{B}^{s+t+1})$. Since $\mu_{s,t}(\{\|x_0\| = 1\}) = 0$ and $0 \notin \{\|x_0\| \geqslant 1\})$, applying Lemma 2.5.20 yields

$$E\left(f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) \,\middle|\, \|X_0\| > x\right)$$
$$\to \int f d\mu_{s,t} \llcorner \{\|x_0\| > 1\} = E(f(Y_{-s}, \ldots, Y_t))$$

as $x \to \infty$, which establishes the weak convergence.

According to the Daniell-Kolmogorov existence theorem (Pollard 2002, Theorem 53), there exists a stochastic process $(Y_t)_{t\in\mathbb{Z}}$ in $\mathbb{B}$ such that, for every $s, t \in \mathbb{Z}$, the distribution of $(Y_{-s}, \ldots Y_t)$ is $\nu_{s,t}$.

We finally have $P(\|Y_0\| > y) = y^{-\alpha}$ because the law Y_0 satisfies $\nu_{0,0}(\lambda A) = \lambda^{-\alpha}\nu_{0,0}(A)$ for every $\lambda > 0$ and $\nu_{0,0}(\{\|x_0\| \geqslant 1\}) = 1$. □

Now Theorem 2.3.2 is proved.

Proof of Theorem 2.3.3 (27). (i) implies (iii). By (iii) of Theorem 2.3.2, we have

$$\mathcal{L}\left(\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) \,\middle|\, \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left(Y_{-s}, \ldots, Y_t\right).$$

By the Continuous mapping theorem (Resnick 2008, p 152) applied with the map $\zeta(x_{-s}, \ldots, x_t) = (x_{-s}/x_0, \ldots, x_t/x_0)$, we obtain that

$$\mathcal{L}\left(\left(\frac{X_{-s}/x}{\|X_0\|/x}, \ldots, \frac{X_t/x}{\|X_0\|/x}\right) \,\middle|\, \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left(\frac{Y_{-s}}{\|Y_0\|}, \ldots, \frac{Y_t}{\|Y_0\|}\right),$$

which is exactly

$$\mathcal{L}\left(\left(\frac{X_{-s}}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left(\Theta_{-s}, \ldots, \Theta_t\right).$$

(i) implies (iii). It suffices to take $s = 0$.

(ii) implies (i). Let Y be Pareto(α) distributed and independent of $(\Theta_t)_{t\geqslant 0}$.

First of all, remark that

$$\mathcal{L}\left(\left(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \dots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| > x\right) \xrightarrow{d} \mathcal{L}\left(Y, \Theta_0, \dots, \Theta_t\right),$$

Indeed, by van der Vaart and Wellner (1996), Lemma 1.4.2, to see this it suffices to prove that:

$$E\left[f\left(\frac{\|X_0\|}{x}\right) g\left(\frac{X_0}{\|X_0\|}, \dots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| > x\right] \\ \to E[f(Y)g(\Theta_0, \dots, \Theta_t)].$$

for every continuous and bounded functions f and g. Using Lebesgue's dominated convergence theorem and the fact that $\|Y\| \geqslant 1$, the functions f can be replaced by the family of functions $1\{\cdot > y\}$ where $y \geqslant 1$, which satisfy the property:

$$\begin{aligned}
&\lim_{x\to\infty} E\left[1\left\{\frac{\|X_0\|}{x} > y\right\} g\left(\frac{X_0}{\|X_0\|}, \dots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| > x\right] \\
&= \lim_{x\to\infty} \frac{P(\|X_0\| \geqslant xy)}{P(\|X_0\| \geqslant x)} E\left[g\left(\frac{X_0}{\|X_0\|}, \dots, \frac{X_t}{\|X_0\|}\right) \,\middle|\, \|X_0\| > xy\right] \\
&= y^{-\alpha} E\left[g(\Theta_0, \dots, \Theta_t)\right] = E\left[1\{Y > y\}\right] E\left[g(\Theta_0, \dots, \Theta_t)\right] \\
&= E\left[1\{Y > y\} g(\Theta_0, \dots, \Theta_t)\right].
\end{aligned}$$

Finally apply the Continuous mapping theorem (Resnick 2008, p 152) with the function $\zeta(y, x_{-s}, \dots, x_t) = y(x_{-s}, \dots, x_t)$ to obtain

$$\mathcal{L}\left(\left(\frac{X_0}{x}, \dots, \frac{X_t}{x}\right) \,\middle|\, \|X_0\| > x\right) \xrightarrow{d} \mathcal{L}\left(Y(\Theta_0, \dots, \Theta_t)\right),$$

so that $Y(\Theta_0, \dots, \Theta_t)$ is the spectral process. $\square$

Here is the proof of the proposition describing how the spectral process changes after a linear transformation.

Proof of Proposition 2.3.4 (page 28). Let $h \in \mathcal{C}_b(\mathbb{B}_2^{s+t+1})$. By Definition 2.2.1 and Lemma 2.5.20, as $x \to \infty$,

$$E\left[h\left(\frac{T(X_{-s})}{x},\dots,\frac{T(X_t)}{x}\right)\;\middle|\; \|T(X_0)\| > x\right]$$
$$= \frac{E\left[h\left(\frac{T(X_{-s})}{x},\dots,\frac{T(X_t)}{x}\right)1\{\|\frac{T(X_0)}{x}\| > 1\}\right]}{P(\|X_0\| > x)}\frac{P(\|X_0\| > x)}{P(\|T(X_0)\| > x)}$$
$$\to \int h\left(T(y_{-s}),\dots,T(y_t)\right)1\{\|T(y_0)\| > 1\}d\mu_{-s,t}(y_{-s},\dots,y_t)$$
$$\cdot\frac{1}{E[\|T(\Theta_0^X)\|^\alpha]}.$$

Write $\varphi := (\varphi_{-s},\dots,\varphi_t)$ and $\Phi := (\Theta_{-s}^X,\dots,\Theta_t^X)$. Thanks to the independence of Theorem 2.3.3 this limit can be rewritten

$$\int_{\substack{r>0\\ \|\varphi\|=1}} h\left(T(r\varphi_{-s}),\dots,T(r\varphi_t)\right)1\{\|T(r\varphi_0)\| > 1\}\frac{P(\Phi\in d\varphi)}{E[\|T(\Theta_0^X)\|^\alpha]}d(-r^{-\alpha}).$$

Applying this result to $h(x_{-s},\dots,x_t) = f\left(\frac{x_{-s}}{\|x_0\|},\dots,\frac{x_t}{\|x_0\|}\right)$ we obtain

$$\int_{\substack{r>0\\ \|\varphi\|=1}} f\left(\frac{T(\varphi_{-s})}{\|T(\varphi_0)\|},\dots,\frac{T(\varphi_t)}{\|T(\varphi_0)\|}\right)1\{r\|T(\varphi_0)\| > 1\}\frac{P(\Phi\in d\varphi)}{E[\|T(\Theta_0^X)\|^\alpha]}d(-r^{-\alpha})$$
$$= \int_{r>0} E[f\left(\frac{T(\Theta_{-s}^X)}{\|T(\Theta_0^X)\|},\dots,\frac{T(\Theta_t^X)}{\|T(\Theta_0^X)\|}\right)1\{r\|T(\Theta_0^X)\| > 1\}]\frac{d(-r^{-\alpha})}{E[\|T(\Theta_0^X)\|^\alpha]}$$

which yields the announced formula after use of Fubini's theorem and the fact that $\int_{r>0} 1\{r\|T(\Theta_0^X)\| > 1\}]d(-r^{-\alpha}) = \|T(\Theta_0^X)\|^\alpha$. □

Finally the time change formula is obtained for functions waiving the recursive part in the construction of the spectral process.

Proof of Theorem 2.3.6 (page 30). The proof of Proposition 2.7.7 gives a recursive way to compute $\nu_{s,t}$. In this proof remark that, if $f(0,\dots) = 0$, then $\int f(0,\ldots)d\nu_{s,t} = 0$ and $g = f$, so

$$E\left[f\left(Y_{-s},\ldots,Y_t\right)\right] = \int_0^\infty E\left[f\left(v\Theta_0,\ldots,v\Theta_{s+t}\right)1\left\{\|\Theta_s\| > \frac{1}{v}\right\}\right]d(-v^{-\alpha}). \quad (2.12)$$

Now set

$$h(x_{-s},\ldots,x_t) = f\left(\frac{x_{-s}}{\|x_0\|},\ldots,\frac{x_t}{\|x_0\|}\right)(\|x_0\| \wedge 1)$$

in order to write $E\left[f(\Theta_{-s},\ldots,\Theta_t)\right] = E\left[h(Y_{-s},\ldots,Y_t)\right]$. We can apply the last formula to obtain, using Fubini's Theorem,

$$\begin{aligned} &E\left[f(\Theta_{-s},\ldots,\Theta_t)\right] \\ &= \int_0^\infty E\left[f\left(\frac{\Theta_0}{\|\Theta_s\|},\ldots,\frac{\Theta_{s+t}}{\|\Theta_s\|}\right)1\left\{\|\Theta_s\| > \frac{1}{v}\right\}\right]d(-v^{-\alpha}) \\ &= E\left[f\left(\frac{\Theta_0}{\|\Theta_s\|},\ldots,\frac{\Theta_{s+t}}{\|\Theta_s\|}\right)\int_0^\infty 1\left\{\|\Theta_s\| > \frac{1}{v}\right\}d(-v^{-\alpha})\right] \\ &= E\left[f\left(\frac{\Theta_0}{\|\Theta_s\|},\ldots,\frac{\Theta_{s+t}}{\|\Theta_s\|}\right)\|\Theta_s\|^\alpha\right], \end{aligned}$$

which is the announced formula.

The weak convergence of Proposition 2.7.7 was proved for f Lipschitz continuous and bounded. According to the Portmanteau Lemma, the same result holds for f continuous and bounded, which can be assumed for f in this proof. Once the result got for these f, there only are a few steps involving the monotone convergence to get the same result for integrable functions. □

2.8 Proofs for Section 2.4

The fact that the law Y_0 is $\mu \llcorner [\|x\| > 1]$ and that the law of $(X_{-s},\ldots,X_t)$ is $\mu_{s,t} \llcorner [\|x_0\| > 1]$ is due to the fact that, given $f \in \mathcal{C}_{bd}(\mathbb{B}^{s+t+1})$, on the one hand

$$\begin{aligned}
& E\left[f\left(\frac{X_{-s}}{x},\ldots,\frac{X_t}{x}\right) \;\Big|\; \|X_0\|>x\right] \\
&= \frac{E\left[f\left(\frac{X_{-s}}{x},\ldots,\frac{X_t}{x}\right)1\{\|X_0\|>x\}\right]}{P(\|X_0\|>x)} \\
&\to \int f d\mu_{s,t} \llcorner [\|x_0\|>1], \qquad x\to\infty,
\end{aligned}$$

according to Definition 2.3.1 with $V_k(x) = W(x) = P(\|X_0\| > x)$ and Lemma 2.5.20 and, on the other hand,

$$E\left[f\left(\frac{X_{-s}}{x},\ldots,\frac{X_t}{x}\right) \;\Big|\; \|X_0\|>x\right] \to E\left[f\left(Y_{-s},\ldots,Y_t\right)\right]$$

according to Theorem 2.3.2. The reader can note that with another $V_k \in RV_{-\alpha}$ in Definition 2.3.1 (for instance $V_k = U_k$) the stated property may do not hold anymore.

The coefficient of tail dependence λ_t is

$$\begin{aligned}
\lambda_t &:= \lim_{x\to\infty} P(\|X_t\|>x \mid \|X_0\|>x) \\
&= P(\|Y_t\|>1) \\
&= P(\|Y_0\|\|\Theta_t\|>1) \\
&= \int_1^\infty P(y\|\Theta_t\|>1)d(-y^{-\alpha}) \\
&= \int_1^\infty P(\|\Theta_t\|^\alpha > y^{-\alpha})d(-y^{-\alpha}) \\
&= \int_0^1 P(\|\Theta_t\|^\alpha > u)du \\
&= \int_0^\infty P[\min(\|\Theta_t\|^\alpha,1)>u]du \\
&= E[\min(\|\Theta_t\|^\alpha,1)].
\end{aligned}$$

The candidate extremal index θ^* is

$$
\begin{aligned}
\theta^* &:= \lim_{t\to\infty}\lim_{x\to\infty} P(\max_{i=1,\dots,t}\|X_i\| \leqslant x \mid \|X_0\| > x) \\
&= \lim_{t\to\infty} P(\max_{i=1,\dots,t}\|Y_i\| \leqslant 1) \\
&= P(\sup_{i\geqslant 1}\|Y_i\| \leqslant 1) = P(\|Y_0\|\sup_{i\geqslant 1}\|\Theta_i\| \leqslant 1) \\
&= \int_1^\infty P(y\sup_{i\geqslant 1}\|\Theta_i\| \leqslant 1)d(-y^{-\alpha}) \\
&= \int_1^\infty P(\sup_{i\geqslant 1}\|\Theta_i\|^\alpha \leqslant y^{-\alpha})d(-y^{-\alpha}) \\
&= \int_0^1 P(\sup_{i\geqslant 1}\|\Theta_i\|^\alpha \leqslant u)du \\
&= 1 - \int_0^1 P(\sup_{i\geqslant 1}\|\Theta_i\|^\alpha > u)du \\
&= E\max(1-\sup_{i\geqslant 1}\|\Theta_i\|^\alpha, 0) \\
&= E\left(\sup_{i\geqslant 0}\|\Theta_i\|^\alpha - \sup_{i\geqslant 1}\|\Theta_i\|^\alpha\right),
\end{aligned}
$$

the last step coming from the equalities

$$\max(1-x,0) = \max(1,x) - x = \max(\|\Theta_0\|, x) - x.$$

Concerning the extremogram, let a_n is a positive sequence satisfying $nP(\|X_0\| > a_n) \to 1$ as $n\to\infty$ and let A and B continuity sets of the distributions of Y_0 and Y_h respectively, such that A or B is included in $\{x \in \mathbb{B} : \|x\| > 1\}$. Then $A\times B \subset \{x\in\mathbb{B} : \|x\| > 1\}$ and $A\times B$ is a continuity set of (Y_0, Y_h) (recall that $\partial(A\times B) = \partial A\times \mathrm{cl}\, B \cup \mathrm{cl}\, A\times\partial B$). By the Continuous mapping theorem, Proposition 2.2.6 and Theorem 2.3.2, we have thus

$$
\begin{aligned}
\rho_{A,B}(h) &:= \lim_{n\to\infty} nP\left(\frac{X_0}{a_n}\in A, \frac{X_h}{a_n}\in B\right) \\
&= \lim_{n\to\infty} nP\left(\frac{1}{a_n}(X_0, X_h)\in A\times B\right) \\
&= P(Y_0\in A, Y_h\in B)
\end{aligned}
$$

which is the claim. Note that, thanks to the homogeneity property of Proposition 2.2.2, $A = \{x \in \mathbb{B} : \|x\| > y_1\}$ and $B = \{x \in \mathbb{B} : \|x\| > y_2\}$ always are continuity sets of Y_0 and Y_h respectively.

Chapter 3

Point processes of extremes

Based on Meinguet and Segers (2010).

Abstract. The point process of extremes yields the distribution of the number of extremes of a heavy tailed functional time series in selected spatial areas. Under the finite-cluster condition and a mixing condition, its limit distribution is given by a compound Poisson process whose expression depends on the tail process, the spectral processes and the extremal index of the original series.

3.1 Introduction

A first application of the concepts of tail and spectral processes developed in Chapter 2 is the study of the limit distribution of the point process

$$N_n = \sum_{i=1}^{n} \delta_{X_i/a_n},$$

where $(X_t)_{t\in\mathbb{Z}}$ is a jointly regularly varying time series in a separable Banach space $\mathbb{B}$ and $a_n \to \infty$ is a normalizing sequence such that

$$nP(\|X_0\| > a_n) \to 1.$$

The point process N_n is an accurate tool to obtain information about extremes. As an example, we have the identity

$$P(\|X\|_{n-k,n} \leqslant a_n r) = P(N_n(\mathbb{B} \setminus B(0,r)) \leqslant k)$$

where $\|X\|_{n-k,n}$ is the $(n-k)$-order statistic of $\|X_1\|, \ldots, \|X_n\|$. Passing to the limit as $n \to \infty$ only leaves the extreme values that the process can have "above its normal level". In other terms, this limit yields the full distribution of the number of extremes in a selected spatial area.

The key step to obtain the limit distribution of N_n is to obtain the limit distribution of the *cluster process* defined in Section 3.2.

The limit distributions of $(\zeta_n)_{n\in\mathbb{N}}$ first, and then of $(N_n)_{n\in\mathbb{N}}$, will be given, under the finite-cluster condition and a mixing condition, in Section 3.2 by Theorem 3.2.1 and Theorem 3.2.2 respectively. The expression of these limit distributions will involve only the tail and spectral processes. The proofs towards these theorems and the intermediate results are given in Sections 3.3 to 3.8. The general idea is to show the convergence of the Laplace transforms.

To conclude, Example 3.2.3 establishes the equality between the candidate extremal index θ^* treated in this thesis and the extremal index θ originally defined in Leadbetter (1983) under the hypotheses of Theorem 3.2.2.

3.2 Definition and properties

Let $(X_t)_{t\in\mathbb{Z}}$ be a jointly regularly varying time series in a separable Banach space $\mathbb{B}$. Let $a_n \to \infty$ be a normalizing sequence such that $nP(\|X_0\| > a_n) \to 1$. The object of interest is the point process

$$N_n = \sum_{i=1}^{n} \delta_{X_i/a_n}.$$

One wishes to determine the limit distribution of N_n as $n \to \infty$. If the X_i are iid the answer is well-known: N_n weakly converges to a Poisson process N whose mean measure is the exponent measure μ of X_0. In the general stationary case, temporal dependence may cause extremes to appear in clusters. These clusters change the nature of the limit of N_n from Poisson to compound Poisson.

Here we assume that $(X_t)_{t\in\mathbb{Z}}$ satisfies two more conditions to obtain the limit distribution of N_n in the stationary case. Choose a sequence $(r_n)_{n\in\mathbb{N}}$ such that $r_n \to \infty$ and let $r_n/n \to 0$ as $n \to \infty$. The first condition is

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\Big(\max_{m\leqslant|t|\leqslant r_n} \|X_t\| > a_n u \mid \|X_0\| > a_n u\Big) = 0. \qquad \text{(C)}$$

The second condition is, denoting $k_n = \lfloor n/r_n \rfloor$, for nonnegative $f \in \mathcal{C}_{bd}$, to have

$$\lim_{n\to\infty} \{E\exp[-\sum_{i=1}^{n} f\left(\frac{X_i}{a_n}\right)]\} - \{E\exp[-\sum_{i=1}^{r_n} f\left(\frac{X_i}{a_n}\right)]\}^{k_n} = 0. \quad (\mathcal{A}(a_n))$$

The *finite-cluster condition* (C) prohibits the length of clusters of extremes of $\|X_t\|$ to be infinite. It was introduced in Smith (1992) and is in particular met for a large number of Markov chains.

The *mixing condition* $(\mathcal{A}(a_n))$ allows us to approach in distribution N_n by a sum of k_n iid point processes with the same distribution as $\sum_{i=1}^{r_n} \delta_{X_i/a_n}$. It was introduced in Davis and Hsing (1995) and for instance used in Davis and Mikosch (1998) to study ARCH models.

The main step to get the expression of the limit of N_n for a stationary sequence $(X_t)_{t\in\mathbb{Z}}$ is to derive an expression of the limit in distribution of the cluster process. Given $u > 0$, the *cluster process* of $(X_t)_{t\in\mathbb{Z}}$ is the sequence of point processes $(\zeta_n)_{n\in\mathbb{N}}$ in $M_0(\mathbb{B})$ having the same distribution as

$$\sum_{i=1}^{r_n} \delta_{X_i/(a_n u)} \qquad \text{conditionally on} \qquad \{\max_{i=1,\dots,r_n} \|X_i\| > a_n u\}.$$

Theorem 3.2.1 (Cluster process). *Let $(X_t)_{t\in\mathbb{Z}}$ be a jointly regularly varying time series in a separable Banach space $\mathbb{B}$. Let $(Y_t)_{t\in\mathbb{Z}}$ be the tail process of $(X_t)_{t\in\mathbb{Z}}$. Under condition* (C)*, for every $u > 0$, the cluster process ζ_n converges in distribution to a point process ζ on $M_0(\mathbb{B})$ having the same distribution as*

$$\sum_{t\in\mathbb{Z}} \delta_{Y_t} \qquad \textit{conditionally on} \qquad \{\sup_{t\leqslant -1} \|Y_t\| \leqslant 1\}.$$

A clarifying remark: according to Proposition 3.6.2, the probability of the conditioning set is equal to the candidate extremal index θ^* (see Section 2.4) of the series $(X_t)_{t\in\mathbb{Z}}$, i.e.

$$P(\sup_{t\leqslant -1} \|Y_t\| \leqslant 1) = \theta^*,$$

which is nonzero thanks to condition (C) (Proposition 3.6.4).

The following theorem yields the weak limit of N_n for every jointly regularly varying time series.

Theorem 3.2.2 (Point process of extremes). *Let $(X_t)_{t\in\mathbb{Z}}$ be a jointly regularly varying time series in a separable Banach space $\mathbb{B}$. Let $(\Theta_t)_{t\in\mathbb{Z}}$ be its spectral process. Let θ^* be the candidate extremal index of the series $(\|X_t\|)_{t\in\mathbb{Z}}$. Under conditions (C) and $(\mathcal{A}(a_n))$, N_n converges in distribution to a point process N in M_0 with Laplace functional $\Psi_N(f) = E\exp(-\int f dN)$ given by*

$$\Psi_N(f) = \exp\{-\int_0^\infty E[\exp(-\sum_{j=1}^\infty f(v\Theta_j)) - \exp(-\sum_{j=0}^\infty f(v\Theta_j))]d(-v^{-\alpha})\}$$

for nonnegative $f \in \mathcal{C}_{bd}(\mathbb{B})$. Furthermore, for every $u > 0$, the restriction $N^{(u)}$ of N to $\{x \in \mathbb{B} : \|x\| > u\}$ admits the representation

$$N^{(u)} \overset{d}{=} \sum_{i=1}^{T^{(u)}} \sum_{j=1}^\infty \delta_{uZ_{ij}} 1\{\|Z_{ij}\| > 1\},$$

where $T^{(u)}$ is a Poisson random variable with mean $\theta^* u^{-\alpha}$ and $\left(\sum_{j=1}^\infty \delta_{Z_{ij}}\right)_{i\geqslant 1}$ are iid copies of the asymptotic cluster process ζ and independent of $T^{(u)}$.

Note that $N^{(u)}$ is also equal to the restriction of $\sum_{i=1}^{T^{(u)}} \sum_{j=1}^\infty \delta_{uZ_{ij}}$ to $\mathbb{B}_0 \setminus B[0,u]$ since $\|Z_{ij}\| > 1$ if and only if $\|uZ_{ij}\| > u$.

Example 3.2.3. Theorem 3.2.2 allows us to prove that the parameter θ^*, defined in Section 3.6 as

$$\begin{aligned}
\theta^* &:= \lim_{t\to\infty} \lim_{x\to\infty} P(\max_{i=1,\dots,t} \|X_i\| \leqslant x \mid \|X_0\| > x) \\
&= P(\sup_{i\geqslant 1} \|Y_i\| \leqslant 1) \\
&= E\left(\sup_{i\geqslant 0} \|\Theta_i\|^\alpha - \sup_{i\geqslant 1} \|\Theta_i\|^\alpha\right),
\end{aligned}$$

is the extremal index of the series $(\|X_t\|)_{t\in\mathbb{Z}}$. Indeed, under the conditions (C) and $(\mathcal{A}(a_n))$, for every $u > 0$, as $n \to \infty$, the last theorem yields

$$\begin{aligned}
P(\max_{1\leqslant t\leqslant r_n} \|X_t\| \leqslant a_n u) &= P(N_n(\mathbb{B}\setminus B(0,u))=0) \\
&\to P(N(\mathbb{B}\setminus B(0,u))=0) \\
&= P(T^{(u)}=0) \\
&= e^{-\theta^* u^{-\alpha}}
\end{aligned}$$

since $T^{(u)}$ is Poisson with parameter $\lambda = \theta u^{-\alpha}$. This is the original definition of the extremal index θ of $(\|X_t\|)_{t\in\mathbb{Z}}$ (Leadbetter 1983).

The structure adopted here is based on the approach of Basrak and Segers (2008) for multivariate time series with an adaptation to spaces that are the not necessarily locally compact thanks to the concept of M_0-convergence of Hult and Lindskog (2006).

3.3 Laplace transform

The main tool used in this chapter to prove the convergence results is the Laplace transform of the distribution of a random object. It is a convenient way to characterize the distributions of point processes so that convergence in distribution becomes pointwise convergence.

First let us recall the construction of the Laplace transform of a random variable.

Given a nonnegative random variable

$$X : (\Omega, \mathcal{A}, P) \to [0, +\infty),$$

its Laplace transform is

$$\Psi_X : [0,+\infty) \to [0,1] : \lambda \mapsto \Psi_X(\lambda) = Ee^{-\lambda X}. \tag{3.1}$$

It uniquely determines the distribution of X (Feller 1971, XIII.1). If

$$X : (\Omega, \mathcal{A}, P) \to \mathbb{R}^d$$

is a random vector with nonnegative components, the Laplace transform of X is

$$\Psi_X : [0,+\infty)^d \to [0,1] : \lambda \mapsto \Psi_X(\lambda) = Ee^{-\sum_{i=1}^{d} \lambda_i X_i} \tag{3.2}$$

and its knowledge is also equivalent to the distribution of X (Feller 1971, XIII.8). If $(\mathbb{D}, d)$ is a complete separable metric space whose points are measures on a set $\mathbb{E}$, the Laplace transform of a random element

$$M : (\Omega, \mathcal{A}, P) \to \mathbb{D}$$

is given by

$$\Psi_M : \mathbb{A}^+ \to [0,1] : f \mapsto \Psi_M(f) = Ee^{-M(f)} \tag{3.3}$$

where $\mathbb{A}^+$ is the set of all nonnegative measurable and bounded functions from $\mathbb{E}$ to $\mathbb{R}_+$. Developing the expression yields

$$\Psi_M(f) = \int_\Omega e^{-\int_{\mathbb{E}} f dM_\omega} dP(\omega) = \int_{\mathbb{D}} e^{-\int_{\mathbb{E}} f d\mu} d\nu_M(\mu).$$

This is a two-level formula. Once more the knowledge of Ψ_M determines the law of M and conversely (Resnick 2008, Propositions 3.4 and 3.5).

Let $\mathcal{C}_{bd}^+$ denote the set of $f \in \mathcal{C}_{bd}$ that satisfy $f(x) \geqslant 0$ for every $x \in \mathbb{B}_0$.

Proposition 3.3.1. *If $\mathbb{D} = M_0$ and $\mathbb{E} = \mathbb{B}_0$, the Laplace functional Ψ_M on $\mathcal{C}_{bd}^+$ uniquely determines the distribution of M.*

Proof. Let $M : (\Omega, \mathcal{A}, P) \to (M_0, \mathcal{B}_{M_0})$ be a random measure. The distribution of M is the probability measure $M_\# P = P \circ M^{-1}$ on M_0. According to Hult and Lindskog (2006), a basis of open sets for $\nu \in M_0$ is given by

$$\begin{aligned} &B_\nu(k, f_1, \ldots, f_k, a, b) \\ &\qquad = \left\{ \mu \in M_0 \;\middle|\; (\forall\, 1 \leqslant i \leqslant k) : a < \int f_i \, d\mu - \int f_i \, d\nu < b \right\} \end{aligned}$$

where $k \in \mathbb{N}$, $f_i \in \mathcal{C}_{bd}$ and $a, b \in \mathbb{Q}$. The family B_ν is closed under finite intersections and hence is a π-system. As any open set in M_0 is a countable union of these elements since M_0 is separable, this kind of set generates the Borel σ-field $\mathcal{B}_{M_0}$. Hence, according to Theorem 2.5.15, it is sufficient to know the value of $M_\# P$ on the sets of the type $B_\phi(k, f_1, \dots, f_k, a, b)$. Since

$$\begin{aligned} &M_\# P\left(\left\{\mu \in M_0 \;\middle|\; (\forall\, 1 \leqslant i \leqslant k):\; a < \int f_i\, d\mu - \int f_i\, d\phi < b\right\}\right) \\ &= P\left(\left\{\omega \in \Omega \;\middle|\; (\forall\, 1 \leqslant i \leqslant k):\; a < \int f_i\, dM_\omega - \int f_i\, d\phi < b\right\}\right), \end{aligned}$$

it suffices to show that Ψ_M determines the joint distribution of the k-dimensional random vector

$$(M(f_1), \dots, M(f_k)).$$

Its distribution is uniquely determined by its Laplace transform given by (3.2):

$$\Psi_{(M(f_1),\dots,M(f_k))}(\lambda) = Ee^{-\sum\limits_{i=1}^{k} \lambda_i M(f_i)},$$

for $\lambda_1, \dots, \lambda_k \geqslant 0$. It is equal to

$$Ee^{-M\left(\sum\limits_{i=1}^{k} \lambda_i f_i\right)} = \Psi_M\left(\sum_{i=1}^{k} \lambda_i f_i\right)$$

and $\sum_{i=1}^{k} \lambda_i f_i$ well belongs $\mathcal{C}_{bd}^+$. Then the knowledge of $\Psi_M(f)$ for $f \in \mathcal{C}_{bd}^+$ only is sufficient. □

There is a useful convergence criterion for weak convergence of point processes.

Proposition 3.3.2. *Let η_n and η be random elements in $(M_0, \mathcal{B}_{M_0})$. Then $\eta_n \xrightarrow{d} \eta$ if and only if their Laplace transforms converge pointwise, i.e. for every $f \in \mathcal{C}_{bd}^+$: $\Psi_{\eta_n}(f) \to \Psi_\eta(f)$.*

Proof. ($\Rightarrow$) Given $f \in \mathcal{C}_{bd}^+$, the map from M_0 to $\mathbb{R}_+$ that takes μ and maps it onto $\mu(f)$ is continuous. Then, if $\eta_n \xrightarrow{d} \eta$, the Continuous mapping theorem (Resnick 2008, p 152) implies that $\eta_n(f) \xrightarrow{d} \eta(f)$ for every $f \in \mathcal{C}_{bd}^+$. Thus $e^{-\eta_n(f)} \xrightarrow{d} e^{-\eta(f)}$ and, by Lebesgue's dominated convergence theorem, $Ee^{-\eta_n(f)} \to Ee^{-\eta(f)}$.

($\Leftarrow$) Assume that for every $f \in \mathcal{C}_{bd}^{+}$ we have $\Psi_{\eta_n}(f) \to \Psi_{\eta}(f)$. We will show that (i) the sequence η_n is tight in M_0 and (ii) every weakly converging subsequence $\eta_{n'}$ of η_n admits η as weak limit. Properties (i) and (ii) together implies that η_n converges in distribution to η.

For (i), we use Proposition 2.5.18 which states that η_n is tight in M_0 if and only if for every $f \in \mathcal{C}_{bd}^{+}$ the sequence of random variables $\eta_n(f)$ is tight in $\mathbb{R}$. Let $f \in \mathcal{C}_{bd}^{+}$. According to (3.1), the Laplace transform of the random variable $\eta_n(f)$ is, for $\lambda > 0$,

$$Ee^{-\lambda\eta_n(f)} = Ee^{-\eta_n(\lambda f)}$$

which converges, since λf lies in $\mathcal{C}_{bd}^{+}$, to

$$Ee^{-\eta(\lambda f)} = Ee^{-\lambda\eta(f)}$$

under the assumption made here. Hence the sequence $\eta_n(f)$ weakly converges in $\mathbb{R}$ and is thus relatively compact. By Prohorov's theorem it is tight. By Proposition 2.5.18, η_n is tight in M_0.

For (ii), let $\eta_{n'}$ be any weakly converging subsequence of η_n. Call $\tilde{\eta}$ its weak limit. By the first part of the proof

$$\Psi_{\eta_{n'}}(f) \to \Psi_{\tilde{\eta}}(f).$$

The assumption

$$\Psi_{\eta_n}(f) \to \Psi_{\eta}(f)$$

implies that $\tilde{\eta} \overset{d}{=} \eta$. Hence $\eta_{n'} \xrightarrow{d} \eta$. □

3.4 Compound Poisson processes

Poisson and compound Poisson processes are the weak limits of N_n when the original series $(X_t)_{t\in\mathbb{Z}}$ is iid or stationary respectively. Here we introduce these processes and compute their Laplace transforms that will be useful these convergences in Sections 3.5 and 3.8.

Poisson processes, or *Poisson Random Measure (PRM)*, are particular random measures.

Definition 3.4.1. *A random measure*

$$N : (\Omega, \mathcal{A}, P) \to (M_0, \mathcal{B}_{M_0})$$

is a point measure *if*

$$N = \sum_{i=1}^{m} \delta_{X_i} \qquad or \qquad N = \sum_{i=1}^{\infty} \delta_{X_i}$$

for some $X_i : \Omega \to \mathbb{B}_0$. The set of all point measures is called $M_p \subset M_0$.

Note that, since $N \in M_0$, for every subset $A \subset \mathbb{B}_0$ bounded away from the origin, we have $N(A) < \infty$.

Definition 3.4.2. *A random measure*

$$N : (\Omega, \mathcal{A}, P) \to (M_0, \mathcal{B}_{M_0})$$

is a Poisson process with mean measure $\mu \in M_0$ *if*
(i) N is a point measure such that for every $A \in \mathcal{B}_{\mathbb{B}_0}$ and $k \in \mathbb{N}$,

$$P(N(A) = k) = \begin{cases} \dfrac{e^{-\mu(A)} \left(\mu(A)\right)^k}{k!} & \text{if } \mu(A) < +\infty, \\ 0 & \text{if } \mu(A) = \infty, \end{cases}$$

(ii) If $A_1, \ldots, A_k \in \mathcal{B}_{\mathbb{B}_0}$ are disjoint, then $N(A_1), \ldots, N(A_k)$ are independent random variables.

The Laplace transform of a Poisson process is a central concept. To obtain it we look at the definition of Section 3.3 in the special case $\mathbb{D} = M_0$ and $\mathbb{E} = \mathbb{B}_0$.

Proposition 3.4.3. *If N is PRM(μ), then its Laplace transform is given by*

$$\Psi_N(f) = e^{-\int_{\mathbb{B}_0} \left(1 - e^{-f(x)}\right) d\mu(x)}$$

for $f \in \mathcal{C}_{bd}^{+}$.

Proof. See Resnick (2007), Theorem 5.1. □

Recall that if τ is a Poisson(λ) random variable and if $X_1, X_2, \ldots$ are iid random elements with distribution ν and independent of τ, then

$$N(\omega, A) = \sum_{i=1}^{\tau(\omega)} \delta_{X_i(\omega)}(A)$$

is a Poisson process with mean measure $\lambda\nu$ (Resnick 2008, p 132).

The compound Poisson process is a generalization of a Poisson process in the following sense.

Definition 3.4.4. *A random measure*

$$Y : (\Omega, \mathcal{A}, P) \to (M_0, \mathcal{B}_{M_0})$$

is a compound Poisson process *if it admits the representation*

$$Y(\omega, A) \stackrel{d}{=} \sum_{i=1}^{\tau(\omega)} D_i(\omega, A)$$

where τ is a Poisson(λ) random variable and $D_1, D_2, \dots$ are nonnegative iid copies independent of τ of a random measure $D : (\Omega, \mathcal{A}, P) \to (M_0, \mathcal{B}_{M_0})$.

Its Laplace transform is the following.

Proposition 3.4.5. *The Laplace transform of a compound Poisson process is given by*

$$\Psi_Y(f) = e^{-\lambda(1 - \Psi_D(f))}$$

for $f \in \mathcal{C}_{bd}^+$.

Proof. Let $f \in \mathcal{C}_{bd}^+$. Remember that the moment-generating function of the Poisson(λ) random variable τ is given by

$$\varphi_\tau(t) := Ee^{\tau t} = e^{-\lambda(1 - e^t)}.$$

As a consequence,

$$
\begin{aligned}
\Psi_Y(f) & \\
&= Ee^{-Y(f)} \\
&= E_\tau\, E_D\, e^{-(\sum_{i=1}^{\tau} D_i)(f)} \\
&= E_\tau\, E_D\, e^{-\sum_{i=1}^{\tau} D_i(f)} \\
&= E_\tau\, E_D \prod_{i=1}^{\tau} e^{-D_i(f)} \\
&= E_\tau \prod_{i=1}^{\tau} E_D\, e^{-D_i(f)} \\
&= E_\tau \prod_{i=1}^{\tau} E_D\, e^{-D(f)}
\end{aligned}
\qquad
\begin{aligned}
&= E_\tau \left(E_D\, e^{-D(f)}\right)^\tau \\
&= E_\tau \left(\Psi_D(f)\right)^\tau \\
&= E_\tau \left(e^{\tau \ln \Psi_D(f)}\right) \\
&= \varphi_\tau \left(\ln \Psi_D(f)\right) \\
&= e^{-\lambda(1 - e^{(\ln \Psi_D(f))})} \\
&= e^{-\lambda(1 - \Psi_D(f))}
\end{aligned}
$$

which is the expected formula. □

3.5 Point process limit for iid trials

This section establishes the fundamental link between regularly varying objects, point processes and Poisson processes. This result is well-known in the finite-dimensional case (Resnick 2008, Proposition 3.22) and in the infinite-dimensional case (de Haan and Ferreira 2006, Theorem 9.3.9). A proof is again given here to see its translation in Banach spaces using the set of measures M_0.

Proposition 3.5.1. *Let X be a random element in a separable Banach space $\mathbb{B}$. Then X is regularly varying with index $\alpha > 0$ and exponent measure $\mu \in M_0$ if and only if, given iid copies $X_1, X_2, \ldots$ of X, there* *exists a sequence* (a_n) *regularly varying with index* $1/\alpha$ *such that* $N_n \xrightarrow{d} N$*, where*

$$N_n = \sum_{i=1}^{n} \delta_{\frac{X_i}{a_n}} \qquad \textit{and} \qquad N \textit{ is } PRM(\mu).$$

Proof. As we have seen in paragraph 3.3, the conclusion $N_n \xrightarrow{d} N$ holds if and only if, for every $f \in \mathcal{C}_{bd}^{+}$, we have $\Psi_{N_n}(f) \to \Psi_N(f)$ as n goes to infinity, Ψ indicating the Laplace transform of the process. Let us compute

$$\begin{aligned}
&\Psi_{N_n}(f) \\
&= Ee^{-N_n(f)} \\
&= Ee^{-\sum_{i=1}^{n}\int f d\delta_{\frac{X_i}{a_n}}} \\
&= Ee^{-\sum_{i=1}^{n} f\left(\frac{X_i}{a_n}\right)} \\
&= E\prod_{i=1}^{n} e^{-f\left(\frac{X_i}{a_n}\right)} \\
&= \prod_{i=1}^{n} Ee^{-f\left(\frac{X_i}{a_n}\right)} \\
&= \prod_{i=1}^{n} Ee^{-f\left(\frac{X}{a_n}\right)}
\end{aligned}
\qquad
\begin{aligned}
&= [Ee^{-f\left(\frac{X}{a_n}\right)}]^n \\
&= [\int_{\mathbb{B}_0} e^{-f\left(\frac{x}{a_n}\right)} d\nu_X(x)]^n \\
&= [1 - \int_{\mathbb{B}_0} (1 - e^{-f\left(\frac{x}{a_n}\right)}) d\nu_X(x)]^n \\
&= [1 - \frac{1}{n}\int_{\mathbb{B}_0} (1 - e^{-f\left(\frac{x}{a_n}\right)}) n d\nu_X(f)]^n \\
&= [1 - \frac{1}{n}\int_{\mathbb{B}_0} (1 - e^{-f\left(\frac{x}{a_n}\right)}) nP\,(X \in dx)]^n \\
&= [1 - \frac{1}{n}\int_{\mathbb{B}_0} (1 - e^{-f\,(x)} n) P\,(X \in a_n dx)]^n \\
&= \left(1 - \frac{c_n(f)}{n}\right)^n .
\end{aligned}$$

with $c_n(f) := \int_{\mathbb{B}_0} (1 - e^{-f\,(x)}) nP\,(X \in a_n dx)$. Moreover

$$\Psi_N(f) = e^{-\int_{\mathbb{B}_0}(1 - e^{-f(x)}) d\mu(x)} = e^{-c(f)}$$

with $c(f) := \int_{\mathbb{B}_0} (1-e^{-f(x)})d\mu(x)$ since N is a Poisson random measure with mean μ.

Now, if X is regularly varying with index $\alpha > 0$ and exponent measure μ, we have seen that an equivalent definition is

$$nP(X \in a_n dx) \xrightarrow{M_0} \mu, \qquad n \to \infty.$$

Hence $c_n(f) \to c(f)$ as $n \to \infty$ according to the fact that $1 - e^f \in \mathcal{C}_{bd}$, because $1 - e^f = 0$ if and only if $f = 0$.

Then

$$\begin{aligned}\lim_{n\to\infty} \Psi_{N_n}(f) &= \lim_{n\to\infty} \left(1 - \tfrac{c_n(f)}{n}\right)^n \\ &= e^{-c(f)} \\ &= \Psi_N(f),\end{aligned}$$

which proves that $N_n \xrightarrow{d} N$ by Proposition 3.3.2.

Conversely, if

$$\Psi_{N_n}(f) \to \Psi_N(f), \qquad n \to \infty,$$

i.e.

$$\left(1 - \frac{c_n(f)}{n}\right)^n \to e^{-c(f)}, \qquad n \to \infty,$$

this implies that $c_n(f) \to c(f)$ as $n \to \infty$. First note that $c_n(f)$ has a limit. If not, $\left(1 - \frac{c_n(f)}{n}\right)^n$ would not have one either since the exponential function is bijective. Thus the only possible value for the limit of $-c_n(f)$ as $n \to \infty$ is $-c(f)$. Finally $nP(X \in a_n dx) \xrightarrow{M_0} \mu$ as $n \to \infty$. □

3.6 Limiting conditions

There are two more conditions that a jointly regularly varying time series $(X_t)_{t\in\mathbb{Z}}$ will be assumed to have to obtain the limit distribution of N_n in the stationary case. They were defined in Section 3.2 but are recall in more precise settings here. First recall that $(a_n)_{n\in\mathbb{N}}$ is a normalizing sequence such that $nP(\|X_0\| > a_n) = 1$ and choose any sequence $(r_n)_{n\in\mathbb{N}}$

such that $r_n \to \infty$ and $r_n/n \to 0$ as $n \to \infty$. Notations that will be used are

$$M_n := \max_{1 \leqslant i \leqslant n} \|X_i\| \qquad \text{and} \qquad M_{s,t} := \max_{s \leqslant i \leqslant t} \|X_i\|.$$

The first condition can be either

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(\max_{m \leqslant |t| \leqslant r_n} \|X_t\| > a_n u \mid \|X_0\| > a_n u) = 0, \tag{C}$$

or

$$\lim_{m\to\infty} \limsup_{n\to\infty} \sum_{i=m+1}^{r_n} P(\|X_i\| > a_n u \mid \|X_1\| > a_n u) = 0, \tag{C'}$$

or

$$\begin{cases} \lim\limits_{m\to\infty} \limsup\limits_{n\to\infty} P(M_{m+1,r_n} > a_n u \mid \|X_1\| > a_n u) = 0, \\ \lim\limits_{m\to\infty} \limsup\limits_{n\to\infty} P(M_{r_n - m} > a_n u \mid \|X_1\| > a_n u) = 0, \end{cases} \tag{C''}$$

Denoting $k_n = \left\lfloor \dfrac{n}{r_n} \right\rfloor$, the second condition is that, for every $f \in \mathcal{C}_{bd}^{+}$,

$$\lim_{n\to\infty} Ee^{-\sum_{i=1}^{n} f\left(\frac{X_i}{a_n}\right)} - \{Ee^{-\sum_{i=1}^{r_n} f\left(\frac{X_i}{a_n}\right)}\}^{k_n} = 0. \tag{$\mathcal{A}(a_n)$}$$

Finite-cluster conditions (C), (C') and (C'') prohibit the size of clusters of large values of $\|X_t\|$ to be infinite. The implications between those conditions are

$$\begin{array}{ccc} \text{(C)} & \leftrightarrow & \text{(C}''\text{)} \\ \uparrow & \nearrow & \\ \text{(C}'\text{)} & & \end{array}.$$

For our purpose we have to refer to one of these three conditions. It will be the weaker, condition (C). Condition (C) was introduced in Smith (1992) and is in particular met for Markov chains.

The *mixing condition* ($\mathcal{A}(a_n)$) allows us to approach N_n by a sum of k_n iid point processes with the same distribution as $\sum_{i=1}^{r_n} \delta_{X_i/a_n}$. It was introduced in Davis and Hsing (1995) and also used in Davis and Mikosch (1998) for the ARCH model.

Recall that candidate extremal index defined is Section 2.4 is

$$\theta^* = \lim_{r\to\infty} \lim_{x\to\infty} P(M_r \leqslant x \mid \|X_0\| > x) \in [0,1].$$

Under conditions (C) and ($\mathcal{A}(a_n)$), θ^* actually is the extremal index θ of the series $(\|X_t\|)_{t\in\mathbb{Z}}$ (see Example 3.2.3) and in this case are thus interchangeable. For this section we also define

$$\theta_n = \frac{P(M_{r_n} > a_n u)}{r_n P(\|X_0\| > a_n u)}.$$

Note that the definition of θ_n is symmetric in time

$$\frac{1}{\theta_n} = E\left[\sum_{i=1}^{r_n} 1\{\|X_i\| > a_n u\} \;\middle|\; M_{r_n} > a_n u\right],$$

which means the expected size of clusters of extremes given that there is an extreme.

The main result here is Proposition 3.6.4 that states that, under condition (C), θ_n converges to θ^* as $n \to \infty$ and $\theta^* > 0$. This we will useful to study the convergence of the cluster process. Besides this shows that $1/\theta^*$ can be considered as the expected cluster size of $(X_t)_{t\in\mathbb{Z}}$.

To get to this result first some elementary properties.

Proposition 3.6.1. *If $(X_t)_{t\in\mathbb{Z}}$ is a jointly regularly varying time series in a separable Banach space $\mathbb{B}$, then $P(M_{r_n} > a_n u) \to 0$ as $n \to \infty$.*

Proof. By stationarity and the fact that $X_0 \in RV_{-\alpha}$,

$$\begin{aligned}
P(M_{r_n} > a_n u) &= P(\{\text{there exists } 1 \leqslant i \leqslant r_n \text{ such that } \|X_i\| > a_n u\}) \\
&\leqslant P(\{\|X_1\| > a_n u\} \cup \ldots \cup \{\|X_{r_n}\| > a_n u\}) \\
&\leqslant \sum_{i=1}^{r_n} P(\{\|X_i\| > a_n u\}) \\
&= \sum_{i=1}^{r_n} P(\{\|X_0\| > a_n u\}) \\
&= r_n P(\{\|X_0\| > a_n u\}) \\
&= \frac{r_n}{n} n P(\{\|X_0\| > a_n u\}) \\
&\to 0 u^{-\alpha} = 0 \quad \text{as } n \to \infty,
\end{aligned}$$

as stated. □

This explains why r_n/n must tend to zero as $n \to \infty$. In particular, the event $\{M_{r_n} > a_n u\}$ can be qualified as *rare*.

Proposition 3.6.2. *If* $(X_t)_{t\in\mathbb{Z}}$ *is a jointly regularly varying time series in a separable Banach space* $\mathbb{B}$ *admitting* $(Y_t)_{t\in\mathbb{Z}}$ *as tail process, then*

$$\theta^* = P(\sup_{i\geqslant 1} \|Y_i\| \leqslant 1)$$

where θ^* *is the candidate extremal index of the series* $(\|X_t\|)_{t\in\mathbb{Z}}$.

Proof. By Theorem 2.3.3, the Continuous mapping theorem (Resnick 2008, p 152) and the Levy monotone convergence theorem,

$$\begin{aligned}
\theta^* &= \lim_{r\to\infty}\lim_{x\to\infty} P(M_r \leqslant x \mid \|X_0\| > x)\\
&= \lim_{r\to\infty}\lim_{x\to\infty} P(\max_{i=1,\dots,r} \|X_i\| \leqslant x \mid \|X_0\| > x)\\
&= \lim_{r\to\infty}\lim_{x\to\infty} P(\max_{i=1,\dots,r} \frac{\|X_i\|}{x} \leqslant 1 \mid \|X_0\| > x)\\
&= \lim_{r\to\infty} P(\max_{i=1,\dots,r} \|Y_i\| \leqslant 1)\\
&= P(\sup_{i\geqslant 1} \|Y_i\| \leqslant 1),
\end{aligned}$$

the statement of the theorem. □

The third asymptotic result is also an immediate consequence of regular variation and condition *(C)*.

Proposition 3.6.3. *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a jointly regularly varying time series in a separable Banach space* $\mathbb{B}$*. Under condition* (C)*, for every* $u, v > 0$*,*

$$\lim_{m\to\infty}\limsup_{n\to\infty} P(M_{-r_n,-m} \vee M_{m,r_n} > a_n uv \mid \|X_0\| > a_n u) = 0.$$

Proof. Let $u > 0$ and $0 < v \leqslant 1$. Then thanks to regular variation and condition *(C)*

$$\begin{aligned}
&\lim_{m\to\infty}\limsup_{n\to\infty} P(M_{-r_n,-m}\vee M_{m,r_n} > a_n uv \mid \|X_0\| > a_n u)\\
&= \lim_{m\to\infty}\limsup_{n\to\infty} \frac{P(\max_{m\leqslant|t|\leqslant r_n}\|X_t\| > a_n uv, \|X_0\| > a_n u)}{P(\|X_0\| > a_n uv)} \cdot \frac{P(\|X_0\| > a_n uv)}{P(\|X_0\| > a_n u)}\\
&\leqslant \lim_{m\to\infty}\limsup_{n\to\infty} \frac{P(\max_{m\leqslant|t|\leqslant r_n}\|X_t\| > a_n uv, \|X_0\| > a_n uv)}{P(\|X_0\| > a_n uv)} \cdot \frac{P(\|X_0\| > a_n uv)}{P(\|X_0\| > a_n u)}\\
&= \lim_{m\to\infty}\limsup_{n\to\infty} P(\max_{m\leqslant|t|\leqslant r_n}\|X_t\| > a_n uv \mid \|X_0\| > a_n uv) \cdot \frac{P(\|X_0\| > a_n uv)}{P(\|X_0\| > a_n u)}\\
&= 0v^{-\alpha} = 0.
\end{aligned}$$

If $v \geqslant 1$ the same result holds thanks to condition (C) only:

$$\begin{aligned}
&\lim_{m\to\infty}\limsup_{n\to\infty} P(M_{-r_n,-m}\vee M_{m,r_n} > a_n uv \mid \|X_0\| > a_n u)\\
&\leqslant \lim_{m\to\infty}\limsup_{n\to\infty} P(M_{-r_n,-m}\vee M_{m,r_n} > a_n u \mid \|X_0\| > a_n u)\\
&= 0.
\end{aligned}$$

Thus the result is true for every $v > 0$. □

Proposition 3.6.4. *Let $(X_t)_{t\in\mathbb{Z}}$ be a jointly regularly varying time series in a separable Banach space $\mathbb{B}$. Let $(Y_t)_{t\in\mathbb{Z}}$ be the tail process of $(X_t)_{t\in\mathbb{Z}}$. Under condition* (C)*, $P(Y_t \to 0$ as $|t| \to \infty) = 1$ ($Y_t \to 0$ meaning $\|Y_t\| \to 0$) and, for every $u > 0$, we have the further equalities:*

$$\begin{aligned}
\theta^* &= \lim_{n\to\infty}\theta_n\\
&= \lim_{n\to\infty} P(M_{r_n} \leqslant a_n u \mid \|X_0\| > a_n u) > 0.
\end{aligned}$$

Proof. According to the preceding result, by Theorem 2.3.3, the Continuous mapping theorem (Resnick 2008, p 152), for every $\varepsilon > 0$ and every $v > 0$, there exists $m \in \mathbb{N}$ such that for every $r \geqslant m$:

$$P(\max_{m \leqslant |t| \leqslant r_n} \|Y_t\| > v) \leqslant \varepsilon \qquad \text{i.e.} \qquad P(\max_{m \leqslant |t| \leqslant r_n} \|Y_t\| < v) \geqslant 1 - \varepsilon.$$

This implies that

$$P(\|Y_t\| \to 0 \text{ as } |t| \to \infty) = 1.$$

To prove that $\theta^* = \lim_{n\to\infty} \theta_n$ and is strictly positive, let

$$\theta_{m,n} := P(M_m \leqslant a_n u \mid \|X_0\| > a_n u)$$

and

$$\theta_m^Y := P(\max_{i=1,\dots,m} \|Y_i\| \leqslant 1).$$

By Segers (2005), Section 2 p 332, for which the results stay valid in Banach spaces, under condition (C) we have $\liminf\limits_{n\to\infty} \theta_n > 0$. So if θ_n has a limit, it is positive. By Theorem 2.3.3 and the Continuous mapping theorem (Resnick 2008, p 152),

$$\lim_{n\to\infty} \theta_{m,n} = \theta_m^Y.$$

By the Levy monotone convergence theorem,

$$\lim_{m\to\infty} \theta_m^Y = P(\sup_{i\geqslant 1} \|Y_i\| \leqslant 1) = \theta^*.$$

By Segers (2005), Theorem 3.1, Equation (5),

$$\lim_{m\to\infty} \limsup_{n\to\infty} |\theta_n - \theta_{m,n}| = 0.$$

Then

$$\limsup_{n\to\infty} |\theta_n - \theta^*| \leqslant \lim_{m\to\infty} \limsup_{n\to\infty} \left(|\theta_n - \theta_{m,n}| + |\theta_{m,n} - \theta_m^Y| + |\theta_m^Y - \theta^*|\right) = 0$$

hence $\limsup\limits_{n\to\infty} |\theta_n - \theta^*| = 0$ i.e. $\lim\limits_{n\to\infty} |\theta_n - \theta^*| = 0$ since $|\theta_n - \theta^*| \geqslant 0$.

It remains to prove that

$$\lim_{n\to\infty} \theta_n = \lim_{n\to\infty} P(M_{r_n} \leqslant a_n u \mid \|X_0\| > a_n u).$$

Notice that this equality is false without the limit as it can be shown computing both expression in the iid case. The elements in the right-hand side are equal to $\theta_{r_n,n}$ so it is equivalent to prove that

$$\lim_{n\to\infty} |\theta^* - \theta_{r_n,n}| = 0.$$

For $r_n \geqslant m$, we have $\theta_{m,n} \geqslant \theta_{r_n,n}$ and

$$\begin{aligned} 0 &\leqslant \theta_{m,n} - \theta_{r_n,n} \\ &= P(M_m \leqslant a_n u < M_{r_n} \mid \|X_0\| > a_n u) \\ &= P(M_m \leqslant a_n u < M_{m+1,r_n} \mid \|X_0\| > a_n u) \\ &\leqslant P(M_{m+1,r_n} \geqslant a_n u \mid \|X_0\| > a_n u). \end{aligned}$$

Thus, by condition (C), we have

$$\lim_{m\to\infty} \limsup_{n\to\infty} |\theta_{m,n} - \theta_{r_n,n}| = 0.$$

Hence

$$\begin{aligned} &\limsup_{n\to\infty} |\theta^* - \theta_{r_n,n}| \\ &\qquad \leqslant \lim_{m\to\infty} \limsup_{n\to\infty} \left(|\theta^* - \theta_n| + |\theta_n - \theta_{m,n}| + |\theta_{m,n} - \theta_{r_n,n}|\right) = 0 \end{aligned}$$

and, finally, $\limsup_{n\to\infty} |\theta^* - \theta_{r_n,n}| = 0$ i.e.

$$\lim_{n\to\infty} |\theta^* - \theta_{r_n,n}| = 0$$

since $|\theta^* - \theta_{r_n,n}| \geqslant 0$. □

3.7 Convergence of the cluster process

A study of the asymptotic distribution of the cluster process is the last step before getting the asymptotic distribution of N_n.

Proof of Theorem 3.2.1 (page 67). First let us show that ζ is well defined, that is $P(\sup_{t\leqslant -1} \|Y_t\| \leqslant 1) \neq 0$ and ζ is a point process in M_0. For the first point, we will at the end of this proof that $P(\sup_{t\leqslant -1} \|Y_t\| \leqslant$

$1) = \theta^* > 0$. For the second, if the random elements Y_t are defined on the probability space Ω', we know that for almost $\omega' \in \Omega'$ $\lim_{t\to\infty} \|Y_t(\omega')\| = 0$ (Proposition 3.6.4). This implies for these ω' that $(\forall r > 0) : \#\{Y_t(\omega') \in \mathbb{B}_0 \setminus B(0,r) \mid t \in \mathbb{Z}\} < +\infty$. Then, possibly eliminating the set of probability zero where this does not hold, we can assume that $\sum_{t\in\mathbb{Z}} \delta_{Y_t} \in M_0$.

Let $u > 0$. In order to prove that $\zeta_n \xrightarrow{d} \zeta$, we will prove that their Laplace transforms converge pointwise, i.e. for every $f \in \mathcal{C}^+_{bd} : \Psi_{\zeta_n}(f) \to \Psi_\zeta(f)$ as $n \to \infty$. Let $f \in \mathcal{C}^+_{bd}$.

For $k \leqslant l$, define

$$c_n(k,l) = e^{-\sum_{j=k}^{l} f\left(\frac{X_j}{a_n u}\right)}.$$

Notice that

$$\Psi_{\zeta_n}(f) = E\left[c_n(1,r_n) \mid M_{r_n} > a_n u\right]$$

and

$$\Psi_\zeta(f) = E[e^{-\sum_{i=-\infty}^{\infty} f(Y_j)} \mid M^Y_{-\infty,-1} \leqslant 1].$$

We will decompose the event $\{M_{r_n} > a_n u\}$ with respect to the smallest $j \in \{1,\ldots,r_n\}$ such that $\|X_j\| > a_n u$. Note that

$$E\left[c_n(1,r_n)\ ;\ M_{r_n} > a_n u\right] = \sum_{j=1}^{r_n} E(c_n(1,r_n)\ ;\ M_{j-1} \leqslant a_n u < X_j).$$

By definition of $\mathcal{C}^+_{bd}$, there exists $v > 0$ such that $f \equiv 0$ on $B(0,v)$. Let m and n such that $2m+1 \leqslant r_n$ (so that the following choice of j makes senses). For $j \in \{m+1,\ldots,r_n-m\}$ we have the following relation "$c_n(1,r_n) = c_n(j-m,j+m)$ except if $M_{1,j-m-1} \vee M_{j+m+1,r_n} > a_n uv$". Indeed $\|X_j\| \leqslant a_n uv$ if and only if $\|X_j\|/(a_n u) \leqslant v$, which implies that

$f\left(\|X_j\|/(a_n u)\right) = 0$. For these j,

$$
\begin{aligned}
\Delta_{m,n}(j) &:= \big| E\left[c_n(1, r_n);\ M_{j-1} \leqslant a_n u < \|X_j\|\right] \\
&\quad - E\left[c_n(j-m, j+m);\ M_{j-m,j-1} \leqslant a_n u < \|X_j\|\right] \big| \\
&= \big| E\big[c_n(1, r_n) 1_{M_{j-1} \leqslant a_n u < \|X_j\|} \\
&\qquad\qquad - c_n(j-m, j+m) 1_{\{M_{j-m,j-1} \leqslant a_n u < \|X_j\|\}}\big] \big| \\
&= \big| E\big[c_n(1, r_n) 1_{M_{j-1} \leqslant a_n u < \|X_j\|} \\
&\qquad\qquad - c_n(j-m, j+m) 1_{\{M_{j-m,j-1} \leqslant a_n u < \|X_j\|\}}\big] \big| \\
&\leqslant \big| E\big[c_n(1, r_n) - c_n(j-m, j+m)\ ;\ a_n u < \|X_j\|\big] \big| \\
&\leqslant P(M_{1,j-m-1} \vee M_{j+m+1,r_n} > a_n u v, \|X_j\| > a_n u).
\end{aligned}
$$

The second inequality follows from the above relation and the first from the fact that

$$
M_j \leqslant a_n u < \|X_j\| \qquad \text{implies} \qquad M_{j-m,j-1} \leqslant a_n u < \|X_j\|,
$$

so that

$$
1_{\{M_j \leqslant a_n u < \|X_j\|\}} \leqslant 1_{\{M_{j-m,j-1} \leqslant a_n u < \|X_j\|\}} \leqslant 1_{\{a_n u < \|X_j\|\}},
$$

together with the fact that $0 \leqslant c_n(j-m, j+m) - c_n(1, r_n) \leqslant 1$.

Using the definition of the conditional expectation, the definition of θ_n, the preceding expansion of $E\left[c_n(1, r_n)\ ;\ M_{r_n} > a_n u\right]$, the stationarity, the estimation of $\Delta_{m,n}(j)$ for $j \in \{m+1, \ldots, r_n - m\}$ and the fact that $0 \leqslant c_n(j-m, j+m) - c_n(1, r_n) \leqslant 1$ for $j \notin \{m+1, \ldots, r_n - m\}$, we obtain

$$
\begin{aligned}
\Delta_{m,n} &:= \big| E\left[c_n(1,r_n) \mid M_{r_n} > a_n u\right] \\
&\qquad - \frac{1}{\theta_n} E\left[c_n(-m,m)\ ;\ M_{-m,-1} \leqslant a_n u \mid \|X_0\| > a_n u\right]\big| \\
&= \frac{1}{P(M_{r_n} > a_n u)} \big| E\left[c_n(1,r_n)\ ;\ M_{r_n} > a_n u\right] \\
&\qquad - r_n E\left[c_n(-m,m)\ ;\ M_{-m,-1} \leqslant a_n u < \|X_0\|\right]\big| \\
&= \frac{1}{P(M_{r_n} > a_n u)} \big|\sum_{j=1}^{r_n} E(c_n(1,r_n); M_{j-1} \leqslant a_n u < X_j) \\
&\qquad - r_n E\left[c_n(-m,m); M_{-m,-1} \leqslant a_n u < \|X_0\|\right]\big| \\
&= \frac{1}{P(M_{r_n} > a_n u)} \big|\sum_{j=1}^{r_n} \big\{ E(c_n(1,r_n); M_{j-1} \leqslant a_n u < \|X_j\|) \\
&\qquad - E\left[c_n(j-m, j+m); M_{j-m,j-1} \leqslant a_n u < \|X_j\|\right]\big\}\big| \\
&\leqslant \frac{1}{P(M_{r_n} > a_n u)} \left|(r_n - 2m)\Delta_{m,n}(0) + 2mP(\|X_0\| > a_n u)\right| \\
&\leqslant \frac{1}{P(M_{r_n} > a_n u)} \big| r_n P(M_{1,-m-1} \vee M_{m+1,r_n} > a_n uv, \|X_0\| > a_n u) \\
&\qquad + 2mP(\|X_0\| > a_n u)\big| \\
&= \frac{1}{\theta_n}\left\{ P(M_{1,-m-1} \vee M_{m+1,r_n} > a_n uv \mid \|X_0\| > a_n u) + \frac{2m}{r_n} \right\}.
\end{aligned}
$$

As $\lim_{n\to\infty} \theta_n = \theta^* > 0$, since the first term in braces tends to 0 and $r_n \to \infty$, we have $\lim_{m\to\infty} \limsup_{n\to\infty} \Delta_{m,n} = 0$.

From Theorem 2.3.3 and the Portmanteau lemma, since $f \in \mathcal{C}_b^+$ implies

$$
e^{-\sum_{i=-m}^{m} f\left(\frac{f_j}{a_n u}\right)} \in \mathcal{C}_b^+,
$$

we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{\theta_n} E\left[c_n(-m,m)\ ;\ M_{-m,-1} \leqslant a_n u \mid \|X_0\| > a_n u\right] \\
&= \lim_{n\to\infty} \frac{1}{\theta_n} E[e^{-\sum_{i=-m}^{m} f\left(\frac{X_j}{a_n u}\right)}\ ;\ M_{-m,-1} \leqslant a_n u \mid \|X_0\| > a_n u]. \\
&= \frac{1}{\theta^*} E[e^{-\sum_{i=-m}^{m} f(Y_j)}\ ;\ M^Y_{-m,-1} \leqslant 1].
\end{aligned}$$

Then, using the result about $\Delta_{m,n}$ and, as $Y_t \to 0$ if $|t| \to \infty$, Lebesgue's dominated convergence theorem,

$$\begin{aligned}
&\lim_{n\to\infty} E\left[c_n(1,r_n) \mid M_{r_n} > a_n u\right] \\
&= \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{\theta_n} E\left[c_n(-m,m)\ ;\ M_{-m,-1} \leqslant a_n u \mid \|X_0\| > a_n u\right] \\
&= \frac{1}{\theta^*} E[e^{-\sum_{i=-\infty}^{\infty} f(Y_j)}\ ;\ M^Y_{-\infty,-1} \leqslant 1] \\
&= E[e^{-\sum_{i=-\infty}^{\infty} f(Y_j)} \mid M^Y_{-\infty,-1} \leqslant 1],
\end{aligned}$$

the last equality coming from the fact that $\theta^* = P(M^Y_{-\infty,-1} \leqslant 1)$. This can be seen by taking $f \equiv 0$ in the preceding equality, which yields

$$\begin{aligned}
&\lim_{n\to\infty} E\left[1 \mid M_{r_n} > a_n u\right] = \frac{1}{\theta^*} E\left[1\{M^Y_{-\infty,-1} \leqslant 1\}\right] \\
&\Leftrightarrow \lim_{n\to\infty} \frac{E\left[1\{M_{r_n} > a_n u\}\right]}{P(M_{r_n} > a_n u)} = \frac{1}{\theta^*} P(M^Y_{-\infty,-1} \leqslant 1) \\
&\Leftrightarrow 1 = \frac{1}{\theta^*} P(M^Y_{-\infty,-1} \leqslant 1).
\end{aligned}$$

Thus $\Psi_{\zeta_n}(f) \to \Psi_{\zeta}(f)$ as $n \to \infty$ for every $f \in \mathcal{C}^+_{bd}$. According to Proposition 3.3.2, this is equivalent to $\zeta_n \xrightarrow{d} \zeta$. □

Proposition 3.7.1. *The Laplace transform of the limit point process ζ is given by*

$$\Psi_\zeta(f) = \frac{1}{\theta^*}\int_0^\infty E[e^{-\sum_{i=0}^{\infty} f(y\Theta_i)} 1\{y \sup_{i\geqslant 0}\|\Theta_i\| > 1\} - e^{-\sum_{i=1}^{\infty} f(y\Theta_i)} 1\{y \sup_{i\geqslant 1}\|\Theta_i\| > 1\}]d(-y^{-\alpha})$$

for every $f \in \mathcal{C}_{bd}^+$. Furthermore if $f \equiv 0$ on $B(0,1)$, this expression becomes

$$\begin{aligned}\Psi_\zeta(f) &= \frac{1}{\theta^*}E[e^{-\sum_{i=0}^{\infty} f(Y_i)} - e^{-\sum_{i=1}^{\infty} f(Y_i)} 1\{M_{1,\infty}^Y > 1\}] \\ &= 1 - \frac{1}{\theta^*}[e^{-\sum_{i=1}^{\infty} f(Y_i)} - e^{-\sum_{i=0}^{\infty} f(Y_i)}].\end{aligned}$$

Proof. Let $f \in \mathcal{C}_{bd}^+$ and let $v > 0$ be such that $f(x) = 0$ if $\|x\| \leqslant v$. Recall that

$$\begin{aligned}\Psi_\zeta(f) &= Ee^{-\zeta(f)} \\ &= \frac{1}{\theta^*}E[e^{-\sum_{i=-\infty}^{\infty} f(Y_j)}\ ;\ M_{-\infty,-1}^Y \leqslant 1] \\ &= \frac{1}{\theta^*}\lim_{m\to\infty} E[e^{-\sum_{i=-m}^{m} f(Y_j)}\ ;\ M_{-\infty,-1}^Y \leqslant 1].\end{aligned}$$

For $m \geqslant 1$ and $q \in \{0, 1, \dots, m\}$, set

$$a_{m,q} = E[e^{-\sum_{i=-q}^{m} f(Y_j)}\ ;\ M_{-q,-1}^Y \leqslant 1].$$

Then

$$E[e^{-\sum_{i=-m}^{m} f(Y_j)}\ ;\ M_{-\infty,-1}^Y \leqslant 1] = a_{m,m} = a_{m,0} + \sum_{q=1}^{m}(a_{m,q} - a_{m,q-1}).$$

For $q \in \{0, 1, \dots, m\}$ we have

$$a_{m,q} - a_{m,q-1} = E[g_{m,q}(Y_{-q}, \dots, Y_m)]$$

where

$$\begin{aligned} g_{m,q}(x_{-q}, \dots, x_m) &= e^{-\sum_{i=-q}^{m} f(x_j)} 1\{\max_{j=-q,\dots,-1} \|x_j\| \leqslant 1\} \\ &\quad - e^{-\sum_{i=-q+1}^{m} f(x_j)} 1\{\max_{j=-q+1,\dots,-1} \|x_j\| \leqslant 1\}. \end{aligned}$$

Note that $g_{m,q}(x_{-q}, \dots, x_m) = 0$ if $\|x_{-q}\| \leqslant 1$ because then both terms become equal. Consequently, the version (2.12) of the time change formula (page 60) implies

$$\begin{aligned} & a_{m,q} - a_{m,q-1} \\ &= \int_0^\infty E[g_{m,q}(r\Theta_0, \dots, r\Theta_{m+q})\ ;\ r\|\Theta_q\|]d(-r^{-\alpha}) \\ &= \int_v^\infty E[e^{-\sum_{i=0}^{m+q} f(r\Theta_j)}\ ;\ rM_{0,q-1}^{\Theta} \leqslant 1 < r\|\Theta_q\|]d(-r^{-\alpha}) \\ &\quad - \int_v^\infty E[e^{-\sum_{i=1}^{m+q} f(r\Theta_j)}\ ;\ rM_{1,q-1}^{\Theta} \leqslant 1 < r\|\Theta_q\|]d(-r^{-\alpha}), \end{aligned}$$

remembering that if $r \leqslant v$ both terms are the same. Let

$$\begin{aligned} b_{m,q} &= \int_v^\infty E[e^{-\sum_{i=0}^{m} f(r\Theta_j)}\ ;\ rM_{0,q-1}^{\Theta} \leqslant 1 < r\|\Theta_q\|]d(-r^{-\alpha}) \\ &\quad - \int_v^\infty E[e^{-\sum_{i=1}^{m} f(r\Theta_j)}\ ;\ rM_{1,q-1}^{\Theta} \leqslant 1 < r\|\Theta_q\|]d(-r^{-\alpha}). \end{aligned}$$

Since $\sum_{j=l}^{m+q} f(r\Theta_j) = \sum_{j=l}^{m} f(r\Theta_j)$ unless, for some $t \geqslant m+1$, $r\|\Theta_t\| > v$, we have

$$
\begin{aligned}
&|(a_{m,q} - a_{m,q-1}) - b_{m,q}| \\
&\qquad \leqslant 2 \int_v^\infty P[rM^\Theta_{m+1,\infty} > v \ , \ rM^\Theta_{1,q-1} \leqslant 1 < r\|\Theta_q\|]d(-r^{-\alpha}),
\end{aligned}
$$

hence, as the events $\{rM^\Theta_{1,q-1} \leqslant 1 < r\|\Theta_q\|\}$ are disjoint and their union is $\{rM^\Theta_{1,m} > 1\}$,

$$
\begin{aligned}
&\left|\sum_{q=1}^m (a_{m,q} - a_{m,q-1}) - \sum_{q=1}^m b_{m,q}\right| \\
&\leqslant 2 \int_v^\infty \sum_{q=1}^m P[rM^\Theta_{m+1,\infty} > v \ , \ rM^\Theta_{1,q-1} \leqslant 1 < r\|\Theta_q\|]d(-r^{-\alpha}) \\
&= 2 \int_v^\infty P[rM^\Theta_{m+1,\infty} > v \ , \ rM^\Theta_{1,m} > 1]d(-r^{-\alpha}) \\
&\leqslant 2 \int_v^\infty P[rM^\Theta_{m+1,\infty} > v]d(-r^{-\alpha}) \\
&= 2v^{-\alpha} \int_1^\infty P[yM^\Theta_{m+1,\infty} > 1]d(-y^{-\alpha}) \\
&= 2v^{-\alpha} P[M^Y_{m+1,\infty} > 1].
\end{aligned}
$$

Without loss of generality, we can assume $v \leqslant 1$. Indeed, if $v \geqslant 1$, then $v = 1$ does also the job for f and, if $v \leqslant 1$, it suffices to keep it itself. In this case, since $\|Y_0\| \sim$Pareto(α) and $\|\Theta_0\| = 1$ almost surely,

$$
\begin{aligned}
a_{m,0} &= E[e^{-\sum_{i=0}^m f(Y_j)}] \\
&= \int_1^\infty E[e^{-\sum_{i=0}^m f(r\Theta_j)}]d(-r^{-\alpha}) \\
&= \int_v^\infty E[e^{-\sum_{i=0}^m f(r\Theta_j)} \ ; \ r\|\Theta_0\| > 1]d(-r^{-\alpha}).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
a_{m,0} + \sum_{q=1}^{m} b_{m,q} &= \int_v^\infty E[e^{-\sum_{j=0}^{m} f(r\Theta_j)}; rM_{0,m}^{\Theta} > 1] d(-r^{-\alpha}) \\
&\qquad - \int_v^\infty E[e^{-\sum_{j=1}^{m} f(r\Theta_j)}; rM_{1,m}^{\Theta} > 1] d(-r^{-\alpha}).
\end{aligned}
$$

Then

$$
\begin{aligned}
\Psi_\zeta(f) &= \frac{1}{\theta^*} \lim_{m\to\infty} E[e^{-\sum_{i=-m}^{m} f(Y_j)};\ M_{-\infty,-1}^{Y} \leqslant 1] \\
&= \frac{1}{\theta^*} \lim_{m\to\infty} \left(a_{m,0} + \sum_{q=1}^{m} (a_{m,q} - a_{m,q-1}) \right) \\
&= \frac{1}{\theta^*} \lim_{m\to\infty} \left(a_{m,0} + \sum_{q=1}^{m} b_{m,q} \right) \\
&= \frac{1}{\theta^*} \lim_{m\to\infty} \int_v^\infty E[e^{-\sum_{j=0}^{m} f(r\Theta_j)}; rM_{0,m}^{\Theta} > 1] d(-r^{-\alpha}) \\
&\qquad - \frac{1}{\theta^*} \lim_{m\to\infty} \int_v^\infty E[e^{-\sum_{j=1}^{m} f(r\Theta_j)}; rM_{1,m}^{\Theta} > 1] d(-r^{-\alpha}) \\
&= \frac{1}{\theta^*} \int_0^\infty E[e^{-\sum_{j=0}^{\infty} f(r\Theta_j)}; rM_{0,\infty}^{\Theta} > 1] d(-r^{-\alpha}) \\
&\qquad - \frac{1}{\theta^*} \int_0^\infty E[e^{-\sum_{j=1}^{\infty} f(r\Theta_j)}; rM_{1,\infty}^{\Theta} > 1] d(-r^{-\alpha}),
\end{aligned}
$$

which is the first part of the statement.

To prove the second formula for which $v = 1$, remark that

$$
E[e^{-\sum_{j=1}^{\infty} f(Y_j)}; M_{1,\infty}^{Y} \leqslant 1] = E[1; M_{1,\infty}^{Y} \leqslant 1] = P(M_{1,\infty}^{Y} \leqslant 1) = \theta^*,
$$

hence

$$
\begin{aligned}
\Psi_\zeta(f) &= \frac{1}{\theta^*}\{E[e^{-\sum_{j=0}^{\infty} f(Y_j)}] - E[e^{-\sum_{j=1}^{\infty} f(Y_j)} 1\{M_{1,\infty}^Y > 1\}]\} \\
&= \frac{1}{\theta^*}\{E[e^{-\sum_{j=0}^{\infty} f(Y_j)}] - E[e^{-\sum_{j=1}^{\infty} f(Y_j)} (1 - 1\{M_{1,\infty}^Y \leqslant 1\})]\} \\
&= \frac{1}{\theta^*}E[e^{-\sum_{j=0}^{\infty} f(Y_j)}] - \frac{1}{\theta^*}E[e^{-\sum_{j=1}^{\infty} f(Y_j)}] + 1 \\
&= 1 - \frac{1}{\theta^*}E[e^{-\sum_{j=1}^{\infty} f(Y_j)} - e^{-\sum_{j=0}^{\infty} f(Y_j)}],
\end{aligned}
$$

which is the expected formula. □

3.8 Point process limit for stationary time series

The goal of this chapter is to obtain the distribution of the point process of extremes for a jointly regularly varying time series.

Proof of Theorem 3.2.2 (page 68). First let us show that $\Psi_{N_n} \to \Psi_N$ pointwise on $\mathcal{C}_{bd}$. Let $f \in \mathcal{C}_{bd}^+$ and let $u > 0$ be such that $f(x) = 0$ if $\|x\| \leqslant u$.

Define

$$
\begin{aligned}
\phi_n(f) &:= k_n(1 - Ee^{-\sum_{j=1}^{r_n} f\left(\frac{X_j}{a_n}\right)}) \\
&= k_n(1 - E[e^{-\sum_{j=1}^{r_n} f\left(\frac{X_j}{a_n}\right)} ;\ M_{r_n} > a_n u]) \\
&= k_n P(M_{r_n} > a_n u) E[1 - e^{-\sum_{j=1}^{r_n} f\left(\frac{X_j}{a_n}\right)} \mid M_{r_n} > a_n u].
\end{aligned}
$$

On the one hand, by the definition of k_n, the regular variation of $\|X_0\|$, the definition and the properties of θ_n (see proof of Proposition 3.6.4

page 81),

$$\begin{aligned} k_n P(M_{r_n} > a_n u) &:= \frac{P(M_{r_n} > a_n u)}{r_n P(\|X_0\| > a_n u)} n P(\|X_0\| > a_n u) + o(1) \\ &= \theta_n n P(\|X_0\| > a_n u) + o(1) \\ &\to \theta^* u^{-\alpha} \quad \text{as} \quad n \to \infty. \end{aligned}$$

On the other hand, thanks to Proposition 3.2.1 and the Portmanteau Lemma,

$$E[1 - e^{-\sum_{j=1}^{r_n} f\left(\frac{X_j}{a_n}\right)} \mid M_{r_n} > a_n u] \to 1 - Ee^{-\sum_{j=1}^{\infty} f(Z_j u)}$$

as $n \to \infty$. Consequently,

$$\lim_{n\to\infty} \phi_n(f) = u^{-\alpha}\theta^*(1 - Ee^{-\sum_{j=1}^{\infty} f(Z_j u)}) =: \phi(f).$$

The map $g(x) := f(ux)$ belongs to $f \in \mathcal{C}_{bd}^+$ and is such that $g(x) = 0$ if $\|f\| \leqslant x$. Then, thanks to Proposition 3.7.1,

$$\begin{aligned} \phi(f) &= \theta^* u^{-\alpha}(1 - \{1 - \frac{1}{\theta^*} E[e^{-\sum_{i=1}^{\infty} f(uY_i)} - e^{-\sum_{i=0}^{\infty} f(uY_i)}]\}) \\ &= u^{-\alpha} E[e^{-\sum_{i=1}^{\infty} f(uY_i)} - e^{-\sum_{i=0}^{\infty} f(uY_i)}] \\ &= u^{-\alpha} \int_1^{\infty} E[e^{-\sum_{i=1}^{\infty} f(uy\Theta_i)} - e^{-\sum_{i=0}^{\infty} f(uy\Theta_i)}] d(-y^{-\alpha}) \\ &= \int_u^{\infty} E[e^{-\sum_{i=1}^{\infty} f(v\Theta_i)} - e^{-\sum_{i=0}^{\infty} f(v\Theta_i)}] d(-v^{-\alpha}) \\ &= \int_0^{\infty} E[e^{-\sum_{i=1}^{\infty} f(v\Theta_i)} - e^{-\sum_{i=0}^{\infty} f(v\Theta_i)}] d(-v^{-\alpha}), \end{aligned}$$

the last equality coming from the fact that, if $v \in (0,u]$, then $f(v\Theta_0)=0$ almost surely, so both terms become equal which makes the integrand equal to zero.

Condition $(\mathcal{A}(a_n))$ says that $\Psi_{N_n}(f)$ can be written as

$$\begin{aligned}\Psi_{N_n}(f) &= E[e^{-\sum_{j=1}^{r_n} f\left(\frac{X_j}{a_n}\right)}]^{k_n} + o(1)\\ &= \{1-\frac{1}{k_n}\phi_n(f)\}^{k_n} + o(1).\end{aligned}$$

with the notation in terms of ϕ_n, hence

$$\lim_{n\to\infty} \Psi_{N_n}(f) = e^{-\phi(f)} =: \Psi(f).$$

This proves the formula for the limit of $\Psi_{N_n}(f)$.

It remains to prove that $\Psi(f)$ is equal to the Laplace transform $\Psi_{N^{(u)}}(f)$. For that consider one of the previous expression of $\Psi(f)$ and, according to the Laplace transform of a compound Poisson random variable (Proposition 3.4.5),

$$\begin{aligned}\Psi(f) &= e^{-\theta^* u^{-\alpha}(1-Ee^{-\sum_{j=1}^{\infty} f(Z_j u)})}\\ &= e^{-\lambda_{N^{(u)}}(1-\Psi_{\sum_{j=1}^{\infty}\delta_{Z_j u}}(f))} = \Psi_N(f).\end{aligned}$$

which completes the proof according to Proposition 3.3.2. □

Remark. About the proof, the convergence of $\Psi_{N_n}(f)$ to $\Psi(f)$ for every $f \in \mathcal{C}_{bd}^+$ does already imply that $(N_n)_{n\in\mathbb{N}}$ converges to some point process N in M_0. Indeed the function $t \mapsto \Psi(tf)$ is continuous at $t=0$ and $t \mapsto \Psi_{N_n}(tf) = E\exp(-tN_n(f))$ is the Laplace transform of the non-negative random variable $N_n(f)$. By the continuity theorem for Laplace transforms (Feller 1971, Theorem VIII.2), each $(N_n(f))_{n\in\mathbb{N}}$ must have a weak limit and so is tight. This implies that N_n is tight and thus has a weak limit $N \in M_0$. Since the set of the point processes of M_0 is weakly closed, N is a point process. However, showing that $\Psi(f) = \Psi_N(f)$ for the right point process N allows us to conclude the same without these steps and gives directly the limit.

Chapter 4

Heavy tailed linear processes

Based on Meinguet and Segers (2010).

Abstract. The theory of heavy tailed time series applies to linear processes in Banach spaces. They are composed of infinite sums of linearly transformed independent random elements whose common distribution is regularly varying. Linear processes in Banach spaces generalize the well-known autoregressive moving average processes to functional data analysis. Explicit expressions of their tail processes, spectral processes and extremal indices are given here. These results are first obtained through a direct computation, and secondly by seeing linear processes as autoregressive processes of order one in a suitable sequence space.

4.1 Introduction

In this chapter we particularize the results of Chapter 2 to linear processes. Examples of linear processes are the well-known autoregressive moving average ARMA(p,q) time series

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where ε_t is the innovation of the process at time t. By developing the expression of X_{t-i}, every ARMA(p,q) process can be represented as a moving average with infinitely many coefficients:

$$X_t = \varepsilon_t + \sum_{i=1}^{\infty} \gamma_i \varepsilon_{t-i}.$$

The extension to Banach spaces of the multiplication $\gamma_i \varepsilon_{t-i}$ is a general linear operation $T_i(\varepsilon_{t-i})$. For instance, thinking of the concentration of polluting gas over a city as a spatial phenomenon, an idea of what could be a functional representation is

$$X_t(x) = \int K(x-y) X_{t-1}(y) dy + \varepsilon_t(x),$$

i.e. $X_t = K * X_{t-1} + \varepsilon_t$ where K is a kernel representing what becomes at time $t+1$ a particle of pollution deposited at the origin at time t. In other terms, the pollution today at location x is the average pollution around x of the day before plus the new contribution of the day.

In standard functional data analysis the innovations ε_t of linear processes are assumed to follow a Gaussian distribution. As in Embrechts et al. (1997), Chapter 7, the normality of the innovations is replaced here by a regularly varying law in the sense of Chapter 2.

To follow this aim, given two separable Banach spaces $\mathbb{B}_1$ and $\mathbb{B}_2$, the class of heavy tailed linear processes studied here takes the general form

$$X_t = \sum_{i \in \mathbb{Z}} T_i(Z_{t-i}), \qquad t \in \mathbb{Z},$$

where $T_i : \mathbb{B}_1 \to \mathbb{B}_2$ are continuous linear maps between $\mathbb{B}_1$ and $\mathbb{B}_2$ whereas the innovations $(Z_t)_{t\in\mathbb{Z}}$ are independent and identically distributed random elements in $\mathbb{B}_1$ whose common law is regularly varying.

In Section 4.2, Lemma 4.2.1 provides a condition for the almost sure convergence of X_t as an infinite sum. Joint regular variation is proved in Theorem 4.2.2 which yields an explicit expression for the distribution of the spectral process $(\Theta_t)_{t\in\mathbb{Z}}$. Its distribution depends on the spectral measure of the innovations and the maps T_i. It reflects the common extreme-value heuristic that large values in the series most likely arise from a single large shock among the innovations that caused the shock. In this case, the total value of the sum is approximately the value of the innovation. Proposition 4.2.3 is an application of Theorem 4.2.2 to the extremal index of the series $(\|X_t\|)_{t\in\mathbb{Z}}$. Proposition 4.2.4 gives a sufficient condition for autoregressive processes of order 1 to be jointly regularly varying.

The discussion ends in Section 4.3 with a study of the *shifted time series* $Z_t = (\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots)$. The series $(\zeta_t)_{t\in\mathbb{Z}}$ is an iid sequence of regularly varying random elements in a separable Banach space $\mathbb{B}_1$ and Z_t is seen as an autoregressive process of order 1 in the sequence space $\mathbb{B}_1^{\mathbb{N}}$. Every causal linear time series from $\mathbb{B}_1$ to $\mathbb{B}_2$ admits a representation as a linear transformation of this AR(1) process.

4.2 Definition and properties

Given two separable Banach spaces $\mathbb{B}_1$ and $\mathbb{B}_2$ and linear continuous maps $T_i : \mathbb{B}_1 \to \mathbb{B}_2$, linear processes have the form

$$X_t = \sum_{i\in\mathbb{Z}} T_i(Z_{t-i}),\ t \in \mathbb{Z},$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid regularly varying innovations in $\mathbb{B}_1$. Say that the Z_t commonly have the distribution of Z with index of regular variation $\alpha > 0$. The first result provides a wide criterion for the random elements X_t to exist and for the series to be stationary.

Lemma 4.2.1. *Let $(Z_t)_{t\in\mathbb{Z}}$ be a sequence of iid regularly varying random elements with index $\alpha > 0$ taking values in a Banach space $\mathbb{B}_1$. Let $\mathbb{B}_2$ be a Banach space and let $T_i : \mathbb{B}_1 \to \mathbb{B}_2$, $i \in \mathbb{Z}$, be linear continuous maps. If there exists $0 < \delta < \min(1, \alpha)$ such that*

$$\sum_{i\in\mathbb{Z}} \|T_i\|^\delta < \infty \tag{4.1}$$

then the sum $X_t = \sum_{i\in\mathbb{Z}} T_i(Z_{t-i})$ is almost surely convergent and the process $(X_t)_{t\in\mathbb{Z}}$ is stationary.

Proof. See page 105. □

If the law of Z is concentrated on a closed linear subspace $\mathbb{B}_{1,0}$ of $\mathbb{B}_1$, then (4.1) may be replaced by the weaker condition

$$\textstyle\sum_n \|T_n\|_0^\delta < \infty, \tag{4.2}$$

with δ as before and $\|T\|_0 = \sup\{\|Tx\| : x \in \mathbb{B}_{1,0}, \|x\| = 1\}$ the operator norm of the restriction of the linear operator $T : \mathbb{B}_1 \to \mathbb{B}_2$ to $\mathbb{B}_{1,0}$ (note that $\|T\|_0 \leqslant \|T\|$). Likewise, in Lemma 4.2.1, $\|T_n\|$ may be replaced by $\|T_n\|_0$. Further note that if $\alpha > 1$ and if the innovations have expectation zero ($\mathbb{B}_1 = \mathbb{B}_2 = \mathbb{R}$), condition (4.1) is not necessary for convergence of the series X_t, although in that case the convergence does not need to hold absolutely; see for instance Mikosch and Samorodnitsky (2000), Lemma A.3.

We have a general result about the joint regular variation of linear series satisfying the assumptions of Lemma 4.2.1.

Theorem 4.2.2. *Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be two separable Banach spaces. Let $(Z_i)_{i\in\mathbb{Z}}$ be a sequence of iid regularly varying random elements with index $\alpha > 0$ in $\mathbb{B}_1$ distributed according to a common random element Z. Let $T_i : \mathbb{B}_1 \to \mathbb{B}_2$, $i \in \mathbb{Z}$, be linear continuous maps satisfying* (4.1)*. If there exists $i \in \mathbb{Z}$ such that $E[\|T_i(\Theta^Z)\|^\alpha] \neq 0$, then the linear time series $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying with index α and spectral process*

$$E\left[f\left(\Theta_{-s},\dots,\Theta_t\right)\right] = \frac{\sum_{i\in\mathbb{Z}} E\left[f\left(\frac{T_{-s+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|},\dots,\frac{T_{t+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}\right)\|T_i(\Theta^Z)\|^\alpha\right]}{\sum_{i\in\mathbb{Z}} E[\|T_i(\Theta^Z)\|^\alpha]},$$

for every $f : \mathbb{B}_2^{t-s+1} \to \mathbb{R}$ continuous and bounded, with Θ^Z a random element in $\mathbb{B}_1$ with the spectral measure of Z as law.

Proof. See page 112. □

This spectral process can be decomposed into two independent components. Indeed,

$$(\Theta_{-s},\dots,\Theta_t) \stackrel{d}{=} \Theta(-s,t;S)$$

where S is a discrete random variable in $\mathbb{Z}$ independent of $(\Theta_t)_{t\in\mathbb{Z}}$, with distribution

$$P(S=i) = \frac{E[\|T_i(\Theta^Z)\|^\alpha]}{\sum_{j=-\infty}^{\infty} E[\|T_j(\Theta^Z)\|^\alpha]}, i \in \mathbb{Z},$$

and $\Theta(-s,t;i)$ is a random element in $\mathbb{B}_2^{s+t+1}$ whose law is defined by

$$E\left[f\left(\Theta(-s,t;i)\right)\right]$$
$$= E\left[f\left(\frac{T_{-s+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|},\dots,\frac{T_{t+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}\right)\|T_i(\Theta^Z)\|^\alpha\right]\frac{1}{E[\|T_i(\Theta^Z)\|^\alpha]}.$$

The following proposition gives an expression of the candidate extremal index θ^* of a linear time series in terms of the spectral process. Recall that the candidate extremal index θ^* is equal to the extremal index θ if, for instance, the finite-cluster condition (C) and the mixing condition ($\mathcal{A}(a_n)$) of Section 3.2 hold (see Example 3.2.3).

Proposition 4.2.3. *Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be two separable Banach spaces. Let $(Z_i)_{i\in\mathbb{Z}}$ be a sequence of iid regularly varying random elements with index $\alpha > 0$ in $\mathbb{B}_1$ distributed according to a common random element Z. Let $T_i : \mathbb{B}_1 \to \mathbb{B}_2$, $i \in \mathbb{Z}$, be linear continuous maps satisfying* (4.1)*. Let $(X_t)_{t\in\mathbb{Z}}$ be the linear process $X_t = \sum_{i\in\mathbb{Z}} T_i(Z_{t-i})$, $t \in \mathbb{Z}$. If there exists $i \in \mathbb{Z}$ such that $E[\|T_i(\Theta^Z)\|^\alpha] \neq 0$, then the candidate extremal index θ^* of the series $(\|X_t\|)_{t\in\mathbb{Z}}$ is given by*

$$\theta^* = \frac{1}{\sum_{t\in\mathbb{Z}} E[\|T_t(\Theta^Z)\|^\alpha]} E\left[\sup_{t\in\mathbb{Z}}\|T_t(\Theta^Z)\|^\alpha\right],$$

where Θ^Z is a random element in $\mathbb{B}_1$ having the spectral measure of Z as law.

Proof. See page 119. □

An example of jointly regularly varying linear series is the class of autoregressive processes with regularly varying iid innovations.

Given an operator $T : \mathbb{B} \to \mathbb{B}$, an AR(1) process is a time series recursively defined by

$$X_t = T(X_{t-1}) + Z_t. \tag{4.3}$$

The convergence condition (4.1) is easily checked for an AR(1) process.

Proposition 4.2.4. *If there exists $j \geqslant 1$ such that $\|T^j\| < 1$, then the AR(1) process* (4.3) *satisfies the convergence condition* (4.1)*.*

Proof. See page 119. □

4.3 Shift operator

Let $(\zeta_t)_{t\in\mathbb{Z}}$ be a sequence of iid regularly varying random elements with index $\alpha > 0$ in a separable Banach space $\mathbb{B}_1$. The *shifted time series* related to $(\zeta_t)_{t\in\mathbb{Z}}$ is the process

$$X_t = (\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \dots)$$

in $\mathbb{B}_1^{\mathbb{N}}$. The goal is to represent $(X_t)_{t\in\mathbb{Z}}$ as jointly regularly varying AR(1) process in the sequence space $\mathbb{B}_1^{\mathbb{N}}$ endowed with an appropriate norm. It will be shown that every causal linear time series from $\mathbb{B}_1$ to $\mathbb{B}_2$ admits a representation as linear transformation of this AR(1) process. The conclusion is that an AR(1) process is a linear time series and any causal linear time series is a transformation of an AR(1) process, both constructions leading of course to the same properties.

The first step of this section is to endow $\mathbb{B}_1^{\mathbb{N}}$ with a norm that will make $(X_t)_{t\in\mathbb{Z}}$ jointly regularly varying. Choose $0 < p < \min(1, \alpha)$ and a sequence $(w_i)_{i\geqslant 0}$ of strictly positive numbers such that $w_0 = 1$ and $\sum_{i=0}^{\infty} w_i^p < +\infty$. Define the norm of $x = (x_0, x_1, \dots) \in \mathbb{B}_1^{\mathbb{N}}$ by

$$\|x\| = \sum_{i=0}^{\infty} w_i \|x_i\|$$

and the set $\mathbb{B}_p = \{x \in \mathbb{B}_1^{\mathbb{N}} : \|x\| < +\infty\}$. First remark that $\mathbb{B}_p$ is a separable Banach space so that we can apply the theory of Chapter 2 to processes in $\mathbb{B}_p$.

Proposition 4.3.1. *The set $\mathbb{B}_p$ is a separable Banach space.*

Proof. First it is clear that $\mathbb{B}_p$ is a normed vector space.

For the completude, let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{B}_p$. The map $L(x_0, x_1, \dots) = (w_0 x_0, w_1 x_1, \dots)$ from $\mathbb{B}_p$ to l^1 is linear, continuous and invertible. Since $l^1(\mathbb{B}_1)$ is complete, the Cauchy sequence $(L(x_n))_{n\in\mathbb{N}}$ converges to an element $y \in l^1$. Consequently $(x_n)_{n\in\mathbb{N}}$ converges to $L^{-1}(y) \in \mathbb{B}_p$ and $\mathbb{B}_p$ is complete.

For the separability, let D be a countable dense subset of $\mathbb{B}_1$. Set

$$D^{\mathbb{N}} := \{(x_1, \dots, x_k, 0, 0, \dots) \mid x_i \in D,\ k \in \mathbb{N}\}.$$

The subset $D^{\mathbb{N}}$ is countable by construction. Let $y = (y_0, y_1, \dots) \in \mathbb{B}_p$ and $\varepsilon > 0$. Choose $k \in \mathbb{N}$ such that $\sum_{i=k}^{\infty} w_i \|y_i\| \leqslant \varepsilon/2$. Since $\mathbb{B}_1$ is separable, there exists $x_i \in D$ which is $(\varepsilon/2)/(2^{i+1}w_i)$-close to $y_i \in \mathbb{B}_1$, $0 \leqslant i < k$. Set $x = (x_1, \ldots, x_k, 0, 0, \ldots)$. Consequently,

$$\|x - y\| = \sum_{i=0}^{\infty} w_i \|x_i - y_i\| \leqslant \sum_{i=0}^{k-1} w_i \frac{\varepsilon/2}{2^{i+1}w_i} + \sum_{i=k}^{\infty} w_i \|y_i\| \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

which proves that $D^{\mathbb{N}}$ is dense in $\mathbb{B}_p$. □

The set $\mathbb{B}_p$ is a subspace of $\mathbb{B}_1^{\mathbb{N}}$ and the members of the shifted time series $(X_t)_{t\in\mathbb{Z}}$ with underlying process $(\zeta_t)_{t\in\mathbb{Z}}$ almost surely belongs to $\mathbb{B}_p$:

$$\begin{aligned} E\|X_t\|^p = E\left[\left(\sum_{i=0}^{\infty} w_i \|\zeta_{t-i}\|\right)^p\right] &\leqslant E\left[\sum_{i=0}^{\infty} w_i^p \|\zeta_{t-i}\|^p\right] \\ &= \sum_{i=0}^{\infty} w_i^p E\left[\|\zeta_{t-i}\|^p\right] = E\left[\|\zeta\|^p\right] \sum_{i=0}^{\infty} w_i^p < +\infty. \end{aligned}$$

The shifted process $(X_t)_{t\in\mathbb{Z}}$ can be viewed as an autoregressive process. If L is defined from $\mathbb{B}_p$ to $\mathbb{B}_p$ by

$$L(x_0, x_1, x_2, \ldots) = (0, x_0, x_1, x_2 \ldots),$$

then

$$X_t = (\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots) = L(X_{t-1}) + (\zeta_t, 0, 0, \ldots).$$

Call $Z_t := (\zeta_t, 0, 0, \ldots)$ the innovations of this process. They are regularly varying with index α and spectral measure $\lambda_Z = \mathcal{L}(\Theta^\zeta, 0, 0, \ldots)$ where Θ^ζ is a random element in $\mathbb{B}_1$ with distribution equal to the spectral measure of ζ. The following proposition gives the distribution of the forward spectral process of $(X_t)_{t\in\mathbb{Z}}$.

Proposition 4.3.2. *Let $(Z_i)_{i\in\mathbb{Z}}$ be a sequence of iid regularly varying random elements with index $\alpha > 0$ in $\mathbb{B}_1$ distributed according to a common random element Z. Then the time series $X_t = (\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots)$ is jointly regularly varying in $\mathbb{B}_p$. Its forward spectral process $(\Theta_t^X)_{t\geqslant 0}$ has the structure*

$$(\Theta_t^X)_{t\geqslant 0} \stackrel{d}{=} (L^t(\Theta_0^X))_{t\in\mathbb{Z}}$$

where Θ_0^X is a random element having the spectral measure of X_0 as law:

$$\Theta_0^X = (0, \ldots, 0, \frac{\Theta^\zeta}{w_I}, 0, \ldots),$$

the position of Θ^ζ / w_I being distributed according to the random variable I whose law is

$$P(I = i) = \frac{w_i^\alpha}{\sum_{j=0}^\infty w_j^\alpha}, \qquad i \in \{0, 1, 2, \ldots\},$$

and Θ^ζ being a random element in $\mathbb{B}_1$ distributed according to the spectral measure of ζ.

Proof. Let

$$\mathbb{B}_{p,0} = \{(x_0, x_1, x_2, \ldots) \in \mathbb{B}_p \mid x_1 = x_2 = \ldots = 0\}$$

be the subspace on which the law of the innovations Z_t are concentrated. Let $\|\cdot\|_0$ be the restriction to $\mathbb{B}_{p,0}$ of the operator norm $\|\cdot\|$ on $\mathbb{B}^{\mathbb{N}}$:

$$\|T\|_0 = \sup\{\|T(x)\| : x \in \mathbb{B}_{p,0}, \|x\| = 1\}.$$

Since $\|L^n\|_0 = w_n$ and $\sum_{n \in \mathbb{N}} w_n^p < +\infty$, there exists $j \in \mathbb{N}$ such that $\|L^j\|_0 < 1$. The hypotheses of Proposition 4.2.4 are thus satisfied which implies that $(X_t)_{t \in \mathbb{Z}}$ is jointly regularly varying. By definition

$$\left(\left(\frac{X_0}{x}, \ldots, \frac{X_t}{x} \right) \,\middle|\, \|X_0\| > x \right) \xrightarrow{d} (Y_0^X, \ldots, Y_t^X), \qquad x \to \infty,$$

where $(Y_t^X)_{t \in \mathbb{N}}$ is the forward tail process of $(X)_{t \in \mathbb{Z}}$. Besides, since

$$X_t = L^t(X_0) + \sum_{i=1}^t L^{t-i}(Z_i),$$

we have

$$\left(\left(\frac{X_0}{x}, \ldots, \frac{X_t}{x} \right) \,\middle|\, \|X_0\| > x \right) \xrightarrow{d} (L^0(Y_0^X), \ldots, L^t(Y_t^X))_{t \in \mathbb{Z}}, \ x \to \infty.$$

By the uniqueness of the limit

$$(Y_t^X)_{t \in \mathbb{Z}} \stackrel{d}{=} (L^t(Y_0^X))_{t \in \mathbb{Z}}$$

hence

$$(\Theta_t^X)_{t\in\mathbb{Z}} \overset{d}{=} (L^t(\Theta_0^X))_{t\in\mathbb{Z}}$$

where $(\Theta_t^X)_{t\in\mathbb{N}}$ is the forward spectral process of $(X)_{t\in\mathbb{Z}}$. The random element Θ_0^X has the distribution of $Y_0^X/\|Y_0^X\|$. Given $f \in \mathcal{C}_b(\mathbb{B}_p)$, using (4.4) and a computation similar to the one in the proof of Theorem 4.2.2 (page 112) , as $x \to \infty$,

$$\begin{aligned}
&E\left[f\left(\frac{1}{x}X_0\right) \,\middle|\, \|X_0\| > x\right] \\
&= E\left[f\left(\frac{1}{x}(\zeta_0, \zeta_{-1}, \ldots)\right) \,\middle|\, \|(\zeta_0, \zeta_{-1}, \ldots)\| > x\right] \\
&= \sum_{i\geqslant 0} E\left[f\left(0, \ldots, 0, \frac{1}{w_i}\frac{\zeta}{x}, 0, \ldots\right) \,\middle|\, \|\zeta\| > x\right] \frac{w_i^\alpha}{\sum_{j=0}^\infty w_j^\alpha} + o_x(1)
\end{aligned}$$

where the nonzero argument of f is located at the i^{th} place for the i^{th} term of the sum. So

$$Y_0^X = (0, \ldots, 0, \frac{Y^\zeta}{w_I}, 0, \ldots)$$

where Y^ζ is the tail representation of ζ and the random position of Y^ζ/w_I is distributed according to the random variable I with

$$P(I = i) = \frac{w_i^\alpha}{\sum_{j=0}^\infty w_j^\alpha}, \qquad i \in \{0, 1, 2, \ldots\}.$$

Finally

$$\begin{aligned}
\Theta_0^X = \frac{Y_0^X}{\|Y_0^X\|} &= \frac{(0, \ldots, 0, \frac{Y^\zeta}{w_I}, 0, \ldots)}{\|(0, \ldots, 0, \frac{Y^\zeta}{w_I}, 0, \ldots)\|} \\
&= (0, \ldots, 0, \frac{1}{w_I}\frac{Y^\zeta}{\|Y^\zeta\|}, 0, \ldots) = (0, \ldots, 0, \frac{\Theta^\zeta}{w_I}, 0, \ldots)
\end{aligned}$$

where Θ^ζ/w_I located at the I^{th} place of the sequence. $\square$

Proposition 4.3.3 the same as Proposition 4.2.2 in Section 4.2 but for causal linear processes only.

Proposition 4.3.3. *Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be two separable Banach spaces. Let $(Z_i)_{i\in\mathbb{Z}}$ be a sequence of iid regularly varying random elements with index $\alpha > 0$ in $\mathbb{B}_1$ distributed according to a common random element Z. Let $T_i : \mathbb{B}_1 \to \mathbb{B}_2$, $i \in \mathbb{Z}$, be linear continuous maps satisfying* (4.1)*. Let $(X_t)_{t\in\mathbb{Z}}$ be the linear process $X_t = \sum_{i\geqslant 0} T_i(Z_{t-i})$, $t \in \mathbb{Z}$. Then $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying and its spectral process $(\Theta_t^X)_{t\in\mathbb{Z}}$ namely satisfies*

$$E[f(\Theta_{-s}^X, \dots \Theta_t^X)] = \frac{E[\sum_{i=0}^{\infty} f\left(\frac{T_{s+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}, \dots, \frac{T_{t+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}\right) \|T_i(\Theta^Z)\|^\alpha]}{\sum_{i=0}^{\infty} E[\|T_i(\Theta^Z)\|^\alpha]},$$

for every $f : \mathbb{B}_2^{t-s+1} \to \mathbb{R}$ continuous and bounded, where Θ^Z is a random element in $\mathbb{B}_1$ having the spectral measure of Z as law.

Proof. Let $0 < p < \min(1, \alpha)$, let $(w_i)_{i\geqslant 0}$ be a sequence of strictly positive numbers such that $\sup_{i\geqslant 0} \|T_i\|/w_i = C < +\infty$ and consider the associated Banach space $\mathbb{B}_p$. Let

$$\mathbb{T} : \mathbb{B}_p \to \mathbb{B}_2 : (x_0, x_{-1}, \dots) \mapsto \sum_{i\geqslant 0} T_i(x_{-i})$$

so that $X_t = \mathbb{T}(\zeta_t)$. The linear map $\mathbb{T}$ is continuous since

$$\begin{array}{l|l} \sup_{\|(x_n)\|_{\mathbb{B}_p}=1} \|\mathbb{T}((x_n))\|_{\mathbb{B}_2} & \leqslant \sup_{\|(x_n)\|_{\mathbb{B}_p}=1} \sum_{i\geqslant 0} \|T_i\| \|x_{-i}\|_{\mathbb{B}_1} \\ = \sup_{\|(x_n)\|_{\mathbb{B}_p}=1} \|\sum_{i\geqslant 0} T_i(x_{-i})\|_{\mathbb{B}_2} & \leqslant C \sup_{\|(x_n)\|_{\mathbb{B}_p}=1} \sum_{i\geqslant 0} w_i \|x_{-i}\|_{\mathbb{B}_1} \\ \leqslant \sup_{\|(x_n)\|_{\mathbb{B}_p}=1} \sum_{i\geqslant 0} \|T_i(x_{-i})\|_{\mathbb{B}_2} & = C \end{array}$$

hence, as a consequence of Proposition 4.3.2,

$$\Theta_t^\zeta = (0, \dots, 0, \frac{\Theta^Z}{w_I}, 0, \dots)$$

where the nonzero component is located at index $I + t$ with probability $P(I = i) = w_i^\alpha / \sum_{j=0}^{\infty} w_j^\alpha$. Let $0 \leqslant s < t$. According to Proposition 2.3.4, for $f \in \mathcal{C}(\mathbb{B}_2^{t-s+1})$ we have

$$
\begin{aligned}
&E\left[f(\Theta_s^X, \ldots \Theta_t^X)\right] \\
&= \frac{1}{E[\|\mathbb{T}(\Theta_0^\zeta)\|^\alpha]} E[f\left(\frac{\mathbb{T}(\Theta_s^\zeta)}{\|\mathbb{T}(\Theta_0^\zeta)\|}, \ldots, \frac{\mathbb{T}(\Theta_t^\zeta)}{\|\mathbb{T}(\Theta_0^\zeta)\|}\right) \|\mathbb{T}(\Theta_0^\zeta)\|^\alpha] \\
&= \frac{\sum_{i=0}^\infty w_j^\alpha}{\sum_{i=0}^\infty E[\|T_i(\Theta^Z/w_i)\|^\alpha] w_i^\alpha} \\
&\qquad .E[f\left(\frac{T_{s+I}(\Theta^Z/w_I)}{\|T_I(\Theta^Z/w_I)\|}, ..., \frac{T_{t+I}(\Theta^Z/w_I)}{\|T_I(\Theta^Z/w_I)\|}\right) \|T_I(\Theta^Z/w_I)\|^\alpha] \\
&= \frac{\sum_{j=0}^\infty w_j^\alpha}{\sum_{i=0}^\infty E[\|T_i(\Theta^Z)\|^\alpha]} E[f\left(\frac{T_{s+I}(\Theta^Z)}{\|T_I(\Theta^Z)\|}, ..., \frac{T_{t+I}(\Theta^Z)}{\|T_I(\Theta^Z)\|}\right) \|T_I(\Theta^Z)\|^\alpha w_I^{-\alpha}] \\
&= \frac{1}{\sum_{i=0}^\infty E[\|T_i(\Theta^Z)\|^\alpha]} E[\sum_{i=0}^\infty f\left(\frac{T_{s+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}, ..., \frac{T_{t+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}\right) \|T_i(\Theta^Z)\|^\alpha]
\end{aligned}
$$

where $(\Theta_t^X)_{t\in\mathbb{Z}}$ is the spectral process of $(X_t)_{t\in Z}$. □

4.4 Proofs for Section 4.2

Proof of Lemma 4.2.1 (page 97). Existence. Using consecutively the triangle inequality, the fact that $(x+y)^\delta \leqslant x^\delta + y^\delta$ given $0 < x$, $0 < y$ and $0 < \delta < 1$, $\|T(x)\| \leqslant \|T\|\|x\|$, the monotone convergence, the hypothesis (4.1) and eventually the fact that, for a positive regularly varying random variable X with index $\alpha > 0$, for every $0 < \beta < \alpha$, we have

$EX^{\beta} < +\infty$, we get

$$
\begin{aligned}
& E\left(\|X_t\|^{\delta}\right) \\
&= E\left(\|\sum_{i=1}^{\infty} T_i(Z_{t-i})\|^{\delta}\right) \\
&\leqslant E\left([\sum_{i=1}^{\infty} \|T_i(Z_{t-i})\|]^{\delta}\right) \\
&\leqslant E\left(\sum_{i=1}^{\infty} \|T_i(Z_{t-i})\|^{\delta}\right) \\
&\leqslant E\left(\sum_{i=1}^{\infty} \|T_i\|^{\delta}\|Z_{t-i}\|^{\delta}\right)
\end{aligned}
\qquad
\begin{aligned}
&= E\left(\lim_{k\to\infty} \sum_{i=1}^{k} \|T_i\|^{\delta}\|Z_{t-i}\|^{\delta}\right) \\
&= \lim_{k\to\infty} E\left(\sum_{i=1}^{k} \|T_i\|^{\delta}\|Z_{t-i}\|^{\delta}\right) \\
&= \lim_{k\to\infty} \left(\sum_{i=1}^{k} \|T_i\|^{\delta}\right) E\left(\|Z\|^{\delta}\right) \\
&= \left(\sum_{i=1}^{\infty} \|T_i\|^{\delta}\right) E\left(\|Z\|^{\delta}\right) \\
&< +\infty
\end{aligned}
$$

so that X_t almost surely exists.

Stationarity. For each $f : \mathbb{B}^k \to \mathbb{R}_+$ continuous and bounded we have to verify that

$$E[f(X_t, \ldots, X_{t+k})] = E[f(X_0, \ldots, X_k)]$$

i.e.

$$
\begin{aligned}
Ef(\sum_{i=1}^{\infty} T_i(Z_{t-i}), \ldots, \sum_{i=1}^{\infty} T_i(Z_{t+k-i})) = \\
Ef(\sum_{i=1}^{\infty} T_i(Z_{-i}), \ldots, \sum_{i=1}^{\infty} T_i(Z_{k-i}))
\end{aligned}
$$

which is the case by dominated convergence. □

Lemma 4.4.1. *Let $(Z_i)_{i\in\mathbb{Z}}$ be a sequence of independent and identically distributed random element taking values in a Banach space $\mathbb{B}_1$. Let $\mathbb{B}_2$ be another Banach space and let $T_i : \mathbb{B}_1 \to \mathbb{B}_2$, $i \in \mathbb{Z}$, be linear continuous maps.*

(i) the map $x \mapsto V(x) := P(\|Z_i\| > x)$ belongs to $RV_{-\alpha}$ for some $\alpha > 0$,

(ii) $\lim_{x\to\infty} P(\|T_i(Z_i)\| > x)/V(x) = a_i \in [0,\infty)$,

(iii) the maps T_i *satisfy* (4.1),

then the series $X_t = \sum_{i\in\mathbb{Z}} T_i(Z_{t-i}), t \in \mathbb{Z}$, *is almost surely absolutely convergent and as* $x \to \infty$,

$$E\big|1\{\|\textstyle\sum_{i\in\mathbb{Z}} T_i(Z_i)\| > x\} - \sum_{i\in\mathbb{Z}} 1\{\|T_i(Z_i)\| > x\}\big| = o\big(V(x)\big). \tag{4.4}$$

In particular,

$$\lim_{x\to\infty} \frac{P(\|\sum_{i\in\mathbb{Z}} T_i(Z_i)\| > x)}{V(x)} = \sum_{i\in\mathbb{Z}} a_i < \infty. \tag{4.5}$$

The proof of Lemma 4.4.1 is based on a number of lemmas.

Lemma 4.4.2. *Let* X *and* Y *be independent random elements of a Banach space* $\mathbb{B}$. *Suppose that there exists* $V \in RV_{-\alpha}$ *with* $\alpha > 0$ *and nonnegative numbers* a, b *such that*

$$\lim_{x\to\infty} \frac{P(\|X\| > x)}{V(x)} = a, \qquad \lim_{x\to\infty} \frac{P(\|Y\| > x)}{V(x)} = b. \tag{4.6}$$

Then as $x \to \infty$,

$$E\big|1\{\|X+Y\| > x\} - 1\{\|X\| > x\} - 1\{\|Y\| > x\}\big| = o\big(V(x)\big). \tag{4.7}$$

Proof. For real x, let $(x)^+ = x \vee 0$. Use the formula $|r| = (r)^+ + (-r)^+$ for real r to decompose the absolute value on the left-hand side of (4.7).

Fix $0 < \varepsilon < 1$. For positive numbers r, s, x, if $r + s > x$, then necessarily $r \vee s > (1-\varepsilon)x$ or $r \wedge s > \varepsilon x$. It follows that

$$\begin{aligned} 1\{\|X+Y\| > x\} &\leqslant 1\{\|X\| + \|Y\| > x\} \\ &\leqslant 1\{\|X\| \vee \|Y\| > (1-\varepsilon)x\} + 1\{\|X\| \wedge \|Y\| > \varepsilon x\} \\ &\leqslant 1\{\|X\| > (1-\varepsilon)x\} + 1\{\|Y\| > (1-\varepsilon)x\} \\ &\qquad + 1\{\|X\| > \varepsilon x\}1\{\|Y\| > \varepsilon x\}. \end{aligned}$$

We find that

$$\begin{aligned} &1\{\|X+Y\| > x\}) - 1\{\|X\| > x\} - 1\{\|Y\| > x\} \\ &\qquad \leqslant 1\{(1-\varepsilon)x < \|X\| \leqslant x\} + 1\{(1-\varepsilon)x < \|Y\| \leqslant x\} \\ &\qquad\qquad + 1\{\|X\| > \varepsilon x\}1\{\|Y\| > \varepsilon x\}. \end{aligned}$$

By regular variation of V and by (4.6),

$$\limsup_{x\to\infty} \frac{1}{V(x)} E\big(1\{\|X+Y\|>x\} - 1\{\|X\|>x\} - 1\{\|Y\|>x\}\big)^+ \leqslant (a+b)\big((1-\varepsilon)^{-\alpha}-1\big).$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$E\big(1\{\|X+Y\|>x\} - 1\{\|X\|>x\} - 1\{\|Y\|>y\}\big)^+ = o\big(V(x)\big) \quad (4.8)$$

as $x \to \infty$.

Since $\|v+w\| \geqslant |\|v\|-\|w\||$ for $v, w \in \mathbb{B}$, we have

$$\begin{aligned} 1\{\|X+Y\|>x\} &\geqslant 1\{|\|X\|-\|Y\||>x\} \\ &= 1\{\|X\|>x+\|Y\|\} + 1\{\|Y\|>x+\|X\|\}. \end{aligned}$$

It follows that

$$\begin{aligned} &1\{\|X\|>x\} + 1\{\|Y\|>y\} - 1\{\|X+Y\|>x\} \\ &\qquad \leqslant 1\{x<\|X\|\leqslant x+\|Y\|\} + 1\{x<\|Y\|\leqslant x+\|X\|\}. \end{aligned}$$

Fix $\varepsilon > 0$. We have

$$\begin{aligned} &1\{x<\|X\|\leqslant x+\|Y\|\} \\ &\qquad \leqslant 1\{x<\|X\|\leqslant (1+\varepsilon)x\} + 1\{\|X\|>x\}1\{\|Y\|>\varepsilon x\}. \end{aligned}$$

By regular variation of V and (4.6),

$$\limsup_{x\to\infty} \frac{P(x<\|X\|\leqslant x+\|Y\|)}{V(x)} \leqslant a\big(1-(1+\varepsilon)^{-\alpha}\big).$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$\lim_{x\to\infty} \frac{P(x<\|X\|\leqslant x+\|Y\|)}{V(x)} = 0.$$

By symmetry, the same is true for X and Y interchanged. We conclude that

$$E\big(1\{\|X\|>x\} + 1\{\|Y\|>y\} - 1\{\|X+Y\|>x\}\big)^+ = o\big(V(x)\big) \quad (4.9)$$

as $x \to \infty$. Finally, combine (4.8) and (4.9) to arrive at (4.7). $\square$

Lemma 4.4.3. *Let $X_1, \ldots, X_n$ be independent random elements of a Banach space $\mathbb{B}$. Suppose that there exists $V \in RV_{-\alpha}$ with $\alpha > 0$ and nonnegative numbers $a_1, \ldots, a_n$ such that*

$$\lim_{x\to\infty} \frac{P(\|X_i\| > x)}{V(x)} = a_i, \qquad i \in \{1, \ldots, n\}. \tag{4.10}$$

Then as $x \to \infty$,

$$E\big|1\big(\|\textstyle\sum_{i=1}^n X_i\| > x\big) - \sum_{i=1}^n 1\{\|X_i\| > x\}\big| = o\big(V(x)\big). \tag{4.11}$$

In particular,

$$\lim_{x\to\infty} \frac{P(\|\sum_{i=1}^n X_i\| > x)}{V(x)} = \sum_{i=1}^n a_i. \tag{4.12}$$

Proof. The proof is by induction on n. The case $n = 2$ was treated in Lemma 4.4.2. So suppose that $n \geqslant 3$. By the triangle inequality,

$$\begin{aligned}
&\big|1\big(\|\textstyle\sum_{i=1}^n X_i\| > x\big) - \sum_{i=1}^n 1\{\|X_i\| > x\}\big| \\
&\quad\leqslant \big|1\big(\|\textstyle\sum_{i=1}^{n-1} X_i + X_n\| > x\big) - 1\big(\|\sum_{i=1}^{n-1} X_i\| > x\big) - 1\{\|X_n\| > x\}\big| \\
&\qquad\qquad + \big|1\big(\|\textstyle\sum_{i=1}^{n-1} X_i\| > x\big) - \sum_{i=1}^{n-1} 1\{\|X_i\| > x\}\big|.
\end{aligned}$$

Take expectations and apply the induction hypothesis twice to arrive at (4.11). Equation (4.12) follows immediately from (4.10) and (4.11). □

Remark 4.4.4. *Under the conditions of Lemma 4.4.1 we also have*

$$E\left|1\{\max_{|i|\leqslant n} \|T_i(Z_i)\| > x\} - \sum_{|i|\leqslant n} 1\{\|T_i(Z)\| > x)\}\right| = o\big(V(x)\big), \qquad x \to \infty,$$

the proof of which is similar to but simpler than the one of (4.11)*: proceed by induction on n, the key step in case $n = 2$ being the identity*

$$\begin{aligned}
&1\{\max(\|X_1\|, \|X_2\|) > x\} \\
&\qquad = 1\{\|X_1\| > x\} + 1\{\|X_2\| > x\} - 1\{\|X_1\| > x\}1\{\|X_2\| > x\}.
\end{aligned}$$

It then follows that

$$E\left|1\left(\left\|\sum_{i\leqslant n} X_i\right\| > x\right) - 1\left(\max_{i\leqslant n} \|X_i\| > x\right)\right| = o\big(V(x)\big).$$

which means that the sum $\sum_{i\leqslant n} X_i$ *is large in norm if and only if one of the terms* X_i *is large in norm. In that case* $\|\sum_{i\leqslant n} X_i\|$ *is approximately the same as* $\max_{i\leqslant n}\|X_i\|$*, the largest term dominating the whole sum.*

Proof of Lemma 4.4.1. Thank to Lemma 4.2.1 the series $\sum_i T_i(Z_i)$ exists almost surely.

Write $X_i = T_i(Z_i)$. In order to prove (4.4), let m be a positive integer. Write

$$\begin{aligned}
&\bigl|1\{\|\textstyle\sum_i X_i\| > x\} - \sum_i 1\{\|X_i\| > x\}\bigr| \\
&\qquad\leqslant \bigl|1\{\|\textstyle\sum_i X_i\| > x\} - 1\{\|\sum_{|i|\leqslant m} X_i\| > x\}\bigr| \\
&\qquad\quad + \bigl|1\{\|\textstyle\sum_{|i|\leqslant m} X_i\| > x\} - \sum_{|i|\leqslant m} 1\{\|X_i\| > x\}\bigr| \\
&\qquad\quad + \textstyle\sum_{|i|>m} 1\{\|X_i\| > x\}. \qquad (4.13)
\end{aligned}$$

The three terms on the right-hand side are treated separately.

Consider the first term on the right-hand side of (4.13). Let $X = \sum_{|i|\leqslant m} X_i$ and $Y = \sum_{|i|>m} X_i$. Using the identity $|1_A - 1_B| = 1_{A\,\triangle\,B}$, it is not difficult to see that for $0 < \varepsilon < 1$,

$$\begin{aligned}
&\bigl|1\{\|X+Y\| > x\} - 1\{\|X\| > x\}\bigr| \\
&\qquad\leqslant 1\{\|Y\| > \varepsilon x\} + 1\{(1-\varepsilon)x < \|X\| \leqslant (1+\varepsilon)x\}.
\end{aligned}$$

By Resnick (2008), Lemma 4.24, since $\|Y\| \leqslant \sum_{|i|>m} \|T_i\|\|Z_i\|$,

$$\limsup_{y\to\infty} \frac{P(\|Y\| > y)}{V(y)} \leqslant \sum_{|i|>m} \|T_i\|^\alpha.$$

By (ii) and (4.12),

$$\lim_{y\to\infty} \frac{P(\|X\| > y)}{V(y)} = \sum_{|i|\leqslant m} a_i.$$

By regular variation of V and from $\|T_i(Z_i)\| \leqslant \|T_i\|\|Z_i\|$, it follows that $a_i \leqslant \|T_i\|^\alpha$ (see Proof of Theorem 4.2.2 on page 112). In view of the three preceding displays, we have

$$\begin{aligned}
&\limsup_{x\to\infty} \frac{1}{V(x)} E\big|1\{\|X+Y\|>x\} - 1\{\|X\|>x\}\big| \\
&\qquad \leqslant \varepsilon^{-\alpha} \sum_{|i|>m} \|T_i\|^\alpha + \big((1-\varepsilon)^{-\alpha} - (1+\varepsilon)^{-\alpha}\big) \sum_{|i|\leqslant m} a_i \\
&\qquad \leqslant \varepsilon^{-\alpha} \sum_{|i|>m} \|T_i\|^\alpha + \big((1-\varepsilon)^{-\alpha} - (1+\varepsilon)^{-\alpha}\big) \sum_{i\in\mathbb{Z}} \|T_i\|^\alpha. \qquad (4.14)
\end{aligned}$$

Next, for the second term on the right-hand side of (4.13), Lemma 4.4.3 implies immediately that

$$E\big|1\{\|\textstyle\sum_{|i|\leqslant m} X_i\| > x\} - \sum_{|i|\leqslant m} 1\{\|X_i\|>x\}\big| = o\big(V(x)\big), \qquad x\to\infty. \qquad (4.15)$$

Finally, consider the third term on the right-hand side of (4.13). Let m be sufficiently large such that $\|T_i\| \leqslant 1$ whenever $|i| > m$. By Potter's theorem (Resnick 2008, p 22), there exists $x_0 > 0$ such that $V(cx)/V(x) \leqslant (1+\alpha-p)c^{-p}$ for every $x \geqslant x_0$ and $c \geqslant 1$. Hence

$$\frac{P(\|X_i\|>x)}{V(x)} \leqslant \frac{P(\|T_i\|\|Z_i\|>x)}{V(x)} \leqslant (1+\alpha-p)\|T_i\|^p, \quad x \geqslant x_0,\ |i|>m. \qquad (4.16)$$

As a consequence,

$$\limsup_{x\to\infty} \frac{1}{V(x)} \sum_{|i|>m} P(\|X_i\|>x) \leqslant (1+\alpha-p) \sum_{|i|>m} \|T_i\|^p. \qquad (4.17)$$

Taking expectations in (4.13) and inserting the upper bounds in (4.14), (4.15) and (4.17) yields the following inequality, valid for all $0<\varepsilon<1$ and for all sufficiently large integer m:

$$\begin{aligned}
&\limsup_{x\to\infty} \frac{1}{V(x)} E\big|1\{\|\textstyle\sum_i X_i\|>x\} - \sum_i 1\{\|X_i\|>x\}\big| \\
&\qquad \leqslant \varepsilon^{-\alpha} \sum_{|i|>m} \|T_i\|^\alpha + \big((1-\varepsilon)^{-\alpha} - (1+\varepsilon)^{-\alpha}\big) \sum_{i\in\mathbb{Z}} \|T_i\|^\alpha + o_m(1).
\end{aligned}$$

First let $m\to\infty$ and then let $\varepsilon \downarrow 0$ to arrive at (4.4).

Equation (4.5) follows from (4.4) and the dominated convergence theorem, the use of which is justified by (iii) and (4.16). □

Proof of Theorem 4.2.2 (page 98). Write μ the exponent measure of Z such that $\mu(\{x \in \mathbb{B} : \|x\| > 1\}) = 1$. First remark that $\mu(\partial[\|T_i(y)\| > 1]) = \mu([\|T_i(y)\| = 1]) = 0$. Otherwise the map

$$u \mapsto \mu([\|T_i(y)\| > u]) = u^{-\alpha}\mu([\|T_i(y)\| > 1])$$

which is continuous would have a discontinuity at $u = 1$. So, according to the definition of exponent measure (Definition 2.2.1), we have

$$P\left(\|T_i(Z)\| > x\right) = P\left(\|T_i(Z/x)\| > 1\right) \sim P\left(\|Z\| > x\right)\mu([\|T_i(y)\| > 1]).$$

Let Θ^Z be a random element in $\mathbb{B}$ the spectral measure of Z as law. Then, by (2.6),

$$\begin{aligned}
\mu([\|T_i(x)\| > 1]) &= \int 1\{[\|T_i(x)\| > 1]\}d\mu \\
&= \int_0^\infty E1\{[\|T_i(x)\| > 1]\}(r\Theta^Z)d(-r^{-\alpha}) \\
&= \int_0^\infty P(\|T_i(r\Theta^Z)\| > 1)d(-r^{-\alpha}) \\
&= \int_0^\infty P(\|T_i(\Theta^Z)\|^\alpha > v)d(v) \\
&= E[\|T_i(\Theta^Z)\|^\alpha]
\end{aligned}$$

According to (4.4) and (4.5), we obtain

$$\sum_{i\in\mathbb{Z}} P\left(\|T_i(Z)\| > x\right) \sim P\left(\|Z\| > x\right)\sum_{i\in\mathbb{Z}} E[\|T_i(\Theta^Z)\|^\alpha].$$

Note that there exists a random element $\Theta(-s,t;i)$ in $\mathbb{B}_2^{s+t+1}$ such that

$$\mathcal{L}\left(\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|}, \ldots, \frac{T_{t+i}(Z)}{\|T_i(Z)\|}\right) \,\middle|\, \|T_i(Z)\| > x\right) \xrightarrow{d} \mathcal{L}\left(\Theta(-s,t;i)\right).$$

Indeed, given $f \in \mathcal{C}_b(\mathbb{B}_2^{s+t+1})$, thanks to Definition 2.2.1 and Lemma 2.5.20, as $x \to \infty$,

$$
\begin{aligned}
&E\left[f\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|}, \dots, \frac{T_{t+i}(Z)}{\|T_i(Z)\|} \right) \,\middle|\, \|T_i(Z)\| > x \right] \\
&= \frac{E\left[f\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|}, \dots, \frac{T_{t+i}(Z)}{\|T_i(Z)\|} \right) 1\{\|T_i(Z)\| > x\} \right]}{P(\|Z\| > x)} \\
&\qquad \cdot \frac{P(\|Z\| > x)}{P(\|T_i(Z)\| > x)} \\
&= \frac{E\left[f\left(\frac{T_{-s+i}(Z/x)}{\|T_i(Z/x)\|}, \dots, \frac{T_{t+i}(Z/x)}{\|T_i(Z/x)\|} \right) 1\{\|T_i(Z/x)\| > 1\} \right]}{P(\|Z\| > x)} \\
&\qquad \cdot \frac{P(\|Z\| > x)}{P(\|T_i(Z)\| > x)} \\
&\to \int f\left(\frac{T_{-s+i}(y)}{\|T_i(y)\|}, \dots, \frac{T_{t+i}(y)}{\|T_i(y)\|} \right) 1\{\|T_i(y)\| > 1\} d\mu(y) \frac{1}{E[\|T_i(\Theta^Z)\|^\alpha]}.
\end{aligned}
$$

By (2.12) this limit can be rewritten

$$
\begin{aligned}
&\int_{r>0, \|\varphi\|=1} f\left(\frac{T_{-s+i}(r\varphi)}{\|T_i(r\varphi)\|}, \dots, \frac{T_{t+i}(r\varphi)}{\|T_i(r\varphi)\|} \right) 1\{\|T_i(r\varphi)\| > 1\} \\
&\qquad \cdot \frac{P(\Theta^Z \in d\varphi)}{E[\|T_i(\Theta^Z)\|^\alpha]} d(-r^{-\alpha}) \\
&= \int_{\|\varphi\|=1} f\left(\frac{T_{-s+i}(\varphi)}{\|T_i(\varphi)\|}, \dots, \frac{T_{t+i}(\varphi)}{\|T_i(\varphi)\|} \right) \|T_i(\varphi)\|^\alpha \frac{P(\Theta^Z \in d\varphi)}{E[\|T_i(\Theta^Z)\|^\alpha]} \\
&= E\left[f\left(\frac{T_{-s+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}, \dots, \frac{T_{t+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|} \right) \|T_i(\Theta^Z)\|^\alpha \right] \frac{1}{E[\|T_i(\Theta^Z)\|^\alpha]} \\
&=: E\left[f\left(\Theta(-s, t; i) \right) \right].
\end{aligned}
$$

Now take $f \in \mathcal{C}_b(\mathbb{B}_2)$ and let us compute the distribution of the tail process according to the above results, Equation (4.4) and the two notes below, that, as $x \to \infty$,

$$
\begin{aligned}
&E\left[f\left(\frac{X_{-s}}{x}, \dots, \frac{X_t}{x}\right) \,\Big|\, \|X_0\| > x \right] \\
&= \frac{E\left[f\left(\frac{X_{-s}}{x}, \dots, \frac{X_t}{x}\right) 1\{\|X_0\| > x\}\right]}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)} \\
&= \sum_{i\in\mathbb{Z}} \frac{E\left[f\left(\frac{X_{-s}}{x}, \dots, \frac{X_t}{x}\right) 1\{\|T_i(Z_{-i})\| > x\}\right]}{P(\|T_i(Z)\| > x)} \\
&\qquad\qquad \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)} + o(1) \\
&= \sum_{i\in\mathbb{Z}} E\left[f\left(\frac{X_{-s}}{x}, \dots, \frac{X_t}{x}\right) \,\Big|\, \|T_i(Z_{-i})\| > x \right] \\
&\qquad\qquad \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)} + o(1) \\
&= \sum_{i\in\mathbb{Z}} E\left[f\left(\frac{T_{-s+i}(Z)}{x}, \dots, \frac{T_{t+i}(Z)}{x}\right) \,\Big|\, \|T_i(Z)\| > x \right] \\
&\qquad\qquad \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)} + o(1) \\
&\to \sum_{i\in\mathbb{Z}} E\left[f(Y\Theta(-s,t;i))\right] E[\|T_i(\Theta^Z)\|^\alpha] \frac{1}{\sum_{j=-\infty}^{\infty} E[\|T_j(\Theta^Z)\|^\alpha]}.
\end{aligned}
$$

where $Y \sim \text{Pareto}(\alpha)$ is independent of $(\Theta(-s,t;i))_{i\in\mathbb{Z}}$. So the spectral process of $(X_t)_{t\in Z}$ is given by

$$
\begin{aligned}
&E\left[f\left(\Theta_{-s}, \dots, \Theta_t\right)\right] \\
&= \sum_{i\in\mathbb{Z}} E\left[f(\Theta(-s,t;i))\right] E[\|T_i(\Theta^Z)\|^\alpha] \frac{1}{\sum_{j\in\mathbb{Z}} E[\|T_j(\Theta^Z)\|^\alpha]} \\
&= \frac{\sum_{i\in\mathbb{Z}} E\left[f\left(\frac{T_{-s+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}, \dots, \frac{T_{t+i}(\Theta^Z)}{\|T_i(\Theta^Z)\|}\right) \|T_i(\Theta^Z)\|^\alpha \right]}{\sum_{i\in\mathbb{Z}} E[\|T_i(\Theta^Z)\|^\alpha]},
\end{aligned}
$$

by the result of the previous paragraph. □

Note 1. Why is the step

$$\sum_{i\geqslant 0} E\left[f\left(\frac{X_0}{x}\right)\,\middle|\,\|T_i(Z_{-i})\|>x\right]\frac{P(\|T_i(Z)\|>x)}{P(\|Z\|>x)}\frac{P(\|Z\|>x)}{P(\|X_0\|>x)}$$

$$=\sum_{i\geqslant 0} E\left[f\left(\frac{T_i(Z)}{x}\right)\middle|\,\|T_i(Z)\|>x\right]\frac{P(\|T_i(Z)\|>x)}{P(\|Z\|>x)}\frac{P(\|Z\|>x)}{P(\|X_0\|>x)}+o(1)$$

true the proof of Theorem 4.2.2?

Decompose X_0 in

$$\sum_{j=0}^{\infty}\frac{T_j(Z_{-j})}{x}=\sum_{j=0,j\neq i}^{\infty}\frac{T_j(Z_{-j})}{x}+\frac{T_i(Z_{-i})}{x}.$$

Put

$$U_x\stackrel{d}{=}\left(\frac{\sum_{j=0}^{\infty}T_j(Z_{-j})}{x}\,\middle|\,\|T_i(Z_{-i})\|>x\right)$$

as well as

$$V_x\stackrel{d}{=}\left(\frac{T_i(Z_{-i})}{x}\,\middle|\,\|T_i(Z_{-i})\|>x\right),$$

U_x-V_x meaning

$$U_x-V_x\stackrel{d}{=}\left(\frac{\sum_{j=0,j\neq i}^{\infty}T_j(Z_{-j})}{x}\,\middle|\,\|T_i(Z_{-i})\|>x\right)\stackrel{d}{=}\frac{\sum_{j=0,j\neq i}^{\infty}T_j(Z_{-j})}{x}.$$

Set $\mu_x=\mathcal{L}(U_x-V_x)$ and let us show that

$$U_x-V_x\xrightarrow{d}0,\qquad x\to\infty.$$

This is based on the fact that $\mu_x(A)=\mu_1(xA)$. Assume that $0\notin\partial A$. Then, if $0\notin A$, we have $0\notin \operatorname{cl} A$ and the limit of $\mu_x(A)$ is $\mu_1(\emptyset)=0$. Else, if $0\in A$, we have $0\in\operatorname{int} A$ and the limit of $\mu_x(A)$ is $\mu_1(\mathbb{B}_2)=1$. So $\mu_x\to\delta_0$ in distribution.

Moreover, V_x has a weak limit, let us say W:

$$V_x\xrightarrow{d}W,\qquad x\to\infty.$$

So Slutsky's theorem implies that

$$U_x = (U_x - V_x) + V_x \xrightarrow{d} W, \qquad x \to \infty.$$

Thus, by the Portmanteau Lemma, we obtain that for all continuous and bounded f:

$$\lim_{x\to\infty} E\left[f\left(\frac{\sum_{j=0}^{\infty} T_j(Z_{-j})}{x} \right) \middle| \|T_i(Z_{-i})\| > x \right] = \lim_{x\to\infty} E\left[f\left(\frac{T_i(Z)}{x} \right) \middle| \|T_i(Z)\| > x \right].$$

Consequently, by the properties of the limits,

$$\lim_{n\to\infty} \sum_{|i|\leqslant n} \lim_{x\to\infty} E\left[f\left(\frac{\sum_{j=0}^{\infty} T_j(Z_{-j})}{x} \right) \middle| \|T_i(Z_{-i})\| > x \right] \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}$$
$$= \lim_{n\to\infty} \sum_{|i|\leqslant n} \lim_{x\to\infty} E\left[f\left(\frac{T_i(Z)}{x} \right) \middle| \|T_i(Z)\| > x \right] \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}$$

hence

$$\lim_{n\to\infty} \lim_{x\to\infty} \sum_{|i|\leqslant n} E\left[f\left(\frac{\sum_{j=0}^{\infty} T_j(Z_{-j})}{x} \right) \middle| \|T_i(Z_{-i})\| > x \right] \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}$$
$$= \lim_{n\to\infty} \lim_{x\to\infty} \sum_{|i|\leqslant n} E\left[f\left(\frac{T_i(Z)}{x} \right) \middle| \|T_i(Z)\| > x \right] \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}$$

Consider the set $E = [1, \infty]$ and the measure $\mu = \delta_\infty$ on E. Set

$$F_n : E \to \mathbb{R} : x \mapsto \sum_{|i| \leqslant n} E\left[f\left(\frac{\sum_{j=0}^{\infty} T_j(Z_{-j})}{x} \right) \middle| \|T_i(Z_{-i})\| > x \right] \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}$$

and

$$G_n : E \to \mathbb{R} : x \mapsto \sum_{|i| \leqslant n} E\left[f\left(\frac{T_i(Z)}{x} \right) \middle| \|T_i(Z)\| > x \right] \cdot \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}$$

so that we have

$$\lim_{n \to \infty} \int F_n d\mu = \lim_{n \to \infty} \int G_n d\mu.$$

The sequences (F_n) and (G_n) are increasing (using $f = f^+ - f^-$, f can be assumed to be nonnegative). Then

$$\int \lim_{n \to \infty} F_n d\mu = \int \lim_{n \to \infty} G_n d\mu,$$

i.e.

$$\lim_{x \to \infty} \lim_{n \to \infty} F_n(x) = \lim_{x \to \infty} \lim_{n \to \infty} G_n(x),$$

i.e.

$$\begin{aligned} &\lim_{x \to \infty} \sum_{i \in \mathbb{Z}} E\left[f\left(\frac{X_0}{x} \right) \middle| \|T_i(Z_{-i})\| > x \right] \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)} \\ &= \lim_{x \to \infty} \sum_{i \in \mathbb{Z}} E\left[f\left(\frac{T_i(Z)}{x} \right) \middle| \|T_i(Z)\| > x \right] \frac{P(\|T_i(Z)\| > x)}{P(\|Z\| > x)} \frac{P(\|Z\| > x)}{P(\|X_0\| > x)}, \end{aligned}$$

which is the statement.

Note 2. Here is the details explaining how we get $f(Y\Theta(-s, t; i))$ at the penultimate display of the proof of Theorem 4.2.2.

We know that

$$\mathcal{L}\left(\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|},\ldots,\frac{T_{t+i}(Z)}{\|T_i(Z)\|}\right)\ \Big|\ \|T_i(Z)\|>x\right)\xrightarrow{d}\mathcal{L}\left(\Theta(-s,t;i)\right).$$

Take $Y\sim$ Pareto(α) and independent of $\Theta(-s,t;i)$.

First of all, remark that

$$\mathcal{L}\left(\left(\frac{\|T_i(Z)\|}{x},\frac{T_{-s+i}(Z)}{\|T_i(Z)\|},\ldots,\frac{T_{t+i}(Z)}{\|T_i(Z)\|}\right)\ \Big|\ \|T_i(Z)\|>x\right)$$
$$\xrightarrow{d}\mathcal{L}\left(Y,\Theta(-s,t;i)\right).$$

Indeed, by van der Vaart and Wellner (1996), Lemma 1.4.2, to see this it suffices to prove that:

$$E\left[f\left(\frac{\|T_i(Z)\|}{x}\right)g\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|},\ldots,\frac{T_{t+i}(Z)}{\|T_i(Z)\|}\right)\ \Big|\ \|T_i(Z)\|>x\right]$$
$$\to E[f(Y)g(\Theta(-s,t;i))].$$

for every continuous and bounded f and g. Using Lebesgue's dominated convergence theorem and the fact that $\|Y\|\geqslant 1$, the functions f can be replaced by the family of functions $1\{\cdot>y\}$ where $y\geqslant 1$, which satisfy the property:

$$\lim_{x\to\infty}E\left[1\left\{\frac{\|T_i(Z)\|}{x}>y\right\}g\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|},\ldots,\frac{T_{t+i}(Z)}{\|T_i(Z)\|}\right)\ \Big|\ \|T_i(Z)\|>x\right]$$
$$=\lim_{x\to\infty}\frac{P(\|T_i(Z)\|\geqslant xy)}{P(\|T_i(Z)\|\geqslant x)}E\left[g\left(\frac{T_{-s+i}(Z)}{\|T_i(Z)\|},\ldots,\frac{T_{t+i}(Z)}{\|T_i(Z)\|}\right)\Big|\ \|T_i(Z)\|>xy\right]$$
$$=y^{-\alpha}E\left[g(\Theta(-s,t;i))\right]=E\left[1\{Y>y\}\right]E\left[g(\Theta(-s,t;i))\right]$$
$$=E\left[1\{Y>y\}g(\Theta(-s,t;i))\right].$$

Finally apply the Continuous mapping theorem (Resnick 2008, p 152) with the function $h(y,x_{-s},\ldots,x_t)=y(x_{-s},\ldots,x_t)$ to obtain

$$\mathcal{L}\left(\left(\frac{T_{-s+i}(Z)}{x},\ldots,\frac{T_{t+i}(Z)}{x}\right)\ \Big|\ \|T_i(Z)\|>x\right)\xrightarrow{d}\mathcal{L}\left(Y\Theta(-s,t;i)\right).$$

Proof of Proposition 4.2.3 (page 99). According to Table 2.1, the candidate extremal index θ^* is

$$\theta^* = 1 - E\left[\min\left\{\|\Theta_0\|^\alpha,\ \sup_{t\geqslant 1}\|\Theta_t\|^\alpha\right\}\right].$$

We have by monotone convergence and Theorem 4.2.2,

$$\begin{aligned}
&E\left[\min\left\{\|\Theta_0\|^\alpha,\ \sup_{t\geqslant 1}\|\Theta_t\|^\alpha\right\}\right]\\
&= \lim_{m\to\infty} E\left[\min\left\{\|\Theta_0\|^\alpha,\ \max_{1\leqslant t\leqslant m}\|\Theta_t\|^\alpha\right\}\right]\\
&= \lim_{m\to\infty} \frac{\sum_{n\in\mathbb{Z}} E[\min\{\|T_n\Theta^Z\|^\alpha, \max_{1\leqslant t\leqslant m}\|T_{n+t}\Theta^Z\|^\alpha\}]}{\sum_{n\in\mathbb{Z}} E[\|T_n\Theta^Z\|^\alpha]}\\
&= \frac{\sum_{n\in\mathbb{Z}} E[\min\{\|T_n\Theta^Z\|^\alpha, \sup_{t\geqslant 1}\|T_{n+t}\Theta^Z\|^\alpha\}]}{\sum_{n\in\mathbb{Z}} E[\|T_{n\in\mathbb{Z}}\Theta^Z\|^\alpha]}.
\end{aligned}$$

For a nonnegative sequence $(a_n)_{n\in\mathbb{Z}}$ such that $\sum_n a_n < \infty$, we have

$$\sum_{n\in\mathbb{Z}} \min\left(a_n,\ \sup_{t\geqslant 1} a_{n+t}\right) = \sum_{n\in\mathbb{Z}} a_n - \sup_{n\in\mathbb{Z}} a_n,$$

which yields the statement. □

Proof of Proposition 4.2.4 (page 99). Let $0 < p < \min(1,\alpha)$. Since $X_t = T(X_{t-1}) + Z_t = Z_t + T(Z_{t-1}) + T^2(Z_{t-2}) + \dots$, the series $(X_t)_{t\in\mathbb{Z}}$ admits the expression of a linear process with $T_i(z) = T^i(z)$, $i \geqslant 0$. Moreover,

$$\begin{aligned}
&\sum_{i=0}^{\infty}\|T_i\|^p\\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{j-1}\|T^{kj+l}\|^p\\
&\leqslant \sum_{k=0}^{\infty}\sum_{l=0}^{j-1}(\|T^j\|^k)^p\|T^l\|^p\\
&= \sum_{k=0}^{\infty}(\|T^j\|^p)^k\sum_{l=0}^{j-1}\|T^l\|^p\\
&< +\infty
\end{aligned}$$

which implies (4.1). □

Chapter 5

Maxima of moving maxima of continuous functions

Based on Meinguet (2010).

Abstract. Maxima of moving maxima of continuous functions (CM3) are max-stable processes aimed at modeling extremes of continuous phenomena over time. They are defined as Smith and Weissman's M4 processes with continuous functions rather than vectors. After standardization of the margins of the observed process into unit-Fréchet, CM3 processes can model the remaining spatio-temporal dependence structure. CM3 processes have the property of joint regular variation. The spectral processes from this class admit particularly simple expressions. Furthermore, depending on the speed with which the parameter functions tend toward zero, CM3 processes fulfill the finite-cluster condition and the strong mixing condition. For instance, these three properties put together have implications for the expression of the extremal index. A method for fitting a CM3 to data is investigated. The first step is to estimate the length of the temporal dependence. Then, by selecting a suitable number of blocks of extremes of this length, clustering algorithms are used to estimate the total number of different profiles. The number of parameter functions to retrieve is equal to the product of these two numbers. They are estimated thanks to the output of the partitioning algorithms in the previous step. The full procedure only requires one parameter which is the range of variation allowed among the different profiles. The dissimilarity between the

original CM3 and the estimated version is evaluated by means of the Hausdorff distance between the graphs of the parameter functions.

5.1 Introduction

After our study in Chapter 4 of heavy tailed linear processes, this chapter is devoted to processes endowed with the same structure but with a maximum instead of a sum. The main characteristic of "max-linear" processes is their capacity to model the cross-sectional tail dependence and the temporal tail dependence of more general process, that do not have to be heavy tailed, after a standardization of the margins into unit-Fréchet. Here we focus on the case of continuous functions as objects of interest. After having derived their main properties, we suggest and compare several estimation methods.

Maxima of moving maxima of continuous functions, in short CM3, are the analogue of Smith and Weissman's M4 processes (Smith and Weissman 1996) with continuous functions rather than vectors. Let $a_i^{(j)}$ ($i \in \mathbb{Z}_+$, $j \in \mathbb{Z}$) be strictly positive, real, continuous functions on a compact domain of $\mathbb{R}^q$, let us say $[0,1]^q$. The functions $a_i^{(j)}$ are the *parameter functions*. They are assumed to satisfy, for every $x \in [0,1]^q$, the equality $\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} a_i^{(j)}(x) = 1$. A CM3 process $(X_t)_{t\in\mathbb{Z}}$ is defined by the expression

$$X_t(x) = \sup_{j\in\mathbb{Z}} \sup_{i\geqslant 0} a_i^{(j)}(x) Z_{t-i}^{(j)} \qquad (x \in [0,1]^q),$$

where the innovations $Z_i^{(j)}$ are independent and identically distributed unit-Fréchet random variables, i.e. $P(Z_t \leqslant z) = \exp(-1/z)$ for $z > 0$. CM3 processes generalize M4 processes in the sense that M4 have the same expression except that the parameters $a_i^{(j)}$ are vectors of $\mathbb{R}^d$.

The fact that, given real numbers $\xi_i > 0$ ($i \in \mathbb{N}$) such that $\xi_1 + \xi_2 + \ldots = 1$, the distribution of $\max(\xi_1 Z_1, \xi_2 Z_2, \ldots)$ stays unit-Fréchet implies that X_t has unit-Fréchet margins. However the transformation from $(Z_t)_{t\in\mathbb{Z}}$ to $(X_t)_{t\in\mathbb{Z}}$ induces a dependence structure over time and in space. Extremes appear in temporal clusters and, at time t, a large value for X_t at location x causes large values at other locations. From this fact, CM3 processes are able to model a wide range of spatio-temporal dependence. The first part of this chapter is a study of some properties:

spectral process, strong mixing condition, finite-cluster condition and extremal index.

The second objective of this chapter is to fit CM3 processes to samples with measurement errors. For that purpose, CM3 will be discretized into D-dimensional M4 selecting D points x_d $(1 \leqslant d \leqslant D)$ in the domain. It will be also assumed that $0 \leqslant i < K$ and $1 \leqslant j \leqslant L$ for finite constants K and L. The practical model studied is thus

$$X_t(d) = \max_{1 \leqslant j \leqslant L} \max_{0 \leqslant i < K} a_i^{(j)}(x_d) Z_{t-i}^{(j)} + \varepsilon_t(x_d) \qquad (1 \leqslant d \leqslant D)$$

where $\varepsilon_t(x_d)$ are independent $N(0, \sigma^2)$ random variables. The parameter K is the length of the temporal dependence and L is the total number of reproducible patterns that we can observe up to a multiplicative constant in the process.

Figure 5.1 shows a realization of a CM3 plotted versus a M4 with $K = 3$ and $L = 2$. The spatio-temporal dependence clearly appears. The two different profiles of extremes are plotted in different colors.

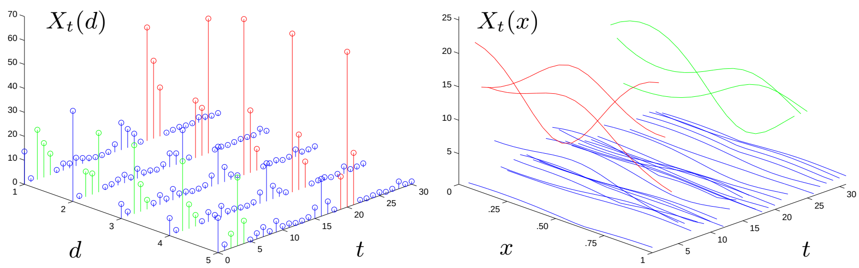


Figure 5.1: Representation of a M4 with $D = 5$ on the left and of a CM3 on $[0, 1]$ on the right (with $K = 3$ and $L = 2$)

In Section 5.2, a coherent set of properties for CM3 is established. The motivation is similar as in Segers (2006) and Smith (2003) but now for random continuous functions. Theorem 5.2.3 is the joint regular variation of those processes in the sense of Chapter 2. The spectral process of a CM3 has a discrete distribution, given by the theorem. Next, depending on the speed with which the parameter functions $a_i^{(j)}$

tend towards zero, Theorem 5.2.4 yields the finite-cluster condition and Theorem 5.2.5 yields the strong mixing condition. These three properties together also have specific implications, for instance the inverse of the extremal index θ becomes the expected size of clusters of extremes in the sense of Segers (2005).

CM3 processes are also examples of max-stable random fields (Davis and Mikosch 2008, de Haan and Ferreira 2006): for every finite space-time subset $A \times T \subset [0,1]^q \times \mathbb{Z}$, the random vector $(X_t(x))_{x \in A, t \in T}$ has a multivariate extreme value distribution. This property of M4 is inherent to CM3 since the law of a continuous random field is characterized by its finite dimensional distributions that are M4 according to Example 5.2.2.

Section 5.3 is a preparation for the estimation of the parameter functions. Extremes will play a central role in identifying the recursive patterns and their relative frequencies. So we need to study the probabilistic properties of the blocks of extremes that can be observed in CM3. The harmonic mean makes convenient the expressions of the frequencies of the reproducible patterns that can be observed.

In Section 5.4, we suggest and compare empirical methods to estimate K, L and the parameter functions without assumptions on these objects. It is a complement to Zhang and Smith (2004), where the case $L = 1$ has only been treated, and to Zhang (2008), where assumptions are made on the parameters.

This study is designed to improve the statistical analyses of extreme events, as done in Süveges (2009) and Süveges and Davison (2010) for instance.

5.2 Definition and properties

Choose a nonempty compact domain of $\mathbb{R}^q$. To not have too many notations, this compact will be taken to be $[0,1]^q$. Given an array $Z_i^{(j)}$ $(i \in \mathbb{Z}_+,\ j \in \mathbb{Z})$ of independent unit-Fréchet random variables, if $a_i^{(j)} : [0,1]^q \to \mathbb{R}_+^*$ are deterministic strictly positive continuous functions, a *maximum of moving maxima of continuous functions* (or *CM3 process*) is a stochastic process defined by

$$X_t(x) = \sup_{j \in \mathbb{Z}} \sup_{i \geqslant 0} (a_i^{(j)}(x) Z_{t-i}^{(j)}). \tag{5.1}$$

If furthermore

$$(\forall x \in [0,1]^q): \sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} a_i^{(j)}(x) = 1$$

we say that $(X_t)_{t\in\mathbb{Z}}$ is a *standardized CM3 process.*

The first result is an necessary condition before any use of CM3 processes. Recall that the sup-norm of a continuous function $f : [0,1]^q \to \mathbb{R}$ is $\|f\|_\infty = \sup_{x\in[0,1]^q} |f(x)|$ and that this supremum is achieved.

Proposition 5.2.1. *If*

$$\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} \|a_i^{(j)}\|_\infty < \infty, \tag{5.2}$$

then, for every $t \in \mathbb{Z}$, X_t in (5.1) *is a random element in $\mathcal{C}([0,1]^q, \mathbb{R}_+)$ and the process $(X_t)_{t\in\mathbb{Z}}$ is stationary.*

Proof. See page 146. □

Example 5.2.2. If $(X_t)_{t\in\mathbb{Z}}$ is a standardized CM3 with $0 \leqslant i < K$ and $1 \leqslant j \leqslant L$, then for each $x_1, \dots, x_D \in \mathbb{R}^q$, the process $(X_t(x_1), \dots, X_t(x_D))_{t\in\mathbb{Z}}$ is a standardized M4. Under (5.2) $(\|X_t\|_\infty)_{t\in\mathbb{Z}}$ is a non-standardized M3. In both cases $0 \leqslant i < K$ and $1 \leqslant j \leqslant L$. This is helpful for the estimation of K and L if $x_1, \dots, x_D$ are far enough separated, see Section 5.4.2.

A CM3 process is an example of jointly regularly varying time series defined in Chapter 2. There exists a process $(\Theta_t)_{t\in\mathbb{Z}}$ in $\mathcal{C}([0,1]^q, \mathbb{R}_+)$, called *spectral process* which is the limit in distribution, as $x \to \infty$,

$$\mathcal{L}\left((X_t/\|X_0\|_\infty)_{t\in\mathbb{Z}} \;\middle|\; \|X_0\|_\infty \geqslant x\right) \xrightarrow{d} \mathcal{L}\left((\Theta_t)_{t\in\mathbb{Z}}\right)$$

in the proper product space. According to Chapter 2, this process captures all aspects of extremal dependence, both within space and over time.

Theorem 5.2.3. *Setting $a_i^{(j)} = 0$ if $i < 0$, under condition* (5.2)*, a CM3 process $(X_t)_{t\in Z}$ is jointly regularly varying with index $\alpha = 1$ and spectral process*

$$(\Theta_{-s}, \dots, \Theta_t) \stackrel{d}{=} \left(\frac{a_{-s+I}^{(J)}}{\|a_I^{(J)}\|_\infty}, \dots, \frac{a_{t+I}^{(J)}}{\|a_I^{(J)}\|_\infty} \right),$$

where (I,J) *is a random vector on* $\mathbb{Z}_+ \times \mathbb{Z}$ *having distribution*

$$P[(I,J)=(i,j)] = \frac{\|a_i^{(j)}\|_\infty}{\sum_{l\in\mathbb{Z}}\sum_{k\geqslant 0}\|a_k^{(l)}\|_\infty},\ i\in\mathbb{Z}_+,\ j\in\mathbb{Z}.$$

Proof. See page 148. □

All CM3 processes satisfying (5.2) also satisfy the finite-cluster condition. This property prevents a sequence of extremes occurring in a CM3 from being infinite over time even if $K=+\infty$ or $L=+\infty$.

Theorem 5.2.4. *Under condition* (5.2)*, a CM3 process* $(X_t)_{t\in Z}$ *satisfies the finite-cluster condition: there exists* $(r_n)_{n\in\mathbb{N}}$ *with* $r_n\to\infty$ *and* $r_n/n\to 0$ *such that*

$$\lim_{m\to\infty}\limsup_{n\to\infty} P(\max_{m\leqslant|t|\leqslant r_n}\|X_t\|_\infty > n \mid \|X_0\|_\infty > n) = 0. \tag{C}$$

Proof. See page 150. □

Together with the finite-cluster condition, the strong mixing property leads to nice properties. To obtain the strong mixing property a sufficient condition is

$$\sum_{j\in\mathbb{Z}}\sum_{i\geqslant 0} i\|a_i^{(j)}\|_\infty < \infty. \tag{5.3}$$

Note that (5.2) and (5.3) are trivial whenever $K<+\infty$ and $L<+\infty$.

Theorem 5.2.5. *Under condition* (5.3)*, a CM3 process* $(X_t)_{t\in Z}$ *satisfies the strong mixing condition:*

$$\lim_{m\to\infty}\sup_{\substack{A\in\sigma(-\infty,-m)\\ B\in\sigma(m,\infty)}} |P(A\cap B)-P(A)P(B)| = 0 \tag{M}$$

where $\sigma(r,s)$ *is the* σ*-field generated by* $\{X_t \mid r\leqslant t\leqslant s\}$.

Proof. See page 152. □

If $(X_t)_{t\in\mathbb{Z}}$ is a regularly varying time series with index 1, the extremal index θ of the univariate time series $(\|X_t\|_\infty)_{t\in\mathbb{Z}}$ is defined as the quantity between 0 and 1 such that

$$P(\max_{1\leqslant t\leqslant n}\|X_t\|_\infty \leqslant nx) \to e^{-\theta/x}$$

as $n\to\infty$. The extremal index of a CM3 process is the following.

Proposition 5.2.6. *Under condition* (5.2)*, if* $(X_t)_{t\in Z}$ *is a CM3 process, the extremal index of* $(\|X_t\|_\infty)_{t\in\mathbb{Z}}$ *is*

$$\theta = \frac{1}{\sum_{j\in\mathbb{Z}}\sum_{i\geqslant 0}\|a_i^{(j)}\|_\infty} \sum_{j\in\mathbb{Z}} \max_{i\geqslant 0}\|a_i^{(j)}\|_\infty.$$

Proof. See page 154. □

Once conditions (C) and (M) are satisfied, which is the case under (5.3) by Theorem 5.2.4 and Theorem 5.2.5, there are further characterizations of the extremal index such that

$$\theta \quad = \quad \lim_{t\to\infty}\lim_{x\to\infty} P(\max_{i=1,\dots,t}\|X_i\|_\infty \leqslant x \mid \|X_0\|_\infty > x) \tag{5.4}$$

and, in this case, $1/\theta$ is the expected size of clusters of extremes in the sense of Segers (2005), which is recalled as follows. Let $u_n \to \infty$ be a thresholding sequence and $r_n \to \infty$ be such that the expected number of exceedances in a sample of size r_n tends towards 0:

$$E\left[\sum_{i=1}^{r_n} 1\{\|X_i\|_\infty > u_n\}\right] = r_n P(\|X_1\|_\infty > u_n) \to 0.$$

Then, denoting $M_n := \max\{\|X_1\|_\infty, \dots, \|X_n\|_\infty\}$, under (C) and (M), we have

$$E\left[\sum_{i=1}^{r_n} 1\{\|X_i\|_\infty > u_n\} \;\middle|\; M_{r_n} > u_n\right] = \frac{r_n P(\|X_1\|_\infty > u_n)}{P(M_{r_n} > u_n)} \to \frac{1}{\theta}$$

as $n \to \infty$.

5.3 Block profiles

In this section we study further probabilistic features of CM3 processes in order to build a method to estimate the parameter functions $a_i^{(j)}$ in the case $0 \leqslant i < K$ and $1 \leqslant j \leqslant L$. The theoretical model for the rest of the chapter is thus

$$X_t(x) = \max_{1\leqslant j\leqslant L}\max_{0\leqslant i<K} a_i^{(j)}(x) Z_{t-i}^{(j)} \qquad (x \in [0,1]^q) \tag{5.5}$$

The estimation method suggested in this chapter is based on the fact that a large value of $Z_i^{(j)}$ causes large values of X_t for $i \leqslant t < i+K$ and the possibility to have in this case

$$(X_t, \ldots, X_{t+K-1}) = Z_i^{(j)}(a_0^{(j)}, \ldots, a_{K-1}^{(j)}). \tag{5.6}$$

By "block profile" we mean a sequence $(X_t, \ldots, X_{t+K-1})$ satisfying (5.6) for some $1 \leqslant j \leqslant L$. The corresponding sequence of functions $(a_0^{(j)}, \ldots, a_{K-1}^{(j)})$ will be called "profile" or "pattern".

In Section 5.3.1 we compute the probability of the events (5.6) and their frequencies of occurrence for the different values of j. In Section 5.3.2 we have a brief look at the correlation between all the possible blocks of length K available in a sample. They are not independent if they overlap. In Section 5.3.3 we give the necessary sample size to expect that an event of the form (5.6) realizes at least once.

To compute the exact value of the probability of occurrence of the different blocks, the knowledge of the parameter functions is needed, which is particular not the case in the estimation. This is the reason why we also give lower and upper bounds for the true values. These bounds only depend on a unique parameter C, which is the maximal variation among the parameter functions $a_i^{(j)}$.

5.3.1 Relative frequencies

To recover the functions $a_i^{(j)}$ from a sample of size T, the first step is to understand how (5.5) works. Consider as an example a simple situation when $K = 3$ and $L = 2$. A finite number of functions $a_i^{(j)}$ is uniformly bounded from below by a positive constant since they are strictly positive. Thus, if for instance $Z_1^{(2)}$ is large enough, the value of

$(X_1, \ldots, X_K)$ at a given position $x \in [0,1]^q$ is

$$X_1(x) = \max \begin{pmatrix} a_2^{(1)}(x)Z_{-1}^{(1)} & a_1^{(1)}(x)Z_0^{(1)} & a_0^{(1)}(x)Z_1^{(1)} \\ a_2^{(2)}(x)Z_{-1}^{(2)} & a_1^{(2)}(x)Z_0^{(2)} & a_0^{(2)}(x)\mathbf{Z_1^{(2)}} \end{pmatrix} = a_0^{(2)}(x)Z_1^{(2)}$$

$$X_2(x) = \max \begin{pmatrix} a_2^{(1)}(x)Z_0^{(1)} & a_1^{(1)}(x)Z_1^{(1)} & a_0^{(1)}(x)Z_2^{(1)} \\ a_2^{(2)}(x)Z_0^{(2)} & a_1^{(2)}(x)\mathbf{Z_1^{(2)}} & a_0^{(2)}(x)Z_2^{(2)} \end{pmatrix} = a_1^{(2)}(x)Z_1^{(2)}$$

$$X_3(x) = \max \begin{pmatrix} a_2^{(1)}(x)Z_1^{(1)} & a_1^{(1)}(x)Z_2^{(1)} & a_0^{(1)}(x)Z_3^{(1)} \\ a_2^{(2)}(x)\mathbf{Z_1^{(2)}} & a_1^{(2)}(x)Z_2^{(2)} & a_0^{(2)}(x)Z_3^{(2)} \end{pmatrix} = a_2^{(2)}(x)Z_1^{(2)} \tag{5.7}$$

so that the second pattern appears:

$$(X_1, X_2, X_3)(x) = (a_0^{(2)}, a_1^{(2)}, a_2^{(2)})(x)Z_1^{(2)}.$$

How likely is this kind of events to occur? To compute their probabilities, first remark that (5.7) is equivalent to

$$\begin{array}{ll}
Z_{-1}^{(1)} \leqslant \min_x[\frac{a_0^{(2)}(x)}{a_2^{(1)}(x)}]Z_1^{(2)} & Z_{-1}^{(2)} \leqslant \min_x[\frac{a_0^{(2)}(x)}{a_2^{(2)}(x)}]Z_1^{(2)} \\
Z_0^{(1)} \leqslant \min_x[\frac{a_0^{(2)}(x)}{a_1^{(1)}(x)}, \frac{a_1^{(2)}(x)}{a_2^{(1)}(x)}]Z_1^{(2)} & Z_0^{(2)} \leqslant \min_x[\frac{a_0^{(2)}(x)}{a_1^{(2)}(x)}, \frac{a_1^{(2)}(x)}{a_2^{(2)}(x)}]Z_1^{(2)} \\
Z_1^{(1)} \leqslant \min_x[\frac{a_0^{(2)}(x)}{a_0^{(1)}(x)}, \frac{a_1^{(2)}(x)}{a_1^{(1)}(x)}, \frac{a_2^{(2)}(x)}{a_2^{(1)}(x)}]Z_1^{(2)} & - \\
Z_2^{(1)} \leqslant \min_x[\frac{a_1^{(2)}(x)}{a_0^{(1)}(x)}, \frac{a_2^{(2)}(x)}{a_1^{(1)}(x)}]Z_1^{(2)} & Z_2^{(2)} \leqslant \min_x[\frac{a_1^{(2)}(x)}{a_0^{(2)}(x)}, \frac{a_2^{(2)}(x)}{a_1^{(2)}(x)}]Z_1^{(2)} \\
Z_3^{(1)} \leqslant \min_x[\frac{a_2^{(2)}(x)}{a_0^{(1)}(x)}]Z_1^{(2)} & Z_3^{(2)} \leqslant \min_x[\frac{a_2^{(2)}(x)}{a_0^{(2)}(x)}]Z_1^{(2)}
\end{array} \tag{5.8}$$

For the general case, let $A(l^*)$ be the event (5.6) with fixed $j = l^*$:

$$A(l^*) = \{\forall x \in [0,1]^q : (X_t, \ldots, X_{t+K-1})(x) = Z_t^{(l^*)}(a_0^{(l^*)}, \ldots, a_{K-1}^{(l^*)})(x)\},$$

i.e. $A(l^*)$ is the event for a K-block starting a time t to be a block profile of type l^*. Generalizing (5.8) shows that the event $A(l^*)$ is the intersection of $(2K-1)L-1$ conditions involving the random variables

$Z_i^{(j)}$ for $t-K+1 \leqslant i \leqslant t+K-1$. Remembering that the density of Z is $f_Z(z) = z^{-2}\exp(-z^{-1})$, the probability $p^{(l^*)}$ of $A(l^*)$ is

$$\begin{aligned} p^{(l^*)} \quad = \quad & \int_0^\infty \prod_{l=1}^{L} P(Z \leqslant \min_x [\frac{a_0^{(l^*)}(x)}{a_{K-1}^{(l)}(x)}] z) \\ & \prod_{l=1}^{L} P(Z \leqslant \min_x [\frac{a_0^{(l^*)}(x)}{a_{K-2}^{(l)}(x)}, \frac{a_1^{(l^*)}(x)}{a_{K-1}^{(l)}(x)}] z) \\ & \cdots \\ & \prod_{\substack{l=1 \\ l \neq l^*}}^{L} \; P(Z \leqslant \min_x [\frac{a_0^{(l^*)}(x)}{a_0^{(l)}(x)}, \ldots, \frac{a_{K-1}^{(l^*)}(x)}{a_{K-1}^{(l)}(x)}] z) \\ & \cdots \\ & \prod_{l=1}^{L} P(Z \leqslant \min_x [\frac{a_{K-2}^{(l^*)}(x)}{a_0^{(l)}(x)}, \frac{a_{K-1}^{(l^*)}(x)}{a_1^{(l)}(x)}] z) \\ & \prod_{l=1}^{L} P(Z \leqslant \min_x [\frac{a_{K-1}^{(l^*)}(x)}{a_0^{(l)}(x)}] z) z^{-2} \exp(-z^{-1}) dz \end{aligned} \tag{5.9}$$

Denoting $m_k^{(l;l^*)}$ the minimum written on line k of (5.9) and by extension $m_k^{(l^*;l^*)} = 1$, the value of $p^{(l^*)}$ is

$$p^{(l^*)} = \frac{1}{1 + \sum_{k=1}^{2K-1} \sum_{\substack{l=1 \\ l \neq l^* \text{ if } k=K}}^{L} \frac{1}{m_k^{(l;l^*)}}} = \frac{\mathrm{harm}(m_1^{(1;l^*)}, \ldots, m_{2K-1}^{(L;l^*)})}{(2K-1)L} \tag{5.10}$$

where harm is the harmonic mean of the $(2K-1)L$ minima.

It is thus possible to compute $p^{(l^*)}$ exactly given the parameter functions. If the parameter functions $a_i^{(j)}$ satisfy

$$\frac{1}{C^{(l^*)}} \leqslant \frac{a_k^{(l^*)}(x)}{a_{k'}^{(l)}(x)} \leqslant C^{(l^*)} \tag{5.11}$$

for all x, k, k', l, l^*, then the probability $p^{(l^*)}$ of success to reveal

$$(a_0^{(l^*)}, \ldots, a_{K-1}^{(l^*)})$$

by picking up a random block satisfies

$$\underline{p}^{(l^*)} := \frac{1}{C^{(l^*)}(2K-1)L} \leqslant p^{(l^*)} \leqslant \frac{C^{(l^*)}}{(2K-1)L} =: \overline{p}^{(l^*)}. \tag{5.12}$$

The harmonic mean being more sensitive to small values, the lower bound is actually closer to the exact probability.

Under the knowledge of K, the probability that a random K-block is a profile differs from pattern to pattern. But under the control condition (5.11), if all $C^{(l^*)}$ are themselves bounded above by a common constant C, the probability $p = p^{(1)} + \ldots + p^{(L)}$ for a random block to be any profile can be estimated by

$$\underline{p} := \frac{1}{C(2K-1)} \leqslant p. \tag{5.13}$$

which has the remarkable property not to depend neither on L nor on the dimension of the ambient space.

As an illustration, Table 5.1 shows the number of found block profiles versus their expectations, knowing and without knowing the parameter functions for five simulations of (5.16). The different patterns are split in columns.

Simulation with parameters $C=5$, $D=20$, $K=5$, $L=5$ and $T=5000$												
Expected value ($p^{(l^*)}T$)						$\underline{p}T$	Really found					
#1	#2	#3	#4	#5	total	total	total	#1	#2	#3	#4	#5
31	31	30	34	33	159	135	147	28	20	25	30	44
31	32	33	29	36	161	135	159	30	20	43	36	30
35	35	31	31	33	166	135	166	31	35	24	30	46
32	32	36	33	33	165	135	160	33	26	43	27	31
30	30	30	31	33	154	135	143	31	28	32	19	33

Table 5.1: CM3: Estimation of the number of block profiles versus the corresponding values obtained in simulation

5.3.2 Correlation

As we have seen in Section 5.3.1, we need

$$(2K-1)L/\,\mathrm{harm}(m_1^{(1;l^*)}, \ldots, m_{2K-1}^{(L;l^*)})$$

independent random blocks from the series $(X_t)_{t\in\mathbb{Z}}$ to expect that at least one is proportioned like the $l^{*\text{th}}$ profile. Practically, given a chain $X_1, \ldots, X_T$ with T observations, we have $T - K + 1$ dependent blocks of length K. The main pieces of information about the dependence structure between these blocks can be summarized in the following way.

i) If a K-block $(X_t, \ldots, X_{t+K-1})$ is a block profile, it does not overlap with another block profile.

ii) Given that the K-block $(X_t, \ldots, X_{t+K-1})$ is not a profile, the probability that one of the $K-1$ next K-blocks $(X_{t+1}, \ldots, X_{t+K}), \ldots$ is a profile is higher than without conditioning.

iii) If consecutive blocks are not profiles, the probability that the next one is a profile stops increasing after K non-profile blocks.

To see i), for instance, have a look at the third matrix in (5.7). If $(X_1, \ldots, X_3)$ is like the second profile, then in particular $a_2^{(2)} Z_1^{(2)} > a_0^{(1)} Z_3^{(1)}$. But for $(X_3, \ldots, X_5)$ to be like the first profile, we must have $a_2^{(2)} Z_1^{(2)} < a_0^{(1)} Z_3^{(1)}$. Then $(X_3, \ldots, X_5)$ cannot be proportioned like the first profile. Proceed similarly for any two non-disjoint blocks and any two different profiles.

To see ii), if for instance $(X_1, \ldots, X_3)$ in (5.8) is not a profile, that means that at least one of the 15 inequalities is not satisfied, although we do not know precisely how many. Some of the reverse inequalities lie in the conditions for the $K-1$ next blocks to be profile and some do not. To get the exact incidence, we need to condition on the number of inequalities not satisfied in (5.8) and whether or not they participate in the conditions for the next blocks to be profile. In any case: the probability that the next blocks are profile increases.

To see iii), simply remark that the process is K-dependent.

5.3.3 Sample size

As a consequence of Section 5.3.2, the expected number of profiles of type l is greater than unity in a chain of length at least

$$T = K - 1 + \frac{(2K-1)L}{\operatorname{harm}(m_1^{(1;l)}, \ldots, m_{2K-1}^{(L;l)})}. \tag{5.14}$$

Given an upper bound C on all the $C^{(l^*)}$ in (5.11), the minimum sample size needed to expect at least M repetitions of a particular profile in the chain is

$$T = K - 1 + CM(2K - 1) \qquad (5.15)$$

if we do not know the parameter functions but only C.

5.4 Estimation

The estimation methodology for the parameter functions $a_i^{(j)}$ starts from a discretization at D points x_d $(1 \leqslant d \leqslant D)$ of the domain. CM3 processes from (5.5) are seen as a high-dimensional M4. Furthermore we may want to consider independent and normally distributed errors with variance σ^2 at each measurement point. Thus the "practical model" studied in this section is

$$X_t(d) = \max_{1\leqslant j\leqslant L} \max_{0\leqslant i<K} a_i^{(j)}(x_d) Z_{t-i}^{(j)} + \varepsilon_t(x_d) \qquad (1 \leqslant d \leqslant D) \qquad (5.16)$$

where $(Z_t)_{t\in\mathbb{Z}}$ are independent unit-Fréchet, $\varepsilon_t(x_d)$ are independent $N(0, \sigma^2)$ random variables, the $a_i^{(j)}(x)$ are positive continuous functions defined on $[0,1]^q$ and for every $x \in [0,1]^q$ we have that

$$\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} a_i^{(j)}(x) = 1.$$

It is important to note that the profiles $(a_0^{(j)}, \ldots, a_{K-1}^{(j)})$ contain not only the information about the shapes of the profile but also, according to Section 5.3.1, their probability of occurrence. The shapes will be denoted $a_{i,0}^{(j)}$ while the frequencies of occurrence $f^{(j)}$. That is

$$a_i^{(j)} = \alpha^{(j)} a_{i,0}^{(j)} \qquad (5.17)$$

where the coefficients $\alpha^{(j)}$ must be chosen so that $f^{(j)} = p^{(j)}$ in Section 5.3.1.

The first step of the procedure is to estimate the length of the tail dependence K. This is done in Section 5.4.1 taking the average size of the clusters of exceedances over a threshold. Next blocks of extremes are selected to estimate number of patterns L, the shapes $a_{i,0}^{(j)}$ of the parameter functions and their frequencies $f^{(j)}$. The algorithm to locate the blocks of extremes explained in Section 5.4.2 is based on a multivariate approach. The value of L is determined in Section 5.4.3 as the number of clusters among the chosen blocks of extremes. The functions $a_{i,0}^{(j)}$ are yielded by the natural output (centroids, medoids, ...) of the partitioning algorithm used to determine L and the values $f^{(j)}$ through the size of the different clusters. Then the solution $a_i^{(j)}$ of (5.17) is obtained in Section 5.4.4 thanks to an iterative algorithm. To measure the quality of the estimation, the quantification of the dissimilarity between the original parameter functions and their estimations is done in Section 5.4.5 in terms of the Hausdorff distance.

5.4.1 Estimation of the length of the tail dependence

Eight estimators of K have been tested for the model (5.16) with $\sigma = 0$ when the knowledge of the parameter functions $a_i^{(j)}$ is replaced by the range C given in (5.11).

The first step is to select the values considered as extremes. A first approach consist of working on $(\|X_t\|_\infty)$ and choosing the values above the threshold $\max_t(\|X_t\|_\infty)/C$. This will be referred as the scalar version. A second approach can be to use all the available information by doing the previous operation for the D components of (X_t) separately. In this last case the threshold also depends on the location. This method will be referred as the multivariate version.

Once the extremes are selected, a runs declustering generates a sequence (s_n) of all sizes of clusters of extremes found in the univariate or multivariate scan. More precisely, only contiguous extremes were considered here to make a cluster. This is the runs declustering with $r = 0$.

From the sequence (s_n), we estimate K through mean(s_n), median(s_n) or mode(s_n) with the nomenclature as follow.

	Average	Time series	Threshold (scalar or vector)
$\hat{K}_\mu$	mean	$(\|X_t\|_\infty)$	$\max_t((\|X_t\|_\infty)/C$
$\hat{K}_\mu^M$	mean	(X_t)	$(\max_t(X_t(d))/C)_{1\leqslant d\leqslant D}$
$\hat{K}_m$	median	$(\|X_t\|_\infty)$	$\max_t(\|X_t\|_\infty)/C$
$\hat{K}_m^M$	median	(X_t)	$(\max_t(X_t(d))/C)_{1\leqslant d\leqslant D}$
$\hat{K}_o$	mode	$(\|X_t\|_\infty)$	$\max_t(\|X_t\|_\infty)/C$
$\hat{K}_o^M$	mode	(X_t)	$(\max_t(X_t(d))/C)_{1\leqslant d\leqslant D}$

The ceil or floor options to get rid of the decimals are used to build the eight following estimators:

$$\hat{K}_1 = \text{ceil}(K_\mu), \quad \hat{K}_3 = \text{ceil}(K_\mu^M), \quad \hat{K}_5 = \text{ceil}(K_m), \hat{K}_7 = \hat{K}_o,$$
$$\hat{K}_2 = \text{round}(K_\mu), \hat{K}_4 = \text{round}(K_\mu^M), \hat{K}_6 = \text{ceil}(K_m^M), \hat{K}_8 = \hat{K}_o^M.$$

Figure 5.2 shows the success rate the eight estimators of K against the length of the simulated chain. The tests were performed with $N = 50000$ trials at each step: for C from 1 to 10, D from 1 to 20, K from 1 to 5, L from 1 to 5, $\sigma = 0$ and, for each of these parameters, 10 different sets of coefficients $a_i^{(j)}$ randomly generated (uniformly, without time or space correlation).

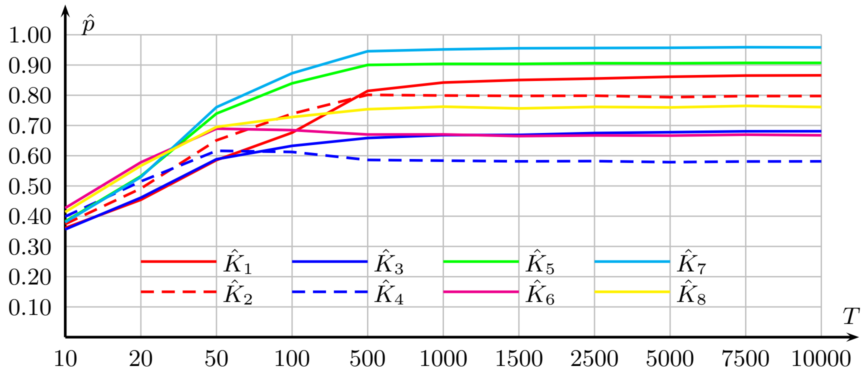


Figure 5.2: CM3: Proportion of success in estimating K

According to these empirical results, the winner for $T \geqslant 35$ is the univariate version and the mode as average cluster size. If $T \leqslant 35$ the best success rate is obtained with the multivariate version and the median.

5.4.2 Extremal clustering

Once we know the length K of the tail dependence thanks to Section 5.4.1, the next step of the procedure studied here to recover the parameter functions $a_i^{(j)}$ of a theoretical CM3 process (5.5) is to locate the blocks of extremes. Indeed, according to Section 5.3.2, the probability that at least one block of length K in the chain is block profile of type l in a sample of size T is greater than

$$1-(1-p^{(l)})^{T-K+1}$$

which tends to 1 as T tends to infinity.

The suggested method locates the positions of blocks of extremes in the practical model (5.16) maximizing the "likelihood" of being a multivariate extreme. This idea comes from the wish not to lose information across the D dimensions. Nevertheless a bad situation can still happen when, for some $0 \leqslant i^* < K$, all $a_{i^*}^{(j)}(x_d)$ are negligible in comparison with the $a_i^{(j)}(x_d)$, $i \neq i^*$, for instance. If the points $x_1, \dots, x_D$ of the discretization are far enough separated to obtain independent-like patterns, it is unlikely that all the $a_{i^*}^{(j)}(x_d)$ are negligible in the same time.

We explain the method on the following example with $D = 2$ and $K = 3$:

$X_t(1)$	5	3	4	14	19	2	7
$X_t(2)$	6	1	10	5	2	1	4

.

First step

Using the order statistics, mark the K largest values in the D chains by 1.

$d=1$	0	0	0	1	1	0	1
$d=2$	1	0	1	1	0	0	0

Second step

Compute the sum of the extremal status for each t.

$d=1$	0	0	0	1	1	0	1
$d=2$	1	0	1	1	0	0	0
$\Rightarrow \lambda$	1	0	1	2	1	0	1

Third step

Compute the moving sum (MS) of order K. This is considered as the likelihood λ to have an large value at time t among the $(Z_t^{(l)})_{1 \leqslant l \leqslant L}$.

λ	1	0	1	2	1	0	1
MS	1	1	2	3	4	3	2

Extract the profile

Find the index t that maximizes the moving sum. Then

$$(X_{t-K+1}, \dots, X_t)/\|X_{t-K+1}\|_\infty \tag{5.18}$$

is the shape of the first block profile to store in the memory:

MS	1	1	2	3	4	3	2
$X_t(1)$	5	3	4	14	19	2	7
$X_t(2)$	6	1	$\underline{10}$	5	2	1	4

$$\begin{array}{lcl} \text{Shape}_1(x_1) = & (& 4/\underline{10}, \quad 14/\underline{10}, \quad 19/\underline{10} \quad) \\ \text{Shape}_1(x_2) = & (& 10/\underline{10}, \quad 5/\underline{10}, \quad 2/\underline{10} \quad) \end{array}$$

To decide between multiple maxima, for instance if

MS	1	1	2	4	2	4	4

we first choose the single maximum. If there are consecutive maxima, as a second criterion, we take the block that maximizes $\sum_{t\in\text{block}} \|X_t\|_\infty$ among those.

Set the block that has just been considered to 0 and repeat this loop until having gathered the desired number Q of temporally disjoint blocks of extremes of length K ($Q = \underline{p}T$ is the suggestion of Section 5.3.1 if we only know a uniform bound C on the variation of the parameter functions).

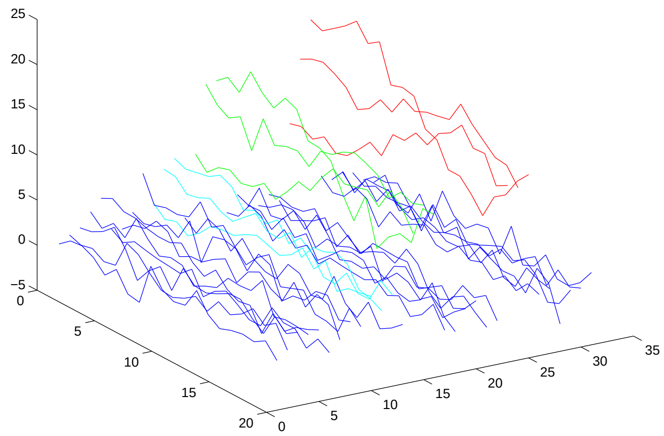


Figure 5.3: Blocks of extremes obtained for a simulated CM3 with measurement errors ($\sigma = 1$)

5.4.3 Estimation of the number of patterns

With Q blocks of extremes of length K normalized as in (5.18), the goal is to estimate the number of reproducible patterns L in the observed process. To do this, we create a $Q \times KD$-table inside which each of the Q lines is made of the D temporal vectors of length K placed successively. We estimate L with the number of clusters for the observations of this table.

Partitioning methods

To break the lines of the table up into groups, we tried several algorithms among which five retained our attention: hierarchical clustering with Ward's aggregation criterion, hierarchical clustering with the Euclidean distance between the centroids (Hastie, Tibshirani, and Friedman 2009, Ward 1963), k-means with the Euclidean squared distance, k-means with Pearson's correlation after standardization (Seber 1984, Spath 1985) and finally Partitioning Around Medoids (PAM) with the classical Euclidean distance (Theodoridis and Koutroumbas 2006, van der Laan, Pollard, and Bryan 2003).

Number of clusters

For each of those algorithms we implemented two criteria to determine the number of clusters. The first: one stops when the percentage of the total variance not explained by the clustering is less than 20%, i.e. when

$$\frac{\text{SSE}}{\text{SStot}} = \frac{\sum_{\text{cluster}}(\#\,\text{obs(cluster)} - 1)\sum_{\text{variable in cluster}} s^2(\text{variable})}{(Q-1)\sum_{\text{variable in table}} s^2(\text{variable})} \leqslant .2.$$

We refer to this method as the *elbow method* (Ketchen and Shoo 1996, Mardia, Kent, and Bibby 1979), see Figure 5.4.

The second method to find the number of clusters here is the first value that yields a total silhouette TtSil for the clustering above 85% of Q. Let $a(q)$ be the average distance between the q^{th} observation and the members of its own cluster. Then repeat this operation between the q^{th} observation and the members of all the other clusters, and set $b(q)$ to the lowest value found. The silhouette $s(q)$ of the q^{th} observation is

$$s(q) = \frac{b(q) - a(q)}{\max\{a(q), b(q)\}}.$$

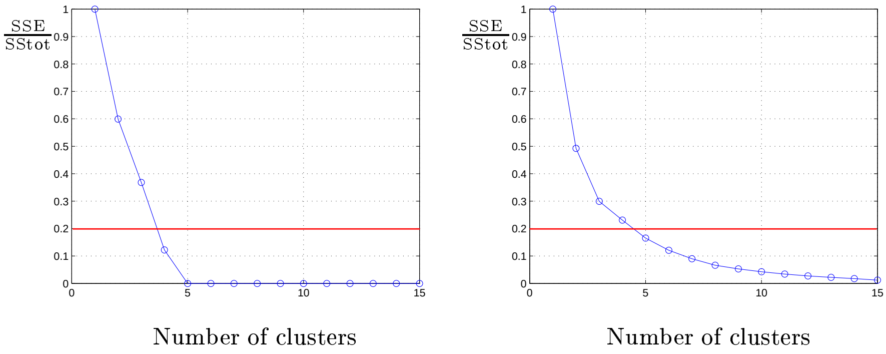


Figure 5.4: Elbow method: perfect clustering on the left, CM3 with errors on the right ($L = 5$)

Thus $-1 \leqslant s(q) \leqslant 1$ and $s(q)$ measures how dissimilar the q^{th} observation is to its own cluster (Kaufman and Rousseeuw 1990, Lleti, Ortiz, L.A., and Sánchez 2004, Rousseeuw 1987). The distance taken into account here is the Euclidean squared distance. We stop the partitioning at the smallest number of clusters satisfying

$$\frac{\text{TtSil}}{Q} = \frac{1}{Q} \sum_{q=1}^{Q} s(q) \geqslant 0.85$$

if this occurs. We refer to this method as the *silhouette method* (see Figure 5.5).

Both methods are unable to detect that $L = 1$. We have thus considered $\hat{L} = 1$ when the estimated variance of each variable is less than 0.005.

<u>Estimation</u>

For $\sigma = 0$, Figure 5.6 shows the success rate of the following eleven estimators of L against the length of the simulated chain.

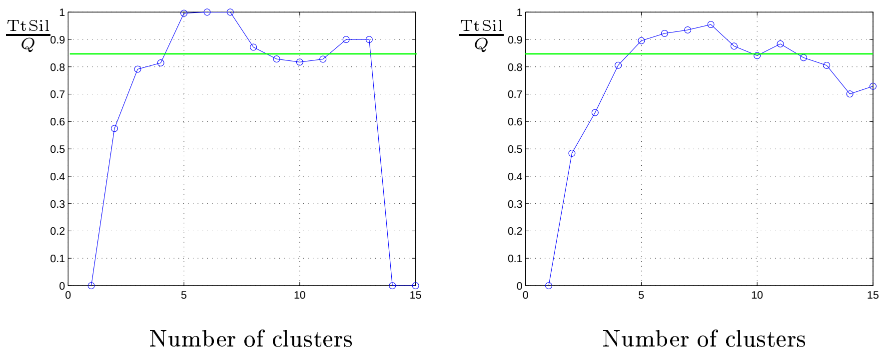


Figure 5.5: Silhouette method: Perfect clustering on the left, CM3 with errors on the right ($L = 5$)

	Clustering	Distance	Number of clusters
$\hat{L}_1$	Hierarchical	Ward	Elbow
$\hat{L}_2$	Hierarchical	Ward	Silhouette
$\hat{L}_3$	Hierarchical	Euclidean / centroids	Elbow
$\hat{L}_4$	Hierarchical	Euclidean / centroids	Silhouette
$\hat{L}_5$	k-means	Euclidean squared	Elbow
$\hat{L}_6$	k-means	Euclidean squared	Silhouette
$\hat{L}_7$	k-means	Pearson's correlation	Elbow
$\hat{L}_8$	k-means	Pearson's correlation	Silhouette
$\hat{L}_9$	PAM	Euclidean	Elbow
$\hat{L}_{10}$	PAM	Euclidean	Silhouette
$\hat{L}_{11}$	Consensus: $\hat{L}_{11} = \mathrm{mode}(\hat{L}_1, \ldots, \hat{L}_{10})$		

The tests were performed with $N = 6000$ trials at each step: for C in $\{2, 4, 6, 8, 10\}$, D in $\{1, 5, 10, 15, 20\}$, K from 2 to 5 and L from 1 to 5. The number of blocks of extremes chosen to build the table is $Q = \min(\mathrm{ceil}(\underline{p}T), 100)$ with $\underline{p}$ from 5.3.1. We added the constraint that it has to be possible to see each profile once, i.e. $(K+1)L \leqslant T$, to exclude challenges such as finding 5 profiles of length 5 in a chain of length 10.

The combination between PAM and elbow yields the best success rate for small sample sizes $T \leqslant 50$. For $T \geqslant 50$ Ward's algorithm with the silhouette is the best. Here the consensus curve (mode) does not

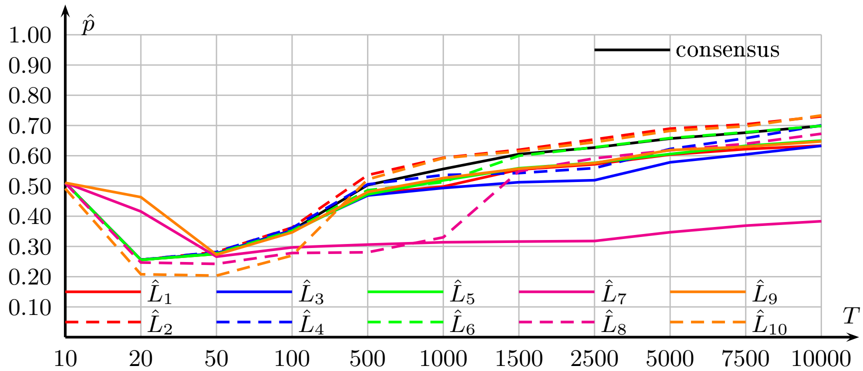


Figure 5.6: CM3: Proportion of success in estimating L

provide a better performance.

The gap between $T = 20$ and $T = 500$ is an intermediate area between a situation where the trivial case $Q = L$ frequently occurs and the apparition of asymptotic properties.

These results are not so excellent but, since the implemented algorithm is the "eye" of the analyst, it only sees what we want it to see. For real applications it probably does not matter if not very frequent profiles are missed. Moreover optimizing C, Q and the thresholds for a precise sample can improve the performances for that particular situation.

5.4.4 Recovering the parameter functions

Once wet got the length K of the tail dependence in Section 5.4.1 and the number L of different patterns in Section 5.4.3, it remains to estimate the parameter functions $a_i^{(j)}$ $(0 \leqslant i < K,\ 1 \leqslant j \leqslant L)$. Depending on the partitioning method to estimate L, we choose for the shapes $a_{i,0}^{(j)}$ of the L profiles the natural output of the algorithm: the centroids for the hierarchical clustering and k-means and the medoids with PAM. Essentially the difference is that in the first case the estimator of a shape is a mean of observations and in the second case a median observation.

The relationship between the parameter functions $a_i^{(j)}$, the shapes $a_{i,0}^{(j)}$ and their frequencies of occurrence is given by (5.10). The theoretical probabilities of occurrence of the different profiles must match the empirical frequencies $(f^{(1)}, \ldots, f^{(L)})$ in the table of Section 5.4.3. Thus the last step is to normalize the $a_{i,0}^{(j)}$ to obtain $\hat{a}_i^{(j)}$ such that on the one hand

$$\operatorname{harm}\left((m_k^{(l;1)})_{k,l}\right)/f^{(1)} = \ldots = \operatorname{harm}\left((m_k^{(l;L)})_{k,l}\right)/f^{(L)} \qquad (5.19)$$

$(1 \leqslant k \leqslant 2K-1,\ 1 \leqslant l \leqslant L)$ and on the other hand

$$\sum_{j\in\mathbb{Z}}\sum_{i\geqslant 0}\hat{a}_i^{(j)}(x_d) = 1 \qquad (1 \leqslant d \leqslant D). \qquad (5.20)$$

We seek a relation of the type $\hat{a}_i^{(l)} = \alpha^{(l)}a_{i,0}^{(l)}$ $(1 \leqslant l \leqslant L)$ which is a problem of rank L. The solution can be obtained iteratively: if $a_{i,n}^{(j)}$ is the n^{th} update of $a_{i,0}^{(j)}$ and $(p_n^{(1)},\ldots,p_n^{(L)})$ the probabilities given by (5.10) at step n, then define $a_{i,n+1}^{(j)}$ by

$$a_{i,n+1}^{(j)}(\cdot) = a_{i,n+1}^{(j)}(\cdot)\frac{f^{(j)}}{p_n^{(j)}} \qquad (5.21)$$

and stop when the error $|(f^{(1)},\ldots,f^{(L)}) - (p_n^{(1)},\ldots,p_n^{(L)})|_\infty$ is small. The iterative algorithm (5.21), if it converges, converges to the solution of (5.19). Indeed, let

$$M_n^{(l^\star)} = \begin{pmatrix} m_{1,n}^{(1:l^\star)} & \ldots & m_{2K-1,n}^{(1:l^\star)} \\ \ldots & \ldots & \ldots \\ m_{1,n}^{(L:l^\star)} & \ldots & m_{2K-1,n}^{(L:l^\star)} \end{pmatrix}$$

be the n^{th} update of the initial collection of numbers $(m_{k,0}^{(l;l^\star)})_{k,l}$ given the $a_{i,0}^{(j)}$. Define

$$T : \mathbb{R}^{(2K-1)L^2} \to \mathbb{R}^{(2K-1)L^2} : (M_n^{(1)},\ldots,M_n^{(L)}) \mapsto (M_{n+1}^{(1)},\ldots,M_{n+1}^{(L)})$$

as the resulting operation of (5.21) on the $m_{k,n}^{(l;l^\star)}$. The operator T acts

n	0	1	2	3	4	5	6	7	8
$M_n^{(1)}$	5.95 2.62	5.95 4.70	5.95 4.09	5.95 4.23	5.95 4.19	5.95 4.20	5.95 4.20	5.95 4.20	5.95 4.20
$p_n^{(1)}$	**3.64**	**5.25**	**4.85**	**4.94**	**4.91**	**4.92**	**4.92**	**4.92**	**4.92**
$M_n^{(2)}$	6.03 7.11	3.36 7.11	3.86 7.11	3.74 7.11	3.77 7.11	3.76 7.11	3.76 7.11	3.76 7.11	3.76 7.11
$p_n^{(2)}$	**6.53**	**4.57**	**5.01**	**4.90**	**4.93**	**4.92**	**4.92**	**4.92**	**4.92**

Table 5.2: Realization of the iterative algorithm (5.21) with parameters $K = 1$, $L = 2$, $f_1 = f_2 = 1$ given the initial values $M_0^{(1)}$ and $M_0^{(2)}$

so that $M_n^{(l^\star)}$ becomes

$$M_{n+1}^{(l^\star)} = \begin{pmatrix} (m_{1,n}^{(1:l^\star)} & \dots & m_{2K-1,n}^{(1:l^\star)}) \dfrac{f^{(l^\star)}}{p_n^{(l^\star)}} \dfrac{p_n^{(1)}}{f^{(1)}} \\ \dots & \dots & \dots \quad \vdots \\ (m_{1,n}^{(l^\star:l^\star)} & \dots & m_{2K-1,n}^{(l^\star:l^\star)}) \quad 1 \\ \dots & \dots & \dots \quad \vdots \\ (m_{1,n}^{(L:l^\star)} & \dots & m_{2K-1,n}^{(L:l^\star)}) \dfrac{f^{(l^\star)}}{p_n^{(l^\star)}} \dfrac{p_n^{(L)}}{f^{(L)}} \end{pmatrix}.$$

Observe that (5.19) is equivalent to

$$T(M^{(1)}, \dots, M^{(L)}) = (M^{(1)}, \dots, M^{(L)}).$$

Consequently, if $(M_n^{(1)}, \dots, M_n^{(L)})_n$ converges, since T is continuous, (5.21) converges to the solution of (5.19). The convergence seems generally fast as shows the example of Table 5.2. This yields

$$\alpha^{(l)} = \alpha \frac{f^{(l)}}{p_0^{(l)}} \frac{f^{(l)}}{p_1^{(l)}} \frac{f^{(l)}}{p_l^{(l)}} \cdots \qquad (1 \leqslant l \leqslant L). \tag{5.22}$$

where any $\alpha > 0$ suits to obtain (5.19). To simultaneously obtain (5.20), we know according to Section 5.3.1 that the solution of (5.19) and (5.20) together exists and is unique, thus, as $n \to \infty$ the value of

$\sum_{j\in\mathbb{Z}}\sum_{i\geqslant 0} a_{i,n}^{(j)}(x)$ cannot depend on x. Although, because of numerical reasons, it may slightly vary with x. Thus we suggest to keep the dependence in x and replace α in (5.22) by

$$\alpha(x) = \lim_{n\to\infty} \frac{1}{\sum_{j\in\mathbb{Z}}\sum_{i\geqslant 0} a_{i,n}^{(j)}(x)}$$

in (5.22). This completes the procedure to estimate the $a_i^{(j)}$ form observations.

Figure 5.7 shows an output of the full algorithm using PAM, with measurement errors but given the true values of K and L.

5.4.5 Distance between two sets of parameter functions

To measure the quality of the estimation, it is necessary to quantify the dissimilarity between the original parameter functions $a_i^{(j)}$ ($0 \leqslant i < K$, $1 \leqslant j \leqslant L$) and their estimations $\hat{a}_i^{(j)}$ ($0 \leqslant i < \hat{K}$, $1 \leqslant j \leqslant \hat{L}$). If $\hat{K} \neq K$ the estimation can certainly be qualified as bad, so that only the case $\hat{K} = K$ requires a discussion.

The order in which the different patterns are retrieved can change and their total numbers can differ. Consequently the Hausdorff distance between the L graphs of $(a_i^{(j)}(x))_i$ ($1 \leqslant j \leqslant L$) and the $\hat{L}$ graphs of $(\hat{a}_i^{(j)}(x))_i$ ($1 \leqslant j \leqslant \hat{L}$), that are compact in $[0,1]^q \times \mathbb{R}^K$ (or in $\{1,\ldots,D\}\times \mathbb{R}^K \cong \mathbb{R}^{DK}$ for the discrete version), perfectly suits. Recall that the Hausdorff distance between nonempty compacts A and B is

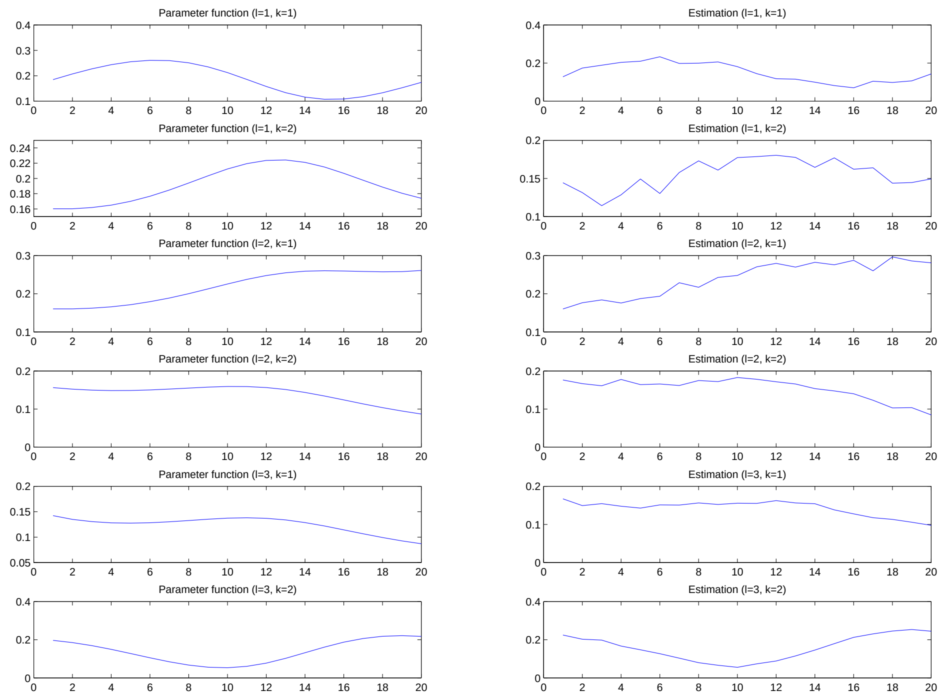
$$d_{\mathcal{H}}(A,B) = \max\{\sup_{a\in A} d(a,B), \sup_{b\in b} d(b,A)\},$$

here considered with the Euclidean distance d. We have thus to compute $d_{\mathcal{H}}(A,B)$ with

$$A = \bigcup_{j=1}^{L} \operatorname{graph}[(a_i^{(j)}(x))_{0\leqslant i<K}] \quad \text{and} \quad B = \bigcup_{j=1}^{\hat{L}} \operatorname{graph}[(\hat{a}_i^{(j)}(x))_{0\leqslant i<K}]$$

to reach the stated goal.

To illustrate Section 5.4.4, Figure 5.8 shows smoothed histograms of the Hausdorff distance with the estimation of the parameter functions with sample sizes $T = 100$, 500, 1000, 5000 given the true values of K



$T = 500$

Figure 5.7: CM3: Recovery of the parameter functions with $D = 20$, $K = 2$, $L = 3$, $\sigma = 1$ and PAM

and L. The test was performed with $N = 5000$ trials for each sample size: C running in $\{2,4,6,8,10\}$, D in $\{1,5,10,15,20\}$, K from 2 to 5, L from 1 to 5 with 10 repetitions of each. The other parameters were $Q = \min(\text{ceil}(\underline{p}T), 100)$ with $\underline{p}$ from Section 5.3.1 and $\sigma = 1$.

5.5 Online package

The Matlab code written for this chapter is available on the Matlab Central File Exchange at
http://www.mathworks.com/matlabcentral/fileexchange/26539.

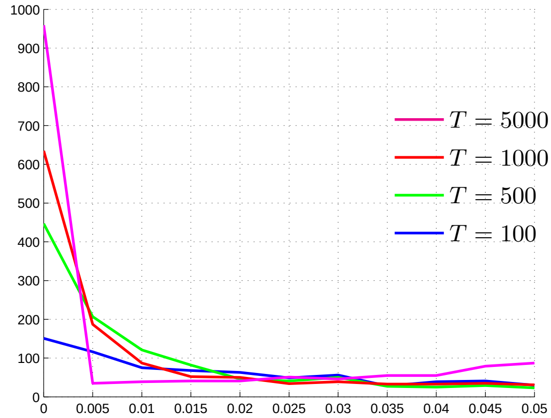


Figure 5.8: CM3: Smoothed histograms of the Hausdorff distance between the original parameter functions and their estimations

5.6 Proofs for Section 5.2

Proof of Proposition 5.2.1 (page 125). CLAIM 1 - *X_t is a random element in the set $\mathcal{C}([0,1]^q, \mathbb{R}_+)$.* Given $t \in \mathbb{Z}$, write $X \stackrel{d}{=} X_t$ and

$$X^{(m)} = \sup_{i,|j|\leqslant m} a_i^{(j)} Z_{t-i}^{(j)}.$$

For $z > 0$, since the $Z_i^{(j)}$ are iid unit-Fréchet,

$$\begin{aligned} P(\|X\|_\infty \leqslant z) &= P(\forall i \geqslant 0, \forall j \in \mathbb{Z} : Z_{t-i}^{(j)} \leqslant z / \|a_i^{(j)}\|_\infty) \\ &= \prod_{\substack{j \in \mathbb{Z} \\ i \geqslant 0}} \exp\bigl(-\|a_i^{(j)}\|_\infty / z\bigr) = \exp\bigl(-\textstyle\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} \|a_i^{(j)}\|_\infty / z\bigr). \end{aligned}$$

If $\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} \|a_i^{(j)}\|_\infty = \infty$, then this probability is zero for all $z > 0$, hence $\|X\|_\infty = \infty$ with probability one.

If $\sum_{j\in\mathbb{Z}} \sum_{i\geqslant 0} \|a_i^{(j)}\|_\infty < \infty$, then

$$X^{(m)} \leqslant X \leqslant X^{(m)} + \sup_{i,|j|>m} \|a_i^{(j)}\|_\infty Z_{t-i}^{(j)}$$

and thus $\|X - X^{(m)}\|_\infty \leqslant \sup_{i,|j|>m} \|a_i^{(j)}\|_\infty Z_{t-i}^{(j)}$. Writing

$$Y_m = \sup_{i,|j|>m} \|a_i^{(j)}\|_\infty Z_i,$$

we have $Y_m \geqslant 0$ and Y_m is decreasing in m. By a similar computation as above, we find that for $y > 0$,

$$P(Y_m \leqslant y) = \exp\bigl(-\textstyle\sum_{i,|j|>m} \|a_i^{(j)}\|_\infty / y\bigr).$$

As a consequence, $Y_m \to 0$ in probability and, by monotonicity, almost surely. Since the uniform limit of a sequence of continuous functions is continuous, by monotonicity, X is continuous with probability one.

Then the fact that, for every $t \in \mathbb{Z}$, the map $\omega \mapsto X_t(\omega, \cdot)$ with values in $\mathcal{C}([0,1]^q, \mathbb{R}_+)$ is measurable follows easily.

Claim 2 - $(X_t)_{t \in Z}$ *is a stationary time series.* In extension, this property is that, for every $n \in \mathbb{N}$, every $h \geqslant 0$ and every measurable set $A \subset \mathcal{C}([0,1]^q, \mathbb{R}^h)$,

$$P((X_0, \ldots, X_n) \in A) = P((X_h, \ldots, X_{n+h}) \in A).$$

The argument is based on two facts:
1) It suffices to pick up $x \in [0,1]^q$, $z_0, \ldots, z_n \in \mathbb{R}$ and verify that

$$P(X_0(x) \leqslant z_0, \ldots, X_n(x) \leqslant z_n) = P(X_h(x) \leqslant z_0, \ldots, X_{n+h}(x) \leqslant z_n),$$

the case with k points $x_1, \ldots x_k \in [0,1]^q$ being similar.

2) The right hand-side of the expression in 1) admits the expansion

$$\begin{aligned}
&P(X_h(x) \leqslant z_0, \ldots, X_{n+h}(x) \leqslant z_n) \\
&= P(\sup_{j\in\mathbb{Z}} \sup_{i\geqslant 0} a_i^{(j)}(x) Z_{h-i}^{(j)} \leqslant z_0, \ldots, \sup_{j\in\mathbb{Z}} \sup_{i\geqslant 0} a_i^{(j)}(x) Z_{n+h-i}^{(j)} \leqslant z_n) \\
&= P(\forall j_0 \in \mathbb{Z}, \forall i_0 \geqslant 0 : a_{i_0}^{(j_0)}(x) Z_{h-i_0}^{(j_0)} \leqslant z_0, \ldots, \\
&\qquad\qquad \forall j_n \in \mathbb{Z}, \forall i_n \in \mathbb{N} : a_{i_n}^{(j_n)}(x) Z_{n+h-i_n}^{(j_n)} \leqslant z_n) \\
&= P\Bigg[\forall j \in \mathbb{Z} : Z_{h+n}^{(j)} \leqslant \frac{z_n}{a_0^{(j)}(x)}, \\
&\qquad Z_{h+n-1}^{(j)} \leqslant \min\left(\frac{z_n}{a_1^{(j)}(x)}, \frac{z_{n-1}}{a_0^{(j)}(x)}\right), \\
&\qquad Z_{h+n-2}^{(j)} \leqslant \min\left(\frac{z_n}{a_2^{(j)}(x)}, \frac{z_{n-1}}{a_1^{(j)}(x)}, \frac{z_{n-2}}{a_0^{(j)}(x)}\right), \\
&\qquad \ldots, \\
&\qquad Z_h^{(j)} \leqslant \min\left(\frac{z_n}{a_n^{(j)}(x)}, \frac{z_{n-1}}{a_{n-1}^{(j)}(x)}, \ldots, \frac{z_0}{a_0^{(j)}(x)}\right), \\
&\qquad Z_{h-1}^{(j)} \leqslant \min\left(\frac{z_n}{a_{n+1}^{(j)}(x)}, \frac{z_{n-1}}{a_n^{(j)}(x)}, \ldots, \frac{z_0}{a_1^{(j)}(x)}\right), \\
&\qquad \ldots\Bigg]
\end{aligned}$$

which allows us to dispose of the variable h by independence. □

Proof of Theorem 5.2.3 (page 125). Given $s, t \geqslant 0$ consider the Dirac masses in $\mathcal{C}([0,1]^q, \mathbb{R})$

$$\begin{aligned}
\delta(-s,t;i,j) &\stackrel{d}{=} \left(\frac{a_{-s+i}^{(j)}}{\|a_i^{(j)}\|_\infty}, \ldots, \frac{a_{t+i}^{(j)}}{\|a_i^{(j)}\|_\infty}\right) \\
&\stackrel{d}{=} \left[\left(\frac{a_{-s+i}^{(j)} Z}{\|a_i^{(j)} Z\|_\infty}, \ldots, \frac{a_{t+i}^{(j)} Z}{\|a_i^{(j)} Z\|_\infty}\right) \;\middle|\; \|a_i^{(j)}\|_\infty Z > x\right]
\end{aligned}$$

with $i \geqslant 0$ and $j \in \mathbb{Z}$.

Remark 4.4.4 (109) with $T_i^{(j)}(z) = a_i^{(j)} z$ leads to the new conclusion that

$$E\left|1\{\sup_{j\in\mathbb{Z}} \sup_{i\geqslant 0} \|a_i^{(j)}\| Z_{-i}^{(j)} > x\} - \sum_{\substack{j\in\mathbb{Z} \\ i\geqslant 0}} 1\{\|a_i^{(j)}\| Z > x)\}\right| = o\big(P(Z > x)\big) \tag{5.23}$$

as $x \to \infty$. Consequently, for $f \in \mathcal{C}_b(\mathcal{C}([0,1]^q, \mathbb{R}^{s+t+1}))$, by virtue of (5.23)

$$
\begin{aligned}
&E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) \,\middle|\, \|X_0\|_\infty > x \right] \\
&= \frac{E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|X_0\|_\infty > x\}\right]}{P(Z > x)} \frac{P(Z > x)}{P(\|X_0\|_\infty > x)} \\
&= \frac{E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\sup\limits_{j\in\mathbb{Z}} \sup\limits_{i\geqslant 0} \|a_i^{(j)}\|_\infty Z_{-i}^{(j)} > x\}\right]}{P(Z > x)} \frac{P(Z > x)}{P(\|X_0\|_\infty > x)} \\
&= \sum_{\substack{j\in\mathbb{Z}\\ i\geqslant 0}} \frac{E\left[f\left(\frac{X_{-s}}{x}, \ldots, \frac{X_t}{x}\right) 1\{\|a_i^{(j)}\|_\infty Z_{-i}^{(j)} > x\}\right]}{P(\|a_i^{(j)}\|_\infty Z_{-i}^{(j)} > x)} \\
&\qquad \cdot \frac{P(\|a_i^{(j)}\|_\infty Z_{-i} > x)}{P(Z > x)} \frac{P(Z > x)}{P(\|X_0\|_\infty > x)} + o(1).
\end{aligned}
$$

Thanks to the dominated convergence and the continuity of f, the last expression is equal to

$$
\begin{aligned}
&\sum_{\substack{j\in\mathbb{Z}\\ i\geqslant 0}} E\left[f\left(\frac{\sup\limits_{l\in\mathbb{Z}} \sup\limits_{k\geqslant 0} (a_k^{(l)} Z_{-s-k}^{(l)})}{x}, \ldots, \frac{\sup\limits_{l\in\mathbb{Z}} \sup\limits_{k\geqslant 0} (a_k^{(l)} Z_{t-k}^{(l)})}{x} \right) \,\middle|\, \|a_i^{(j)}\|_\infty Z_{-i}^{(j)} > x \right] \\
&\qquad \cdot \frac{P(\|a_i^{(j)}\|_\infty Z > x)}{P(Z > x)} \frac{P(Z > x)}{P(\|X_0\|_\infty > x)} + o(1) \\
&= \sum_{\substack{j\in\mathbb{Z}\\ i\geqslant 0}} E\left[f\left(\frac{a_{-s+i}^{(j)} Z}{x}, \ldots, \frac{a_{t+i}^{(j)} Z}{x}\right) \,\middle|\, \|a_i^{(j)}\|_\infty Z > x \right] \\
&\qquad \cdot \frac{P(\|a_i^{(j)}\|_\infty Z > x)}{P(Z > x)} \frac{P(Z > x)}{P(\|X_0\|_\infty > x)} + o(1) \\
&= \sum_{\substack{j\in\mathbb{Z}\\ i\geqslant 0}} E\left[f\left(\frac{Z}{x}(a_{-s+i}^{(j)}, \ldots, a_{t+i}^{(j)}\right) \,\middle|\, \|a_i^{(j)}\|_\infty Z > x \right] \\
&\qquad \cdot \frac{P(\|a_i^{(j)}\|_\infty Z > x)}{P(Z > x)} \frac{P(Z > x)}{P(\|X_0\|_\infty > x)} + o(1).
\end{aligned}
$$

Using the Continuous mapping theorem and the regular variation of Z, the last term converges to

$$\sum_{\substack{j\in\mathbb{Z}\\ i\geqslant 0}} E\left[f(Y\delta(-s,t;i,j))\right] \|a_i^{(j)}\|_\infty \frac{1}{\sum_{l\in\mathbb{Z}}\sum_{k\geqslant 0}\|a_k^{(l)}\|_\infty}$$

where $Y\sim$ Pareto(1). The factors

$$\pi_{ij} := \frac{\|a_i^{(j)}\|_\infty}{\sum_{l\in\mathbb{Z}}\sum_{k\geqslant 0}\|a_k^{(l)}\|_\infty}$$

form a probability distribution on $\mathbb{Z}_+\times\mathbb{Z}$, let us say of a random vector S.

According to Theorem 2.3.3, the spectral process of the CM3 process $(X_t)_{t\in Z}$ is

$$(\Theta_{-s},\dots,\Theta_t) \stackrel{d}{=} \delta(-s,t;S)$$

which has the announced distribution. □

Proof of Theorem 5.2.4 (page 126). It suffices to check that

$$\lim_{m\to\infty}\limsup_{n\to\infty} P(\exists t\in\{2m,\dots,r_n\}:\|X_t\|>n \mid \|X_0\|>n)=0.$$

For the convenience of the proof, set $p=t-i$ so that

$$X_t=\sup_{j\in\mathbb{Z}}\sup_{p\leqslant t} a_{t-p}^{(j)}Z_p^{(j)}.$$

Let $b_i^{(j)} := \|a_i^{(j)}\|_\infty$ to have

$$\|X_t\|=\sup_{j\in\mathbb{Z}}\sup_{p\leqslant t} b_{t-p}^{(j)}Z_p^{(j)}.$$

Given $m>0$, write $A_n := \{\|X_0\|>n\}$ so that

$$A_n=\{\exists j\in\mathbb{Z},\exists p\leqslant 0: b_{-p}^{(j)}Z_p^{(j)}>n\}.$$

We have, as $n\to\infty$,

$$\begin{aligned}
&nP(A_n)\\
&= n[1-P(\forall j\in\mathbb{Z},\forall p\leqslant 0: b_{-p}^{(j)}Z_p^{(j)}\leqslant n)]\\
&= n[1-\exp(-\tfrac{1}{n}\textstyle\sum_{j\in\mathbb{Z}}\sum_{p\leqslant 0} b_{-p}^{(j)})]\\
&\to \textstyle\sum_{j\in\mathbb{Z}}\sum_{p\leqslant 0} b_{-p}^{(j)} = \sum_{j\in\mathbb{Z}}\sum_{p\geqslant 0} b_p^{(j)} < +\infty.
\end{aligned}$$

For $m > 0$, decompose $\{\exists t \in \{2m, \ldots, r_n\} : \|X_t\| > n\} = C_{m,n} \dot\cup D_{m,n}$ where

$$C_{m,n} = \{\exists t \in \{2m, \ldots, r_n\}, \exists j \in \mathbb{Z}, \exists p > m : b_{t-p}^{(j)} Z_p^{(j)} > n\},$$
$$D_{m,n} = \{\exists t \in \{2m, \ldots, r_n\}, \exists j \in \mathbb{Z}, \exists p \leqslant m : b_{t-p}^{(j)} Z_p^{(j)} > n\}.$$

We have

$$\begin{aligned} P(C_{m,n}) &= 1 - P(\forall t \in \{2m, \ldots, r_n\}, \forall j \in \mathbb{Z}, \forall p > m : b_{t-p}^{(j)} Z_p^{(j)} \leqslant n) \\ &= 1 - P(\forall j \in \mathbb{Z}, \forall p > m : [\max_{2m \leqslant t \leqslant r_n} b_{t-p}^{(j)}] Z_p^{(j)} \leqslant n) \\ &= 1 - \exp(-\frac{1}{n} \sum_{j \in \mathbb{Z}} \sum_{p > m} \max_{2m \leqslant t \leqslant r_n} b_{t-p}^{(j)}) \\ &\leqslant \frac{1}{n} \sum_{j \in \mathbb{Z}} \sum_{p > m} \max_{2m \leqslant t \leqslant r_n} b_{t-p}^{(j)} \\ &\leqslant \frac{m_n}{n} \sum_{j \in \mathbb{Z}} \sum_{k \geqslant 0} b_k^{(j)} \to 0 \end{aligned}$$

as $n \to \infty$.

Setting $p = 2m - q$ write

$$A_n = \{\exists j \in \mathbb{Z}, \exists q \geqslant 2m : b_{q-2m}^{(j)} Z_q^{(j)} > n\}$$

and

$$D_{m,n} = \{\exists j \in \mathbb{Z}, \exists q \geqslant m : \max_{q \leqslant t \leqslant q-2m+r_n} b_t^{(j)} Z_q^{(j)} > n\}$$

hence

$$\begin{aligned} nP(A_n \cap D_{m,n}) &\leqslant n \sum_{j \in \mathbb{Z}} \sum_{q \geqslant 2m} P(\{b_{q-2m}^{(j)} Z_q^{(j)} > n\} \cap D_{m,n}) \\ &\leqslant n \sum_{j \in \mathbb{Z}} \sum_{q \geqslant 2m} P(\min[b_{q-2m}^{(j)}, \max_{t \geqslant q} b_t^{(j)}] Z_q^{(j)} > n) \\ &\quad + n \sum_{j \in \mathbb{Z}} \sum_{q \geqslant 2m} P(b_{q-2m}^{(j)} Z_q^{(j)} > n) P(D_{m,n}). \end{aligned}$$

The first term of the sum is bounded above by

$$\sum_{j \in \mathbb{Z}} \sum_{q \geqslant 0} \min[b_q^{(j)}, \max_{t \geqslant 2m} b_t^{(j)}]$$

and the second term tends to 0 as $n \to \infty$ since

$$n \sum_{j\in\mathbb{Z}} \sum_{q\geqslant 2m} P(b_{q-2m}^{(j)} Z_q^{(j)} > n) \leqslant \sum_{j\in\mathbb{Z}} \sum_{q\geqslant 0} b_q^{(j)} < \infty$$

and

$$P(D_{m,n}) = \frac{1}{n} \sum_{j\in\mathbb{Z}} \sum_{q\geqslant m} \max_{q\leqslant t\leqslant q-2m+r_n} b_t^{(j)} \leqslant \frac{m_n}{n} \sum_{j\in\mathbb{Z}} \sum_{q\geqslant 0} b_q^{(j)} \to 0.$$

Consequently

$$\limsup_{n\to\infty} nP(A_n \cap D_{m,n}) \leqslant \sum_{j\in\mathbb{Z}} \sum_{p\in\mathbb{Z}} \min[b_p^{(j)}, \max_{t\geqslant 2m} b_t^{(j)}].$$

Since A_n and $C_{m,n}$ are independent,

$$\begin{aligned}
&P(\exists t \in \{2m,\ldots,r_n\} : \|X_t\| > n \mid \|X_0\| > n)\\
&= P(C_{m,n} \cup D_{m,n} \mid A_n)\\
&= \frac{P([C_{m,n} \cup D_{m,n}] \cap A_{m,n})}{P(A_n)}\\
&\leqslant \frac{P([C_{m,n} \cap A_n] \cup [D_{m,n} \cap A_n]\})}{P(A_n)}\\
&\leqslant P(C_{m,n}) + \frac{P(D_{m,n} \cap A_n)}{P(A_n)}
\end{aligned}$$

so that

$$\begin{aligned}
&\limsup_{n\to\infty} P(\exists t \in \{2m,\ldots,r_n\} : \|X_t\| > n \mid \|X_0\| > n)\\
&\leqslant \frac{1}{\sum_{j\in\mathbb{Z}} \sum_{p\geqslant 0} b_p^{(j)}} \left[\sum_{j\in\mathbb{Z}} \sum_{p\in\mathbb{Z}} \min[b_p^{(j)}, \max_{t\geqslant 2m} b_t^{(j)}] \right]
\end{aligned}$$

which tends to 0 if $m \to \infty$. □

Proof of Theorem 5.2.5 (page 126). Again set $p = t - i$ to have

$$X_t = \sup_{j\in\mathbb{Z}} \sup_{p\leqslant t} a_{t-p}^{(j)} Z_p^{(j)}.$$

For large $t \geqslant 0$, we will prove that we can approach X_t by

$$X_t^{(+)} = \sup_{j \in \mathbb{Z}} \sup_{1 \leqslant p \leqslant t} a_{t-p}^{(j)} Z_p^{(j)}$$

in the sense of

$$\lim_{m \to +\infty} P(\forall t \geqslant m : X_t^{(+)} = X_t) = 1.$$

Then the conclusion will follow from the fact that the processes $(X_t^{(+)})_{t \geqslant 1}$ and $(X_t)_{t \leqslant -1}$ are independent.

As the Borel σ-field on $\mathcal{C}([0,1]^q)$ is generated by the finite dimensional sets, it is sufficient to check that the limit is 1 on every subset $\{x_1, \ldots, x_k\} \subset [0,1]^q$. Passing to the complementary, compute

$$\begin{aligned} p_m &:= P(\exists t \geqslant m, \exists 1 \leqslant i \leqslant k \, : \, X_t^{(+)}(x_i) < X_t(x_i)) \\ &= P(\exists t \geqslant m, \exists 1 \leqslant i \leqslant k \, : \, \sup_{j \in \mathbb{Z}} \sup_{1 \leqslant p \leqslant t} a_{t-p}^{(j)}(x_i) Z_p^{(j)} < \\ &\qquad\qquad\qquad \textstyle\sup_{j \in \mathbb{Z}} \sup_{p \leqslant t} a_{t-p}^{(j)}(x_i) Z_p^{(j)}) \\ &= P(\exists t \geqslant m, \exists 1 \leqslant i \leqslant k \, : \, \sup_{j \in \mathbb{Z}} \sup_{1 \leqslant p \leqslant t} a_{t-p}^{(j)}(x_i) Z_p^{(j)} < \\ &\qquad\qquad\qquad \textstyle\sup_{j \in \mathbb{Z}} \sup_{p \leqslant 0} a_{t-p}^{(j)}(x_i) Z_p^{(j)}). \end{aligned}$$

If $a, b > 0$ and Z_1, Z_2 are standard unit-Fréchet, as in (5.10),

$$P(aZ_1 < bZ_2) = \frac{1}{1 + \frac{a}{b}} < \frac{b}{a}$$

so that

$$p_m \leqslant \sum_{t \geqslant m} \sum_{i=1}^{k} \frac{\sum_{j \in \mathbb{Z}} \sum_{p \leqslant 0} a_{t-p}^{(j)}(x_i)}{\sum_{j \in \mathbb{Z}} \sum_{1 \leqslant p \leqslant t} a_{t-p}^{(j)}(x_i)} =: \sum_{t \geqslant m} \sum_{i=1}^{k} \frac{N_t(x_i)}{D_t(x_i)}.$$

According to (5.2), $D_t(x)$ uniformly converges to a continuous function $D(x)$ as $t \to \infty$. Without loss of generality, we can assume that $D(t) \equiv 1$. Write $D_t(x) = 1 - \varepsilon_t(x)$ with $\|\varepsilon_t\|_\infty \to 0$ as $t \to \infty$. Consequently,

$$\frac{1}{D_t(x)} = 1 + \frac{\varepsilon_t(x)}{D_t(x)}$$

hence

$$p_m \leqslant \sum_{t \geqslant m} \sum_{i=1}^{k} N_t(x_i) + \sum_{t \geqslant m} \sum_{i=1}^{k} N_t(x_i) \frac{\varepsilon_t(x_i)}{D_t(x_i)}.$$

Next, use the fact that for every $\eta > 0$, there exists m_η such that for every $t \geqslant m_\eta$ and every $1 \leqslant i \leqslant k$

$$0 \leqslant \frac{\varepsilon_t(x_i)}{D_t(x_i)} \leqslant \eta$$

to see that

$$\frac{\displaystyle\sum_{t \geqslant m} \sum_{i=1}^{k} N_t(x_i) \frac{\varepsilon_t(x_i)}{D_t(x_i)}}{\displaystyle\sum_{t \geqslant m} \sum_{i=1}^{k} N_t(x_i)} = o(1)$$

and so obtain

$$p_m \leqslant (1 + o(1)) \sum_{t \geqslant m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} \sum_{p \leqslant 0} a_{t-p}^{(j)}(x_i).$$

Since

$$\sum_{t \geqslant m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} \sum_{p \leqslant 0} a_{t-p}^{(j)}(x_i) = \sum_{k \geqslant m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} (k - m + 1) a_k^{(j)}(x_i),$$

thanks to (5.3) we have that p_m vanishes as $m \to \infty$, which proves the result. □

Proof of Proposition 5.2.6 (page 127). Since

$$\|X_t\|_\infty = \| \sup_{j \in \mathbb{Z}} \sup_{i \geqslant 0} a_i^{(j)} Z_{t-i}^{(j)} \|_\infty = \sup_{j \in \mathbb{Z}} \sup_{i \geqslant 0} \|a_i^{(j)}\|_\infty Z_{t-i}^{(j)}$$

it suffices to transcribe the formula for a M3 from Smith and Weissman (1996). □

Further work

Besides the linear processes in function spaces with heavy tailed innovations of Chapter 4, CM3 processes of Chapter 5 are a second example of jointly regularly varying time series in function spaces. Linear processes are aimed at modelling time series when the innovations are heavy tailed, whereas CM3 processes appear after a standardization of the margins into unit-Fréchet. While linear processes are models for the observations themselves, CM3 processes are models for the spatio-temporal dependence structure between extremes. Both tracks could be deepened.

About linear processes, the next step could be to derive the spectral process of the AR(1) model $X_t = K * X_{t-1} + \varepsilon_t$, ie.

$$X_t(x) = \int K(x-y)X_{t-1}(y)dy + \varepsilon_t(x),$$

given the knowledge of the convolution kernel K. This is to be discussed in an appropriate normed function space. The problem of fitting such a kernel $\hat{K}$ belongs to the category of deconvolution problems and is a topic of current research. The model above could be applied for instance to the trivariate spatio-temporal pollution dataset with information about the CO, NO and NO_2 levels in the state of California, which has been used for instance in Genton and Apanasovich (2010). Extremes of functional linear processes may yields important pieces of information for this phenomenon.

Further studies about CM3 processes could be to determine whether or not the approximation theorems for M3 processes (Deheuvels 1983) and for M4 processes (Smith and Weissman 1996) also hold for CM3, that is to say if max-stable processes in function spaces, after having removed the deterministic components, can be arbitrary closely approached by a CM3. From these papers also arises the question of a generalization of the multivariate extremal index. Since such an object becomes hard to figure out in function spaces, the right way to proceed might be to replace the extremal index by the spectral process. The expression of the spectral process of a CM3 process is known. If a CM3 models well the extremal behavior of the data, its spectral process should be close to the one of the original series. A measure of this distance can be introduced and the conditions under which the "true" spectral process can be approached in this way should be investigated.

Finally, a third project could be to derive a nonparametric estimator of the spectral process, whose importance has been described in Chapter 2. According to Chapter 5, CM3 processes provide a simple way to approach the spectral process in distribution. A way to skip the CM3 processes could be to use estimators for the spectral measure in metric spaces, for instance developed in Liu (2009). As we have seen in Chapter 2, the existence of a spectral process (joint regular variation) is equivalent to the existence of a sequence of exponent measures or a sequence of spectral measures. Estimating these spectral measures then leads to an estimation of the spectral process.

List of Symbols

Number sets

$\mathbb{N}$	naturals numbers $\{0, 1, 2, \ldots\}$
$\mathbb{Z}$	integers
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	nonnegative real numbers

Operations on sets

A^c	complement of A
$\cap$	intersection
$\cup$	union
$\setminus$	difference
$\#$	cardinal

The set $\mathbb{R}^n$

$\lvert \cdot \rvert_2$	Euclidean norm on $\mathbb{R}^n$
$\lvert \cdot \rvert_\infty$	infinity norm on $\mathbb{R}^n$

Sets of continuous functions

$\mathcal{C}$	$\mathcal{C}([0,1]^d, \mathbb{R}^q)$ endowed with $\lvert \cdot \rvert_\infty$ on $\mathbb{R}^q$
$\lVert \cdot \rVert_\infty$	sup-norm on $\mathcal{C}$

Topology

$\cong$	homeomorphism
$\operatorname{int} A$	interior of A
$\operatorname{cl} A$	closure of A
∂A	border of A

Banach spaces

$\mathbb{B}$	a Banach space
$E \subset \mathbb{B}$	a subset of $\mathbb{B}$
$K \subset \mathbb{B}$	a compact subset of $\mathbb{B}$

Metric spaces

$B(x, r)$	open ball centered at x with radius r
$B[x, r]$	closed ball centered at x with radius r
$\mathbb{S}$	unit sphere

Functions

dom	domain
Im	image
spt	support
graph	graph
$\mathcal{C}(X)$	continuous functions from X to $\mathbb{R}$
$\mathcal{C}_b(X)$	$\{f \in \mathcal{C}(X) \mid f$ is bounded$\}$
$\mathcal{C}_{bd}(X)$	$\{f \in \mathcal{C}(X) \mid f$ is bounded and $(\exists r > 0)(\forall x \in B(x_0, r)) : f(x) = 0\}$
$\mathcal{C}^+_{bd}(X)$	$\{f \in \mathcal{C}_{bd}(X) \mid f \geqslant 0\}$

Probabilities

$\mathcal{B}_E$	Borelian σ-field of the set E
$(\Omega, \mathcal{A}, P)$	probability space
P	probability of the right probability space

Random elements

X	random element
ν_X	distribution measure of X ($P(X \in \cdot\,)$)
$\mathcal{L}(X)$	ν_X
$\mathcal{L}(X\|A)$	conditional distribution ($\nu_X(\,\cdot \cap A)/\nu_X(A)$)
E	expected value
$E[X;A]$	$E[X1_A]$
$E[X\|A]$	conditional expectation ($E[X;A]/P(A)$)
$\Psi_X(f)$	Laplace transform of X (Section 3.3)
$\varphi_X(t)$	moment-generating function of X (Section 3.4)

Stochastic processes

(X_t)	stochastic process
ε_t or Z_t	innovations in a stochastic process
iid	independent and identically distributed

Sets of measures

$M_b(X)$	set of all measure μ on $\mathcal{B}_X$ such that $\mu(X) < +\infty$
$M_+(X)$	set of all nonnegative Radon measure on X
$M_0(X, x_0)$	set of all measure μ on $\mathcal{B}_{X_0}$ such that $\mu(X_0 \setminus B(x_0, r)) < +\infty$ for every $r > 0$
$M_p(X)$	set X in M_0

Operations on measures

$h_\#\mu$	pushforward of μ by h (i.e. $h_\#\mu(A) = \mu(h^{-1}(A))$)
$\mu \llcorner A$	restriction of μ to A (i.e. $\mu \llcorner A(B) = \mu(A \cap B)$)
$P(X \in dx)$	$d\nu_X(x)$

Convergence of measures

$\xrightarrow{d}$	convergence in distribution
$\xrightarrow{w}$	weak convergence
$\xrightarrow{v}$	vague convergence
$\xrightarrow{M_0}$	M_0-convergence (Section 2.5.1)

Extremal features

(Y_t)	tail process (Theorem 2.3.2)
(Θ_t)	spectral process (Theorem 2.3.3)
(N_n)	point process of extremes (Section 3.2)
(ζ_n)	cluster process (Section 3.2)
λ	spectral measure (Proposition 2.2.3)
θ	extremal index (Section 2.4)

List of Figures

List of Tables

Bibliography

Abdelazziz Allam and Tahar Mourid. Geometric absolute regularity of Banach space-valued autoregressive processes. *Statist. Probab. Lett.*, 60(3):241–252, 2002. ISSN 0167-7152.

Anestis Antoniadisa and Theofanis Sapatinas. Wavelet methods for continuous-time prediction using hilbert-valued autoregressive processes. 87:133—-158, 2003.

Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic Process. Appl.*, doi:10.1016/j.spa.2008.05.004, 2008.

Bojan Basrak, Richard A. Davis, and Thomas Mikosch. Regular variation of GARCH processes. *Stochastic Process. Appl.*, 99(1):95–115, 2002. ISSN 0304-4149.

Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. ISBN 0-471-00710-2. A Wiley-Interscience Publication.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.

Denis Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. ISBN 0-387-95052-4. Theory and applications.

Denis Bosq. Limit theorems for Banach-valued autoregressive processes. Applications to real continuous time processes. *Bull. Belg. Math. Soc. Simon Stevin*, 3(5):537–555, 1996. ISSN 1370-1444.

Daniel Cooley, Douglas Nychka, and Philippe Naveau. Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.*, 102(479):824–840, 2007. ISSN 0162-1459.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer Series in Statistics. Springer-Verlag, New York, 1988. ISBN 0-387-96666-8.

Richard A. Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.*, 23(2):879–917, 1995. ISSN 0091-1798.

Richard A. Davis and Thomas Mikosch. Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Process. Appl.*, 118(4):560–584, 2008. ISSN 0304-4149.

Richard A. Davis and Thomas Mikosch. The extremogram: a correlogram for extreme events. *Bernoulli*, 15(4):977–1009, 2009. ISSN 1350-7265. doi: 10.3150/09-BEJ213. URL `http://dx.doi.org/10.3150/09-BEJ213`.

Richard A. Davis and Thomas Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.*, 26 (5):2049–2080, 1998. ISSN 0090-5364.

Laurens de Haan and Ana Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. ISBN 978-0-387-23946-0; 0-387-23946-4. An introduction.

Laurens de Haan and Tao Lin. On convergence toward an extreme value distribution in $C[0,1]$. *Ann. Probab.*, 29(1):467–483, 2001. ISSN 0091-1798.

Paul Deheuvels. Point processes and multivariate extreme values. *J. Multivariate Anal.*, 13(2):257–272, 1983. ISSN 0047-259X. doi: 10.1016/0047-259X(83)90025-8. URL `http://dx.doi.org/10.1016/0047-259X(83)90025-8`.

Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. ISBN 3-540-60931-8. For insurance and finance.

William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

R.A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.

Marc Genton and Tatiyana Apanasovich. Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, 97 (1):15–30, 2010.

Evarist Giné, Marjorie G. Hahn, and Pirooz Vatan. Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Related Fields*, 87(2):139–165, 1990. ISSN 0178-8051. doi: 10.1007/BF01198427. URL `http://dx.doi.org/10.1007/BF01198427`.

B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453, 1943. ISSN 0003-486X.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning (2nd edition)*. Springer, New York, 2009. ISBN 0-387-84857-6.

Henrik Hult and Filip Lindskog. Extremal behavior of regularly varying stochastic processes. *Stochastic Process. Appl.*, 115(2):249–274, 2005. ISSN 0304-4149.

Henrik Hult and Filip Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006. ISSN 0350-1302.

Olav Kallenberg. *Random measures.* Akademie-Verlag, Berlin, fourth edition, 1986. ISBN 0-12-394960-2.

Leonard Kaufman and Peter J. Rousseeuw. *Finding groups in data.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1990. ISBN 0-471-87876-6. An introduction to cluster analysis, A Wiley-Interscience Publication.

Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131:207–248, 1973. ISSN 0001-5962.

David J. Jr Ketchen and Christopher L. Shoo. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management J.*, 17(6):441–458, 1996. doi: 10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.

M. R. Leadbetter. Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete*, 65(2):291–306, 1983. ISSN 0044-3719.

Shuyan Liu. *Lois stables et procesessus ponctuels: liens et estimations des paramètres.* PhD thesis, Université des Sciences et Technologies de Lille, France, 2009.

R. Jr Lleti, M.C. Ortiz, Sarabia L.A., and M.S Sánchez. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimic Acta*, 515:87–100, 2004. doi: doi:10.1016/j.aca.2003.12.020.

Kantilal Varichand Mardia, John T. Kent, and John M. Bibby. *Multivariate analysis.* Academic Press [Harcourt Brace Jovanovich Publishers], London, 1979. ISBN 0-12-471250-9. Probability and Mathematical Statistics: A Series of Monographs and Textbooks.

André Mas. Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Process. Appl.*, 99(1):117–135, 2002. ISSN 0304-4149.

André Mas. Weak convergence in the functional autoregressive model. *J. Multivariate Anal.*, 98(6):1231–1261, 2007. ISSN 0047-259X.

André Mas and Ludovic Menneteau. Large and moderate deviations for infinite-dimensional autoregressive processes. *J. Multivariate Anal.*, 87(2):241–260, 2003. ISSN 0047-259X.

André Mas and Besnik Pumo. The ARHD model. *J. Statist. Plann. Inference*, 137(2):538–553, 2007. ISSN 0378-3758.

Thomas Meinguet. Maxima of moving maxima of continuous functions. *Submitted, arXiv:1003.4130*, 2010.

Thomas Meinguet and Johan Segers. Regularly varying time series in Banach spaces. *Submitted, arXiv:1001.3262*, 2010.

Thomas Mikosch and Gennady Samorodnitsky. The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Appl. Probab.*, 10(3):1025–1064, 2000. ISSN 1050-5164.

David Pollard. *A user's guide to measure theoretic probability*, volume 8 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-80242-3; 0-521-00289-3.

Besnick Pumo. Prediction of continuous time processes by $c[0,1]$-valued autoregressive process. 1:297–309, 1998.

Sidney I. Resnick. *Heavy-tail phenomena.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. ISBN 978-0-387-24272-9; 0-387-24272-4. Probabilistic and statistical modeling.

Sidney I. Resnick. *Extreme values, regular variation and point processes.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. ISBN 978-0-387-75952-4. Reprint of the 1987 original.

Peter J. Rousseuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Comp. and Appl. Math.*, 20:53–65, 1987. doi: doi:10.1016/0377-0427(87)90125-7.

G. A. F. Seber. *Multivariate observations.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1984. ISBN 0-471-88104-X.

Johan Segers. Approximate distributions of clusters of extremes. *Statist. Probab. Lett.*, 74(4):330–336, 2005. ISSN 0167-7152.

Johan Segers. Rare events, temporal dependence, and the extremal index. *J. Appl. Probab.*, 43(2):463–485, 2006. ISSN 0021-9002. doi: 10.1239/jap/1152413735. URL `http://dx.doi.org/10.1239/jap/1152413735`.

Richard L. Smith. Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of, Extreme Values in Finance, Telecommunications and the Environment, edited by B. Finkenstadt and H. Rootzen, Chapman and Hall/CRC Press, London. pages 1–78, 2003.

Richard L. Smith. The extremal index for a Markov chain. *J. Appl. Probab.*, 29(1):37–45, 1992. ISSN 0021-9002.

Richard L. Smith and Ishay Weissman. Characterization and estimation of the multivariate extremal index. *Dept. Stat. Oper. Res., Univ. North Carolina, Chapel Hill, NC*, 1996.

H. Spath. *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples, translated by J. Goldschmidt.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Halsted Press, 1985.

M. Süveges. *Statistical Analysis of Clusters of Extreme Events.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2009.

M. Süveges and A. C. Davison. A dirichlet mixture approach for clusters of extreme events. *Submitted*, 2010.

Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Third Edition.* Academic Press, Inc., Orlando, FL, USA, 2006. ISBN 0123695317.

Mark J. van der Laan, Katherine S. Pollard, and Jennifer Bryan. A new partitioning around medoids algorithm. *J. Stat. Comput. Simul.*, 73(8):575–584, 2003. ISSN 0094-9655. doi: 10.1080/0094965031000136012. URL `http://dx.doi.org/10.1080/0094965031000136012`.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-49603-9; 0-521-78450-6.

Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. With applications to statistics.

Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.*, 58:236–244, 1963. ISSN 0162-1459.

Zhengjun Zhang. The estimation of M4 processes with geometric moving patterns. *Ann. Inst. Statist. Math.*, 60(1):121–150, 2008. ISSN 0020-3157. doi: 10.1007/s10463-006-0078-0. URL `http://dx.doi.org/10.1007/s10463-006-0078-0`.

Zhengjun Zhang and Richard L. Smith. The behavior of multivariate maxima of moving maxima processes. *J. Appl. Probab.*, 41(4):1113–1123, 2004. ISSN 0021-9002.