WHEN IS A TRUNCATED HEAVY TAIL HEAVY?

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Arijit Chakrabarty

May 2010

WHEN IS A TRUNCATED HEAVY TAIL HEAVY?

Arijit Chakrabarty, Ph.D.

Cornell University 2010

This dissertation addresses the important question of the extent to which random variables and vectors with truncated power tails retain the characteristic features of random variables and vectors with power tails.

We define two truncation regimes - soft truncation and hard truncation, based on the growth rate of the truncating threshold. We study the central limit theorem and the large deviations behavior of the model with truncated power laws in both regimes. The central limit theorem is studied for random vectors taking values in a separable Banach space, while for the large deviations, the random vectors are assumed to be $\mathbb{R}^d$-valued. It turns out that, in the soft truncation regime, truncated power tails behave, in important respects, as if no truncation took place. On the other hand, in the hard truncation regime much of "heavy tailedness" is lost. Based on this observation, we set before ourselves two tasks. The first one is to suggest statistical tests to decide on whether the truncation is soft or hard. The second task is to devise an estimator for the tail exponent from the truncated data, which is consistent regardless of the truncation regime. Finally, we apply our methods to two recent data sets arising from computer networks.

BIOGRAPHICAL SKETCH

Arijit Chakrabarty was born on March 14, 1982 in Calcutta (now Kolkata). After completing high school from Ramakrishna Mission Residential College Narendrapur in 2000, he joined the B. Stat (Bachelor of Statistics) program of the Indian Statistical Institute, Kolkata. He finished the program in 2003, after which he joined the M. Stat (Master of Statistics) program in the same institute. In 2004, he moved to the Delhi center of the institute for the final year of the M Stat training. In 2005, he joined the School of Operations Research and Industrial Engineering (now known as the School of Operations Research and Information Engineering) as a Ph. D. student, with concentration in Applied Probability and Statistics.

Upon completion of his Ph. D., he will join the Department of Mathematics at the Indian Institute of Science, Bengaluru, as a postdoctoral fellow.

to G.C.

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to my adviser, Professor Gennady Samorodnitsky, for his support throughout the duration of the Ph. D. program. Apart from guiding me on my thesis, he has helped me grow in almost every aspect of professional life. His inputs proved invaluable in improving myself as a teacher, a speaker and a researcher. To put it in a sentence, he taught me what it means to be an academician.

I am thankful to Professor Sidney Resnick for the innumerable discussions regarding my research. He is one of those professors who make you feel that there is no distance between a teacher and a student. Talking to him has always been a pleasure.

I thank Professor Richard Durrett and Professor Adrian Lewis for agreeing to serve on my committee. I am grateful to Professor Thomas Mikosch, Professor Michael Todd, Professor Xin Guo, Professor S.R.Srinivasa Varadhan, Professor James Kuelbs and Professor Sourav Chatterjee for being generous to me with their advice. I specially thank Professor Samorodnitsky, Professor Resnick, Professor Mikosch and Professor Durrett for writing letters for me during my job applications.

I am grateful to every member of the School of Operations Research and Information Engineering at Cornell University for making this department a lively place. I thank my teachers at Indian Statistical Institute and Cornell University. Special thanks are due to Professor B. V. Rao for introducing me to the wonderful world of probability. I acknowledge the hospitality I enjoyed at the Indian Statistical Institute, Kolkata and the University of Copenhagen during my visits in Fall 2008.

I am thankful to my friends who made my stay in Ithaca very enjoyable. A partial list (in a random order) is Udayan, Parthanil, Souvik, Romita, Anandarup, Bikramjit, Amrita, Trina, Shirshendu, Abhimanyu, Neelanjan, Ria, Joyjit. I reiterate that this list is by no means complete; I wish there were enough space for a complete list.

The contributions of my family in my life, particularly those of Baba (my dad), Maa (my mom), Sahana (my wife) and Bonti (my sister), go far beyond what can possibly be expressed in words; I choose not to attempt that.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Models with truncated power tails

Power tails are characteristic of models in which probability of one single large value, that can impact the whole system, is relatively big. Probability laws with power tails are ubiquitous in applications. A good fit between empirical distribution of various quantities of interest and distributions with power tails has been reported in such diverse areas as human travel (Brockmann et al. (2006)), earthquake analysis (Corral (2006)), animal science (Bartumeus et al. (2005)) and even in language (Serrano et al. (2009)). The modelling and analysis of such phenomena differ a lot from their classical counterparts, where the tail decays much faster than a power tail, for example exponentially fast. The behavior of models with power tails is governed by the large values that shock the system every now and then, as opposed to the systems which exhibit some stability in the sense that their behavior is determined largely by an averaging effect. In the literature, probability laws with power tails are often referred to as "heavy-tailed distributions"; in this dissertation also both the terms will be used interchangeably.

In many situations there is a "physical" limit that prevents a quantity of interest from taking an arbitrarily large value. The File Allocation Table (FAT) used on most computer systems allows the largest file size to be 4GB (minus one byte) (Microsoft Knowledge Base Article 154997 (2007)); the greatest loss an insurance company is exposed to by an single covered event is limited by its reinsurance contract (see e.g. Mikosch (2009)). Even the number of the atoms in the universe is widely considered to be finite. It is common in practice to combine these two facts together and use a model that features power tails only in a truncated form; such models are often referred to as *truncated Lévy flights*, see e.g. Scholtz and Contreras (1998), Maruyama and Murakami (2003) or Zaninetti and Ferraro (2008). At the first glance this leads to a situation where the power tails, in a sense, completely disappear. The truncation may change dramatically the behavior of the cumulative sums of observations and it always changes dramatically the behavior of the cumulative maxima of the observations. Yet it is precisely such patterns of behavior for which a model with power tails is chosen in the first place. This leads one to ask the natural question: **to what extent, if any, do phenomena well described by models with truncated power tails retain the characteristic features of power tails**?

Answering this question is not straightforward. We start by pointing out that the level of truncation is linked to the amount of observations one has at hand. This can be thought of in different ways. First of all, finiteness of the sample is sometimes taken as the source of the truncation, see e.g. Burrooughs and Tebbens (2001) or Barthelemy et al. (2008). Secondly, both the physical nature of the truncation bound and the available data can be linked to a technological level. This is particularly transparent when one models a phenomenon related to computer or communications systems; see e.g. Jelenković (1999) or Gomez et al. (2000). We describe this situation as a sequence of models, each one with truncated power tails or, in other words, as a triangular array system, which we now proceed to define formally.

There are situations where a realistic model demands that one go beyond finite-dimensional spaces. For example, to model any quantity that changes with time, one needs to consider a suitable function space. Hence, for the underlying space, it makes sense to use some nice infinite dimensional space. For this dissertation, we have chosen separable Banach spaces for that purpose. Formally, a random variable H that takes values in a separable Banach space B is said to be heavy-tailed or to have a power tail if there is a non-null measure μ on $B \setminus \{0\}$ so that for every $\varepsilon > 0$, $\mu(B \setminus B_\varepsilon) < \infty$ and there is a sequence a_n going to infinity so that

$$nP(a_n^{-1} H \in \cdot) \xrightarrow{w} \mu(\cdot) \tag{1.1}$$

on $B \setminus B_\varepsilon$ for all $\varepsilon > 0$. Here, for all $r \geq 0$, B_r denotes the closed ball of radius r centered at the origin:

$$B_r := \{x \in B : \|x\| \leq r\}\,.$$

If $B = \mathbb{R}^d$, then the above is equivalent to

$$nP(a_n^{-1} H \in \cdot) \xrightarrow{v} \mu(\cdot) \tag{1.2}$$

in $\overline{\mathbb{R}^d} \setminus \{0\}$. Here $\overline{\mathbb{R}^d}$ is the compactification of $\mathbb{R}^d$ obtained by adding to the latter a ball of infinite radius centered at origin. The measure μ necessarily satisfies $\mu(\overline{\mathbb{R}^d} \setminus \mathbb{R}^d) = 0$. It also has a scaling property: there exists $\alpha > 0$ such that for any Borel set $A \subset B$ and $c > 0$,

$$\mu(cA) = c^{-\alpha}\mu(A)\,.$$

The assumption (1.1) implies that for $x > 0$,

$$P(\|H\| > x) = x^{-\alpha} l(x)\,,$$

where l is a slowly varying function at infinity. This justifies the usage of the term "power tail". Denote $\mathcal{S} := \{x \in B : \|x\| = 1\}$ and let $\mathcal{B}(\mathcal{S})$ denote the Borel σ-field on $\mathcal{S}$. Define a probability measure σ on $(\mathcal{S}, \mathcal{B}(\mathrm{S}))$ by

$$\sigma(A) := \frac{1}{\mu(B_1^c)} \mu\left(\left\{x \in B : \frac{x}{\|x\|} \in A, \|x\| \geq 1\right\}\right)\,.$$

It's easy to see that (1.1) implies

$$P\left(\frac{H}{\|H\|}\in\cdot\middle|\|H\|>t\right)\xrightarrow{w}\sigma(\cdot) \tag{1.3}$$

in $\mathcal{B}(\mathrm{S})$ as $t\longrightarrow\infty$.

For $n=1,2,\ldots$ (regarded both as the number of observations in the nth row of the triangular array and the number of the model) let $M_n>0$ denote the truncation level. The nth row of the triangular array will consist of observations $X_{nj},\ j=1,\ldots,n$, which we view as generated according to the following mechanism:

$$X_{nj}:=H_j\mathbf{1}\left(\|H_j\|\le M_n\right)+\frac{H_j}{\|H_j\|}(M_n+L_j)\mathbf{1}\left(\|H_j\|>M_n\right), \tag{1.4}$$

$j=1,\ldots,n,\ n=1,2,\ldots$. Here $H_1,H_2,\ldots$ are i.i.d. copies of H that satisfies (1.1), and $(L,L_1,L_2,\ldots)$ is a sequence of i.i.d. nonnegative random variables independent of $(H,H_1,H_2,\ldots)$. For each $n=1,2,\ldots$ we view the observation X_{nj}, $j=1,\ldots,n$ as having power tails that are truncated at level M_n.

We need to comment, at this point, on the role of the random variables $L_1,L_2,\ldots$. One should view them as possessing light tails, even exponentially decaying tails. In many cases taking these random variables to be equal to zero with probability 1 is appropriate; in other applications exponentially fast tapering off of the tails beyond the truncation point has been observed (see e.g. Hong et al. (2008)). The results of this dissertation hold whenever the tails of the random variables $L_1,L_2,\ldots$ are only light enough, not necessarily exponentially light. We have chosen to formulate our results in this way in order to increase their generality, even though we are thinking of their role in the model (1.4) as representing the exponentially fast decaying tails.

Our approach to addressing the question "to what extent do models with truncated power tails retain the characteristic features of power tails?" lies in studying the effect of the rate of growth of the truncation level M_n on the asymptotic properties of the triangular array defined in (1.4). Specifically, we introduce the following definition. We will say that the tails in the model (1.4) are

$$\begin{aligned} &\text{truncated softly} \quad \text{if } \lim_{n\to\infty} nP\left(\|H\| > M_n\right) = 0\,, \\ &\text{truncated hard} \quad \text{if } \lim_{n\to\infty} nP\left(\|H\| > M_n\right) = \infty\,. \end{aligned} \tag{1.5}$$

Clearly, an intermediate regime exists as well. In this dissertation, we shall study the behavior of the partial sums of the triangular array (1.4) in both the regimes defined above. We shall see that there are reasons to say that when the tails of the power law model are truncated softly, much of the power law behavior is preserved, while when they are truncated hard, the same is lost.

We finish this section by pointing out that some of the issues related to models with truncated power tails have been addressed in the literature, but from different angles. The paper Asmussen and Pihlsgard (2005) discusses an application of distributions with truncated power tails in queuing, and addresses the question whether light tailed approximations or heavy approximations work better in this situation. On the other hand, a maximum likelihood estimation procedure of the tail exponent α in a parametric model of truncated power tails (specifically, the truncated Pareto distribution) is given in Aban et al. (2006). Finally, estimation of the tail exponent in randomly censored power models (where the tails are not so much truncated, as contaminated) is discussed in Beirlant et al. (2007) and Einmahl et al. (2008).

1.2 The Central Limit Theorem

Suppose that $X, X_1, X_2, \ldots$ are i.i.d. random variables taking values in a separable Banach space B. X is said to satisfy a central limit theorem if there is non-degenerate B-valued random variable $\mathcal{V}$ and sequences a_n and b_n so that

$$b_n^{-1} \sum_{j=1}^{n} X_j - a_n \Longrightarrow \mathcal{V} \,, \tag{1.6}$$

and in that case, X is said to be in the domain of attraction of $\mathcal{V}$ or the law of $\mathcal{V}$. It has been shown in Araujo and Giné (1980) that if a random variable $\mathcal{V}$ has a domain of attraction, then its characteristic function must necessarily be of one of the two forms described below. Either there is a finite measure Γ on $\mathcal{S} := \{x \in B : \|x\| = 1\}$, $x_0 \in B$ and $0 < \alpha < 2$ so that for all $f \in B'$ which is the dual space of B,

$$E \exp(if(\mathcal{V})) =$$

$$\exp\left[if(x_0) + \int_{\mathcal{S}} \int_0^\infty \left\{ e^{irf(s)} - 1 - irf(s)\mathbf{1}(0 < r \le 1) \right\} r^{-\alpha-1} dr \Gamma(ds) \right] \,, \tag{1.7}$$

or there is $x_0 \in B$ and a function $\Phi : B' \times B' \longrightarrow \mathbb{R}$ satisfying for all $f, g \in B'$,

1. $\Phi(f, f) \ge 0$
2. $\Phi(f, g) = \Phi(g, f)$
3. $\Phi(f, \cdot)$ is linear,

so that for all $f \in B'$,

$$E \exp(if(\mathcal{V})) = \exp\left\{ if(x_0) + \frac{1}{2}\Phi(f, f) \right\} \,. \tag{1.8}$$

If (1.7) holds, then $\mathcal{V}$ is said to be an α-stable random variable with location x_0 and spectral measure Γ, while if (1.8) holds, then $\mathcal{V}$ is said to be a Gaussian random variable with mean x_0 and covariance Φ. It is easy to see that in the latter case, for all $f, g \in B'$,

$$\mathrm{Cov}(f(\mathcal{V}), g(\mathcal{V})) = \Phi(f, g) \,,$$

and hence the name covariance for Φ. Theorem 6.5 in Araujo and Giné (1980) states that if $\mathcal{V}$ is a B-values Gaussian random variable, then there exists $t_0 > 0$ so that for every $t < t_0$,

$$E \exp\left(t\|\mathcal{V}\|^2\right) < \infty\,.$$

On the other hand, if $\mathcal{V}$ is an α-stable random variable with $0 < \alpha < 2$, then there exists $C \in (0, \infty)$ so that

$$P(\|\mathcal{V}\| > x) \sim Cx^{-\alpha} \text{ as } x \longrightarrow \infty.$$

If $B = \mathbb{R}^d$, then nice characterizations of the domains of attraction for stable and Gaussian laws are known. When $0 < \alpha < 2$, it has been shown in Rvačeva (1962) that H is in the domain of attraction of a $\mathbb{R}^d$-valued α-stable random variable $\mathcal{V}$ with location zero and spectral measure Γ if and only if

$$nP(c_n^{-1}H \in \cdot) \xrightarrow{v} \mu \tag{1.9}$$

on $\overline{\mathbb{R}^d}\backslash\{0\}$ for some sequence c_n going to infinity, where the measure μ is defined by

$$\mu(A) = \int_{\mathcal{S}} \int_0^\infty \mathbf{1}(rs \in A) r^{-\alpha-1} dr \Gamma(ds)\,.$$

In that case (1.6) holds with

$$b_n = \sup\{t : nP(\|H\| > t) \geq \alpha^{-1}\Gamma(\mathcal{S})\}\,,$$

and

$$a_n := \begin{cases} 0, & \alpha < 1 \\ nb_n^{-1} \int_{\{\|x\| \leq b_n\}} xP(H \in dx), & \alpha = 1 \\ nb_n^{-1} E(H), & 1 < \alpha < 2\,. \end{cases}\,.$$

In the same paper, it has been shown that a $\mathbb{R}^d$-valued random variable H is in the domain of attraction of a Gaussian random variable $\mathcal{V}$ with mean zero and covariance matrix Φ if and only if

$$\lim_{R\to\infty} \frac{R^2 P(\|H\| > R)}{\int_{\{\|x\|\le R\}} \|x\|^2 P(H \in dx)} = 0\,, \tag{1.10}$$

where $\|\cdot\|$ denotes the L^2 norm, and

$$\lim_{R\to\infty} \frac{\int_{\{\|x\|\le R\}} \langle t, x\rangle^2 P(H \in dx)}{\int_{\{\|x\|\le R\}} \langle u, x\rangle^2 P(H \in dx)} = \frac{\langle t, \Phi t\rangle}{\langle u, \Phi u\rangle}$$

for all $t, u \in \mathbb{R}^d$. In this case, (1.6) holds with

$$b_n = \sup\left\{t : nt^{-2} \int_{\{\|x\|\le t\}} \|x\|^2 P(H \in dx) \ge E\|\mathcal{V}\|^2\right\}\,,$$

where $\|\cdot\|$ denotes the L^2-norm, and

$$a_n = nb_n^{-1} E(H)\,.$$

Unfortunately, nice characterizations of the domains of attraction are not known when B is a general separable Banach space. However, when $0 < \alpha < 2$, it is known that (1.1) which is the equivalent condition of (1.9) on Banach spaces, is a necessary condition for H to be in the domain of attraction of an α-stable random variable; see Theorem 6.18 in Araujo and Giné (1980). Corollary 6.21 in the same book states that if H is a B-valued random variable in the domain of attraction of a Gaussian law, then (1.10) holds.

In this dissertation, we shall study the effect of truncation on the CLT behavior of the partial sums of the triangular array (1.4). Specifically, we answer the question when is the limiting law after suitable centering and scaling α-stable with $0 < \alpha < 2$ and when is the same Gaussian.

1.3 Large Deviations

Suppose $X, X_1, X_2, \ldots$ are i.i.d. $\mathbb{R}^d$-valued random variables and λ_n is a sequence of numbers going to infinity so that as $n \longrightarrow \infty$,

$$\lambda_n^{-1} \sum_{j=1}^{n} X_j \xrightarrow{P} 0 .$$

Studying the large deviations behavior of X means analyzing the decay rate of $P(\lambda_n^{-1} \sum_{j=1}^{n} X_j \in A)$ for a set $A \subset \mathbb{R}^d$ that is bounded away from zero. This analysis, however, is very different for the following two situations - one when X has a power tail and the other when X has some finite exponential moment. Study of the former situation in one dimension dates back to Heyde (1968), Nagaev (1969a), Nagaev (1969b), Nagaev (1979) and Cline and Hsing (1991), among others; see Section 8.6 in Embrechts et al. (1997) and Mikosch and Nagaev (1998) for a survey on this. More recently, the functional version of large deviation principles for heavy-tailed $\mathbb{R}^d$ valued random variables has been studied by Hult et al. (2005). There, it is shown among other things, that if $H_1, H_2, \ldots$ are i.i.d. copies of H that satisfies (1.2), then

$$\frac{P\left(\lambda_n^{-1} \sum_{j=1}^{n} H_j \in \cdot\right)}{nP(\|H\| > \lambda_n)} \xrightarrow{v} \frac{\mu(\cdot)}{\mu(B_1^c)} , \tag{1.11}$$

where λ_n is a sequence satisfying $\lambda_n^{-1} \sum_{j=1}^{n} H_j \longrightarrow 0$ in probability and in addition

$$\begin{aligned} \lambda_n &\gg \sqrt{n^{1+\gamma}} \text{ for some } \gamma > 0, \text{ if } \alpha = 2 \\ \lambda_n &\gg \sqrt{n \log n}, \text{ if } \alpha > 2 , \end{aligned}$$

and B_r denotes the closed ball of radius r centered at the origin. The idea leading to the proof of (1.11) is that for $\sum_{j=1}^{n} H_j$ to be large, it is "necessary and sufficient" that exactly one of the summands is large.

For the situation where X has some finite exponential moment, Cramér's Theorem states that for $A \subset \mathbb{R}^d$,

$$- \inf_{x \in int(A)} \Lambda^*(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log P\left(n^{-1} \sum_{j=1}^{n} X_j \right)$$
$$\leq \limsup_{n\to\infty} \frac{1}{n} \log P\left(n^{-1} \sum_{j=1}^{n} X_j \right) \leq - \inf_{x \in cl(A)} \Lambda^*(x)\,, \tag{1.12}$$

where Λ^* is the Fenchel-Legendre dual of Λ, defined by

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x\rangle - \Lambda(\lambda)\},\ x \in \mathbb{R}^d\,, \tag{1.13}$$

and

$$\Lambda(\lambda) := \log E \exp(\langle \lambda, X\rangle)\,.$$

Though (1.12) is valid even without the assumption that $\Lambda(\lambda) < \infty$ for all λ in an open ball around zero, it is useful only when this is true. A detailed treatment of the theory of large deviations for the situation where finite exponential moments exist can be found in Varadhan (1984) and Dembo and Zeitouni (1998), among others.

In this dissertation, we study the large deviation behavior for the row sums of the triangular array (1.4), separately for the hard truncation and soft truncation regimes, as defined in (1.5). We show that in the soft truncation regime, the large deviation behavior is similar to (1.11) in important respects, while in the hard truncation regime, the same is similar to (1.12).

1.4 The Hill Estimator

Any statistical inference based on data which are believed to come from a probability law with power tails, revolves around estimating the tail index α. One of the most widely used estimator of the tail exponent is the Hill estimator, introduced by Hill (1975). Given a one-dimensional non-negative sample $X_1, \ldots, X_n$, the Hill statistic is defined by

$$h(k,n) = \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k)}}\,, \tag{1.14}$$

where $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$ are the order statistics from the sample $X_1, \ldots, X_n$, and $k = 1, \ldots, n$ is a user-determined parameter, the number of the upper order statistics to use in the estimator. The consistency result for the Hill estimator says that, if $X_1, \ldots, X_n$ are i.i.d. with regularly varying right tail with exponent $\alpha > 0$, and $k = k_n \to \infty$, $k_n/n \to 0$ as $n \to \infty$, then $h(k_n, n) \to 1/\alpha$ in probability as $n \to \infty$; see e.g. Theorem 3.2.2 in de Haan and Ferreira (2006). The heuristic idea behind such a result is the following. If the sample $X_1, \ldots, X_n$ came from a Pareto(α) distribution, *i.e.*, have the density $\alpha x^{-\alpha-1}$ for $x \geq 1$, then $E \log X_1 = 1/\alpha$. Hence, by the weak law of large numbers $n^{-1}\sum_{j=1}^{n} \log X_j$ would be a consistent estimator of α^{-1}. If however, the sample comes from a model which is assumed only to have power tails, then the distribution of $(X_{(j)}/X_{(k)} : 1 \leq j \leq k-1)$ approximates that of the order statistics of a sample of size $k-1$ from Pareto(α) whenever k is a sequence of integers satisfying $1 \ll k \ll n$. Hence, one would expect that

$$\frac{1}{k-1}\sum_{i=1}^{k-1} \log \frac{X_{(i)}}{X_{(k)}}$$

consistently estimates α^{-1}. Clearly, this quantity is asymptotically equivalent to the Hill estimator in probability.

Under some additional distributional assumptions, the Hill estimator is known to be asymptotically normal, as stated in the following result; see, for example, Proposition 9.3 in Resnick (2007).

Theorem 1.4.1. *Suppose $X_1, X_2, \ldots$ are i.i.d. $[0, \infty)$ valued random variables with c.d.f. F, such that $\bar{F} := 1 - F$ is regularly varying with index $-\alpha$, $\alpha > 0$. Suppose that k_n is a sequence of integers satisfying $1 \ll k_n \ll n$. In addition, assume that*

$$\lim_{n\to\infty} \sqrt{k_n} \left(\frac{n}{k_n} \bar{F}(b(n/k_n)y) - y^{-\alpha} \right) = 0$$

locally uniformly in $(0, \infty]$ and

$$\lim_{n\to\infty} \sqrt{k_n} \int_1^\infty \left(\frac{n}{k_n} \bar{F}(b(n/k_n)s) - s^{-\alpha} \right) \frac{ds}{s} = 0\,,$$

where

$$b(\cdot) := \left(\frac{1}{\bar{F}(\cdot)} \right)^{\leftarrow}.$$

Then,

$$\sqrt{k_n} \left(h(k_n, n) - \frac{1}{\alpha} \right) \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,,$$

where for $1 \leq k \leq n$, $h(k, n)$ is the Hill estimator, as defined in (1.14)*, for the random variables $X_1, \ldots, X_n$.*

In this dissertation, the behavior of the Hill estimator, when the sample arises from the model with truncated power law (1.4), is studied.

1.5 Outline of Dissertation

As mentioned earlier, the behavior of the model with truncated power tails is studied in this dissertation, with the goal of understanding the differences between the regimes of hard and soft truncation, as defined in (1.5). On this note, in Chapters 2 and 3, we investigate respectively the central limit theorem and the large deviations behavior of the partial sums of the observations coming from (1.4). Our results for these two chapters make it clear that in important respects, observations with softly truncated tails behave like heavy tailed random variables, while those with hard truncated tails behave like light tailed random variables. Thus, figuring out the truncation regime from given data is an important problem. This task is addressed in Chapter 4, where we suggest statistical procedures for testing the hypothesis of the soft (correspondingly, hard) truncation regime against the appropriate alternative. Finally, we consider the problem of estimating the tail exponent α based on a sample of observations with truncated power tails without knowing the truncation level or, even, if the truncation is soft or hard. We show how this can be accomplished in Chapter 5. In Chapter 6 we apply the statistical techniques of Chapter 4 to two recent data sets related to TCP connections in a large computer network.

CHAPTER 2

CENTRAL LIMIT THEOREM

2.1 Introduction

In this chapter, we study the central limit behavior of the sum of truncated heavy-tailed random vectors. By "heavy-tailed", we shall mean random vectors in the domain of attraction of some α-stable law. Let B be a separable Banach space and $\mathcal{V}$ an α-stable law on that with $0 < \alpha < 2$. Suppose that $H, H_1, H_2, \ldots$ are B-valued random variables in the domain of attraction of $\mathcal{V}$. This means that there are deterministic sequences b_n and c_n so that

$$b_n^{-1} \sum_{j=1}^{n} H_j - c_n \Longrightarrow \mathcal{V} . \tag{2.1}$$

This assumption about H exceeds that defined in Chapter 1, *i.e.*, there exist μ and (a_n) satisfying (1.1). Let $L, L_1, L_2, \ldots$ be i.i.d. $[0, \infty)$-valued random variables such that the families $(H, H_1, H_2, \ldots)$ and $(L, L_1, L_2, \ldots)$ are independent. The triangular array $\{X_{nj} : 1 \le j \le n\}$ is generated according to (1.4). We consider the row sum

$$S_n := \sum_{j=1}^{n} X_{nj} .$$

In this chapter we study the limiting distribution of S_n after suitable centering and scaling.

The aim of such an investigation is to decide, based on this, what the truncated model resembles more - the untruncated case or the case with bounded support. Clearly, if the thresholding sequence M_n is identically equal to infinity, then the limit law of S_n will be $\mathcal{V}$, while if M_n is identically equal to a finite constant, then on $\mathbb{R}^d$, the limit law will be a Gaussian one. Thus, one would expect that there is a dichotomy based on the growth rate of M_n. It turns out that (1.5) is exactly that dichotomy.

Section 2.2 gives the results in general Banach spaces. Section 2.3 specializes to Banach spaces of type 2. A counter-example is discussed in section 2.4.

2.2 General Banach Spaces

We start with the situation where the truncation level M_n grows sufficiently fast with the sample size, so that the truncated power tails model (1.4) is in the soft truncation regime. Theorem 2.2.1 below shows that, in this case, the partial sums of the random vectors with truncated heavy tails converge, when properly centered and scaled, to the same α-stable limit as without truncation.

Theorem 2.2.1. *In the soft truncation regime we have*

$$b_n^{-1} S_n - c_n \Longrightarrow \mathcal{V} \,. \tag{2.2}$$

Proof. By (2.1) it is enough to show that

$$b_n^{-1} \left\| S_n - \sum_{j=1}^{n} H_j \right\| \xrightarrow{p} 0 \,.$$

However, for any $\varepsilon > 0$,

$$P\left(b_n^{-1} \left\| S_n - \sum_{j=1}^{n} H_j \right\| > \varepsilon \right) \leq P\left(\|H_j\| > M_n \text{ for some } j = 1, \ldots, n \right)$$

$$\leq nP\left(\|H_1\| > M_n \right) \to 0 \,,$$

and the claim follows. □

Next, we consider the intermediate regime

$$\lim_{n\to\infty} nP(\|H\| > M_n) = \delta \in (0,\infty)\,. \tag{2.3}$$

It turns out that in this case, the limit is an infinitely divisible law, which is obtained by a certain truncation of the jumps of the α-stable law $\mathcal{V}$ in (2.1).

We start with some preliminaries. As mentioned in Section 1.2, since the limiting law $\mathcal{V}$ in (2.1) is α-stable, the Lévy-Khinchine formula for its characteristic function has the form

$$\hat{\mathcal{V}}(f) = \exp\left[if(x_0) + \int_B \left\{e^{if(x)} - 1 - if(x)\mathbf{1}(\|x\| \le 1)\right\}\mu(dx)\right],\, f \in B' \tag{2.4}$$

for some $x_0 \in B$ and a unique measure μ on $B \setminus \{0\}$ where B' denotes the dual space of B. μ is called the Lévy measure of $\mathcal{V}$. There is a finite measure Γ on $S := \{x \in B : \|x\| = 1\}$ so that

$$\mu(A) = \int_S \left\{\int_0^\infty \mathbf{1}_A(rs) r^{-(1+\alpha)} dr\right\} \Gamma(ds) \tag{2.5}$$

for every Borel set $A \subset B \setminus \{0\}$. Γ is known as the spectral measure of $\mathcal{V}$. The normalized spectral measure of $\mathcal{V}$, denoted by σ, is defined as

$$\sigma(\cdot) := \Gamma(\cdot)/\Gamma(S)\,.$$

We shall denote by B_r the closed ball of radius r, *i.e.*,

$$B_r := \{x \in B : \|x\| \le r\}\,.$$

For further details on stable laws on separable Banach spaces and their domain of attraction, the reader is referred to Araujo and Giné (1980).

Theorem 2.2.2. *If* (2.3) *holds, then*

$$b_n^{-1}\left(S_n - n\int_{\{\|x\| \le M_n\}} xP(H \in dx)\right) \Longrightarrow \mathcal{V}_\delta$$

where

$$\hat{\mathcal{V}}_\delta(f) \;:=\; \exp\Bigg[\int_{\{\|x\|\le C\}} \left\{e^{if(x)} - 1 - if(x)\right\} \mu(dx) + \delta \int_S \left(e^{iCf(x)} - 1\right) \sigma(dx)\Bigg],\ f \in B'$$

and

$$C := \delta^{-1/\alpha}(\alpha^{-1}\Gamma(S))^{1/\alpha}.$$

For the proof, we shall need the following result, which can be proved by similar arguments as in Theorem 5.9, page 129 of Araujo and Giné (1980).

Theorem 2.2.3. *Let $X, X_1, X_2, \dots$ be B-valued random variables in the domain of attraction of some α-stable law with $0 < \alpha < 2$. Assume that there is a sequence d_n and a measure ν on $B \setminus \{0\}$ satisfying*

$$nP(X/d_n \in \cdot) \longrightarrow \nu$$

weakly on B_ε^c for all $\varepsilon > 0$. For $n \ge 1$, let $\{Y_{nj} : 1 \le j \le n\}$ be a family of i.i.d. random variables with

$$P(Y_{n1} \in A) = P\left(X \in A \,\middle|\, \|X\| \le d_n\right),\ A \subset B.$$

Then,

1. $d_n^{-1}\left(\sum_{j=1}^n X_j - a_n\right)$ *converges weakly to the law with characteristic function*

$$\exp\left[\int \left\{e^{if(x)} - 1 - if(x)\mathbf{1}(\|x\| \le 1)\right\} \nu(dx)\right],\ f \in B',$$

where

$$a_n := n\int_{\{\|x\|\le d_n\}} xP(X \in dx).$$

2. $d_n^{-1}\left(\sum_{j=1}^n Y_{nj} - a_n\right)$ *converges weakly to the law with characteristic function*

$$\exp\left[\int_{\{\|x\|\le 1\}} \left\{e^{if(x)} - 1 - if(x)\right\} \nu(dx)\right],\ f \in B'.$$

Proof of Theorem 2.2.2. Write

$$S_n := \sum_{j=1}^{n} H_j \mathbf{1}(\|H_j\| \le M_n)$$

$$+M_n \sum_{j=1}^{n} \frac{H_j}{\|H_j\|}\mathbf{1}(\|H_j\| > M_n) + \sum_{j=1}^{n} L_j \frac{H_j}{\|H_j\|}\mathbf{1}(\|H_j\| > M_n)$$

Clearly, the last sum above is stochastically bounded and hence vanishes on scaling by M_n. Thus, without loss of generality we assume that L is identically zero. It follows by Theorem 6.18, page 150 in Araujo and Giné (1980) that

$$\lim_{n\to\infty} nP(H/b_n \in B_1^c) = \mu(B_1^c) = \alpha^{-1}\Gamma(S)\,,$$

the second equality following from (2.5). By (2.3), it follows that

$$\lim_{n\to\infty} \frac{M_n}{b_n} = \delta^{-1/\alpha}(\alpha^{-1}\Gamma(S))^{1/\alpha} = C\,. \tag{2.6}$$

Hence,

$$nP(H/M_n \in \cdot) \longrightarrow \gamma$$

weakly on B_ε^c for all $\varepsilon > 0$, where $\gamma(dx) = \mu(Cdx)$. Define

$$K_n := \sum_{j=1}^{n} \mathbf{1}(\|H_j\| > M_n)\,.$$

Fix an integer $k \ge 0$. Clearly, for $n \ge k$, the conditional distribution of S_n given that $K_n = k$ is same as that of

$$\sum_{j=1}^{n-k} Y_{nj} + M_n \sum_{j=1}^{k} Z_{nj} =: S_{n,k}^{(1)} + S_{n,k}^{(2)}$$

where $\{Y_{nj} : 1 \le j \le n\}$ is a family of i.i.d. random variables with law given by

$$P(Y_{n1} \in A) = P\left(H \in A \,\middle|\, \|H\| \le M_n\right)$$

and $\{Z_{nj} : 1 \le j \le n\}$ is a family of i.i.d. random variables taking values in $S := \{x \in B : \|x\| = 1\}$, independent of $\{Y_{nj} : 1 \le j \le n\}$ with

$$P(Z_{n1} \in A) = P\left(\frac{H}{\|H\|} \in A \,\middle|\, \|H\| > M_n\right)\,, A \subset S\,.$$

By Theorem 2.2.3, it follows that as $n \longrightarrow \infty$,

$$M_{n-k}^{-1}\left(S_{n,k}^{(1)} - (n-k)\int_{\{\|x\|\le M_n\}} xP(H \in dx)\right)$$

converges weakly to a law with characteristic function

$$\exp\left[\int_{\{\|x\|\le 1\}} \left\{e^{if(x)} - 1 - if(x)\right\}\gamma(dx)\right],\, f \in B'\,. \tag{2.7}$$

This immediately implies that

$$M_n^{-1}\left(S_{n,k}^{(1)} - n\int_{\{\|x\|\le M_n\}} xP(H \in dx)\right)$$

converges weakly to the same law, *i.e.*, the one with characteristic function (2.7). H being in the domain of attraction of $\mathcal{V}$ means that

$$\frac{P\left(\|H\| > r,\, \frac{H}{\|H\|} \in \cdot\right)}{P\left(\|H\| > r\right)} \xrightarrow{w} \sigma(\cdot) \tag{2.8}$$

on S; see e.g. Corollary 6.20 (b) of Araujo and Giné (1980). This implies that

$$M_n^{-1} S_{n,k}^{(2)} \Longrightarrow \sum_{j=1}^{k} U_j$$

where $U_1, U_2, \ldots$ are i.i.d. S-valued random variables distributed as σ. Hence, the conditional distribution of

$$M_n^{-1}\left(S_n - n\int_{\{\|x\|\le M_n\}} xP(H \in dx)\right)$$

given that $K_n = k$ converges weakly to

$$A + \sum_{j=1}^{k} U_j\,,$$

where A is distributed as (2.7), independent of $(U_j : j \ge 1)$. Note that (2.3) implies that K_n converges weakly to a Poisson limit with mean δ. Thus,

$$M_n^{-1}\left(S_n - n\int_{\{\|x\|\le M_n\}} xP(H \in dx)\right)$$

$$\Longrightarrow A + \sum_{j=1}^{N} U_j$$

where N is a Poisson random variable with mean δ, independent of $(A, U_1, U_2, \ldots)$. Thus, for $f \in B'$,

$$\begin{aligned} E \exp\left\{ if \left(\sum_{j=1}^{N} U_j \right) \right\} &= \exp\left\{ \delta \int_S e^{if(x)} \sigma(dx) - 1 \right\} \\ &= \exp\left[\delta \int_S \left(e^{if(x)} - 1 \right) \sigma(dx) \right] . \end{aligned}$$

This shows that the characteristic function of $A + \sum_{j=1}^{N} U_j$ is given by

$$\exp\left[\int_{\{\|x\| \le 1\}} \left\{ e^{if(x)} - 1 - if(x) \right\} \gamma(dx) + \delta \int_S \left(e^{if(x)} - 1 \right) \sigma(dx) \right] , f \in B' .$$

Thus, by (2.6),

$$b_n^{-1} \left(S_n - n \int_{\{\|x\| \le M_n\}} x P(H \in dx) \right)$$

converges weakly to a law with characteristic function

$$\exp\left[\int_{\{\|x\| \le 1\}} \left\{ e^{iCf(x)} - 1 - iCf(x) \right\} \gamma(dx) + \delta \int_S \left(e^{iCf(x)} - 1 \right) \sigma(dx) \right] , f \in B' .$$

A change of variable completes the proof. □

The above result shows that in the intermediate regime, the scaling constant is same as that in the untruncated case. We show furthermore in the following result that the centering constant can also be chosen to be the same as that in (2.1).

Theorem 2.2.4. *In the intermediate regime, $b_n^{-1}(S_n - c_n)$ converges weakly to the law with characteristic function*

$$\exp\left[if(x_\delta) + \int_{\{\|x\| \le C\}} \left\{ e^{if(x)} - 1 - if(x) \right\} \mu(dx) + \delta \int_S \left(e^{iCf(x)} - 1 \right) \sigma(dx) \right], \ f \in B' ,$$

where

$$x_\delta := x_0 - \int_{\{\|x\| \le 1\}} x \mu(dx) + \int_{\{\|x\| \le C\}} x \mu(dx) .$$

Proof. In view of Theorem 2.2.2, all that needs to be shown is

$$\lim_{n\to\infty} b_n^{-1}\left(n\int_{\{\|x\|\le M_n\}} xP(H\in dx)-c_n\right)=x_\delta\,. \tag{2.9}$$

By Theorem 2.2.3 and (2.6), it follows that

$$b_n^{-1}\left(\sum_{j=1}^n H_j - n\int_{\{\|x\|\le M_n\}} xP(H\in dx)\right)$$

converges weakly to the law with characteristic function

$$\exp\left[\int \left\{e^{if(x)}-1-if(x)\mathbf{1}(\|x\|\le C)\right\}\mu(dx)\right]\,,\, f\in B'\,.$$

This in view of (2.1) and (2.4) shows (2.9) and thus completes the proof. □

Finally, we consider the situation where the truncation level M_n grows relatively slowly with the sample size, and that the truncated power tails model (1.4) is in the hard truncation regime. Also assume that $EL^2<\infty$. These assumptions will be in force for the rest of this chapter. As we will see, in this case the partial sums of the random vectors with truncated heavy tails are no longer asymptotically α-stable but, rather, converge in law (under some additional assumptions), after suitable centering and scaling, to a Gaussian limit. Therefore, at least from the point of view of the behavior of partial sums, a model with power tails that have been truncated hard does not behave anymore as a heavy tailed model.

We start with a one-dimensional result.

Theorem 2.2.5. *For every f in B',*

$$B_n^{-1}\left(f(S_n)-Ef(S_n)\right)\Longrightarrow N\left(0,\frac{2}{2-\alpha}\int_S f^2(s)\sigma(ds)\right)\,,$$

where

$$B_n := [nM_n^2P(\|H\|>M_n)]^{1/2},\; n=1,2,\dots\,.$$

The proof is using the following lemma.

Lemma 2.2.1. *For every $f \in B'$,*

$$\lim_{n\to\infty} nB_n^{-2}\int_S\int_0^{M_n} f(s)r^2 P\left(\|H\|\in dr, \frac{H}{\|H\|}\in ds\right) = \frac{\alpha}{2-\alpha}\int_S f(s)\,\sigma(ds).$$

Proof. By (2.8),

$$\begin{aligned}
&\int_S\int_0^{M_n} f(s)r^2 P\left(\|H\|\in dr, \frac{H}{\|H\|}\in ds\right)\\
&= \int_0^{M_n} 2y\left(\int_S f(s)P\left(\|H\|>y,\ \frac{H}{\|H\|}\in ds\right)\right)dy\\
&\quad -M_n^2\int_S f(s)P\left(\|H\|>M_n,\ \frac{H}{\|H\|}\in ds\right)\\
&\sim \int_S f(s)\,\sigma(ds)\left[\int_0^{M_n} 2yP\left(\|H\|>y\right)\,dy - M_n^2 P\left(\|H\|>M_n\right)\right]\\
&\sim \int_S f(s)\,\sigma(ds)\left(\frac{2}{2-\alpha}-1\right)M_n^2 P\left(\|H\|>M_n\right) = n^{-1}B_n^2\int_S f(s)\,\sigma(ds)
\end{aligned}$$

as $n \longrightarrow \infty$, where the second asymptotic equivalence follows from the Karamata theorem (see e.g. Resnick (1987)). □

Proof of Theorem 2.2.5. We shall use the Central Limit Theorem for triangular arrays; see e.g. Theorem 2.4, page 345 in Gut (2005). We need to prove that

$$\lim_{n\to\infty}\frac{n}{B_n^2}\mathrm{Var}\left(f(X_{n1})\right) = \frac{2}{2-\alpha}\int_S f(s)^2\sigma(ds) \tag{2.10}$$

and that for every $\varepsilon > 0$,

$$\frac{n}{B_n^2}E\left(\left|f(X_{n1})-E\left(f(X_{n1})\right)\right|^2\mathbf{1}\left(\left|f(X_{n1})-E\left(f(X_{n1})\right)\right|>\varepsilon B_n\right)\right)\to 0 \tag{2.11}$$

as $n \longrightarrow \infty$. In order to prove (2.10), we will show that

$$\lim_{n\to\infty}\frac{n}{B_n^2}E\left(f(X_{n1})^2\right) = \frac{2}{2-\alpha}\int_S \left(f(s)\right)^2\sigma(ds) \tag{2.12}$$

while

$$\lim_{n\to\infty} \frac{n^{1/2}}{B_n} |E(f(X_{n1}))| = 0 . \tag{2.13}$$

The former claim follows easily from Lemma 2.2.1 and the weak convergence (2.8) by writing

$$\begin{aligned} E\left((f(X_{n1}))^2\right) &= E\left(f(H))^2 \mathbf{1}(\|H\| \le M_n)\right) \\ &+ E\left(\frac{(f(H))^2}{\|H\|^2}(M_n + L_1)^2 \mathbf{1}(\|H\| > M_n)\right) \\ \sim n^{-1}B_n^2 \frac{\alpha}{2-\alpha} \int_S (f(s))^2\, \sigma(ds) &+ (1+o(1))\, M_n^2 E\left(\frac{(f(H))^2}{\|H\|^2}\mathbf{1}(\|H\| > M_n)\right) \\ \sim n^{-1}B_n^2 \frac{\alpha}{2-\alpha} \int_S (f(s))^2\, \sigma(ds) &+ M_n^2 P(\|H\| > M_n) \int_S (f(s))^2\, \sigma(ds) \\ = n^{-1}B_n^2 \left(\frac{\alpha}{2-\alpha} + 1\right) \int_S (f(s))^2\, \sigma(ds) &= n^{-1}B_n^{\alpha} \frac{\alpha}{2-\alpha} \int_S (f(s))^2\, \sigma(ds) . \end{aligned}$$

For (2.13) we write

$$|E(f(X_{n1}))| \le \|f\| \left[E(\|H\|\mathbf{1}(\|H\| \le M_n)) + M_n P(\|H\| > M_n)\right] .$$

Since

$$M_n P(\|H\| > M_n) \ll M_n (P(\|H\| > M_n))^{1/2} = n^{-1/2} B_n ,$$

the claim (2.13) will follow once we check that

$$\lim_{n\to\infty} n^{1/2} B_n^{-1} E[\|H\|\mathbf{1}(\|H\| \le M_n)] = 0 . \tag{2.14}$$

We give separate arguments for the cases $\alpha \le 1$ and $\alpha > 1$.

Case 1 ($\alpha \le 1$): Letting C be a positive constant whose value may change from line to line, by the Karamata theorem,

$$\begin{aligned} E[\|H\|\mathbf{1}(\|H\| \le M_n)] &\le \left(E\left[\|H\|^{3/2}\mathbf{1}(\|H\| \le M_n)\right]\right)^{2/3} \\ &\sim C M_n (P(\|H\| > M_n))^{2/3} \\ &= C n^{-1/2} B_n (P(\|H\| > M_n))^{1/6} \\ &\ll n^{-1/2} B_n . \end{aligned}$$

Case 2 ($1 < \alpha < 2$): Here (2.14) follows trivially from the fact that $E\left[\|H\|\mathbf{1}(\|H\| \le M_n)\right]$ has a finite limit, while $B_n \gg n^{1/2}$ as $\alpha < 2$.

We have now proved (2.10). By (2.13), the remaining condition (2.11) will follow once we check that for every $\varepsilon > 0$,

$$\frac{n}{B_n^2} E\left(|f(X_{n1})|^2 \, \mathbf{1}\left(|f(X_{n1})| > \varepsilon B_n\right)\right) \to 0\,.$$

This is, however, an immediate consequence of the fact that the hard truncation implies that $B_n \gg M_n$ as $n \to \infty$. $\square$

Theorem 2.2.5 immediately shows by the Cramér-Wold device that if $B = \mathbb{R}^d$, then

$$B_n^{-1}\left(S_n - ES_n\right) \Longrightarrow \eta\,,$$

where η is a centered Gaussian law on $\mathbb{R}^d$ whose covariance matrix has the entries

$$\frac{2}{2-\alpha} \int_S s_i s_j \sigma(ds),\ i, j = 1, \ldots, d\,.$$

It also follows that in general Banach spaces, if there is any hope of weak convergence to a non-degenerate limit, then the right scaling constant for $S_n - ES_n$ is B_n.

Recall that a $\mathbb{R}^d$-valued random variable X is in the domain of attraction of some non-Gaussian stable law if and only if there exists a sequence b_n and a non null Radon measure μ on $\bar{\mathbb{R}}^d$ with $\mu(\bar{\mathbb{R}}^d \setminus \mathbb{R}^d) = 0$ such that

$$nP(X/b_n \in \cdot) \longrightarrow \mu$$

vaguely on $\bar{\mathbb{R}}^d \setminus \{0\}$, see Rvačeva (1962). In other words, being in the domain of attraction or not depends only upon the tail. However on general Banach spaces this condition is only necessary and far from sufficient. The problem with general Banach spaces is that even the usual Central Limit Theorem is not well understood. There are examples of random variables with bounded support which do not satisfy the Central Limit Theorem. The difficulty stems from the fact that on Banach spaces, the mere convergence of characteristic functions to that of the limit is not sufficient for weak convergence. One needs to check in addition some tightness condition as stated in the following result, which follows from Theorem 2.1 in Ledoux and Talagrand (1991) and the paragraph preceding that.

Theorem 2.2.6. *A sequence of probability measures $\{\mu_n : n \geq 1\}$ on B converges weakly to another probability measure μ if and only if*

1. for every $f \in B'$,

$$\lim_{n\to\infty} \int_B \exp\left(if(x)\right) \mu_n(dx) = \int_B \exp\left(if(x)\right) \mu(dx)\,,$$

2. for every $\epsilon > 0$, there is a compact set K in B such that

$$\inf_{n\geq 1} \mu_n(K) \geq 1 - \epsilon\,.$$

As a result, unlike as in $\mathbb{R}^d$, a Banach space valued random variable being in the domain of attraction of some α-stable law means more than an assumption about the tail. Easy-to-check criteria for satisfying the Central Limit Theorem on Banach spaces are not known. The following result is an example of the not-so-easy-to-check ones, known as the "small ball criterion"; see Theorem 10.13, page 289 in Ledoux and Talagrand (1991).

Theorem 2.2.7. *Let X be a zero mean B-valued random variable. Then X satisfies the Central Limit Theorem if and only if*

1. $\lim_{t\to\infty} t^2 P(\|X\| > t) = 0$,

2. *for each* $\epsilon > 0$*,*

$$\liminf_{n\to\infty} P\left(\|S_n\|/\sqrt{n} < \epsilon\right) > 0\,.$$

Our next result is an analogue of the above theorem in the truncated setting under hard truncation.

Theorem 2.2.8. *There is a Gaussian measure* γ *on* B *such that*

$$B_n^{-1}(S_n - ES_n) \Rightarrow \gamma \tag{2.15}$$

if and only if the following hold:

1. *(small ball criterion) For every* $\epsilon > 0$

$$\liminf_{n\to\infty} P(B_n^{-1}\|S_n - ES_n\| < \epsilon) > 0$$

2. $\sup_{n\geq 1} B_n^{-1} E\|S_n - ES_n\| < \infty$.

In that case, γ *is given by*

$$\hat{\gamma}(f) = \exp\left(-\frac{2}{2-\alpha}\int_S f^2(s)\sigma(ds)\right)\,,\, f \in B'\,. \tag{2.16}$$

For the proof, we shall need the following lemma, which is an easy consequence of the contraction principle.

Lemma 2.2.2. *Suppose that* $\{(X_j, \alpha_j) : 1 \leq j \leq N\}$ *is a family of* N *independent* $(B \times \mathbb{R})$*-valued random variables. Assume that for each* j*, both* X_j *and* $\alpha_j X_j$ *are symmetric random variables with* $E\|X_j\| < \infty$*. Further, if* $|\alpha_j| \leq 1$ *almost surely for all* j*, then*

$$E\left\|\sum_{j=1}^N \alpha_j X_j\right\| \leq E\left\|\sum_{j=1}^N X_j\right\|\,.$$

Proof. Fix i.i.d. Rademacher random variables $\varepsilon_1, \ldots, \varepsilon_N$ that are independent of $\{(X_j, \alpha_j) : 1 \le j \le N\}$. Denote by E_ε the expectation with respect to the filtration generated by $\{(X_j, \alpha_j) : 1 \le j \le N\}$. Note that

$$\begin{aligned} E\left\| \sum_{j=1}^N \alpha_j X_j \right\| &= E\left\| \sum_{j=1}^N \varepsilon_j \alpha_j X_j \right\| \\ &= EE_\varepsilon \left\| \sum_{j=1}^N \varepsilon_j \alpha_j X_j \right\| \\ &\le EE_\varepsilon \left\| \sum_{j=1}^N \varepsilon_j X_j \right\| \\ &= E\left\| \sum_{j=1}^N \varepsilon_j X_j \right\| \\ &= E\left\| \sum_{j=1}^N X_j \right\| , \end{aligned}$$

where the inequality follows from Theorem 4.4 in Ledoux and Talagrand (1991), which is the contraction principle. This completes the proof. □

Proof of Theorem 2.2.8. First we prove the direct part, *i.e.*, we assume that 1. and 2. hold. We first show that it suffices to check that $\{\mathcal{L}(Z_n)\}$ is relatively compact where

$$\begin{aligned} Z_n &:= B_n^{-1} \sum_{j=1}^n Y_{nj} , \\ Y_{nj} &:= X_{nj} - X'_{nj} \end{aligned}$$

for every n and $X'_{n1}, X'_{n2}, \ldots$ are i.i.d. copies of X_{n1} so that $(X'_{nj} : j \ge 1)$ and $(X_{nj} : j \ge 1)$ are independent families. To see this, note that by Corollary 4.11 in Araujo and Giné (1980), the relative compactness of $\{\mathcal{L}(Z_n)\}$ implies that the sequence $\{\mathcal{L}(B_n^{-1} S_n)\}$ is relatively shift compact, *i.e.*, there exists some sequence $\{v_n\} \subset B$ such that $\{\mathcal{L}(B_n^{-1} S_n - v_n)\}$ is relatively compact. Theorem 4.1 in de Acosta and Giné (1979) states that if $\{V_{nj} : 1 \le j \le n\}$ is a triangular array such that $\{\mathcal{L}(\sum_{j=1}^n V_{nj})\}$ is relatively shift compact and

$$\lim_{t\to\infty} \sup_{n\ge 1} \sum_{j=1}^n E\left[\|V_{nj}\|\mathbf{1}(\|V_{nj}\| > t)\right] = 0\,,$$

then $\{\mathcal{L}[\sum_{j=1}^n (V_{nj} - EV_{nj})]\}$ is relatively compact. In view of that, once we check the following, showing relative compactness of $\{\mathcal{L}(Z_n)\}$ will show that $\{\mathcal{L}[B_n^{-1}(S_n - ES_n)]\}$ is relatively compact:

$$\lim_{t\to\infty} \limsup_{n\to\infty} nE\left[\|U_n\|\mathbf{1}(\|U_n\| > t)\right] = 0\,, \tag{2.17}$$

where

$$U_n := B_n^{-1}\left[H\mathbf{1}(\|H\| \le M_n) + \frac{H}{\|H\|}(M_n + L)\mathbf{1}(\|H\| > M_n))\right]\,.$$

Fix $t > 0$. Since $B_n \gg M_n$, for n large enough,

$$\begin{aligned} nE\left[\|U_n\|\mathbf{1}(\|U_n\| > t)\right] &= nB_n^{-1}E\left[(M_n + L)\mathbf{1}(\|H\| > M_n)\mathbf{1}(L > B_n t - M_n)\right] \\ &\le Ct^{-1}nM_n^2B_n^{-2}P(\|H\| > M_n)\,. \end{aligned}$$

Thus,

$$\limsup_{n\to\infty} nE\left[\|U_n\|\mathbf{1}(\|U_n\| > t)\right] \le Ct^{-1}\,.$$

This shows (2.17) and hence that $\{\mathcal{L}[B_n^{-1}(S_n - ES_n)]\}$ is relatively compact. Thus, in view of Theorem 2.2.5, this will complete the proof of the direct part.

First we record some properties of the random variables defined above, which shall be used in the proof. The hypotheses immediately implies that for all $\epsilon > 0$

$$\liminf_{n\to\infty} P(\|Z_n\| < \epsilon) > 0 \tag{2.18}$$

and that

$$\sup_{n\ge 1} E\|Z_n\| < \infty\,. \tag{2.19}$$

Let $\{F_k\}$ be any sequence of increasing finite-dimensional subspaces so that

$$\text{closure}\left(\bigcup_{k=1}^{\infty} F_k\right) = B\,. \tag{2.20}$$

For any subspace F of B, denote by T_F the canonical map from B to the quotient space B/F. By Corollary 6.19 (page 151) in Araujo and Giné (1980), it follows that for every k $T_{F_k}(H)$ is in the domain of attraction of some α-stable law with the same scaling constant (b_n) as that of H, and that

$$\lim_{k\to\infty}\sup_{n\geq 1} nP\left(\|T_{F_k}(H)\| > b_n\right) = 0\,. \tag{2.21}$$

Clearly, for every k, there is $C_k \in [0,\infty)$ so that as $t \longrightarrow \infty$,

$$P(\|T_{F_k}(H)\| > t) \sim C_k P(\|H\| > t)\,.$$

It follows by (2.21) that

$$\lim_{k\to\infty} C_k = 0\,.$$

Thus,

$$\begin{aligned} &\limsup_{n\to\infty}[M_n^2 P(\|H\| > M_n)]^{-1} E\|T_{F_k}(X_{n1})\|^2 \\ \leq\;& \limsup_{n\to\infty}[M_n^2 P(\|H\| > M_n)]^{-1} E\left(\|T_{F_k}(H)\|^2 \mathbf{1}\left(\|T_{F_k}(H)\| \leq M_n\right)\right) \\ &+ \limsup_{n\to\infty} M_n^{-2}\frac{P(\|T_{F_k}(H)\| > M_n)}{P(\|H\| > M_n)} E(M_n + L)^2\,. \end{aligned}$$

By the Karamata theorem,

$$\lim_{n\to\infty}[M_n^2 P(\|H\| > M_n)]^{-1} E\left(\|T_{F_k}(H)\|^2 \mathbf{1}\left(\|T_{F_k}(H)\| \leq M_n\right)\right) = \frac{\alpha}{2-\alpha} C_k\,,$$

while clearly

$$\lim_{n\to\infty} M_n^{-2}\frac{P(\|T_{F_k}(H)\| > M_n)}{P(\|H\| > M_n)} E(M_n + L)^2 = C_k\,.$$

Thus,

$$\lim_{k\to\infty}\limsup_{n\to\infty}[M_n^2 P(\|H\| > M_n)]^{-1} E\|T_{F_k}(X_{n1})\|^2 = 0$$

which in turn shows,

$$\lim_{k\to\infty} \limsup_{n\to\infty} [M_n^2 P(\|H\| > M_n)]^{-1} E\|T_{F_k}(Y_{n1})\|^2 = 0\,. \tag{2.22}$$

Coming to the proof, in view of the criterion for relative compactness discussed in Ledoux and Talagrand (1991) (page 40-41), it suffices to show that given $\epsilon > 0$, there is a finite dimensional subspace F with

$$\limsup_{n\to\infty} P\left[\|T_F(Z_n)\| > \epsilon\right] \le \epsilon\,. \tag{2.23}$$

Let $\varepsilon_1, \varepsilon_2, \ldots$ be an i.i.d. sequence of Rademacher random variables, independent of $(X_n, X_n', n \ge 1)$, and let E_ε denote the conditional expectation given $\{Y_{nj}\}$. Observing that

$$(\varepsilon_j Y_{nj} : j \ge 1) \stackrel{d}{=} (Y_{nj} : j \ge 1),$$

it suffices to show that for all $\eta > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} P\left[\left|\left\|\sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj})\right\| - E_\varepsilon \left\|\sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj})\right\|\right| > B_n \eta\right] = 0\,, \tag{2.24}$$

and that there is a numerical constant $C > 0$ so that for every $\delta > 0$,

$$\limsup_{k\to\infty} \limsup_{n\to\infty} P\left[E_\varepsilon \left\|\sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj})\right\| > B_n C \delta\right] < \delta\,, \tag{2.25}$$

whenever $\{F_k\}$ is a sequence of finite-dimensional subspaces satisfying (2.20).

To establish (2.24), it suffices to check that

$$\lim_{k\to\infty} \limsup_{n\to\infty} P\left[\left|\left\|\sum_{j=1}^n \varepsilon_j T_{F_k}(u_{nj})\right\| - E_\varepsilon \left\|\sum_{j=1}^n \varepsilon_j T_{F_k}(u_{nj})\right\|\right| > B_n \eta\right] = 0$$

where

$$u_{nj} := Y_{nj} \mathbf{1}(\|Y_{nj}\| \le \beta B_n)\,,$$

$\beta > 0$ is to be specified later. This is because for n large enough,

$$\begin{aligned}
& B_n^{-1} E \left\| \sum_{j=1}^n Y_{nj} \mathbf{1} \left(\|Y_{nj}\| > \beta B_n \right) \right\| \\
\leq \; & B_n^{-1} \sum_{j=1}^n E \left((M_n + L) \mathbf{1} \left(\|H\| > M_n \right) \mathbf{1} \left(L > \beta B_n - M_n \right) \right) \\
= \; & B_n^{-1} n P(\|H\| > M_n) O(B_n^{-1}) \\
= \; & o(1) \,.
\end{aligned}$$

The proof is via the concentration property (Theorem 4.7, page 100 in Ledoux and Talagrand (1991)) of Rademacher averages, which says for $t > 0$ and any subspace F,

$$\begin{aligned}
P_\varepsilon \left[\left| B_n^{-1} \left\| \sum_{j=1}^n \varepsilon_j T_F(u_{nj}) \right\| - M \right| > t \right] \; & \leq \; 4 \exp\left(-\frac{t^2}{8\sigma_{n,F}^2} \right) \\
& \leq \; 32 \frac{\sigma_{n,F}^2}{t^2} \,,
\end{aligned}$$

where P_ε denotes the conditional probability given $\{Y_{nj}\}$,

$$\sigma_{n,F} := B_n^{-1} \sup_{f \in (B/F)', \|f\| \leq 1} \left[\sum_{j=1}^n f^2(T_F(u_{nj})) \right]^{1/2} ,$$

and M is the P_ε-median of $B_n^{-1} \left\| \sum_{j=1}^n \varepsilon_j T_F(u_{nj}) \right\|$. Though it has been suppressed in the notation, $\sigma_{n,F}$ does depend on β. Fix n and a subspace F and note that

$$\begin{aligned}
\left| B_n^{-1} E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_F(u_{nj}) \right\| - M \right| \; & \leq \; E_\varepsilon \left| B_n^{-1} \left\| \sum_{j=1}^n \varepsilon_j T_F(u_{nj}) \right\| - M \right| \\
& \leq \; \int_0^{\sigma_{n,F}} dt + \int_{\sigma_{n,F}}^\infty 32 \frac{\sigma_{n,F}^2}{t^2} dt \\
& = \; 33 \sigma_{n,F} \,.
\end{aligned}$$

Hence, on the set $\{\sigma_{n,F} \leq \eta/66\}$,

$$\left| B_n^{-1} E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_F(u_{nj}) \right\| - M \right| \leq \frac{\eta}{2}$$

and thus, on that set,

$$
\begin{aligned}
& P_\varepsilon\left[\left|\left\|\sum_{j=1}^n \varepsilon_j T_F(u_{nj})\right\| - E_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j T_F(u_{nj})\right\|\right| > B_n\eta\right] \\
\leq \;& P_\varepsilon\left[\left|B_n^{-1}\left\|\sum_{j=1}^n \varepsilon_j T_F(u_{nj})\right\| - M\right| > \frac{\eta}{2}\right] \\
\leq \;& 32\frac{\sigma_{n,F}^2}{(\eta/2)^2} \\
= \;& 128\frac{\sigma_{n,F}^2}{\eta^2}\,,
\end{aligned}
$$

hence, proving that

$$
\begin{aligned}
& P\left[\left|\left\|\sum_{j=1}^n \varepsilon_j T_F(u_{nj})\right\| - E_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j T_F(u_{nj})\right\|\right| > B_n\eta\right] \\
\leq \;& P\left[\sigma_{n,F} > \frac{\eta}{66}\right] + 128\frac{E\sigma_{n,F}^2}{\eta^2} \\
\leq \;& \frac{10^4}{\eta^2}E\sigma_{n,F}^2\,.
\end{aligned}
$$

Thus all that needs to be shown is that given any $\delta > 0$, there is a choice of β depending only on δ, so that

$$\limsup_{k\to\infty}\limsup_{n\to\infty} E\sigma_{n,F_k}^2 \leq \delta\,.$$

Using Lemma 6.6 (page 154) in Ledoux and Talagrand (1991), it follows that for any n, F

$$E\sigma_{n,F}^2 \leq nB_n^{-2}\sup_{f\in(B/F)',\|f\|\leq 1} Ef^2(T_F(u_{n1})) + 8B_n^{-2}E\left\|\sum_{j=1}^n u_{nj}\|u_{nj}\|\right\|\,.$$

Clearly,

$$nB_n^{-2}\sup_{f\in(B/F_k)',\|f\|\leq 1} Ef^2(T_{F_k}(u_{n1})) \leq [M_n^2 P(\|H\| > M_n)]^{-1}E(\|T_{F_k}(Y_{n1})\|^2)$$

which can be made as small as needed by (2.22). For the other part, note that

$$\begin{aligned} B_n^{-2} E \left\| \sum_{j=1}^n u_{nj} \|u_{nj}\| \right\| &\leq \beta B_n^{-1} E \left\| \sum_{j=1}^n u_{nj} \right\| \\ &\leq \beta B_n^{-1} E \left\| \sum_{j=1}^n Y_{nj} \right\| \\ &= \beta E \|Z_n\| \end{aligned}$$

where both the inequalities follow from Lemma 2.2.2. Thus, choosing β smaller than $\delta/(16 \sup_{n\geq 1} E\|Z_n\|)$ (which is positive because of (2.19)) does the trick.

For the proof of (2.25) we shall show that there is an universal constant $C > 0$ so that whenever F is a subspace satisfying

$$\liminf_{n\to\infty} P\left[E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_F(Y_{nj}) \right\| \leq 2B_n\delta \right] > 0\,, \tag{2.26}$$

it follows that

$$\limsup_{n\to\infty} P\left[E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_F(Y_{nj}) \right\| > CB_n\delta \right] \leq \delta\,. \tag{2.27}$$

The reason that this suffices is the following. Fix $\delta > 0$ and a sequence of finite-dimensional subspaces $\{F_k\}$ satisfying (2.20). Note that for all $n, k \geq 1$,

$$\begin{aligned} &P\left[B_n^{-1} E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj}) \right\| > 2\delta \right] \\ \leq\ & P(\|Z_n\| > \delta) + P\left[\left| \left\| \sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj}) \right\| - E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj}) \right\| \right| > B_n\delta \right]. \end{aligned}$$

By (2.24) and (2.18), it follows that

$$\liminf_{k\to\infty} \liminf_{n\to\infty} P\left[E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_{F_k}(Y_{nj}) \right\| \leq 2B_n\delta \right] > 0\,.$$

By (2.27), (2.25) follows.

The proof of (2.27) uses the following isoperimetric inequality; see Theorem 1.4 (page 26) in Ledoux and Talagrand (1991).

Theorem 2.2.9. *Given a probability space (E, Σ, μ) and a fixed, but arbitrary, integer $N \geq 1$, denote by P the product measure $\mu^{\otimes N}$ on E^N. There is a universal positive constant K, independent of N, satisfying for all $A \in \Sigma^{\otimes N}$ and $q, k \geq 1$,*

$$P_*(H_N(A, k, q)) \geq 1 - \left[K\left(\frac{1/P(A)}{k} + \frac{1}{q}\right)\right]^k,$$

where P_* denotes the inner probability associated with P,

$$\begin{aligned} H_N(A, q, k) \;\; := \;\; & \Big\{ x \in B^N : \text{ there exist } x^1, \ldots, x^q \in A \\ & \qquad \text{such that } \# \left\{ i \leq n : x_i \notin \left\{x_i^1, \ldots, x_i^q\right\}\right\} \leq k \Big\}, \end{aligned} \qquad (2.28)$$

and for $x \in E^N$, the coordinates are denoted by $x_1, \ldots, x_N$ respectively.

In what follows, we adopt the notation according to Theorem 2.2.9, that is, for any $u \in B^n$, the coordinates will be denoted by $u_1, \ldots, u_n$ respectively. Fix a subspace F satisfying (2.26). Let $T = T_F$ and define

$$A := \left\{ x \in B^n : E_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i T(x_i) \right\| \leq 2B_n\delta \right\}.$$

Fix $n, q, k \geq 1$ and let $x \in H_n(A, q, k)$, where H_n is as defined in (2.28). Then there exist $u \leq k$, $x^1, \ldots, x^q$ in A and integers $1 \leq i_1 < \ldots < i_u \leq n$ such that

$$\left\{ i \leq n : x_i \notin \left\{x_i^1, \ldots, x_i^q\right\}\right\} = \{i_1, \ldots, i_u\}.$$

Thus,

$$\{1, \ldots, n\} = \{i_1, \ldots, i_u\} \cup I,$$

where

$$I := \cup_{l=1}^q \{ i \leq n : x_i = x_i^l \}.$$

To see this, fix $v \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_u\}$. Then, $x_v \in \{x_v^1, \ldots, x_v^q\}$ and hence

$v \in I$. Thus,

$$\begin{aligned} E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T(x_j) \right\| &\le k \max_{j\le n} \|x_j\| + E_\varepsilon \left\| \sum_{j\in I} \varepsilon_j T(x_j) \right\| \\ \text{(by the contraction principle)} &\le k \max_{j\le n} \|x_j\| + \sum_{l=1}^q E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T(x_j^l) \right\| \\ \text{(since } x^l \in A \text{ for all } l) &\le k \max_{j\le n} \|x_j\| + 2q\delta B_n \,. \end{aligned}$$

This shows: for $n, q, k \ge 1$

$$\begin{aligned} &\left\{ x \in B^n : E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T(x_j) \right\| > (2q+1) B_n \delta \right\} \\ \subset\ & (H_n(A,q,k))^c \cup \left\{ x \in B^n : k \max_{j\le n} \|x_j\| > B_n \delta \right\} . \end{aligned}$$

Let

$$\theta := \liminf_{n\to\infty} P\left[E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_F(Y_{nj}) \right\| \le 2B_n\delta \right] > 0 \,.$$

By Theorem 2.2.9, it follows that

$$\limsup_{n\to\infty} P\left[E_\varepsilon \left\| \sum_{j=1}^n \varepsilon_j T_F(Y_{nj}) \right\| > (2q+1)B_n\delta \right]$$

$$\le \left[K \left(\frac{\log(1/\theta)}{k} + \frac{1}{q} \right) \right]^k + P\left[B_n^{-1} \max_{j\le n} \|Y_{nj}\| > \frac{\delta}{k} \right] ,$$

where K is the universal constant in the isoperimetric inequality. Choose $q = 2K$ and k to be large enough (depending only on θ) so that

$$\left[K \left(\frac{\log(1/\theta)}{k} + \frac{1}{q} \right) \right]^k \le \frac{\delta}{2} \,.$$

All that remains to be shown is

$$\lim_{n\to\infty} P\left[B_n^{-1} \max_{j\le n} \|Y_{nj}\| > \frac{\delta}{k} \right] = 0 \,. \tag{2.29}$$

Note that

$$\max_{j\le n} \|Y_{nj}\| \le \max_{j\le n} \|X_{nj}\| + \max_{j\le n} \|\tilde{X}_{nj}\|$$

and that

$$\max_{j\le n}\|X_{nj}\|\le M_n+\max_{j\le n}L_j\,.$$

Since $EL_1^2<\infty$, $\{n^{-1/2}\max_{j\le n}L_j\}$ is a tight family. This shows (2.29) and thus establishes (2.27) with $C=4q+1$ and hence completes the proof of the direct part.

The converse is straightforward. For **1.**, note that if (2.15) holds, by the continuous mapping theorem,

$$\lim_{n\to\infty}P(B_n^{-1}\|S_n-ES_n\|\le\epsilon)=\gamma\left(B_\epsilon\right)\,,$$

the right hand side being positive because in a separable Banach space a centered Gaussian law puts positive mass on any ball with positive radius centered at origin, see the discussion on page 60-61 in Ledoux and Talagrand (1991). **2.** follows from Theorem 2.1 in de Acosta and Giné (1979). □

2.3 Type 2 Banach Spaces

It would be nice if the statement of Theorem 2.2.8 were true without the assumption of the small ball criterion. Unfortunately, that is not the case, as a counter-example is shown in the next section. However, if a condition named type 2 is imposed on the space, then the claim of Theorem 2.2.8 holds without any further assumption, in the hard truncation regime. Showing that is precisely the content of this section. We start with defining the type of a Banach space.

Definition 1. *A Banach space is said to be of type p, $1\le p\le 2$, if there exists a finite* *constant C such that*

$$E\left\|\sum_{i=1}^{n} X_i\right\|^p \leq C\sum_{i=1}^{n} E\|X_i\|^p$$

for all $n \geq 1$ and zero mean independent B-valued random variables $X_1, \ldots, X_n$ with finite p-th moment.

The following result, which follows from Theorem 7.2 in Araujo and Giné (1980), gives an equivalent definition of a space of type p.

Theorem 2.3.1. *A Banach space B is of type p with $1 \leq p \leq 2$ if and only if the following is true. Whenever $\{\varepsilon_n\}_{n=1}^{\infty}$ is a sequence of independent Rademacher random variables and $\{x_n\}_{n=1}^{\infty} \subset B_1$ is such that $\sum_{n=1}^{\infty} \|x_n\|^p < \infty$, the infinite sum $\sum_{n=1}^{\infty} \varepsilon_n x_n$ converges almost surely.*

This shows that if $1 \leq p \leq p' \leq 2$ and B is of type p', then B is also of type p. Clearly, every space is of type 1.

As commented in the previous section, the Central Limit Theorem on general Banach spaces is not well understood. Banach spaces of type 2 are nice in the sense that every random variable X taking values there with $E||X\|^2 < \infty$ satisfies the Central Limit Theorem. In fact these are the only spaces where this is true. This is the statement of Theorem 10.5 (page 281) in Ledoux and Talagrand (1991). We would like to mention at this point that while the assumption of type 2 is a rather restrictive one, this is a fairly large class. For example, every Hilbert space and l_p for $p \geq 2$ is a Banach space of type 2. We show in the following result that (2.15) can be extended on these spaces.

Theorem 2.3.2. *If B is of type 2 and the model with power law tails* (1.4) *is in the hard truncation regime, then there is a Gaussian measure γ on B such that*

$$B_n^{-1}(S_n - ES_n) \Rightarrow \gamma$$

The characteristic function of γ is given by (2.16).

Proof. In view of Theorem 2.2.5 and using similar arguments as in the proof of Theorem 2.2.8, it suffices to prove that $\{\mathcal{L}(Z_n)\}$ is relatively compact where the definition of Z_n (and Y_{nj}) is exactly the same as in the proof of the latter theorem. Choose a sequence $\{F_k\}$ of finite dimensional subspaces satisfying (2.20). By the type 2 property, there is $C \in (0, \infty)$ so that

$$\begin{aligned} E\|T_{F_k}(Z_n)\|^2 &\leq CB_n^{-2}nE\|T_{F_k}(Y_{n1})\|^2 \\ &= C[M_n^2 P(\|H\| > M_n)]^{-1}E\|T_{F_k}(Y_{n1})\|^2 \,. \end{aligned}$$

Using (2.22), it follows that

$$\lim_{k\to\infty} \limsup_{n\to\infty} E\|T_{F_k}(Z_n)\|^2 = 0$$

which shows (2.23) and thus completes the proof. □

2.4 A Counter-example

The following is a counter-example where for a random variable Y in the domain of attraction of some α-stable law, $B_n^{-1}(S_n - ES_n)$ is not weakly convergent for some choice of M_n in the hard truncation regime and the light tailed random variable L.

Example 1. *Let $1 \leq \alpha < p < 2$ and suppose that B is of type α and not of type p. In the proof of Theorem 9.21 in Ledoux and Talagrand (1991), a symmetric bounded random variable X is constructed, so that $n^{-1/p}\sum_{i=1}^n X_i$ does not converge to 0 in probability, where $X_1, X_2, \ldots$ are i.i.d. copies of X. By the same result,*

$$n^{-1/\alpha}\sum_{i=1}^n X_i \xrightarrow{P} 0 \,.$$

Fix $x \in B \setminus \{0\}$ and define

$$Y := X + xS$$

where S is a $\mathbb{R}$-valued $S\alpha S$ random variable, independent of X. Thus, Y is in the domain of attraction of an α-stable law on B. Let $Y_1, Y_2, \ldots$ denote i.i.d. copies of Y. For a positive number M_n,

$$Y_{ni} := Y_i \mathbf{1}\left(\|Y_i\| \le M_n\right) + M_n \frac{Y_i}{\|Y_i\|} \mathbf{1}\left(\|Y_i\| > M_n\right)$$

is the truncation of Y_i to the ball of radius M_n, as defined in (1.4) *with L identically equal to zero. Let*

$$S_n := \sum_{i=1}^{n} Y_{ni}\,.$$

Clearly $n^{-1/p} S_n$ does not converge to 0 in probability whenever $M_n \longrightarrow \infty$. Note that if in addition $M_n^{1-\frac{\alpha}{2}} \ll n^{\frac{1}{p}-\frac{1}{2}}$, then

$$\begin{aligned} B_n^2 &= nM_n^2 P(\|Y\| > M_n) \\ &= O(nM_n^{2-\alpha}) \\ &= o(n^{2/p})\,. \end{aligned}$$

Thus, under that upper bound on growth rate of M_n, $B_n^{-1} S_n$ is not weakly convergent.

CHAPTER 3

LARGE DEVIATIONS

3.1 Introduction

In this chapter we study the behavior of the large deviation probabilities for sums of truncated heavy-tailed random variables. Let H be a $\mathbb{R}^d$ valued random variable satisfying (1.2) for some sequence a_n going to infinity and a non-null Radon measure μ on $\overline{\mathbb{R}^d}$ with $\mu(\overline{\mathbb{R}^d} \setminus \mathbb{R}^d) = 0$. We further assume that if $\alpha = 1$ then H has a symmetric distribution and if $\alpha > 1$ then $E(H) = 0$. The triangular array $\{X_{nj} : 1 \leq j \leq n\}$ is as defined in (1.4), where $H_1, H_2, \ldots$ are i.i.d. copies of H, M_n is a sequence of positive numbers going to ∞, $L, L_1, L_2, \ldots$ are i.i.d. $[0, \infty)$ valued random variables independent of $H, H_1, H_2, \ldots$ and $\|\cdot\|$ denotes the L^2 norm on $\mathbb{R}^d$, *i.e.*, for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\|x\| := \left(\sum_{j=1}^{d} x_j^2 \right)^{1/2} . \tag{3.1}$$

We shall study large deviations for

$$S_n := \sum_{j=1}^{n} X_{nj} .$$

The motivation for this chapter is similar to that for Chapter 2, deciding based on the growth rate of the truncating threshold, what the model with truncated power tails resemble more - the untruncated one or the one with finite exponential moments. In other words, we want to link the behavior of the large deviation probabilities associated with S_n to the hard and soft truncation regimes defined in (1.5). The large deviations for the soft truncation and hard truncation regimes are studied in Sections 3.2 and 3.3 respectively.

3.2 Soft truncation case

For this section, we assume that M_n goes to ∞ fast enough so that

$$\lim_{n\to\infty} nP(\|H\| > M_n) = 0\,.$$

We assume in addition that if $\alpha = 2$, then

$$\lim_{n\to\infty} M_n/\sqrt{n^{1+\gamma}} = \infty \tag{3.2}$$

for some $\gamma > 0$, and if $\alpha > 2$, then

$$\lim_{n\to\infty} M_n/\sqrt{n\log n} = \infty\,.$$

Define

$$b_n := \begin{cases} \inf\{x : P(\|H\| > x) \le n^{-1}\}, & \alpha < 2 \\ \sqrt{n^{1+\gamma}}, & \alpha = 2 \\ \sqrt{n \log n}, & \alpha > 2\,, \end{cases} \tag{3.3}$$

where γ is same as that in (3.2). Clearly, $1 \ll b_n \ll M_n$ and $\mathcal{L}(b_n^{-1}S_n)$ is a tight sequence. The following result, which is an easy consequence of Lemma 2.1 in Hult et al. (2005), describes the large deviation behavior of $\lambda_n^{-1}S_n$ where $b_n \ll \lambda_n \ll M_n$.

Theorem 3.2.1. *If λ_n is any sequence of positive numbers satisfying $b_n \ll \lambda_n \ll M_n$, then*

$$\frac{P(\lambda_n^{-1}S_n \in \cdot)}{nP(\|H\| > \lambda_n)} \xrightarrow{v} \frac{\mu(\cdot)}{\mu(B_1^c)}$$

on $\overline{\mathbb{R}^d} \setminus \{0\}$. Recall that for all $r \ge 0$, B_r denotes the closed ball of radius r, centered at the origin.

Proof. Fix a sequence λ_n satisfying the hypotheses. The assumption that $\lambda_n \gg b_n$ implies

$$\lambda_n^{-1}S_n \xrightarrow{P} 0\,.$$

By Lemma 2.1 in Hult et al. (2005), it follows that

$$\frac{P\left(\lambda_n^{-1}\sum_{j=1}^n H_j \in \cdot\right)}{P(\|H\| > \lambda_n)} \xrightarrow{v} \frac{\mu(\cdot)}{\mu(B_1^c)}$$

on $\overline{\mathbb{R}^d} \setminus \{0\}$. Note that

$$\begin{aligned}
&\sup_{A \subset \mathbb{R}^d} \left| P(\lambda_n^{-1} S_n \in A) - P\left(\lambda_n^{-1}\sum_{j=1}^n H_j \in A\right)\right| \\
\leq\ & P(\|H_j\| > M_n \text{ for some } 1 \leq j \leq n) \\
\leq\ & nP(\|H\| > M_n) \\
=\ & o(nP(\|H\| > \lambda_n)),
\end{aligned}$$

the last equality following from the assumption that $\lambda_n \ll M_n$. This completes the proof. □

The next result, Theorem 3.2.2, describes the large deviation behavior of $M_n^{-1} S_n$. The reason we call this a large deviation result is the following. This result, for example, shows that for all $r \in (k-1, k)$ such that $\nu^{(k)}(\{x \in \mathbb{R}^d : \|x\| = r\}) = 0$ (which is in fact true for all but countably many r's in $(k-1,k)$),

$$P(\|S_n\| > rM_n) \sim C_r\{nP(\|H\| > M_n)\}^k$$

for some $C_r \in (0, \infty)$.

Theorem 3.2.2. *Suppose $k \geq 1$ and that*

$$P(L > x) = o(P(\|H\| > x)^{k-1}) \tag{3.4}$$

as $x \longrightarrow \infty$. Then, as $n \longrightarrow \infty$,

$$\frac{P(M_n^{-1} S_n \in \cdot)}{\{nP(\|H\| > M_n)\}^k} \xrightarrow{v} \frac{1}{k!}\nu^{(k)}$$

on $\mathbb{R}^d \setminus B_{k-1}$, where

$$\nu^{(k)}(A) := \int \cdots \int 1\left(\sum_{j=1}^k x_j \in A\right) \nu(dx_1) \ldots \nu(dx_k),$$

and

$$\nu(A) := \frac{\mu(A \cap B_1)}{\mu(B_1^c)} + \sigma(A \cap \mathcal{S})\,. \tag{3.5}$$

For the proof of Theorem 3.2.2, we shall need the following lemmas.

Lemma 3.2.1. *As* $t \longrightarrow \infty$,

$$\frac{P(X^t/t \in \cdot)}{P(\|H\| > t)} \xrightarrow{v} \nu$$

on $\overline{\mathbb{R}^d} \setminus \{0\}$, *where, for* $t > 0$,

$$X^t := H\mathbf{1}\left(\|H\| \le t\right) + (t+L)\frac{H}{\|H\|}\mathbf{1}\left(\|H\| > t\right)\,.$$

Proof. Since for all $\epsilon > 0$, ν restricted to B_ϵ^c is a finite measure, it suffices to show that for $\epsilon \in (0,1)$,

$$\lim_{t\to\infty} \frac{P(X^t/t \in B_\epsilon^c)}{P(\|H\| > t)} = \nu(B_\epsilon^c)\,, \tag{3.6}$$

and that for $A \subset \mathbb{R}^d$ which is closed and bounded away from zero,

$$\limsup_{t\to\infty} \frac{P(X^t/t \in A)}{P(\|H\| > t)} \le \nu(A)\,. \tag{3.7}$$

For(3.6), note that

$$\begin{aligned} \lim_{t\to\infty} \frac{P(X^t/t \in B_\epsilon^c)}{P(\|H\| > t)} &= \lim_{t\to\infty} \frac{P(H/t \in B_\epsilon^c)}{P(\|H\| > t)} \\ &= \frac{\mu(B_\epsilon^c)}{\mu(B_1^c)} P(\|H\| > t) \\ &= \nu(B_\epsilon^c)\,, \end{aligned}$$

where the second equality follows from the fact that

$$\frac{P(H/t \in \cdot)}{P(\|H\| > t)} \xrightarrow{v} \frac{\mu(\cdot)}{\mu(B_1^c)}$$

in $\overline{\mathbb{R}^d} \setminus \{0\}$, which is a consequence of (1.2), and that B_ϵ^c is a μ-continuous set. For (3.7), fix an $A \subset \mathbb{R}^d$ which is closed and bounded away from zero. Define a function

$$T : \mathbb{R}^d \setminus \{0\} \longrightarrow \mathcal{S}$$

by

$$T(x) = \frac{x}{\|x\|}\,.$$

Since A is closed,

$$\bigcap_{\epsilon>0} T(A \cap (B_{1+\epsilon} \setminus int(B_{1-\epsilon}))) = A \cap \mathcal{S}\,.$$

To see this, note that the right hand side is trivially contained in the left hand side. Suppose x belongs to the left hand side. Then, for every $n \geq 1$, there is $y_n \in A$ with $|\|y_n\| - 1| \leq 1/n$ and $y_n/\|y_n\| = x$. Clearly, then $y_n \longrightarrow x$ as $n \longrightarrow \infty$. Since A is closed, $x \in A$. That $x \in \mathcal{S}$ is trivial. Thus, for fixed $\delta > 0$ there is $\epsilon > 0$ so that

$$\sigma\left(T(A \cap (B_{1+\epsilon} \setminus int(B_{1-\epsilon})))\right) \leq \sigma(A \cap \mathcal{S}) + \delta\,.$$

Define

$$\tilde{A} := T(A \cap (B_{1+\epsilon} \setminus int(B_{1-\epsilon})))\,.$$

Since $A \cap (B_{1+\epsilon} \setminus int(B_{1-\epsilon}))$ is compact and T is continuous, $\tilde{A}$ is compact and hence closed. Note that

$$P(X^t/t \in A) \leq$$

$$P(X^t/t \in A \cap B_{1-\epsilon}) + P(X^t/t \in A \cap (B_{1+\epsilon} \setminus int(B_{1-\epsilon})) + P(\|X^t\| \geq (1+\epsilon)t)\,.$$

Clearly

$$P(X^t/t \in A \cap B_{1-\epsilon}) = P(H/t \in A \cap B_{1-\epsilon})$$

and hence

$$\limsup_{t\to\infty} \frac{P(X^t/t \in A \cap B_{1-\epsilon})}{P(\|H\| > t)} \leq \frac{\mu(A \cap B_1)}{\mu(B_1^c)}\,.$$

Also,

$$P(\|X^t\| \geq (1+\epsilon)t) \ll P(\|H\| > t)\,.$$

Note that

$$P(X^t/t \in A \cap (B_{1+\epsilon} \setminus int(B_{1-\epsilon})) \quad \leq \quad P(H/\|H\| \in \tilde{A}, \|H\| > (1-\epsilon)t)\,.$$

Since $\tilde{A}$ is closed,

$$\limsup_{t\to\infty} \frac{P(H/\|H\| \in \tilde{A}, \|H\| > (1-\epsilon)t)}{P(\|H\| > t)} \le (1-\epsilon)^{-\alpha}\sigma(\tilde{A}) \le (1-\epsilon)^{-\alpha}(\sigma(A)+\delta)\,.$$

Since ϵ and δ can be chosen to be arbitrarily small, this shows (3.7) and hence completes the proof. □

Lemma 3.2.2. *Suppose that* (3.4) *holds. Then,*

$$\frac{P\left(M_n^{-1}\sum_{j=1}^k X_{nj} \in \cdot\right)}{P(\|H\| > M_n)^k} \quad \stackrel{v}{\longrightarrow} \quad \nu^{(k)}(\cdot)\,,$$

on $\overline{\mathbb{R}^d} \setminus B_{k-1}$.

Proof. First note that for any $r > k-1$,

$$\begin{aligned}
&\nu^{(k)}(B_r^c)\\
= &\int \dots \int \mathbf{1}\left(\sum_{j=1}^k x_j \in B_r^c\right) \nu(dx_1)\dots\nu(dx_k)\\
= &\int_{\{(r-k+1)<\|x_1\|\le 1\}} \dots \int_{\{(r-k+1)<\|x_k\|\le 1\}} \mathbf{1}\left(\sum_{j=1}^k x_j \in B_r^c\right) \nu(dx_1)\dots\nu(dx_k)\,,
\end{aligned}$$

the last equality following from the fact that $\nu(B_1^c) = 0$. Since ν puts finite measure on the set $\{\|x\| > (r-k+1)\}$, it follows that $\nu^{(k)}(B_r^c) < \infty$.

Fix a $\nu^{(k)}$ continuity set $A \subset B_\delta^c$ for some $k-1 < \delta < k$. Fix $\epsilon > 0$ so that $(k-1)(1+\epsilon) < \delta$. Clearly,

$$\begin{aligned}
&P\left(M_n^{-1}\sum_{j=1}^k X_{nj} \in A, \|X_{nj}\| \le (1+\epsilon)M_n, 1 \le j \le k\right)\\
\le\ &P\left(M_n^{-1}\sum_{j=1}^k X_{nj} \in A\right)\\
\le\ &P\left(M_n^{-1}\sum_{j=1}^k X_{nj} \in A, \|X_{nj}\| \le (1+\epsilon)M_n, 1 \le j \le k\right)\\
&+kP(L > \epsilon M_n)P(\|H\| > M_n)\,.
\end{aligned}$$

By the assumption on L,

$$\begin{aligned} P(L > \epsilon M_n) &= o(P(\|H\| > \epsilon M_n)^{k-1}) \\ &= o(P(\|H\| > M_n)^{k-1}) . \end{aligned}$$

Since $A \subset B_\delta^c$ where $\delta > (k-1)(1+\epsilon)$,

$$\begin{aligned} & P\left(M_n^{-1} \sum_{j=1}^{k} X_{nj} \in A, \|X_{nj}\| \le (1+\epsilon) M_n, 1 \le j \le k \right) \\ = & \int_{\{\eta < \|x_1\| \le 1+\epsilon\}} \cdots \int_{\{\eta < \|x_k\| \le 1+\epsilon\}} \mathbf{1}\left(\sum_{j=1}^{k} x_j \in A \right) \\ & P(M_n^{-1} X_{n1} \in dx_1) \ldots P(M_n^{-1} X_{nk} \in dx_k) \\ = & \int_{\{\|x_1\| \le 1+\epsilon\}} \cdots \int_{\{\|x_k\| \le 1+\epsilon\}} \mathbf{1}\left(\sum_{j=1}^{k} x_j \in A \right) \\ & P_n(dx_1) \ldots P_n(dx_k) , \end{aligned} \tag{3.8}$$

where $\eta := \delta - (k-1)(1+\epsilon) > 0$ and $P_n(\cdot)$ denotes the restriction of $P(M_n^{-1} X_{n1} \in \cdot)$ to $\mathbb{R}^d \setminus B_\eta$. Let $\tilde{\nu}$ denote the restriction of ν to $\mathbb{R}^d \setminus B_\eta$. Then, by Lemma 3.2.1, as $n \longrightarrow \infty$,

$$\frac{P_n}{P(\|H\| > M_n)} \xrightarrow{w} \tilde{\nu}$$

on $\mathbb{R}^d \setminus B_\eta$. Thus,

$$\frac{P_n(dx_1) \ldots P_n(dx_k)}{P(\|H\| > M_n)^k} \xrightarrow{w} \tilde{\nu}(dx_1) \ldots \tilde{\nu}(dx_k)$$

on $(\mathbb{R}^d \setminus B_\eta)^k$, as $n \longrightarrow \infty$. Consider the function $f : \mathbb{R}^{d \times k} \longrightarrow \mathbb{R}$ defined by

$$f(x_1, \ldots, x_k) = \mathbf{1}(\|x_1\| \le 1+\epsilon) \ldots \mathbf{1}(\|x_k\| \le 1+\epsilon) \mathbf{1}\left(\sum_{j=1}^{k} x_j \in A \right) .$$

The set of discontinuities of f is contained in

$$\bigcup_{j=1}^{k} \{(x_1, \ldots, x_k) : \|x_j\| = 1+\epsilon\} \cup \left\{ (x_1, \ldots, x_k) : \sum_{j=1}^{k} x_j \in \partial A \right\} .$$

The product measure $\tilde{\nu}^k$ gives zero measure to this set because ν (and hence $\tilde{\nu}$) does not charge anything outside B_1 and the set A has been chosen to satisfy

$$\int \dots \int \mathbf{1}\left(\sum_{j=1}^{k} x_j \in \partial A\right) \nu(dx_1)\dots\nu(dx_k) = 0\,.$$

Thus, as $n \longrightarrow \infty$, the right hand side of (3.8) is asymptotically equivalent to

$$\begin{aligned} & P(\|H\| > M_n)^k \int_{\{\|x_1\| \le 1+\epsilon\}} \cdots \int_{\{\|x_k\| \le 1+\epsilon\}} \mathbf{1}\left(\sum_{j=1}^{k} x_j \in A\right) \tilde{\nu}(dx_1)\dots\tilde{\nu}(dx_k) \\ = \;& P(\|H\| > M_n)^k \nu^{(k)}(A)\,. \end{aligned}$$

This completes the proof. □

We shall also need the following result, which has been proved in Prokhorov (1959).

Lemma 3.2.3. *If $X_1, \dots, X_N$ are i.i.d. $\mathbb{R}$-valued independent random variables with $|X_i| \le C$ a.s. where $0 < C < \infty$, then, for $\lambda > 0$,*

$$P(S_N - ES_N > \lambda) \le \exp\left\{-\frac{\lambda}{2C}\sinh^{-1}\frac{C\lambda}{2\mathrm{Var}(S_N)}\right\}\,,$$

where

$$S_N := \sum_{i=1}^{n} X_i\,.$$

The proof of Theorem 3.2.2 is based on the idea that for $M_n^{-1}S_n$ to belong to a set A that is bounded away from B_{k-1} and is not entirely contained in B_k^c, it is "necessary and sufficient" that $M_n^{-1}\sum_{u=1}^{k} X_{nj_u}$ belongs to A for exactly one tuple $1 \le j_1 < \ldots < j_k \le n$. This idea is similar to the idea in the proof of Lemma 2.1 in Hult et al. (2005), that S_n is large "if and only if" exactly one of the summands is large.

Proof of Theorem 3.2.2. We shall show that for every $\nu^{(k)}$-continuous set $A \subset \mathbb{R}^d \setminus B_\delta$ for some $\delta > k-1$,

$$\lim_{n\to\infty} \frac{P(M_n^{-1} S_n \in A)}{\{nP(\|H\| > M_n)\}^k} = \frac{1}{k!}\nu^{(k)}(A)\,. \tag{3.9}$$

We first show the lower bound, *i.e.*, the lim inf in (3.9) is at least the right hand side. Fix a set A as described above. Define for $\epsilon > 0$

$$A^{-\epsilon} := \{x \in A : \text{ for all } y \in \mathbb{R}^d \text{ with } \|y-x\| < \epsilon, y \in A\}\,.$$

Clearly,

$$\lim_{\epsilon\downarrow 0} \nu^{(k)}(A^{-\epsilon}) = \nu^{(k)}(int(A)) = \nu^{(k)}(A)\,,$$

where the second equality is true because A is $\nu^{(k)}$-continuous. Thus, for the lower bound, it suffices to show that for all $\epsilon > 0$ so that $A^{-\epsilon}$ is a $\nu^{(k)}$-continuity set (which is true for all but countably many ϵ's),

$$\liminf_{n\to\infty} \frac{P(M_n^{-1} S_n \in A)}{\{nP(\|H\| > M_n)\}^k} \geq \frac{1}{k!}\nu^{(k)}(A^{-\epsilon})\,. \tag{3.10}$$

Fix $\epsilon > 0$ so that $A^{-\epsilon}$ is a $\nu^{(k)}$-continuity set. Since we want to show (3.10), we can assume without loss of generality that $\nu^{(k)}(A^{-\epsilon}) > 0$. Fix $n \geq k$ and define for $1 \leq j_1 < \ldots < j_k \leq n$

$$C_{j_1\ldots j_k} := \left\{ M_n^{-1} \sum_{u=1}^k X_{nj_u} \in A^{-\epsilon}, \left\| \sum_{i\in\{1,\ldots,n\}\setminus\{j_1,\ldots,j_k\}} X_{ni} \right\| < \epsilon M_n \right\}\,.$$

Though the above definition also depends on n, we suppress that to keep the notation simple. Clearly,

$$P(M_n^{-1} S_n \in A) \geq P\left(\bigcup C_{j_1\ldots j_k}\right)\,,$$

where the union is taken over all subsets of $\{1,\ldots,n\}$, and

$$\begin{aligned} P\left(C_{1,\ldots,k}\right) &= P\left(M_n^{-1}\sum_{j=1}^k X_{nj} \in A^{-\epsilon}\right) P\left(\left\|\sum_{i=1}^{n-k} X_{ni}\right\| < M_n\epsilon\right) \\ &\sim P\left(M_n^{-1}\sum_{j=1}^k X_{nj} \in A^{-\epsilon}\right) \\ &\sim P(\|H\| > M_n)^k \nu^{(k)}(A^{-\epsilon})\,, \end{aligned}$$

as $n \longrightarrow \infty$, where the first equivalence is true because

$$M_n^{-1} \sum_{i=1}^{n-k} X_{ni} \xrightarrow{P} 0$$

and the second equivalence follows from Lemma 3.2.2. Thus, for (3.10), all that remains to show is

$$P\left(\bigcup C_{j_1 \ldots j_k}\right) \sim \sum P\left(C_{j_1 \ldots j_k}\right) , \tag{3.11}$$

where the union and the sum are both taken over all subsets of $\{1, \ldots, n\}$. Fix $\eta > 0$ so that $(k-1)(1+\eta) < \delta$ and subsets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_k\}$ of $\{1, \ldots, n\}$ so that

$$\#\left(\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_k\}\right) = l < k\,. \tag{3.12}$$

Note that,

$$\begin{aligned}
& P\left(C_{i_1 \ldots i_k} \cap C_{j_1 \ldots j_k}\right) \\
\leq\ & P\left(M_n^{-1} \left\| \sum_{u=1}^{k} X_{nj_u} \right\| > \delta, M_n^{-1} \left\| \sum_{u=1}^{k} X_{ni_u} \right\| > \delta \right) \\
\leq\ & P\Bigg(M_n^{-1} \left\| \sum_{u=1}^{k} X_{nj_u} \right\| > \delta, M_n^{-1} \left\| \sum_{u=1}^{k} X_{ni_u} \right\| > \delta, \\
& \|X_{nu}\| \leq (1+\eta) M_n \text{ for } u \in \{i_1, \ldots, i_k\} \cup \{j_1, \ldots, j_k\}\Bigg) \\
& + 2kP(L > \eta M_n) P(\|H\| > M_n) \\
\leq\ & P\left(\|X_{nj}\| > [\delta - (k-1)(1+\eta)] M_n \text{ for } 1 \leq j \leq 2k - l\right) + o(P(\|H\| > M_n)^k) \\
=\ & O(P(\|H\| > M_n)^{2k-l})\,.
\end{aligned}$$

Clearly, there are at most $O(n^{2k-l})$ pairs of subsets satisfying (3.12). Thus,

$$\begin{aligned}
\sum P\left(C_{i_1 \ldots i_k} \cap C_{j_1 \ldots j_k}\right) &= \sum_{l=0}^{k-1} O(n^{2k-l} P(\|H\| > M_n)^{2k-l}) \\
&= o(n^k P(\|H\| > M_n)^k)\,,
\end{aligned}$$

where the sum in the left hand side of the first line is taken over all pairs of distinct subsets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_k\}$ of $\{1, \ldots, n\}$. This shows (3.11) and thus completes the proof of the lower bound.

For the upper bound, choose a sequence z_n satisfying

$$\{nP(\|H\| > M_n)\}^{\frac{k+1}{k+2}} \ll nP(\|H\| > z_n) \ll \{nP(\|H\| > M_n)\}^{\frac{k}{k+1}} \tag{3.13}$$

if $\alpha < 2$,

$$nP\left(\|H\| > \frac{M_n}{\log n}\right) \ll nP(\|H\| > z_n)$$
$$\ll \min\left(\{nP(\|H\| > M_n)\}^{\frac{k}{k+1}}, nP\left(\|H\| > \frac{n}{M_n}\right)\right) \tag{3.14}$$

if $\alpha > 2$, and

$$nP\left(\|H\| > \frac{M_n}{\log n}\right) \ll nP(\|H\| > z_n)$$
$$\ll \min\left(\{nP(\|H\| > M_n)\}^{\frac{k}{k+1}}, nP\left(\|H\| > \left(\frac{n}{M_n}\right)^{1+\eta}\right)\right) \tag{3.15}$$

for some $\eta > 0$ if $\alpha = 2$. Note that if u_n and v_n are sequences satisfying $u_n \ll v_n \ll 1$, then there exists a sequence w_n with

$$u_n \ll P(\|H\| > w_n) \ll v_n \,,$$

see Resnick (2007) for example. Thus, existence of z_n satisfying (3.13) is immediate from the assumption that $nP(\|H\| > M_n)$ goes to zero as $n \longrightarrow \infty$. A sequence satisfying (3.14) will exist if the following are verified:

$$\frac{M_n}{\log n} \gg \frac{n}{M_n}\,, \tag{3.16}$$
$$nP\left(\|H\| > \frac{M_n}{\log n}\right) \ll \{nP(\|H\| > M_n)\}^{\beta}\,, \tag{3.17}$$

where $\beta = k/(k+1)$. From the fact that $M_n \gg \sqrt{n \log n}$, (3.16) follows. For (3.17), letting $\epsilon \in (0, \alpha - 2)$ and c to be a finite constant whose value may change from line to line, note that

$$\begin{aligned}
& \frac{nP\left(\|H\| > \frac{M_n}{\log n}\right)}{\{nP(\|H\| > M_n)\}^\beta} \\
= \;& n^{1-\beta} M_n^{-\alpha(1-\beta)} (\log n)^\alpha l(M_n/\log n) l(M_n)^{-\beta} \\
\ll \;& n^{1-\beta} M_n^{-\alpha(1-\beta)} (\log n)^\alpha (M_n/\log n)^{\epsilon(1-\beta)/2} M_n^{\epsilon(1-\beta)/2} \\
= \;& n^{1-\beta} M_n^{-\alpha(1-\beta)} (\log n)^c M_n^{\epsilon(1-\beta)} \\
= \;& n^{1-\beta} M_n^{(\epsilon-\alpha)(1-\beta)} (\log n)^c \\
\ll \;& n^{1-\beta} n^{(\epsilon-\alpha)(1-\beta)/2} (\log n)^c \\
= \;& n^{(1-\beta)(2-\alpha+\epsilon)/2} (\log n)^c \\
\to \;& 0 \text{ by choice of } \epsilon\,.
\end{aligned}$$

To establish that a sequence z_n satisfying (3.15) exists, it suffices to check (3.17) and that

$$\frac{M_n}{\log n} \gg \left(\frac{n}{M_n}\right)^{1+\eta} . \tag{3.18}$$

For (3.17), let $0 < \epsilon < 2\{1-(1+\gamma)^{-1}\}$. Once again, letting c to be a finite constant whose value may change from line to line, by similar calculations as above,

$$\begin{aligned}
& \frac{nP\left(\|H\| > \frac{M_n}{\log n}\right)}{\{nP(\|H\| > M_n)\}^\beta} \\
\ll \;& n^{1-\beta} M_n^{(\epsilon-2)(1-\beta)} (\log n)^c \\
\ll \;& n^{1-\beta} n^{(\epsilon-2)(1-\beta)(1+\gamma)/2} (\log n)^c \\
\to \;& 0 \text{ by choice of } \epsilon\,.
\end{aligned}$$

It's easy to check that (3.18) holds for $\eta < 2\gamma$.

Write

$$\tilde{S}_n := \sum_{j=1}^n X_{nj} \mathbf{1}(\|X_{nj}\| \le z_n)\,.$$

Fix $0 < \epsilon < \delta - k + 1$ and define

$$A^\epsilon := \{y \in \mathbb{R}^d : \|y - x\| < \epsilon \text{ for some } x \in A\}\,.$$

Assume that ϵ is chosen so that A^ϵ is also a $\nu^{(k)}$-continuity set. Define the events

$$
\begin{aligned}
D_n &:= \left\{ M_n^{-1} \sum_{u=1}^{l} X_{nj_u} \in A^\epsilon \text{ for at least one tuple} \right. \\
&\qquad \left. 1 \le j_1 < j_2 < \ldots < j_l \le n, 1 \le l < k \right\}, \\
E_n &:= \left\{ M_n^{-1} \sum_{u=1}^{k} X_{nj_u} \in A^\epsilon \text{ for at least one tuple} \right. \\
&\qquad \left. 1 \le j_1 < j_2 < \ldots < j_k \le n \right\}, \\
F_n &:= \{ \|X_{nj}\| > z_n \text{ for at least } (k+1) \text{ many } j\text{'s} \le n \}, \\
G_n &:= \{ \|\tilde{S}_n\| > \epsilon M_n \}.
\end{aligned}
$$

Clearly,

$$
P\left(M_n^{-1} S_n \in A\right) \le P(D_n) + P(E_n) + P(F_n) + P(G_n).
$$

Also,

$$
\begin{aligned}
P(E_n) &\le \frac{n^k}{k!} P\left(M_n^{-1} \sum_{j=1}^{k} X_{nj} \in A^\epsilon \right) \\
&\sim \frac{1}{k!} \{nP(\|H\| > M_n)\}^k \int \cdots \int \mathbf{1}\left(\sum_{j=1}^{k} x_j \in A^\epsilon \right) \nu(dx_1) \ldots \nu(dx_k)
\end{aligned}
$$

by Lemma 3.2.2. By the fact that $A^\epsilon \subset B_\delta^c$,

$$
\begin{aligned}
P(D_n) &\le \sum_{l=1}^{k-1} n^l P\left(\| \sum_{j=1}^{l} X_{nj} \| > \delta M_n \right) \\
&\le \sum_{l=1}^{k-1} n^l l P(L > (\delta/l - 1) M_n) P(\|H\| > M_n) \\
&\ll n^k P(\|H\| > M_n)^k,
\end{aligned}
$$

the last inequality following from (3.4). By the choice of z_n,

$$
P(F_n) \le \{nP(\|H\| > z_n)\}^{k+1} \ll \{nP(\|H\| > M_n)\}^k.
$$

All that remains is to show that

$$P(G_n) \ll \{nP(\|H\| > M_n)\}^k \,. \tag{3.19}$$

Recall that $\|\cdot\|$ denotes the L^2 norm as defined in (3.1). Denoting the coordinates of a $\mathbb{R}^d$-valued random variable Y by $Y^{(j)}$ for $1 \le j \le d$, note that

$$P(G_n) \;\; \le \;\; \sum_{j=1}^{d} P\left(|\tilde{S}_n^{(j)}| > \epsilon M_n/\sqrt{d}\right) \,.$$

In view of this, to show (3.19), It suffices to prove that for $1 \le j \le d$,

$$ES_n^{(j)} \;\; = \;\; o(M_n) \tag{3.20}$$

$$P\left(|\tilde{S}_n^{(j)} - E\tilde{S}_n^{(j)}| > \theta M_n\right) \;\; = \;\; o\left(\{nP(\|H\| > M_n)\}^k\right) \,, \tag{3.21}$$

for all $\theta > 0$. By the assumption that H has a symmetric law when $\alpha = 1$, (3.20) is trivially true in that case. We shall show (3.20) separately for the cases $\alpha < 1$ and $\alpha > 1$. We start with the case $\alpha > 1$. Note that for n large enough so that $z_n < M_n$,

$$\begin{aligned}
|ES_n^{(j)}| \;\; &= \;\; n|E[X_{n1}^{(j)}\mathbf{1}(\|X_{n1}\| \le z_n)]| \\
&= \;\; n|E[H^{(j)}\mathbf{1}(\|H\| \le z_n)]| \\
(\text{since } EH = 0 \text{ when } \alpha > 1) \;\; &= \;\; n|E[H^{(j)}\mathbf{1}(\|H\| > z_n)]| \\
&\le \;\; nE[|H^{(j)}|\mathbf{1}(\|H\| > z_n)] \\
&\le \;\; nE[\|H\|\mathbf{1}(\|H\| > z_n)] \\
&= \;\; O(nz_nP(\|H\| > z_n)) \\
&= \;\; o(M_n)\,.
\end{aligned}$$

where the last step follows from the fact that the choice of z_n implies that $z_n \ll M_n$ and that $nP(\|H\| > z_n) \ll 1$, which are true, in fact, for all α. For the case

$\alpha < 1$, note that for n large enough,

$$\begin{aligned} |ES_n^{(j)}| &= n|E[X_{n1}^{(j)}\mathbf{1}(\|X_{n1}\| \le z_n)]| \\ &= n|E[H^{(j)}\mathbf{1}(\|H\| \le z_n)]| \\ &\le nE[|H^{(j)}|\mathbf{1}(\|H\| \le z_n)] \\ &\le nE[\|H\|\mathbf{1}(\|H\| \le z_n)] \\ &= O(nz_nP(\|H\| > z_n)) \\ &= o(M_n)\,. \end{aligned}$$

Note that by Lemma 3.2.3,

$$P\left(|\tilde{S}_n^{(j)} - E\tilde{S}_n^{(j)}| > \theta M_n\right) \le K_1 \exp\left\{-K_2\frac{M_n}{z_n}\sinh^{-1} K_3\frac{M_n z_n}{\mathrm{Var}(\tilde{S}_n^{(j)})}\right\},$$

for finite positive constants K_1, K_2 and K_3. Thus, all that needs to be shown is

$$\exp\left\{-K_2\frac{M_n}{z_n}\sinh^{-1} K_3\frac{M_n z_n}{\mathrm{Var}(\tilde{S}_n^{(j)})}\right\} \ll \{nP(\|H\| > M_n)\}^k\,. \tag{3.22}$$

We shall show this separately for the cases $\alpha < 2$, $\alpha \ge 2$. We start with the case $\alpha \ge 2$. For (3.22), it suffices to show that

$$\frac{M_n}{z_n} \gg \log n\,, \tag{3.23}$$

$$M_n z_n \gg \mathrm{Var}(\tilde{S}_n)\,. \tag{3.24}$$

It follows directly from choice of z_n that (3.23) is true. If $\alpha > 2$, then

$$\begin{aligned} \frac{\mathrm{Var}(\tilde{S}_n^{(j)})}{M_n z_n} &= O(n/M_n z_n) \\ &= o(1) \end{aligned}$$

by choice of z_n. If $\alpha = 2$, then there is a slowly varying function $m : [0,\infty) \to \mathbb{R}$ at ∞ so that

$$\begin{aligned} \frac{\mathrm{Var}(\tilde{S}_n^{(j)})}{M_n z_n} &= O(nm(z_n)/M_n z_n) \\ &= O\left(n/M_n z_n^{1/(1+\eta)}\right) \\ &= o(1)\,. \end{aligned}$$

Finally, let us come to the case $\alpha < 2$. Note that there is a slowly varying function $m : [0, \infty) \to \mathbb{R}$ at ∞ (which is possibly different from the one chosen just above), so that

$$\begin{aligned} \frac{M_n}{z_n} &\sim \left(\frac{P(\|H\| > z_n)}{P(\|H\| > M_n)} \right)^{1/\alpha} \frac{m(M_n)}{m(z_n)} \\ &\gg \left(\frac{P(\|H\| > z_n)}{P(\|H\| > M_n)} \right)^{1/\alpha} \frac{z_n}{M_n} \\ &\gg \{nP(\|H\| > M_n)\}^{-\frac{k+1}{\alpha(k+2)}} \frac{z_n}{M_n} \,. \end{aligned}$$

This shows that

$$\frac{M_n}{z_n} \gg \{nP(\|H\| > M_n)\}^{-u}$$

for some $u > 0$. Also, note that

$$\begin{aligned} \mathrm{Var}(\tilde{S}_n^{(j)}) &= O(nz_n^2 P(\|H\| > z_n)) \\ &= o(z_n M_n) \,, \end{aligned}$$

the last step following from the facts that $z_n \ll M_n$ and $nP(\|H\| > z_n) \ll 1$. Thus,

$$\frac{M_n}{z_n} \sinh^{-1} K_3 \frac{M_n z_n}{\mathrm{Var}(\tilde{S}_n^{(j)})} \gg \{nP(\|H\| > M_n)\}^{-u} \,,$$

and hence,

$$\begin{aligned} \exp\left\{ -K_2 \frac{M_n}{z_n} \sinh^{-1} K_3 \frac{M_n z_n}{\mathrm{Var}(\tilde{S}_n^{(j)})} \right\} &\ll \exp\left\{ -K_2 \{nP(\|H\| > M_n)\}^{-u} \right\} \\ &\ll \{nP(\|H\| > M_n)\}^k \,. \end{aligned}$$

This shows (3.22) and thus completes the proof. □

Theorem 3.2.2 clearly excludes the boundary cases, *i.e.*, it does not give the decay rate of $P(\|S_n\| > kM_n)$ when k is a positive integer. For stating the results for the boundary case, we need some preliminaries. By Rvačeva (1962), it follows that when $\alpha = 2$, H is in the domain of attraction of a Gaussian random variable G_0. In view of the assumption that $EH = 0$ whenever $\alpha > 1$, this means that there is a sequence a_n going to infinity so that

$$a_n^{-1} \sum_{j=1}^n H_j \Longrightarrow \mathcal{L}(G_0)\,.$$

By the same paper, it follows from (1.2) that if $0 < \alpha < 2$, there is a α-stable random variable S on $\mathbb{R}^d$ so that

$$b_n^{-1} \sum_{j=1}^n H_j \Longrightarrow \mathcal{L}(S)\,,$$

where b_n is as defined in (3.3). Clearly, if $\alpha > 2$, then

$$n^{-1/2} \sum_{j=1}^n H_j \Longrightarrow \mathcal{L}(G)\,,$$

where G is a Gaussian random variable on $\mathbb{R}^d$ with mean zero and covariance matrix same as that of H. Thus,

$$B_n^{-1} \sum_{j=1}^n H_j \Longrightarrow \mathcal{L}(\mathcal{V})\,, \tag{3.25}$$

where

$$B_n := \begin{cases} b_n, & \alpha < 2 \\ a_n, & \alpha = 2 \\ n^{1/2}, & \alpha > 2\,, \end{cases} \tag{3.26}$$

and

$$\mathcal{V} := \begin{cases} S, & \alpha < 2 \\ G_0, & \alpha = 2 \\ G, & \alpha > 2\,. \end{cases} \tag{3.27}$$

Note that

$$\begin{aligned} P\left(S_n \neq \sum_{j=1}^n H_j\right) &\leq P(\|H_j\| > M_n \text{ for some } 1 \leq j \leq n) \\ &\leq nP(\|H\| > M_n) \\ &\to 0\,. \end{aligned}$$

Thus, it follows from (3.25) that

$$B_n^{-1} S_n \Longrightarrow \mathcal{L}(\mathcal{V}) \,. \tag{3.28}$$

The next two results, which are the last two main results of this section, describe the behavior of the large deviation probability for the boundary cases. Specifically, Theorem 3.2.3 gives the decay rate of $P(\|S_n\| > M_n)$ and Theorem 3.2.4 gives the decay rate of $P(\|S_n\| > kM_n)$ for $k \geq 2$.

Theorem 3.2.3. (*The boundary case:* $k = 1$) *For all closed set* $F \subset \mathcal{S}$*,*

$$\limsup_{n\to\infty} \frac{P\left(\|S_n\| > M_n, \frac{S_n}{\|S_n\|} \in F\right)}{nP(\|H\| > M_n)} \leq \Gamma_1(F) \,,$$

where,

$$\Gamma_1(A) := \int_A P(\langle x, \mathcal{V}\rangle \geq 0)\sigma(dx) \,,$$

for $A \subset \mathcal{S}$*, and* $\mathcal{V}$ *is as defined in* (3.27)*. If, in addition,*

$$\int_{\mathcal{S}} P(\langle x, \mathcal{V}\rangle = 0)\sigma(dx) = 0 \,, \tag{3.29}$$

then, as $n \longrightarrow \infty$*,*

$$\frac{P\left(\|S_n\| > M_n, \frac{S_n}{\|S_n\|} \in \cdot\right)}{nP(\|H\| > M_n)} \xrightarrow{w} \Gamma_1(\cdot)$$

weakly on $\mathcal{S}$*.*

Theorem 3.2.4. (*The boundary case:* $k \geq 2$*) Suppose* $k \geq 2$ *and assume that* (3.4) *holds. Then,*

$$\limsup_{n\to\infty} \frac{P\left(\|S_n\| > kM_n, \frac{S_n}{\|S_n\|} \in F\right)}{\{nP(\|H\| > M_n)\}^k} \leq \Gamma_k(F) \,,$$

for all closed set $F \subset \mathcal{S}$*, where for all* $A \subset \mathcal{S}$*,*

$$\Gamma_k(A) := \frac{1}{k!} \sum_{s\in A} P(\langle s, \mathcal{V}\rangle \geq 0)\sigma(\{s\})^k \,.$$

If, in addition, for every $s \in \mathcal{S}$,

$$\liminf_{t \to \infty} \frac{P\left(\|H\| > t, \frac{H}{\|H\|} = s\right)}{P(\|H\| > t)} \geq \sigma(\{s\}) \tag{3.30}$$

and

$$P(\langle s, \mathcal{V} \rangle = 0)\sigma(\{s\}) = 0\,, \tag{3.31}$$

then,

$$\frac{P\left(\|S_n\| > kM_n, \frac{S_n}{\|S_n\|} \in \cdot\right)}{\{nP(\|H\| > M_n)\}^k} \xrightarrow{w} \Gamma_k(\cdot)\,,$$

weakly on $\mathcal{S}$.

It is easy to see that for all $k \geq 1$, $\Gamma_k(\mathcal{S}) \leq \sigma(\mathcal{S}) = 1$, which in particular implies that Γ_k is a finite measure. However, Γ_k might be the null measure, and if that is the case, the statements of Theorems 3.2.3 and 3.2.4 just mean that $P(\|S_n\| > kM_n)$ decays faster than $\{nP(\|H\| > M_n)\}^k$. For the proofs, we shall need the following lemma, which in fact, proves the first parts of both theorems.

Lemma 3.2.4. *Suppose $k \geq 1$ and assume that* (3.4) *holds. Then, as $n \longrightarrow \infty$,*

$$\limsup_{n \to \infty} \frac{P\left(\|S_n\| > kM_n, \frac{S_n}{\|S_n\|} \in F\right)}{\{nP(\|H\| > M_n)\}^k} \leq \Gamma_k(F)\,,$$

for all closed set $F \subset \mathcal{S}$.

Proof. It is easy to see that for all $k \geq 1$ and $A \subset \mathcal{S}$,

$$\Gamma_k(A) = \frac{1}{k!} \int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \mathbf{1}\left(\left\|\sum_{j=1}^k x_j\right\| = k, \frac{\sum_{j=1}^k x_j}{\|\sum_{j=1}^k x_j\|} \in A\right) P\left(\sum_{j=1}^k \langle x_j, \mathcal{V}\rangle \geq 0\right)$$
$$\sigma(dx_1)\ldots\sigma(dx_k)\,.$$

Fix $k \geq 1$ and a closed set $F \subset \mathcal{S}$. Let $0 < \eta < 1$ and define

$$E_n := \left\{\left\|\sum_{u=1}^k X_{nj_u}\right\| > (k-\eta)M_n \text{ for at least one tuple } 1 \leq j_1 < j_2 < \ldots < j_k \leq n\right\}.$$

By similar arguments as in the proof of Theorem 3.2.2, it follows that

$$P\left(\{\|S_n\| > kM_n\} \cap E_n^c\right) = o(\{nP(\|H\| > M_n)\}^k)$$

as $n \longrightarrow \infty$. Thus, for the upper bound, it suffices to show that

$$\limsup_{\eta \downarrow 0} \limsup_{n\to\infty} \frac{P\left(\left\{\|S_n\| > kM_n, \frac{S_n}{\|S_n\|} \in F\right\} \cap E_n\right)}{\{nP(\|H\| > M_n)\}^k}$$
$$\leq \frac{1}{k!}\int_{\mathcal{S}} \dots \int_{\mathcal{S}} \mathbf{1}\left(\|\sum_{j=1}^k x_j\| = k, \frac{\sum_{j=1}^k x_j}{\|\sum_{j=1}^k x_j\|} \in F\right) P\left(\sum_{j=1}^k \langle x_j, \mathcal{V}\rangle \geq 0\right)$$
$$\sigma(dx_1)\dots\sigma(dx_k)\,.$$

and for that it suffices to show

$$\limsup_{\eta \downarrow 0} \limsup_{n\to\infty} \frac{P\left(\|S_n\| > kM_n, \frac{S_n}{\|S_n\|} \in F, \|\sum_{j=1}^k X_{nj}\| > (k-\eta)M_n\right)}{P(\|H\| > M_n)^k}$$
$$\leq \int_{\mathcal{S}} \dots \int_{\mathcal{S}} \mathbf{1}\left(\|\sum_{j=1}^k x_j\| = k, \frac{\sum_{j=1}^k x_j}{\|\sum_{j=1}^k x_j\|} \in F\right) P\left(\sum_{j=1}^k \langle x_j, \mathcal{V}\rangle \geq 0\right)$$
$$\sigma(dx_1)\dots\sigma(dx_k)\,. \tag{3.32}$$

Fix a sequence ϵ_n satisfying $M_n^{-1} \ll \epsilon_n \ll M_n^{-1}B_n$, which is possible because B_n goes to infinity, where B_n is as defined in (3.26). Also $B_n = O(b_n) = o(M_n)$, where b_n is as in (3.3), thus showing that ϵ_n goes to zero as n goes to infinity. Set

$$F^\eta := \{x \in \mathcal{S} : \|x - s\| \leq \eta \text{ for some } s \in F\}\,.$$

Define the events

$$\begin{aligned} U_n \quad &:= \quad \Bigg\{\|\sum_{j=1}^k X_{nj}\| > (k-\eta)M_n, \frac{\sum_{j=1}^k X_{nj}}{\|\sum_{j=1}^k X_{nj}\|} \in F^\eta, \\ &\qquad \left\langle \frac{\sum_{j=1}^k X_{nj}}{\|\sum_{j=1}^k X_{nj}\|}, B_n^{-1}\sum_{j=k+1}^n X_{nj}\right\rangle \geq -\eta\Bigg\}\,, \\ V_n \quad &:= \quad \Bigg\{k-\eta < M_n^{-1}\|\sum_{j=1}^k X_{nj}\| \leq \sqrt{k^2+\epsilon_n}, \|S_n\| > kM_n, \\ &\qquad \left\langle \frac{\sum_{j=1}^k X_{nj}}{\|\sum_{j=1}^k X_{nj}\|}, B_n^{-1}\sum_{j=k+1}^n X_{nj}\right\rangle < -\eta\Bigg\}\,, \end{aligned}$$

$$
\begin{aligned}
W_n &:= \left\{ \| \sum_{j=1}^{k} X_{nj} \| > (k-\eta) M_n, \|S_n\| > M_n, \frac{\sum_{j=1}^{k} X_{nj}}{\| \sum_{j=1}^{k} X_{nj} \|} \notin F^{\eta}, \frac{S_n}{\|S_n\|} \in F \right\}, \\
Y_n &:= \left\{ \| \sum_{j=1}^{k} X_{nj} \| > (k-\eta) M_n, \min_{1 \le j \le k} \|X_{nj}\| < \frac{1-\eta}{2} M_n \right\}, \\
Z_n &:= \left\{ \min_{1 \le j \le k} \|X_{nj}\| \ge \frac{1-\eta}{2} M_n, \| \sum_{j=1}^{k} X_{nj} \| > \sqrt{k^2 + \epsilon_n} M_n \right\}.
\end{aligned}
$$

Note that

$$
\left\{ \|S_n\| > k M_n, \frac{S_n}{\|S_n\|} \in F, \| \sum_{j=1}^{k} X_{nj} \| > (k-\eta) M_n \right\} \subset U_n \cup V_n \cup W_n \cup Y_n \cup Z_n .
$$

Let $k - 1 < r < k - \eta$ be such that

$$
\nu^{(k)} \left(\{ x \in \mathbb{R}^d : \|x\| = r \} \right) = 0 .
$$

For $n \ge 1$, let $P_n(\cdot)$ and $\tilde{\nu}^{(k)}$ denote the restrictions of $P\left(M_n^{-1} \sum_{j=1}^{k} X_{nj} \in \cdot \right)$ and $\nu^{(k)}$ respectively to $\mathbb{R}^d \setminus B_r$, *i.e.*, for $A \subset \mathbb{R}^d$,

$$
\begin{aligned}
P_n(A) &:= P\left(M_n^{-1} \sum_{j=1}^{k} X_{nj} \in A \cap B_r^c \right) \\
\tilde{\nu}^{(k)}(A) &:= \nu^{(k)} \left(A \cap B_r^c \right) .
\end{aligned}
$$

Then, by Lemma 3.2.2, it follows that

$$
\frac{P_n}{P(\|H\| > M_n)^k} \xrightarrow{w} \tilde{\nu}^{(k)} .
$$

By (3.28), it follows that

$$
\frac{P_n(dx)}{P(\|H\| > M_n)^k} P\left(B_n^{-1} \sum_{j=k+1}^{n} X_{nj} \in dy \right) \xrightarrow{w} \tilde{\nu}^{(k)}(dx) P(\mathcal{V} \in dy)
$$

on $\mathbb{R}^d \times \mathbb{R}^d$. Note that

$$
\begin{aligned}
& P(U_n) \\
= & \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}\left(\|x\| > k - \eta, \frac{x}{\|x\|} \in F^{\eta} \right) \mathbf{1}(\langle x, y \rangle \ge -\eta) P_n(dx) \\
& P\left(B_n^{-1} \sum_{j=k+1}^{n} X_{nj} \in dy \right) .
\end{aligned}
$$

Since F^η is a closed set,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{P(U_n)}{P(\|H\| > M_n)^k} \\
\le\;& \int \mathbf{1}\left(\|x\| \ge k-\eta, \frac{x}{\|x\|} \in F^\eta\right) P(\langle x, \mathcal{V}\rangle \ge -\eta)\tilde{\nu}^{(k)}(dx) \\
=\;& \int \mathbf{1}\left(\|x\| \ge k-\eta, \frac{x}{\|x\|} \in F^\eta\right) P(\langle x, \mathcal{V}\rangle \ge -\eta)\nu^{(k)}(dx)\,.
\end{aligned}$$

Letting $\eta \downarrow 0$, we get using the fact that F is a closed set,

$$\begin{aligned}
&\limsup_{\eta\downarrow 0}\limsup_{n\to\infty} \frac{P(U_n)}{P(\|H\| > M_n)^k} \\
\le\;& \int_{\mathbb{R}^d} \mathbf{1}\left(\|x\| \ge k, \frac{x}{\|x\|} \in F\right) P(\langle x, \mathcal{V}\rangle \ge 0)\nu^{(k)}(dx) \\
=\;& \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d} \mathbf{1}\left(\|\sum_{j=1}^k x_j\| \ge k, \frac{\sum_{j=1}^k x_j}{\|\sum_{j=1}^k x_j\|} \in F\right) P\left(\sum_{j=1}^k \langle x_j, \mathcal{V}\rangle \ge 0\right) \\
&\nu(dx_1)\ldots\nu(dx_k) \\
=\;& \int_{\mathcal{S}}\cdots\int_{\mathcal{S}} \mathbf{1}\left(\left\|\sum_{j=1}^k x_j\right\| = k, \frac{\sum_{j=1}^k x_j}{\|\sum_{j=1}^k x_j\|} \in F\right) P\left(\sum_{j=1}^k \langle x_j, \mathcal{V}\rangle \ge 0\right) \\
&\sigma(dx_1)\ldots\sigma(dx_k)\,,
\end{aligned}$$

the last equality being true because $\nu(B_1^c) = 0$ and the restriction of ν to $\mathcal{S}$ is σ. Thus, in order to show (3.32), all that remains is to prove

$$P(V_n) + P(W_n) + P(Y_n) + P(Z_n) \ll P(\|H\| > M_n)^k\,.$$

Note that on the set V_n,

$$\begin{aligned}
k^2M_n^2 \;&<\; \|S_n\|^2 \\
&= \left\|\sum_{j=1}^k X_{nj}\right\|^2 + \left\|\sum_{j=k+1}^n X_{nj}\right\|^2 + 2\left\langle \sum_{j=1}^k X_{nj}, \sum_{j=k+1}^n X_{nj}\right\rangle \\
&\le (k^2+\epsilon_n)M_n^2 + \left\|\sum_{j=k+1}^n X_{nj}\right\|^2 - 2B_n\eta\left\|\sum_{j=1}^k X_{nj}\right\| \\
&\le (k^2+\epsilon_n)M_n^2 + \left\|\sum_{j=k+1}^n X_{nj}\right\|^2 - 2\eta(k-\eta)B_nM_n\,,
\end{aligned}$$

and hence,

$$\begin{aligned}
& P(V_n) \\
\leq\ & P\left(\left\|\sum_{j=1}^{k} X_{nj}\right\| \geq (k-\eta)M_n\right) \\
& \times P\left(\left\|\sum_{j=k+1}^{n} X_{nj}\right\|^2 > 2\eta(k-\eta)B_nM_n - \epsilon_n M_n^2\right) \\
\ll\ & P(\|H\| > M_n)^k \,,
\end{aligned}$$

the last step following from the fact that by the choice of ϵ_n, $\epsilon_n M_n^2 + B_n^2 = o(B_nM_n)$ showing that $2\eta(k-\eta)B_nM_n - \epsilon_n M_n^2$ is much larger than B_n^2 which is the growth rate of $\left\|\sum_{j=k+1}^{n} X_{nj}\right\|^2$. Since for any $u, v \in \mathbb{R}^d$,

$$\begin{aligned}
\left\|\frac{u+v}{\|u+v\|} - \frac{u}{\|u\|}\right\| &\leq \left\|\frac{u+v}{\|u+v\|} - \frac{u}{\|u+v\|}\right\| + \left\|\frac{u}{\|u+v\|} - \frac{u}{\|u\|}\right\| \\
&= \frac{\|v\|}{\|u+v\|} + \left|\frac{\|u+v\| - \|u\|}{\|u+v\|}\right| \\
&\leq 2\frac{\|v\|}{\|u+v\|} \,,
\end{aligned}$$

it follows that

$$\begin{aligned}
P(W_n) &\leq P\left(\left\|\sum_{j=1}^{k} X_{nj}\right\| \geq (k-\eta)M_n\right) P\left(\left\|\sum_{j=k+1}^{n} X_{nj}\right\| > \frac{\eta}{2}M_n\right) \\
&\ll P(\|H\| > M_n)^k \,.
\end{aligned}$$

Clearly,

$$\begin{aligned}
P(Y_n) &\leq \sum_{j=1}^{k} P\left(\|X_{nj}\| > \frac{2k-1-\eta}{2(k-1)}M_n\right) \\
&\leq kP(\|H\| > M_n)P\left(L > \frac{1-\eta}{2(k-1)}M_n\right) \\
&\ll P(\|H\| > M_n)^k \,,
\end{aligned}$$

the last step following by (3.4). Finally,

$$\begin{aligned}
P(Z_n) &\leq kP\left(\|H\| > \frac{1-\eta}{2}M_n\right)^k P\left(L > \left(\frac{\sqrt{k^2+\epsilon_n}}{k} - 1\right)M_n\right) \\
&\ll P(\|H\| > M_n)^k \,,
\end{aligned}$$

the last step being true because by the choice of ϵ_n, it follows that

$$\begin{aligned} 1 &\ll \epsilon_n M_n \\ &= O\left(\left(\frac{\sqrt{k^2+\epsilon_n}}{k}-1\right)M_n\right). \end{aligned}$$

This completes the proof. □

Proof of Theorem 3.2.3. In view of Lemma 3.2.4, it suffices to show that

$$\liminf_{n\to\infty} \frac{P(\|S_n\| > M_n)}{nP(\|H\| > M_n)} \geq \Gamma_1(\mathcal{S}) . \tag{3.33}$$

We assume without loss of generality that $\Gamma_1(\mathcal{S}) > 0$. For $1 \leq j \leq n$, define

$$C_j := \left\{ \|X_{nj}\| \geq M_n, \sum_{1\leq i\leq n, i\neq j} \langle X_{ni}, X_{nj}\rangle > 0 \right\} .$$

Note that

$$P(\|S_n\| > M_n) \geq P\left(\bigcup_{j=1}^n C_j\right) , \tag{3.34}$$

and that

$$\begin{aligned} &P(C_j) \\ = &\int_{\mathcal{S}}\int_{\mathbb{R}^d} \mathbf{1}(\langle x,y\rangle > 0) P\left(\|X_{n1}\| \geq M_n, \frac{X_{n1}}{\|X_{n1}\|} \in dx\right) P\left(B_n^{-1}\sum_{j=2}^n X_{nj} \in dy\right) \\ = &\int_{\mathcal{S}}\int_{\mathbb{R}^d} \mathbf{1}(\langle x,y\rangle > 0) P\left(\|H\| \geq M_n, \frac{H}{\|H\|} \in dx\right) P\left(B_n^{-1}\sum_{j=2}^n X_{nj} \in dy\right) \end{aligned}$$

By (1.3) and (3.28), it follows that

$$\begin{aligned} \liminf_{n\to\infty} \frac{P(C_j)}{P(\|H\| > M_n)} &\geq \int_{\mathcal{S}}\int_{\mathbb{R}^d} \mathbf{1}(\langle x,y\rangle > 0)\sigma(dx)P\left(\mathcal{V} \in dy\right) \\ &= \Gamma_1(\mathcal{S}) , \end{aligned} \tag{3.35}$$

the equality in the last line following from (3.29). In view of (3.34) and (3.35), all that needs to be shown is that

$$n^2 P(C_1 \cap C_2) = o(nP(\|H\| > M_n)) ,$$

but that follows from similar arguments as in the proof of Theorem 3.2.2. This completes the proof. □

Proof of Theorem 3.2.4. In view of Lemma 3.2.4, it suffices to show that if (3.30) and (3.31) hold, then for $k \geq 2$ and $s_1, \ldots, s_r \in \mathcal{S}$,

$$\liminf_{n\to\infty} \frac{P(\|S_n\| > M_n)}{\{nP(\|H\| > M_n)\}^k} \geq \frac{1}{k!}\sum_{i=1}^r P(\langle s_i, \mathcal{V}\rangle \geq 0)\sigma(\{s_i\})^k \,. \tag{3.36}$$

Denote for $1 \leq j_1 < \ldots < j_k \leq n$,

$$C_{j_1\ldots j_k} := \bigcup_{i=1}^r \left\{ \|H_{j_u}\| \geq M_n, \frac{H_{j_u}}{\|H_{j_u}\|} = s_i \text{ for } 1 \leq u \leq k, \sum_{v\neq j_1,\ldots,j_k} \langle s_i, X_{nv}\rangle > 0 \right\} .$$

Note that,

$$P(\|S_n\| > kM_n) \geq P\left(\bigcup C_{j_1\ldots j_k}\right) ,$$

where the union is taken over all tuples $1 \leq j_1 < \ldots < j_k \leq n$. It follows by (3.30) and (3.31) that for any $1 \leq j_1 < \ldots < j_k \leq n$ and $1 \leq i \leq r$,

$$\liminf_{n\to\infty} \frac{P\left(\|H_{j_u}\| \geq M_n, \frac{H_{j_u}}{\|H_{j_u}\|} = s_i \text{ for } 1 \leq u \leq k, \sum_{v\neq j_1,\ldots,j_k} \langle s_i, X_{nv}\rangle > 0\right)}{P(\|H\| > M_n)^k}$$

$$\geq \sigma(\{s_i\})^k P(\langle s_i, \mathcal{V}\rangle \geq 0) ,$$

and hence for $1 \leq j_1 < \ldots < j_k \leq n$,

$$\liminf_{n\to\infty} \frac{P\left(C_{j_1\ldots j_k}\right)}{P(\|H\| > M_n)^k} \geq \sum_{i=1}^r \sigma(\{s_i\})^k P(\langle s_i, \mathcal{V}\rangle \geq 0) .$$

Thus, in order to show (3.36), it suffices to prove that as $n \longrightarrow \infty$,

$$P\left(\bigcup C_{j_1\ldots j_k}\right) \sim \sum P\left(C_{j_1\ldots j_k}\right) ,$$

where the sum and the union are taken over all tuples $1 \leq j_1 < \ldots < j_k \leq n$. That follows from similar arguments leading to the proof of (3.11). This completes the proof. □

3.3 Hard truncation case

For this section, we assume that M_n goes to ∞ slowly enough so that $nP(\|H\| > M_n)$ goes to ∞ as $n \longrightarrow \infty$. Moreover, if $\alpha = 2$, we assume that $E\|H\|^2 < \infty$. We further assume that $Ee^{\epsilon L} < \infty$ for some $\epsilon > 0$.

We shall say that a sequence of random variables Z_n follows the Large Deviations Principle (LDP) with speed a_n and rate function I if for any Borel set A,

$$
\begin{aligned}
-\inf_{x \in int(A)} I(x) &\le \liminf_{n\to\infty} \frac{1}{a_n} \log P(Z_n \in A) \\
&\le \limsup_{n\to\infty} \frac{1}{a_n} \log P(Z_n \in A) \le -\inf_{x \in cl(A)} I(x)\,.
\end{aligned}
$$

The first result of this section is an analogue of Cramér's Theorem because of the following reason. Recall that Cramér's Theorem gives the LDP for $n^{-1}\sum_{i=1}^n Z_i$ where $Z_1, Z_2, \ldots$ are i.i.d. random variables with finite exponential moments. Note that the normalizing constant is n, the rate at which $E\sum_{i=1}^n \|Z_i\|$ grows. The following result gives the LDP for the sequence $S_n/\{nM_nP(\|H\| > M_n)\}$. By the Karamata's Theorem, it is easy to see that if $\alpha < 1$,

$$
E\sum_{i=1}^n \left\| H_i \mathbf{1}\left(\|H_i\| \le M_n\right) + \frac{H_i}{\|H_i\|}(M_n + L_i)\mathbf{1}\left(\|H_i\| > M_n\right) \right\|
$$

grows like $nM_nP(\|H\| > M_n)$ up to a constant, and hence we consider this to be an analogue of Cramér's Theorem, at least for that case. This result, however, is valid for $\alpha < 2$.

Theorem 3.3.1 (Large Deviations ($\alpha < 2$))**.** *$S_n/\{nM_nP(\|H\| > M_n)\}$ follows LDP with speed $nP(\|H\| > M_n)$ and rate function Λ^*, which is the Fenchel-Legendre dual* *(defined in* (1.13)*) of the function* Λ *given by*

$$\Lambda(\lambda) := \begin{cases} \int \left(e^{\langle \lambda, x\rangle} - 1\right) \nu(dx), & 0 < \alpha < 1 \\ \int \left(e^{\langle \lambda, x\rangle} - 1 - \langle \lambda, x\rangle\right) \nu(dx), & \alpha = 1 \\ \int \left(e^{\langle \lambda, x\rangle} - 1 - \langle \lambda, x\rangle\right) \nu(dx) - \frac{1}{\alpha-1} \int_{\mathcal{S}} \langle \lambda, s\rangle \sigma(ds), & 1 < \alpha < 2 \end{cases} .$$

where the measure ν *is as defined in (3.5).*

For the proof, we shall use the following theorem (Theorem 2.3.6 (page 44) in Dembo and Zeitouni (1998)):

Theorem 3.3.2 (Gärtner-Ellis)**.** *If* Y_n *is a sequence of* $\mathbb{R}^d$*-valued random variables with for every* $t \in \mathbb{R}^d$*,*

$$\lim_{n\to\infty} \frac{1}{a_n} \log E e^{\langle t, Y_n\rangle} = \Lambda(t)\,,$$

where $a_n \longrightarrow \infty$*,* $\Lambda(t)$ *is a finite number for every* $t \in \mathbb{R}^d$ *and* Λ *is a differentiable function, then* Y_n/a_n *follows LDP with speed* a_n *and rate function* Λ^* *which is the Fenchel-Legendre dual of* Λ*.*

Proof of Theorem 3.3.1. Define

$$X_n \;:=\; H\mathbf{1}(\|H\| \le M_n) + \frac{H}{\|H\|}(M_n + L)\mathbf{1}(\|H\| > M_n)\,.$$

Since Λ is clearly a differentiable function, using the Gärtner-Ellis theorem, it suffices to show that for all $\lambda \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \frac{1}{P(\|H\| > M_n)} \log E\exp(\langle \lambda, M_n^{-1} X_n\rangle) = \Lambda(\lambda)\,. \tag{3.37}$$

Case: $0 < \alpha < 1$. Note that

$$\begin{aligned} E\exp(\langle \lambda, M_n^{-1} X_n\rangle) &= \int_{\mathbb{R}^d} \exp(\langle \lambda, x\rangle) P(M_n^{-1} X_n \in dx) \\ &= 1 + \int_{\mathbb{R}^d\setminus\{0\}} \left(e^{\langle \lambda, x\rangle} - 1\right) P(M_n^{-1} X_n \in dx)\,. \end{aligned}$$

By Lemma 3.2.1 and the fact that ν charges only $\{x : 0 < \|x\| \le 1\}$, for all $0 < \epsilon < 1$, it follows that

$$\int_{\{\epsilon \le \|x\| \le 3\}} \left(e^{\langle \lambda, x\rangle} - 1\right) P(M_n^{-1} X_n \in dx)$$
$$\sim P(\|H\| > M_n) \int_{\{\|x\| \ge \epsilon\}} \left(e^{\langle \lambda, x\rangle} - 1\right) \nu(dx)\,. \tag{3.38}$$

Since, for $\alpha < 1$, $e^{\langle \lambda, x\rangle} - 1$ is ν-integrable and hence,

$$\lim_{\epsilon \downarrow 0} \int_{\{\|x\| \ge \epsilon\}} \left(e^{\langle \lambda, x\rangle} - 1\right) \nu(dx) = \int \left(e^{\langle \lambda, x\rangle} - 1\right) \nu(dx)\,.$$

Also,

$$\begin{aligned}
& \frac{1}{P(\|H\| > M_n)} \int_{\{\|x\| > 3\}} \left|e^{\langle \lambda, x\rangle} - 1\right| P(M_n^{-1} X_n \in dx) \\
= \;& \frac{1}{P(\|H\| > M_n)} E\left[\left|e^{\langle \lambda, M_n^{-1} X_n\rangle} - 1\right| \mathbf{1}(\|M_n^{-1} X_n\| > 3)\right] \\
\le \;& \frac{1}{P(\|H\| > M_n)} E\left[\exp\left(\langle \lambda, M_n^{-1} X_n\rangle\right) \mathbf{1}(\|M_n^{-1} X_n\| > 3)\right] + P(L > 2M_n)\,.
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
& \frac{1}{P(\|H\| > M_n)} E\left[\exp\left(\langle \lambda, M_n^{-1} X_n\rangle\right) \mathbf{1}(\|M_n^{-1} X_n\| > 3)\right] \\
\le \;& \left[E \exp\left(\langle 2\lambda, M_n^{-1} X_n\rangle\right)\right]^{1/2} \frac{P(\|X_n\| > 3M_n)^{1/2}}{P(\|H\| > M_n)} \\
\le \;& \left[E \exp\left(2M_n^{-1} \|\lambda\| \|X_n\|\right)\right]^{1/2} \frac{P(\|X_n\| > 3M_n)^{1/2}}{P(\|H\| > M_n)}\,.
\end{aligned}$$

Choose n large enough so that $M_n > \max(1, 2\|\lambda\|/\epsilon)$ where ϵ is such that $Ee^{\epsilon L} < \infty$. Also, observe that

$$M_n^{-1} \|X_n\| \le (2 + M_n^{-1} L)\,.$$

Thus,

$$E \exp\left(2M_n^{-1} \|\lambda\| \|X_n\|\right) \le \exp(4C\|\lambda\|) Ee^{\epsilon L} < \infty\,,$$

while,

$$\frac{P(\|X_n\| > 3M_n)^{1/2}}{P(\|H\| > M_n)} = \frac{P(L > 2M_n)^{1/2}}{P(\|H\| > M_n)^{1/2}} \le \frac{e^{-\epsilon M_n}}{P(\|H\| > M_n)^{1/2}} \left[Ee^{\epsilon L}\right]^{1/2} \longrightarrow 0\,.$$

This shows

$$\lim_{n\longrightarrow\infty}\frac{1}{P(\|H\|>M_n)}\int_{\{\|x\|>3\}}\left|e^{\langle\lambda,x\rangle}-1\right|P(M_n^{-1}X_n\in dx)=0\,. \tag{3.39}$$

By Karamata's theorem and the fact that $e^{\langle\lambda,x\rangle}=1+O(\|x\|)$, one can show that there is $C<\infty$ so that,

$$\limsup_{n\to\infty}\frac{1}{P(\|H\|>M_n)}\int_{\{\|x\|<\epsilon\}}\left|e^{\langle\lambda,x\rangle}-1\right|P(M_n^{-1}X_n\in dx)\le C\epsilon^{1-\alpha}\,,$$

thus proving that

$$\lim_{\epsilon\downarrow 0}\limsup_{n\longrightarrow\infty}\frac{1}{P(\|H\|>M_n)}\int_{\{\|x\|<\epsilon\}}\left|e^{\langle\lambda,x\rangle}-1\right|P(M_n^{-1}X_n\in dx)=0\,. \tag{3.40}$$

Clearly, (3.38), (3.39) and (3.40) show (3.37) and hence complete the proof for the case $\alpha<1$.

For the case $\alpha=1$, note that

$$\begin{aligned}E\exp(\langle\lambda,M_n^{-1}X_n\rangle) &= \int_{\mathbb{R}^d}\exp(\langle\lambda,x\rangle)P(M_n^{-1}X_n\in dx)\\ &= 1+\int_{\mathbb{R}^d\setminus\{0\}}\left(e^{\langle\lambda,x\rangle}-1\right)P(M_n^{-1}X_n\in dx)\\ &= 1+\int_{\mathbb{R}^d\setminus\{0\}}\left(e^{\langle\lambda,x\rangle}-1-\langle\lambda,x\rangle\right)P(M_n^{-1}X_n\in dx)\end{aligned}$$

where the last equality follows from the fact that when $\alpha=1$, H (and hence X_n) has a symmetric distribution. Note that $\alpha=1$ implies that

$$\int\left|e^{\langle\lambda,x\rangle}-1-\langle\lambda,x\rangle\right|\nu(dx)<\infty\,. \tag{3.41}$$

By arguments similar to those for the case $\alpha<1$, it follows that as $n\longrightarrow\infty$,

$$\begin{aligned}&\int_{\mathbb{R}^d\setminus\{0\}}\left(e^{\langle\lambda,x\rangle}-1-\langle\lambda,x\rangle\right)P(M_n^{-1}X_n\in dx)\\ &\sim P(\|H\|>M_n)\int\left(e^{\langle\lambda,x\rangle}-1-\langle\lambda,x\rangle\right)\nu(dx)\,.\end{aligned} \tag{3.42}$$

This completes the proof for the case $\alpha=1$.

For the case $1 < \alpha < 2$, note that

$$E \exp(\langle \lambda, M_n^{-1} X_n \rangle)$$

$$= 1 + \int_{\mathbb{R}^d \setminus \{0\}} \left(e^{\langle \lambda, x \rangle} - 1 - \langle \lambda, x \rangle \right) P(M_n^{-1} X_n \in dx) + \int \langle \lambda, x \rangle P(M_n^{-1} X_n \in dx) \,.$$

For this case also, (3.41) clearly holds and similar arguments as those for the case $\alpha < 1$ show (3.42). Thus, all that needs to be shown is as $n \longrightarrow \infty$,

$$\int \langle \lambda, x \rangle P(M_n^{-1} X_n \in dx) \sim -\frac{1}{\alpha - 1} P(\|H\| > M_n) \int_{\mathcal{S}} \langle \lambda, s \rangle \sigma(ds) \,. \tag{3.43}$$

For this, note that by the assumption that $EH = 0$,

$$\begin{aligned}
& \int \langle \lambda, x \rangle P(M_n^{-1} X_n \in dx) \\
= \;& -M_n^{-1} \int_{\{\|x\| > M_n\}} \langle \lambda, x \rangle P(H \in dx) \\
& + \left\{ 1 + M_n^{-1} E(L) \right\} \int_{\mathcal{S}} \langle \lambda, s \rangle P\left(\frac{H}{\|H\|} \in ds, \|H\| > M_n \right) \\
= \;& -M_n^{-1} \int_{M_n}^{\infty} \left\{ \int_{\mathcal{S}} \langle \lambda, s \rangle P\left(\|H\| > r, \frac{H}{\|H\|} \in ds \right) \right\} dr \\
& + M_n^{-1} E(L) \int_{\mathcal{S}} \langle \lambda, s \rangle P\left(\frac{H}{\|H\|} \in ds, \|H\| > M_n \right) \\
= \;& -M_n^{-1} \int_{M_n}^{\infty} \left\{ \int_{\mathcal{S}} \langle \lambda, s \rangle P\left(\|H\| > r, \frac{H}{\|H\|} \in ds \right) \right\} dr \\
& + o(P(\|H\| > M_n)) \,,
\end{aligned}$$

as $n \longrightarrow \infty$. By (1.3), it follows that

$$\begin{aligned}
& \int_{M_n}^{\infty} \left\{ \int_{\mathcal{S}} \langle \lambda, s \rangle P\left(\|H\| > r, \frac{H}{\|H\|} \in ds \right) \right\} dr \\
\sim \;& \int_{M_n}^{\infty} \left\{ \int_{\mathcal{S}} \langle \lambda, s \rangle \sigma(ds) \right\} P(\|H\| > r) dr \\
\sim \;& \frac{1}{\alpha - 1} M_n P(\|H\| > M_n) \int_{\mathcal{S}} \langle \lambda, s \rangle \sigma(ds) \,,
\end{aligned}$$

where the second equivalence follows from the Karamata's Theorem. This shows (3.43) and thus completes the proof. □

Similar calculations as above, for the case $\alpha \geq 2$, will show that $S_n/(nM_n^{-1})$ follows LDP with speed nM_n^{-2} and rate function that is the Fenchel-Legendre dual of $\frac{1}{2}\langle \lambda, D\lambda \rangle$, D being the dispersion matrix of H. This is, however, covered in much more generality in Theorem 3.3.4 below, and hence we chose not to include this case in Theorem 3.3.1. That the statement of Theorem 3.3.4 implies the above, is clear because when $\alpha \geq 2$, it's easy to see that $n^{1/2} \ll nM_n^{-1} \ll n$ and that

$$\begin{aligned} ES_n &= O\left(nM_n P(\|H\| > M_n)\right) \\ &= o(nM_n^{-1}) \,. \end{aligned}$$

Cramér's Theorem deals with $n^{-1}\sum_{i=1}^n Z_i$ where $Z_1, Z_2, \ldots$ are i.i.d. random variables. On a finer scale, $n^{-1/2}\sum_{i=1}^n [Z_i - E(Z_i)]$ possesses a limiting Normal distribution by the central limit theorem. For $\beta \in (1/2, 1)$, the renormalized quantity $n^{-\beta}\sum_{i=1}^n [Z_i - E(Z_i)]$ satisfies an LDP but always with a quadratic rate function. The precise statement for this is the following result, known as moderate deviations. See Theorem 3.7.1 in Dembo and Zeitouni (1998).

Theorem 3.3.3. *Let $Z_1, Z_2, \ldots$ be i.i.d. $\mathbb{R}^d$-valued random variables with finite exponential moments in a ball around the origin. Let the covariance matrix C of Z_1 be invertible. Then for $n^{1/2} \ll a_n \ll n$, $\sqrt{a_n/n}\sum_{i=1}^n [Z_i - E(Z_i)]$ follows LDP with rate function $\frac{1}{2}\langle x, C^{-1}x \rangle$.*

The last result of this chapter is an analogue of the above result, in the setting of truncated heavy-tailed random variables.

Theorem 3.3.4 (Moderate Deviations)**.** *Suppose the sequence a_n satisfies*

$$n^{1/2} M_n P(\|H\| > M_n)^{1/2} \ll a_n \ll nM_n P(\|H\| > M_n), \text{ if } \alpha < 2$$

and

$$n^{1/2} \ll a_n \ll n, \text{ if } \alpha \geq 2 .$$

Then, $a_n^{-1}(S_n - ES_n)$ follows LDP with speed β_n and rate Λ^* where

$$\beta_n := \begin{cases} \frac{a_n^2}{nM_n^2 P(\|H\|>M_n)}, & \text{if } \alpha < 2 \\ \frac{a_n^2}{n}, & \text{if } \alpha \geq 2 , \end{cases}$$

and

$$\Lambda(\lambda) := \frac{1}{2}\langle \lambda, D\lambda \rangle .$$

Here, D is the $d \times d$ matrix with

$$D_{ij} := \frac{2}{2-\alpha} \int_{\mathcal{S}} s_i s_j \sigma(ds)$$

if $\alpha < 2$ and the dispersion matrix of H is $\alpha \geq 2$, which is well defined even when $\alpha = 2$ because it has been assumed in that case, that $E\|H\|^2 < \infty$. If, in addition, D is invertible, then Λ^* is given by

$$\Lambda^*(x) = \frac{1}{2}\langle x, D^{-1}x \rangle .$$

Proof. It is easy to see that $\beta_n \longrightarrow \infty$ as $n \longrightarrow \infty$. Thus, in view of Theorem 3.3.2, it suffices to show that for all $\lambda \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \beta_n^{-1} \log E \exp\left(\langle \lambda, (M_n b_n)^{-1}(S_n - ES_n)\rangle\right) = \frac{1}{2}\langle \lambda, D\lambda \rangle , \tag{3.44}$$

where

$$b_n := \begin{cases} nM_n P(\|H\| > M_n)/a_n, & \alpha < 2 \\ n/(a_n M_n), & \alpha \geq 2 . \end{cases}$$

The assumptions on a_n in the hypotheses imply that

$$1 \ll b_n \ll \{nP(\|H\| > M_n)\}^{1/2}, \text{ if } \alpha < 2$$

and

$$M_n^{-1} \ll b_n \ll M_n^{-1} n^{1/2}, \text{ if } \alpha \geq 2 .$$

Define

$$X_n \quad := \quad H\mathbf{1}(\|H\| \le M_n) + \frac{H}{\|H\|}(M_n + L)\mathbf{1}(\|H\| > M_n)\,.$$

Let ξ_n be defined by

$$\exp(\langle \lambda, (b_n M_n)^{-1}(X_n - EX_n)\rangle)$$

$$= 1 + (b_n M_n)^{-1}\langle \lambda, X_n - EX_n\rangle + \frac{1}{2}(b_n M_n)^{-2}\langle \lambda, (X_n - EX_n)(X_n - EX_n)^T\lambda\rangle + \xi_n\,.$$

Note that

$$\begin{aligned} & E\exp(\langle \lambda, (b_n M_n)^{-1}(X_n - EX_n)\rangle) \\ = \; & 1 + \frac{1}{2}(b_n M_n)^{-2}\langle \lambda, \mathcal{D}(X_n)\lambda\rangle + E\xi_n \\ = \; & 1 + \frac{1}{2}\gamma_n\langle \lambda, D\lambda\rangle(1 + o(1)) + E\xi_n\,, \end{aligned} \tag{3.45}$$

where

$$\gamma_n := \begin{cases} b_n^{-2}P(\|H\| > M_n), & \alpha < 2 \\ b_n^{-2}M_n^{-2}, & \alpha \ge 2\,. \end{cases}$$

To see how (3.45) follows, we first consider the case $\alpha \ge 2$. For any $\mathbb{R}^d$ valued random variable Y, denote the coordinates by $Y^{(1)}, \ldots, Y^{(d)}$ respectively. Note that for $1 \le j, k \le d$,

$$\begin{aligned} & E\left[X_n^{(j)}X_n^{(k)}\right] \\ = \; & E\left[H^{(j)}H^{(k)}\mathbf{1}(\|H\| \le M_n)\right] + O(M_n^2 P(\|H\| > M_n)) \\ = \; & E\left[H^{(j)}H^{(k)}\right](1 + o(1))\,, \end{aligned}$$

in the last equality we used the fact that if $\alpha = 2$, then by assumption $E\|H\|^2 < \infty$ and hence $P(\|H\| > x) = o(x^{-2})$ as $x \longrightarrow \infty$. Similarly, it will follow that for $1 \le j \le d$,

$$\lim_{n\to\infty} E[X_n^{(j)}] = E\left[H^{(j)}\right]\,,$$

and hence, for $1 \le j, k \le d$,

$$\lim_{n\to\infty} \mathrm{Cov}(X_n^{(j)}, X_n^{(k)}) = \mathrm{Cov}(H^{(j)}, H^{(k)})\,,$$

and thus justifying (3.45) in the case $\alpha \ge 2$. For the case $\alpha < 2$, in the proof of Theorem 2.2.5, it has been shown that as $n \longrightarrow \infty$,

$$\mathrm{Var}(\langle \lambda, X_n \rangle) \sim M_n^2 P(\|H\| > M_n) \frac{2}{2-\alpha} \int_{\mathcal{S}} \langle \lambda, s \rangle^2 \sigma(ds)\,,$$

see (2.10). Since the left hand side is same as $\langle \lambda, \mathcal{D}(X_n)\lambda \rangle$, it follows that

$$\begin{aligned}
(b_n M_n)^{-2} \langle \lambda, \mathcal{D}(X_n)\lambda \rangle &\sim b_n^{-2} P(\|H\| > M_n) \frac{2}{2-\alpha} \int_{\mathcal{S}} \langle \lambda, s \rangle^2 \sigma(ds) \\
&= \gamma_n \frac{2}{2-\alpha} \int_{\mathcal{S}} \langle \lambda, s \rangle^2 \sigma(ds) \\
&= \gamma_n \sum_{j=1}^{d} \sum_{k=1}^{d} \lambda_j \lambda_k \frac{2}{2-\alpha} \int_{\mathcal{S}} s_j s_k \sigma(ds) \\
&= \gamma_n \langle \lambda, D\lambda \rangle\,,
\end{aligned}$$

where for $x \in \mathbb{R}^d$, the coordinates are denoted by $x_1, \ldots, x_d$ respectively. This shows (3.45) for the case $\alpha < 2$.

Clearly, $n\gamma_n = \beta_n$, and hence all that needs to be shown for (3.44) is $E\xi_n = o(\gamma_n)$ as $n \longrightarrow \infty$. By Taylor's Theorem, there exists $C < \infty$ so that

$$\begin{aligned}
|\xi_n| &\le C(b_n M_n)^{-3} \|X_n - EX_n\|^3 \exp\left\{C(b_n M_n)^{-1} \|X_n - EX_n\|\right\} \\
&\le C(b_n M_n)^{-3} \|X_n - EX_n\|^3 \exp\left\{C b_n^{-1} \left(4 + \frac{L + E(L)}{M_n}\right)\right\} \\
&\le 8C(b_n M_n)^{-3} \left(\|X_n\|^3 + \|EX_n\|^3\right) \exp\left\{C b_n^{-1} \left(4 + \frac{L + E(L)}{M_n}\right)\right\}\,,
\end{aligned}$$

the last inequality following from the fact that $(A+B)^3 \le 8(A^3 + B^3)$ for all $A, B \ge 0$. Thus,

$$E|\xi_n| = O\left((b_n M_n)^{-3} E\left[\left(\|X_n\|^3 + \|EX_n\|^3\right) \exp(CL/b_n M_n)\right]\right)\,.$$

Note that

$$
\begin{aligned}
& E\left[\|X_n\|^3 \exp(CL/b_nM_n)\right] \\
= \; & E\left[\|H\|^3\mathbf{1}(\|H\| \le M_n)\right] E\left[\exp(CL/b_nM_n)\right] \\
& +P(\|H\| > M_n)E\left[(M_n+L)^3\exp(CL/b_nM_n)\right] \\
= \; & O(1)E\left[\|H\|^3\mathbf{1}(\|H\| \le M_n)\right] + O\left(M_n^3P(\|H\| > M_n)\right) \,.
\end{aligned}
$$

Also,

$$
\begin{aligned}
& \|EX_n\|^3E\left[\exp(CL/b_nM_n)\right] \\
= \; & O(1)\|EX_n\|^3 \\
= \; & O(E(\|X_n\|^3)) \\
= \; & O\left(E\left[\|H\|^3\mathbf{1}(\|H\| \le M_n)\right] + M_n^3P(\|H\| > M_n)\right) \,,
\end{aligned}
$$

the last step following by similar calculations as above. Thus,

$$
E\xi_n = O\left\{(b_nM_n)^{-3}\left(E\left[\|H\|^3\mathbf{1}(\|H\| \le M_n)\right] + M_n^3P(\|H\| > M_n)\right)\right\} \,. \tag{3.46}
$$

If $\alpha < 3$, then by the Karamata's Theorem,

$$
E\left[\|H\|^3\mathbf{1}(\|H\| \le M_n)\right] = O(M_n^3P(\|H\| > M_n)) \,.
$$

It is not hard to see that for all α,

$$
M_n^3P(\|H\| > M_n) = o(b_n^3M_n^3\gamma_n) \,, \tag{3.47}
$$

where we have used the fact that if $\alpha = 2$, then $P(\|H\| > x) = O(x^{-2})$ as $x \longrightarrow \infty$, which follows from the assumption that $E\|H\|^2 < \infty$ in that case. Hence by (3.46), it follows that $E\xi_n = o(\gamma_n)$ for the case $\alpha < 3$. If $\alpha \ge 3$, then

$$
E\left[\|H\|^3\mathbf{1}(\|H\| \le M_n)\right] = o(M_n) = o(b_n^3M_n^3\gamma_n) \,.
$$

Using (3.46) and (3.47), this proves that $E\xi_n = o(\gamma_n)$ for the case $\alpha \ge 3$, and thus completes the proof. □

3.4 Conclusion

The idea behind the investigation of the large deviation principle in the soft truncation regime is clearly similar to that in the untruncated situation considered in Hult et al. (2005), for example. On the other hand, in the hard truncation regime, the large deviation behavior is via the Gärtner-Ellis Theorem, which is the basis of the Cramér's Theorem and the Moderate Deviations result for random variables with finite exponential moments. Thus, from the point of view of large deviation principle also, the conclusion of Chapter 2 holds. That is, when the power tails are truncated softly, the model resembles one with untruncated power tail, while when the same is truncated hard, the model resembles one whose tails decay exponentially fast.

CHAPTER 4

APPLICATION 1: TESTING FOR SOFT AND HARD TRUNCATION

4.1 Introduction

Chapters 2 and 3 of this thesis provide, among other things, evidence that, in certain important respects, random variables with truncated heavy tails retain "most of the tail heaviness" if the truncation is soft, but loose "much of the tail heaviness" if the truncation is hard. Since the truncation level is not observed, how does one decide if the tails of observed data have been truncated softly or hard? In this chapter we construct statistical tests for testing each of the two hypothesis against the corresponding alternative. We restrict ourselves to the case of one-dimensional observations. This is no loss of generality because of the following reason. Our interest lies primarily in deciding the truncation regime which is stated in terms of the norm of the heavy-tailed random variable; see (1.5). Hence, even if the observations are from a higher dimensional space, one can look at the norm of the observations and thus bring it to this setting. In a subsequent chapter, we shall illustrate our methods by applying them to two data sets arising from internet traffic.

The formal setup in this chapter is as follows. We are given a sample $X_1, \ldots, X_n$ of **one-dimensional observations** from the model with truncated power tails (1.4), *i.e.*,

$$X_j := H_j \mathbf{1}\left(|H_j| \le M_n\right) + \frac{H_j}{|H_j|}(M_n + L_j)\mathbf{1}\left(|H_j| > M_n\right) , \tag{4.1}$$

where $H, H_1, H_2, \ldots$ are assumed to be i.i.d. $\mathbb{R}$-valued random variables that have power tails with exponent α. We emphasize a slight change in notation from (1.4): whereas the latter used the notation $X_{n1}, \ldots, X_{nn}$ to emphasize the triangular array nature of the model, in a statistical procedure, when a single sample (*i.e.*, a particular row of the triangular array) is given, the notation $X_1, \ldots, X_n$ is more natural. In this chapter, α can take any positive value. Neither the precise value of α, nor the exact distribution of the random variables L_j are assumed to be known. However, an upper bound on α is assumed to be known.

This chapter is split into three sections, describing, correspondingly, testing the hypothesis of soft truncation, testing the hypothesis of hard truncation, and testing a slightly stronger version of the latter.

4.2 Testing the hypothesis of soft truncation

We consider the following problem of testing a null hypothesis against a simple alternative:

$$\left.\begin{array}{ll} H_0: & P(|H_1| > M) \ll n^{-1} \quad \text{(soft truncation)} \\ H_1: & P(|H_1| > M) \gg n^{-1} \quad \text{(hard truncation)} \end{array}\right\}. \tag{4.2}$$

We assume the tail exponent α satisfies

$$\alpha < A < \infty\,, \tag{4.3}$$

i.e. an upper bound on the tail exponent is available. As a test statistic we will use

$$Z_n(A) := \frac{\sum_{i=1}^n |X_i|^A}{\max_{1\le i\le n} |X_i|^A}\,. \tag{4.4}$$

The following proposition describes the asymptotic distribution of $Z_n(A)$ under the null hypothesis and under the alternative.

Proposition 4.2.1. *(i) Under the hypothesis H_0 of soft truncation,*

$$Z_n(A) \Rightarrow \Gamma_1^{A/\alpha} \sum_{j=1}^{\infty} \Gamma_j^{-A/\alpha}\,, \tag{4.5}$$

where $(\Gamma_j,\ j \geq 1)$ are the arrival times of a unit rate Poisson process on $(0, \infty)$.

(ii) Assume that $EL_1^A < \infty$. Then under the hypothesis H_1 of hard truncation, $Z_n(A) \xrightarrow{P} \infty$.

Proof. For part (i), we define

$$b_n = \inf\bigl\{x > 0 :\ P(|H_1|^A > x) \leq n^{-1}\bigr\},\ n = 1, 2, \ldots .$$

Note that, for any $x > 0$,

$$nP\bigl(b_n^{-1}|X_1|^A > x\bigr) \sim nP\bigl(b_n^{-1}|H_1|^A > x\bigr) \to x^{-\alpha}$$

as $n \to \infty$. It follows from Proposition 3.21 (page 154) in Resnick (1987) that we have the following weak convergence of a sequence of point processes on $(0, \infty]$:

$$N_n := \sum_{j=1}^{n} \delta_{b_n^{-1}|X_1|^A} \Rightarrow N := \sum_{j=1}^{\infty} \delta_{\Gamma_j^{-A/\alpha}} \tag{4.6}$$

as $n \to \infty$. Here δ_a is a point mass at a, and the weak convergence takes place in the space of Radon point measures on $(0, \infty]$ endowed with the topology of vague convergence; see Section 3.4 in Resnick (1987). We would like to use the continuous mapping theorem to deduce (4.5) from (4.6), but a preliminary truncation step is necessary.

For $\varepsilon > 0$ we define

$$Z_n(A; \varepsilon) := \frac{\sum_{i=1}^n |X_i|^A \mathbf{1}\bigl(b_n^{-1}|X_i|^A > \varepsilon\bigr)}{\max_{1\leq i\leq n} |X_i|^A}\,.$$

Notice that $Z_n(A;\varepsilon) = h(N_n)$, where for a Radon point measure $\eta = \sum_j \delta_{r_j}$ on $(0,\infty]$,

$$h(\eta) = \frac{\eta\big((\varepsilon,\infty]\big)}{\max_j r_j} \,.$$

It is standard (and easy) to check that h is continuous with probability 1 at the Poisson random measure N in (4.6), so by the continuous mapping theorem,

$$Z_n(A;\varepsilon) \Rightarrow \Gamma_1^{A/\alpha} \sum_{j=1}^{\infty} \Gamma_j^{-A/\alpha} \mathbf{1}\big(\Gamma_j^{-A/\alpha} > \varepsilon\big) \,.$$

Therefore, the convergence (4.5) will follow once we check that for every $\delta > 0$,

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} P\big(Z_n(A) - Z_n(A;\varepsilon) > \delta\big) = 0 \,. \tag{4.7}$$

To this end, notice that, for any $0 < \theta < 1$ we can select $\tau > 0$ so small that $P\big(\max_{1\le i\le n} |X_i|^A \le \tau b_n\big) \le \theta$ for all n large enough. Then, for all n large enough,

$$\begin{aligned}
P\big(Z_n(A) - Z_n(A;\varepsilon) > \delta\big) &\le \theta + \delta^{-1} E\left(\tau^{-1} b_n^{-1} \sum_{i=1}^{n} |X_i|^A \mathbf{1}\big(b_n^{-1}|X_i|^A \le \varepsilon\big)\right) \\
&= \theta + \delta^{-1}\tau^{-1} n b_n^{-1} E\Big(|X_1|^A \mathbf{1}\big(b_n^{-1}|X_1|^A \le \varepsilon\big)\Big) \\
&= \theta + \delta^{-1}\tau^{-1} n b_n^{-1} E\Big(|H_1|^A \mathbf{1}\big(b_n^{-1}|H_1|^A \le \varepsilon\big)\Big) \\
&\sim \theta + \delta^{-1}\tau^{-1} n b_n^{-1} \Big((1-\alpha/A)^{-1}(\varepsilon b_n) P\big(|H_1|^A > \varepsilon b_n\big) \\
&\sim \theta + \delta^{-1}\tau^{-1} n b_n^{-1} (1-\alpha/A)^{-1}(\varepsilon b_n)\big(\varepsilon^{-\alpha/A} n^{-1}\big) \\
&= \theta + \delta^{-1}\tau^{-1}(1-\alpha/A)^{-1}\varepsilon^{1-\alpha/A}.
\end{aligned}$$

where the second equality holds because of soft truncation, and the first asymptotic equivalence follows from the Karamata theorem. Since $A > \alpha$, we obtain (4.7) by first letting $\varepsilon \to 0$ and then $\theta \to 0$. This completes the proof of part (i).

For part (ii), we start with observing that

$$\frac{\sum_{i=1}^{n} |X_i|^A}{n M_n^A P(|H_1| > M_n)} \ge \frac{\sum_{i=1}^{n} |H_i|^A \mathbf{1}\big(M_n/2 \le |H_i| \le M_n\big)}{n M_n^A P(|H_1| > M_n)} \tag{4.8}$$

$$\geq (M_n/2)^A \frac{\sum_{i=1}^n \mathbf{1}\big(M_n/2 \leq |H_i| \leq M_n\big)}{nM_n^A P(|H_1| > M_n)} \sim 2^{-A}(2^\alpha - 1)$$

in probability. On the other hand, for some constant $c > 0$, by the assumption $EL_1^A < \infty$,

$$\max_{1\leq i\leq n} |X_i|^A \leq c\big(M_n^A + \max_{1\leq j\leq n} L_j^A\big) = cM_n^A + o(1)n$$

a.s. as $n \to \infty$. Since the truncation is hard, and $A > \alpha$, we see that

$$\frac{\max_{i=1,\dots,n} |X_i|^A}{nM_n^A P(|H_1| > M_n)} \to 0 \tag{4.9}$$

a.s. as $n \to \infty$ as well. The claim of part (ii) follows from (4.8) and (4.9). □

Based on Proposition 4.2.1, we suggest the following test for the problem (4.2).

$$\text{reject } H_0 \text{ at significance level } p \in (0,1) \text{ if } \; Z_n(A) > c_p(\alpha/A)\,, \tag{4.10}$$

with $c_p(\theta)$ such that $P(Z(\theta) > c_p(\theta)) = p$, where for $0 < \theta < 1$,

$$Z(\theta) = \Gamma_1^{1/\theta} \sum_{j=1}^\infty \Gamma_j^{-1/\theta}\,. \tag{4.11}$$

The random variable $Z(\theta)$ does not seem to have one of the standard distributions, and we are not aware of any previous studies of the distribution of $Z(\theta)$. The following proposition lists some of the properties of this distribution.

Proposition 4.2.2. *The random variable $Z(\theta)$ is an infinitely divisible random variable. It has a density with respect to the Lebesgue measure, and the Laplace transform*

$$Ee^{-\gamma Z(\theta)} = \left(1 + \gamma e^\gamma \int_0^1 e^{-\gamma x} x^{-\theta}\, dx\right)^{-1}, \tag{4.12}$$

$\gamma > \gamma_0$, where $\gamma_0 < 0$ is the number satisfying

$$1 + \gamma_0 e^{\gamma_0} \int_0^1 e^{-\gamma_0 x} x^{-\theta}\, dx = 0\,.$$

Proof. For $\delta > 0$ let

$$W_\delta = \sum_{j=1}^{\infty} \big(\delta + \Gamma_j\big)^{-1/\theta} .$$

Then W_δ is an infinitely divisible random variable with the Laplace transform

$$Ee^{-\gamma W_\delta} = \exp\left\{ - \int_0^{\delta^{-1/\theta}} \big(1 - e^{-\gamma y}\big) \theta y^{-(1+\theta)} \, dy \right\} \tag{4.13}$$

for all $\gamma \in \mathbb{R}$ because the Lévy measure of W_δ has a compact support; see Rosiński (1990) and Sato (1999). Since

$$Z(\theta) \stackrel{d}{=} 1 + T^{1/\theta} W_T$$

where T is a standard exponential random variable independent of $\big(\Gamma_j : j \geq 1\big)$, it follows that

$$\begin{aligned} Ee^{-\gamma Z(\theta)} &= \int_0^\infty e^{-t} e^{-\gamma} E e^{-\gamma t^{1/\theta} W_t} \, dt \qquad (4.14) \\ &= \int_0^\infty e^{-t} e^{-\gamma} \exp\left\{ - \int_0^{t^{-1/\theta}} (1 - e^{-\gamma t^{1/\theta} y}) \theta y^{-(1+\theta)} \, dy \right\} dt \\ &= e^{-\gamma} \int_0^\infty e^{-t} \exp\left\{ -t \int_0^1 \big(1 - e^{-\gamma x}\big) \theta x^{-(1+\theta)} \, dx \right\} dt \\ &= e^{-\gamma} \int_0^\infty e^{-t} \exp\left[-t \left\{ -(1 - e^{-\gamma}) + \gamma \int_0^1 x^{-\theta} e^{-\gamma x} \, dx \right\} \right] dt \\ &= e^{-\gamma} \int_0^\infty \exp\left[-t \left\{ e^{-\gamma} + \gamma \int_0^1 e^{-\gamma x} x^{-\theta} \, dx \right\} \right] dt \, . \end{aligned}$$

Since the exponent under the integral is positive if and only if $\gamma > \gamma_0$, we obtain (4.12). Additionally, it follows from (4.14) that

$$Z(\theta) \stackrel{d}{=} 1 + Y(T) \, , \tag{4.15}$$

where $\big(Y(t), \, t \geq 0\big)$ is a subordinator satisfying

$$Ee^{-\gamma Y(t)} = \exp\left\{ -t \int_0^1 \big(1 - e^{-\gamma x}\big) \theta x^{-(1+\theta)} \, dx \right\}, \ \ t \geq 0 \, , \tag{4.16}$$

independent of T. Since a Lévy process stopped at an independent infinitely divisible random time is, obviously, infinitely divisible, so is $Z(\theta)$. Furthermore, the characteristic function of $Y(t)$ is integrable on the real line for every $t > 0$, so each $Y(t)$ has a density, and then the same is true for any mixture of $(Y(t))$. Therefore, $Z(\theta)$ has a density. $\square$

Even though we know, by Proposition 4.2.2, that the random variable $Z(\theta)$ has a density, at present we do not know ways to compute this density. One possibility is to estimate the critical values $c_p(\alpha/A)$ to perform the test (4.10). For values of α not too close to the upper bound A (or, equivalently, for the values of θ not too close to 1), it is possible to estimate the critical values $c_p(\theta)$ by the Monte-Carlo method. However, the infinite series has to be truncated at a sufficiently large finite number of terms. The following Proposition shows that a conservative upper bound on the quantile can be obtained even by the truncated series.

Proposition 4.2.3. *Let for $N \geq 1$,*

$$T_N := \Gamma_1^{1/\theta} \sum_{j=1}^{N} \Gamma_j^{-1/\theta} ,$$

where $0 < \theta < 1$. Let C be such that

$$P(T_N > C) = \gamma$$

where $0 < \gamma < 1$. Then,

$$\gamma < P(Z(\theta) > C) ,$$

and for all $\delta > 0$ and $0 < r < 1$,

$$P(Z(\theta) > C + \delta) < \gamma + O\left(e^{-rN^{\lambda}}\right) ,$$

where

$$\lambda := \frac{1-\theta}{1+\theta} .$$

Proof. The first inequality is trivially true. For the second one, note that

$$\begin{aligned}P(Z(\theta) > C+\delta) &\le P(T_N > C) + P\left(\Gamma_1^{1/\theta}\sum_{j=N+1}^{\infty}\Gamma_j^{-1/\theta} > \delta\right)\\ &\le \gamma + P\left(\Gamma_1 > N^\lambda\right) + P\left(\Gamma_N \le N/2\right)\\ &\quad + P\left(N^{\lambda/\theta}\sum_{j=N+1}^{\infty}\left(N/2+\Gamma_j-\Gamma_N\right)^{-1/\theta} > \delta\right)\end{aligned}$$

Clearly for all $r\in(0,1)$,

$$P\left(\Gamma_1 > N^\lambda\right) \le \exp(-rN^\lambda)E\left[\exp(r\Gamma_1)\right] = O\left(\exp(-rN^\lambda)\right).$$

Also,

$$\begin{aligned}P\left(\Gamma_N \le N/2\right) &\le e^{N/2}E\left(e^{-\Gamma_N}\right)\\ &= e^{N/2}\left[E\left(e^{-\Gamma_1}\right)\right]^N\\ &= e^{N/2}2^{-N}\\ &= \exp\left\{-(\log 2 - 1/2)N\right\}\\ &= o\left(e^{-rN^\lambda}\right).\end{aligned}$$

Since

$$\left(\Gamma_j - \Gamma_N : j \ge N+1\right) \stackrel{d}{=} \left(\Gamma_j : j \ge 1\right),$$

it follows that

$$\begin{aligned}&P\left(N^{\lambda/\theta}\sum_{j=N+1}^{\infty}\left(N/2+\Gamma_j-\Gamma_N\right)^{-1/\theta} > \delta\right)\\ =\ & P\left(N^{\lambda/\theta}\sum_{j=1}^{\infty}\left(N/2+\Gamma_j\right)^{-1/\theta} > \delta\right)\\ =\ & P\left(\delta^{-1}N^u\sum_{j=1}^{\infty}\left(N/2+\Gamma_j\right)^{-1/\theta} > N^{u-\lambda/\theta}\right)\\ \le\ & \exp\left(-N^{u-\lambda/\theta}\right)E\left[\exp\left\{\delta^{-1}N^u\sum_{j=1}^{\infty}\left(N/2+\Gamma_j\right)^{-1/\theta}\right\}\right]\\ =:\ & \exp\left(-N^{u-\lambda/\theta}\right)E_n,\end{aligned}$$

where

$$u := \frac{1}{\theta} - 1\,.$$

Clearly, $u - \lambda/\theta = \lambda$ and hence it suffices to show that $E_n = O(1)$. By (4.13), it follows that

$$\begin{aligned} E_n &= \exp\left[-\int_0^{(N/2)^{-1/\theta}} \left\{1-\exp\left(\delta^{-1}N^u y\right)\right\}\theta y^{-(1+\theta)}dy\right] \\ &= \exp\left[K_1 N^{u\theta}\int_0^{K_2/N}\left(e^z-1\right)z^{-(1+\theta)}dz\right]\,, \end{aligned}$$

for some finite positive constants K_1 and K_2. Note that as $x \downarrow 0$,

$$\begin{aligned} &\int_0^x \left(e^z-1\right)z^{-(1+\theta)}dz \\ &= O\left(\int_0^x z^{-\theta}dz\right) \\ &= O\left(x^{1-\theta}\right)\,. \end{aligned}$$

Hence,

$$\begin{aligned} \int_0^{K_2/N}\left(e^z-1\right)z^{-(1+\theta)}dz &= O\left(N^{\theta-1}\right) \\ &= O\left(N^{-u\theta}\right)\,. \end{aligned}$$

This shows that $E_n = O(1)$ and thus completes the proof. □

Using $N = 10^5$ number of terms in the series and generating the (truncated) random variable 10^5 times, we have estimated some quantiles, for a range of values θ, as mentioned in Table 4.1.

For θ closer to 1, the rate of convergence of the truncated sum $\sum_{j=1}^{N}\Gamma_j^{-1/\theta}$ as $N \to \infty$ is very slow, and in order to obtain upper bounds on the quantiles of the random variable $Z(\theta)$ we used Proposition 4.2.2 as described below. Such upper

Table 4.1: $c_p(\theta)$ for various p and θ

p \ θ	0.5	0.6	0.7
.05	4.3	5.8	8.2
.025	5.1	6.9	9.8
.01	6.2	8.4	12.1

bounds lead to conservative versions of the test (4.10). We use the exponential Markov inequality: for $0 < r < -\gamma_0$,

$$P(Z(\theta) \geq z) \leq e^{-rz} E e^{rZ} = e^{-rz} \left(1 - re^{-r} \int_0^1 e^{rx} x^{-\theta}\, dx\right)^{-1},$$

and estimate the integral from above by

$$\int_0^1 e^{rx} x^{-\theta} dx \leq e^{r/k} \frac{k^{\theta-1}}{1-\theta} + \frac{1}{k} \sum_{j=2}^{k} e^{rj/k} \left(\frac{j-1}{k}\right)^{-\theta},$$

$k > 1$. Using $r = .05$ and $k = 10^7$ we computed numbers $\tilde{c}_p(\theta)$ satisfying

$$P\big(Z(\theta) \geq \tilde{c}_p(\theta)\big) \leq p\,.$$

These numbers $\tilde{c}_p(\theta)$ are reported in Table 4.2.

Since we are only assuming that the tail exponent α has a known upper bound as in (4.3), but the exact value of α may be unknown, a possible way to obtain a conservative estimate of the critical value $c_p(\alpha/A)$ in (4.11) is to choose a number $A_1 > A$ and use the statistic $Z_n(A_1)$ instead of $Z_n(A)$ in (4.4). By Proposition 4.2.1, under the null hypothesis, the test statistic converges weakly to $Z(\alpha/A_1)$, which is stochastically smaller than $Z(A/A_1)$, and we obtain a conservative test by modifying (4.10) as follows:

$$\text{reject } H_0 \text{ at significance level } p \in (0,1) \text{ if } \; Z_n(A_1) > c_p(A/A_1)\,. \tag{4.17}$$

Table 4.2: $\tilde{c}_p(\theta)$ for various p and θ

p \ θ	0.8	0.9	0.95
.05	65.43	73.12	127.37
.025	79.29	86.98	141.23
.01	97.62	105.31	159.56

4.3 Testing the hypothesis of hard truncation

In this section we consider the following problem of testing a null hypothesis against a simple alternative:

$$\left. \begin{array}{ll} H_0: & P(|H_1| > M) \gg n^{-1} \quad \text{(hard truncation)} \\ H_1: & P(|H_1| > M) \ll n^{-1} \quad \text{(soft truncation)} \end{array} \right\}. \tag{4.18}$$

We still assume that an upper bound (4.3) on the tail exponent is known. For a test statistic in this case we choose a number $\gamma \in (0,1)$ and define

$$Z_n(A;\gamma) = \frac{\left(\sum_{j=1}^{[\gamma n]} (-1)^j X_j^{\langle A/2 \rangle}\right)^2}{\sum_{j=[\gamma n]+1}^{n} |X_j|^A}. \tag{4.19}$$

Here $a^{\langle b \rangle} = |a|^b \text{sign}(a)$ for real a, b is the signed power. The asymptotic distribution of $Z_n(A;\gamma)$ under the null hypothesis and under the alternative in (4.18) is described in Proposition 4.3.1 below. Recall the standard notation of $S_\alpha(\sigma, \beta, \mu)$ for (the distribution of) an α-stable random variable with the scale σ, skewness β and location μ; see Samorodnitsky and Taqqu (1994). For a symmetric α-stable random variable, $\beta = \mu = 0$. For a positive strictly α-stable random variable with $0 < \alpha < 1$, one has $\beta = 1$ and $\mu = 0$. Finally, for $0 < \alpha < 2$, let

$$C_\alpha = \begin{cases} (\Gamma(1-\alpha)\cos(\pi\alpha/2))^{-1} & \text{if } \alpha \neq 1, \\ 2/\pi & \text{if } \alpha = 1, \end{cases}$$

Proposition 4.3.1. *(i) Assume that $EL_1^{2A} < \infty$. Then under the hypothesis H_0 of hard truncation,*

$$Z_n(A;\gamma) \Rightarrow C_1(\gamma)\chi_1^2\,, \tag{4.20}$$

where $C_1(\gamma) = 2\gamma/(1-\gamma)$, and χ_1^2 is the standard chi-square random variable with one degree of freedom.

(ii) Under the hypothesis H_1 of soft truncation,

$$Z_n(A;\gamma) \Rightarrow C_2(A;\gamma)\frac{S_1^2}{S_2}\,, \tag{4.21}$$

where

$$C_2(A;\gamma) = \left(\frac{\gamma}{1-\gamma}\frac{C_{\alpha/A}}{C_{2\alpha/A}}\right)^{A/\alpha}\,,$$

and S_1 and S_2 are independent random variables, such that S_1 is a symmetric $2\alpha/A$-stable random variable with unit scale, and S_2 is a positive strictly α/A-stable random variable with unit scale.

Proof. The claim of part (i) will follow from the following two statements.

$$\frac{1}{(nM_n^A P(|H_1| > M_n))^{1/2}}\sum_{j=1}^{[\gamma n]}(-1)^j X_j^{\langle A/2\rangle} \Rightarrow \left(\frac{2A\gamma}{A-\alpha}\right)^{1/2} N(0,1)\,, \tag{4.22}$$

and

$$\frac{1}{nM_n^A P(|H_1| > M_n)}\sum_{j=[\gamma n]+1}^{n}|X_j|^A \to \frac{A(1-\gamma)}{A-\alpha} \tag{4.23}$$

in probability. We prove (4.23) first, and it is enough to show that

$$\frac{1}{nM_n^A P(|H_1| > M_n)} E\left(\sum_{j=[\gamma n]+1}^{n}|X_j|^A\right) \to \frac{A(1-\gamma)}{A-\alpha} \tag{4.24}$$

and

$$\frac{1}{\big(nM_n^A P(|H_1| > M_n)\big)^2} \operatorname{Var}\left(\sum_{j=[\gamma n]+1}^{n} |X_j|^A\right) \to 0\,. \tag{4.25}$$

Note that by the Karamata theorem,

$$E\left(\sum_{j=[\gamma n]+1}^{n} |X_j|^A\right) \sim (1-\gamma)n\, E\big(|X_1|^A\big)$$

$$= (1-\gamma)n\Big[E\big(|H_1|^A \mathbf{1}(|H_1| \le M_n)\big) + E(M_n + L_1)^A P(|H_1| > M_n)\Big]$$

$$\sim (1-\gamma)n\left[\frac{\alpha}{A-\alpha} M_n^A P(|H_1| > M_n) + M_n^A P(|H_1| > M_n)\right]$$

$$= \big(nM_n^A P(|H_1| > M_n)\big)\frac{A(1-\gamma)}{A-\alpha}\,,$$

proving (4.24). A similar calculation gives us

$$\operatorname{Var}\left(\sum_{j=[\gamma n]+1}^{n} |X_j|^A\right) \sim (1-\gamma)n \operatorname{Var}\big(|X_1|^A\big)$$

$$\le n\, E\big(|X_1|^{2A}\big) \sim \big(nM_n^{2A} P(|H_1| > M_n)\big)\frac{2A}{2A-\alpha}\,,$$

and (4.25) follows because the truncation is hard. Therefore, we have established (4.23).

In order to prove (4.22), note that the triangular array

$$\tilde{X}_{nj} := H_j^{\langle A/2\rangle} \mathbf{1}\big(|H_j|^{A/2} \le M_n^{A/2}\big) + \frac{H_j}{|H_j|}(M_n^{A/2} + L_j^{A/2})\mathbf{1}\big(|H_j|^{A/2} > M_n^{A/2}\big)\,,$$

$j = 1, \ldots, n$, $n = 1, 2, \ldots$, satisfies the assumptions of Theorem 2.2.5 (with α replaced by $2\alpha/A$), and, therefore,

$$\frac{1}{\big(nM_n^A P(|H_1| > M_n)\big)^{1/2}}\left(\sum_{j=1}^{n} \tilde{X}_{nj} - E\Big(\sum_{j=1}^{n} \tilde{X}_{nj}\Big)\right) \Rightarrow \left(\frac{2A}{A-\alpha}\right)^{1/2} N(0,1)\,.$$

The random variables $\big(X_j^{\langle A/2\rangle}\big)$ form a somewhat different triangular array, namely

$$X_{nj}^{\langle A/2\rangle} = H_j^{\langle A/2\rangle} \mathbf{1}\big(|H_j|^{A/2} \le M_n^{A/2}\big) + \frac{H_j}{|H_j|}(M_n + L_j)^{A/2}\mathbf{1}\big(|H_j|^{A/2} > M_n^{A/2}\big)\,,$$

$j=1,\dots,n,\ n=1,2,\dots$, but an inspection of the proof of Theorem 2.2.5 shows that the argument applies equally well to the latter triangular array, so that

$$\frac{1}{\big(nM_n^A P(|H_1|>M_n)\big)^{1/2}}\left(\sum_{j=1}^n X_{nj}^{\langle A/2\rangle} - E\Big(\sum_{j=1}^n X_{nj}^{\langle A/2\rangle}\Big)\right)$$
$$\Rightarrow \left(\frac{2A}{A-\alpha}\right)^{1/2} N(0,1)\,.$$

In particular, (extending the length of the rows of the triangular array) we see that

$$\frac{1}{(nM_n^A P(|H_1|>M_n))^{1/2}}\left(\sum_{j=1}^n X_{nj}^{\langle A/2\rangle} - \sum_{j=n+1}^{2n} X_{nj}^{\langle A/2\rangle}\right)$$
$$\Rightarrow \left(\frac{4A}{A-\alpha}\right)^{1/2} N(0,1)\,.$$

Replacing n with $[n\gamma/2]$, we obtain (4.22) and, hence, finish the proof of part (i).

For part (ii), we define

$$b_n = \inf\big\{x>0:\ P(|H_1|^{A/2}>x)\le n^{-1}\big\},\ n=1,2,\dots\,.$$

Then for some centering sequence (c_n) we have

$$b_n^{-1}\left(\sum_{j=1}^n H_j^{\langle A/2\rangle} - c_n\right) \Rightarrow Y$$

with Y having a $S_{2\alpha/A}(\sigma,\beta,\mu)$ distribution with $\sigma^{2\alpha/A}=(C_{2\alpha/A})^{-1}$ and some β,μ; see Feller (1971). Because of the soft truncation, the triangular array $\big(X_{nj}^{\langle A/2\rangle}\big)$ satisfies Theorem 2.2.1, and so

$$b_n^{-1}\left(\sum_{j=1}^n X_{nj}^{\langle A/2\rangle} - c_n\right) \Rightarrow Y$$

with the same Y. Extending the rows of the triangular array gives us

$$b_n^{-1}\left(\sum_{j=1}^n X_{nj}^{\langle A/2\rangle} - \sum_{j=n+1}^{2n} X_{nj}^{\langle A/2\rangle}\right) \Rightarrow \left(\frac{2}{C_{2\alpha/A}}\right)^{A/(2\alpha)} S_1\,,$$

where S_1 is a symmetric $2\alpha/A$-stable random variable with unit scale. Replacing n with $[n\gamma/2]$ we obtain

$$\sum_{j=1}^{[\gamma n]} (-1)^j X_j^{\langle A/2 \rangle} \Rightarrow \left(\frac{\gamma}{C_{2\alpha/A}} \right)^{A/(2\alpha)} S_1 \,. \tag{4.26}$$

Next, we also have

$$b_n^{-2} \sum_{j=1}^{n} |H_j|^A \Rightarrow \left(\frac{1}{C_{\alpha/A}} \right)^{A/\alpha} S_2 \,,$$

where S_2 is a positive strictly α/A-stable random variable with unit scale; see once again Feller (1971). As before, because of the soft truncation, Theorem 2.2.1 applies, and we obtain

$$b_n^{-2} \sum_{j=1}^{n} |X_{nj}|^A \Rightarrow \left(\frac{1}{C_{\alpha/A}} \right)^{A/\alpha} S_2 \,.$$

Replacing n with $(1-\gamma)n$, shows that

$$b_n^{-2} \sum_{j=[\gamma n]+1}^{n} |X_j|^A \Rightarrow \left(\frac{1-\gamma}{C_{\alpha/A}} \right)^{A/\alpha} S_2 \,. \tag{4.27}$$

Since the numerator and the denominator of the statistic $Z_n(A;\gamma)$ in (4.19) are independent, the claim of part (ii) of the proposition follows from (4.26) and (4.27). $\square$

Interestingly, the asymptotic distribution of the test statistic $Z_n(A;\gamma)$, under the null hypothesis, does not depend on the choice of the parameter A (as long as it an upper bound on the tail exponent α). Furthermore, under the null hypothesis this asymptotic distribution of the test statistic is light-tailed (e.g. some exponential moments are finite). On the other hand, the asymptotic distribution of the test statistic under the alternative is, clearly, heavy tailed, as even the second moment is infinite. Therefore, a reasonable test will reject the null hypothesis in favor of the alternative if the test statistic is too large. That is, we suggest the following test for the problem (4.18).

$$\text{reject } H_0 \text{ at significance level } p \in (0,1) \text{ if } \; Z_n(A;\gamma) > \frac{2\gamma}{1-\gamma} c_p \,, \tag{4.28}$$

with c_p such that $P(\chi_1^2 > c_p) = p$.

4.4 Testing a stronger version of the hypothesis of hard truncation

The test statistics $Z_n(A;\gamma)$ we used in the previous subsection for the problem (4.18) has a nondegenerate asymptotic distribution under both the null hypothesis and the alternative. This restricts the sensitivity of the resulting test. In order to obtain a more sensitive test we strengthen the null hypothesis. Specifically, in this section we consider the following problem of testing a null hypothesis against a simple alternative:

$$\left.\begin{array}{ll} H_0: & n^{1-\epsilon} P(|H_1| > M) \gg 1 \\ H_1: & nP(|H_1| > M) \ll 1 \end{array}\right\}, \tag{4.29}$$

where ϵ is a fixed number in $(0,1)$.

For this problem one can use the same test statistic $Z_n(A)$ defined in (4.4) as we used for the problem (4.2) of testing the hypothesis of soft truncation. Proposition 4.2.1 tells us that this test statistic diverges in probability to infinity under the hypothesis of hard truncation. The strengthened hypothesis of hard truncation in (4.29) allows us to quantify how fast this divergence takes place. This, in turn, can be used to build a test. The asymptotic distribution of $Z_n(A)$ under the hypothesis of soft truncation is described in Proposition 4.2.1. The next result provides an asymptotic distributional lower bound on the test statistic under the null hypothesis in the problem (4.29). As in the previous sections, we assume that an upper bound (4.3) on the tail exponent is known.

Proposition 4.4.1. *Assume that $EL_1^{2A} < \infty$. Then under the strengthened hypothesis H_0 of hard truncation,*

$$\liminf_{n\to\infty} P\Big(n^{-\epsilon/2} Z_n(A) > x\Big) \geq e^{-x^2} \tag{4.30}$$

for every $x > 0$.

Proof. In the notation of the triangular array (1.4), consider the binomial random variable $N_n = \sum_{j=1}^n \mathbf{1}\big(|H_j| > M_n\big)$. The strengthened hypothesis of hard truncation implies that $P(N_n \geq n^\epsilon) \to 1$ as $n \to \infty$. Notice that, on the event $\{|H_j| > M_n$ for at least one $j \leq n\}$, whose probability increases to 1,

$$\begin{aligned} Z_n(A) &\geq \frac{\sum_{j=1}^n (M_n + L_j)^A \mathbf{1}\big(|H_j| > M_n\big)}{\max_{j=1,\dots,n}(M_n + L_j)^A \mathbf{1}\big(|H_j| > M_n\big)} \\ &= \sum_{i=1}^n \left\{ \frac{(M_n + L_i)\mathbf{1}\big(|H_i| > M_n\big)}{\max_{j=1,\dots,n}(M_n + L_j)\mathbf{1}\big(|H_j| > M_n\big)} \right\}^A \\ &\geq \sum_{i=1}^n \left\{ \frac{L_i \mathbf{1}\big(|H_i| > M_n\big)}{\max_{j=1,\dots,n} L_j \mathbf{1}\big(|H_j| > M_n\big)} \right\}^A \\ &= \frac{\sum_{j=1}^n L_j^A \mathbf{1}\big(|H_j| > M_n\big)}{\max_{j=1,\dots,n} L_j^A \mathbf{1}\big(|H_j| > M_n\big)} . \end{aligned}$$

Therefore, for $x > 0$, using the assumption $EL_1^{2A} < \infty$, we have

$$\begin{aligned} \liminf_{n\to\infty} P\Big(n^{-\epsilon/2} Z_n(A) > x\Big) &\geq \liminf_{n\to\infty} P\Big(\max_{j=1,\dots,N_n} L_j^A < n^{-\epsilon/2} N_n \frac{EL_1^A}{2} x^{-1}\Big) \\ &\geq \liminf_{n\to\infty} E\left[\left(1 - \frac{x^2}{N_n}\right)^{N_n} \mathbf{1}\Big(N_n \geq n_\epsilon\Big)\right] \to e^{-x^2} , \end{aligned}$$

as required. □

Proposition 4.4.1 tells us that under the hypothesis H_0, $n^{-\epsilon/2}Z_n(A)$ is, asymptotically, stochastically larger than the square root of the standard exponential random variable (independently of the parameter A). Therefore, we suggest the following test for the problem (4.29).

$$\text{reject } H_0 \text{ at significance level } p \in (0,1) \text{ if } \; Z_n(A) \leq \left|\log(1-p)\right|^{1/2} n^{\epsilon/2}\,. \tag{4.31}$$

CHAPTER 5

APPLICATION 2: ESTIMATING THE TAIL INDEX FROM TRUNCATED DATA

5.1 Introduction

Estimating the tail exponent α is one of the main statistical issues one faces when working with data for which a model with power tails is contemplated. This is a difficult statistical problem because one attempts to estimate a parameter governing the tail behavior in an otherwise nonparametric model. The situation is even trickier when one tries to estimate the tail exponent in a sample of observations with truncated power tails. This is the task we address in this chapter.

Suppose that we have a one-dimensional non-negative sample from the truncated power law model (1.4). Once again, we use notation similar to that in Chapter 4, instead of the triangular array notation. That is, we assume that our sample $X_1, \ldots, X_n$ is given by

$$X_j = H_j \mathbf{1}(H_j \leq M_n) + (M_n + L_j)\mathbf{1}(H_j > M_n)\,, \tag{5.1}$$

where $H, H_1, H_2, \ldots$ are i.i.d. $[0, \infty)$ valued random variables whose tails are regularly varying with index α. As in Chapter 4, the sequence M_n, the true value of α, the exact form of the distribution of H and the distribution of L are all assumed to be unknown and α can take any positive value.

In order to accomplish the task of estimating the tail exponent from the truncated data, we analyze the behavior of the Hill estimator introduced by Hill (1975), which is one of the best known and widely used estimators of the tail exponent of distributions with non-truncated power tails, in the truncated setting. There are, however, a number of other estimators of the same; see Chapter 4 of de Haan and Ferreira (2006) for a thorough discussion. The discussion in Chapters 2 and 3 makes it intuitive that estimating the tail exponent α should be easier if the tails are truncated softly, than in the case when the tails are truncated hard. That turns out, indeed, to be the case. Specifically, we will show in Section 5.2 that when the model with truncated power tails is in the soft truncation regime, the Hill estimator $h(k_n, n)$, as defined in (5.2) below, is a consistent estimator of the inverse of the tail exponent as long as k_n is a sequence of integers satisfying $1 \ll k_n \ll n$, which is also the range where k_n should belong in the untruncated case for the consistency to hold. However, in the hard truncation regime, it turns out that k_n should satisfy $nP(H > M_n) \ll k_n \ll n$ in order for the consistency to hold. More than anything, this makes it difficult to choose apriori a k so that the Hill estimator is consistent for α^{-1} because of the lack of knowledge about M_n and the exact tail of H. Therefore, we set before ourselves the following task in this chapter - devising a method to choose a (random) k_n based on the data so that $h(k_n, n)$ consistently estimates α^{-1} regardless of the truncation regime. This will be done in Section 5.3. This, in particular, complements the methods suggested in Chapter 4 in the following way. All the tests discussed in that chapter assumes that an upper bound on α is known, which of course, will rarely be the case in practice. However, if one can estimate α regardless of the truncation regime, one can suggest an upper bound for α based on that estimate.

5.2 Effect of truncation on the Hill estimator

Given a sample $X_1, \ldots, X_n$, recall from (1.14) that the Hill statistic is defined by

$$h(k,n) = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k)}} , \tag{5.2}$$

where $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$ are the order statistics from the sample $X_1, \ldots, X_n$, and $k = 1, \ldots, n$ is a user-determined parameter, the number of the upper order statistics to use in the estimator. Also recall that if $X_1, \ldots, X_n$ are i.i.d. random variables whose right tail is regularly varying with exponent $\alpha > 0$, and the sequence k_n satisfies $1 \ll k_n \ll n$ as $n \to \infty$, then $h(k_n, n) \to 1/\alpha$ in probability as $n \to \infty$; see e.g. Theorem 3.2.2 in de Haan and Ferreira (2006).

In spite of the simplicity of the statement of the consistency of the Hill estimator, selecting the number k of the upper order statistics for a given sample with nontruncated power tails remains a daunting problem; see e.g. pp. 192-193 in Embrechts et al. (1997). In the main result of this section, Theorem 5.2.1 below, we will see that one has to be particularly careful when using the Hill estimator on a sample with truncated power tails. Nonetheless, a consistent estimator can still be obtained.

Notice that the next theorem does not impose any conditions on the random variables $L_1, L_2, \ldots$ in the model (5.1).

Theorem 5.2.1. *Suppose that the number k_n of the upper order statistics satisfies*

$$nP(H > M_n) + 1 \ll k_n \ll n \, . \tag{5.3}$$

Then $h(k_n, n) \to 1/\alpha$ in probability as $n \to \infty$.

Note that Theorem 5.2.1 says that, in the soft truncation regime, the Hill estimator is consistent under the same assumption, $k_n/n \to 0$, as in the non-truncated case. The proof is along similar lines as those in the proof of Theorem 4.2 of Resnick (2007). In what follows, we adopt the following notations:

$$M_+(0,\infty] = \text{ Space of all non-negative Radon measures on } (0,\infty]\,,$$

for $x \in \mathbb{R}$, δ_x denotes the measure on $\mathbb{R}$ defined by

$$\delta_x(A) := \begin{cases} 1, & x \in A \\ 0, & x \notin A\,, \end{cases}$$

and for any function $f : [0,\infty) \longrightarrow \mathbb{R}$,

$$f^{\leftarrow}(x) := \inf\{y : f(y) \geq x\}\,.$$

The proof of Theorem 5.2.1 needs the following sequence of lemmas.

Lemma 5.2.1. *If* (5.3) *holds, then* $\nu_n \Rightarrow \nu_\alpha$ *as* $M_+(0,\infty]$*-valued random variables, where* ν_α *is a measure on* $(0,\infty]$ *defined by*

$$\nu_\alpha((x,\infty]) = x^{-\alpha} \textit{ for all } x > 0,$$

$$\nu_n := \frac{1}{k_n}\sum_{i=1}^{n} \delta_{X_i/b(n/k_n)}$$

and

$$b(\cdot) := \left(\frac{1}{P(H > \cdot)}\right)^{\leftarrow}\,. \tag{5.4}$$

Proof. We shall write just k for k_n. First we shall show that as measures on $(0,\infty]$,

$$\frac{n}{k}P\left[\frac{X_1}{b(n/k)} \in \cdot\right] \xrightarrow{v} \nu_\alpha(\cdot)\,. \tag{5.5}$$

The hypothesis implies that

$$\lim_{n\to\infty}\frac{n}{k}P(H > M_n) = 0$$

and hence

$$b(n/k) \ll M_n \,.$$

Thus, for any $x > 0$ and for n large enough so that $xb(n/k) < M_n$,

$$\begin{aligned} P\left[\frac{X_1}{b(n/k)} > x\right] &= P[H > xb(n/k)] \\ &\sim \frac{k}{n} x^{-\alpha} \,, \end{aligned}$$

where the last line follows from the hypothesis that $k \ll n$. This shows (5.5).

In order to complete the claim, it suffices to check that for all non-negative continuous functions h with compact support, defined on $(0, \infty]$,

$$E \exp\left(-\int h d\nu_n\right) \longrightarrow E \exp\left(-\int h d\nu_\alpha\right) \,. \tag{5.6}$$

Note that

$$\begin{aligned} E \exp\left(-\int h d\nu_n\right) &= \left[E \exp\left\{-\frac{1}{k} h(X_1/b(n/k))\right\}\right]^n \\ &= \left[1 - \int_{(0,\infty]} \left\{1 - e^{-\frac{1}{k}h(x)}\right\} P\left(\frac{X_1}{b(n/k)} \in dx\right)\right]^n \\ &= \left[1 - \frac{1}{n}\int_{(0,\infty]} \left\{1 - e^{-\frac{1}{k}h(x)}\right\} nP\left(\frac{X_1}{b(n/k)} \in dx\right)\right]^n \,. \end{aligned}$$

Since h is compactly supported, there is $\epsilon > 0$ so that h is zero on $(0, \epsilon)$. Thus,

$$\int_{(0,\infty]} \left\{1 - e^{-\frac{1}{k}h(x)}\right\} nP\left(\frac{X_1}{b(n/k)} \in dx\right) = \int_{[\epsilon,\infty]} \left\{1 - e^{-\frac{1}{k}h(x)}\right\} nP\left(\frac{X_1}{b(n/k)} \in dx\right) \,.$$

Since $k \gg 1$, using estimates provided by Taylor's expansion, one can show that

$$\begin{aligned} &\int_{[\epsilon,\infty]} \left\{1 - e^{-\frac{1}{k}h(x)}\right\} nP\left(\frac{X_1}{b(n/k)} \in dx\right) \\ &= \int_{[\epsilon,\infty]} h(x) \frac{n}{k} P\left(\frac{X_1}{b(n/k)} \in dx\right) + o\left(k^{-2} nP(X_1 > \epsilon b(n/k))\right) \\ &= (1 + o(1)) \int_{[\epsilon,\infty]} h(x) \nu_\alpha(dx) + o(1) \,, \end{aligned}$$

the last step following from (5.5). Thus,

$$\begin{aligned}\int_{(0,\infty]} \left\{1-e^{-\frac{1}{k}h(x)}\right\} nP\left(\frac{X_1}{b(n/k)} \in dx\right) &\longrightarrow \int_{[\epsilon,\infty]} h(x)\nu_\alpha(dx) \\ &= \int_{(0,\infty]} h(x)\nu_\alpha(dx)\,.\end{aligned}$$

This shows (5.6) and hence completes the proof. □

Lemma 5.2.2. *Under the hypotheses of Theorem 5.2.1,* $X_{(k)}/b(n/k) \xrightarrow{P} 1$.

Proof. Fix $\epsilon > 0$. Note that

$$\begin{aligned}P\left[\left|\frac{X_{(k)}}{b(n/k)}-1\right| > \epsilon\right] &= P[X_{(k)} > (1+\epsilon)b(n/k)] + P[X_{(k)} < (1-\epsilon)b(n/k)] \\ &\le P\left[\left(\frac{1}{k}\sum_{i=1}^n \delta_{X_i/b(n/k)}\right)((1+\epsilon,\infty]) \ge 1\right] \\ &\quad +P\left[\left(\frac{1}{k}\sum_{i=1}^n \delta_{X_i/b(n/k)}\right)((1-\epsilon,\infty]) < 1\right]\,.\end{aligned}$$

But, by Lemma 5.2.1,

$$\left(\frac{1}{k}\sum_{i=1}^n \delta_{X_i/b(n/k)}\right)((1+\epsilon,\infty]) \xrightarrow{P} (1+\epsilon)^{-\alpha} < 1\,,$$

and hence

$$P\left[\left(\frac{1}{k}\sum_{i=1}^n \delta_{X_i/b(n/k)}\right)((1+\epsilon,\infty]) \ge 1\right] \longrightarrow 0\,.$$

Similarly,

$$\left(\frac{1}{k}\sum_{i=1}^n \delta_{X_i/b(n/k)}\right)((1-\epsilon,\infty]) \xrightarrow{P} (1-\epsilon)^{-\alpha} > 1\,,$$

and hence

$$P\left[\left(\frac{1}{k}\sum_{i=1}^n \delta_{X_i/b(n/k)}\right)((1-\epsilon,\infty]) < 1\right] \longrightarrow 0\,.$$

This completes the proof. □

Lemma 5.2.3. *Under the assumptions of Theorem 5.2.1,*

$$\hat{\nu}_n \xrightarrow{P} \nu_\alpha\,,$$

in $M_+(0,\infty]$, *where*

$$\hat{\nu}_n := \frac{1}{k_n}\sum_{i=1}^{n}\delta_{X_i/X_{(k)}}\,.$$

Proof. It suffices to show that

$$\hat{\nu}_n \Rightarrow \nu_\alpha$$

because if the limit is degenerate, weak convergence and convergence in probability are equivalent.

Define an operator $T : M_+(0,\infty] \times (0,\infty) \mapsto M_+(0,\infty]$ by

$$T(\mu, x)(A) = \mu(xA)$$

By Slutsky's theorem and Lemmas 5.2.1 and 5.2.2, it follows that

$$\left(\nu_n, \frac{X_{(k)}}{b(n/k)}\right) \Rightarrow (\nu_\alpha, 1)$$

in $M_+(0,\infty] \times (0,\infty)$. Since

$$\hat{\nu}_n(\cdot) = \nu_n\left(\frac{X_{(k)}}{b(n/k)}\cdot\right) = T\left(\nu_n, \frac{X_{(k)}}{b(n/k)}\right)\,,$$

the conclusion will follow if T is continuous at $(\nu_\alpha, 1)$, which has been shown in the proof of Theorem 4.2 of Resnick (2007). □

Proof of Theorem 5.2.1. Observe that

$$\begin{aligned}
&\int_{[1,\infty)} \hat{\nu}_n((x,\infty])x^{-1}dx \\
(\text{convention: } X_{(0)} = \infty) \quad &= \sum_{i=1}^{k}\int_{\left[\frac{X_{(i)}}{X_{(k)}}, \frac{X_{(i-1)}}{X_{(k)}}\right)} \hat{\nu}_n((x,\infty])x^{-1}dx \\
&= \sum_{i=1}^{k}\frac{i-1}{k}\int_{\left[\frac{X_{(i)}}{X_{(k)}}, \frac{X_{(i-1)}}{X_{(k)}}\right)} x^{-1}dx \\
&= \sum_{i=1}^{k}\frac{i-1}{k}\left(\log\frac{X_{(i-1)}}{X_{(k)}} - \log\frac{X_{(i)}}{X_{(k)}}\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\log\frac{X_{(i)}}{X_{(k)}} = h(k,n)\,,
\end{aligned}$$

and that

$$
\begin{aligned}
& \int_{[1,\infty)} \nu_\alpha((x,\infty])x^{-1}dx \\
= & \int_{[1,\infty)} x^{-\alpha-1}dx \\
= & \frac{1}{\alpha}\,.
\end{aligned}
$$

If only the map $T : M_+(0,\infty] \longrightarrow \mathbb{R}$ defined by

$$T(\mu) := \int_{[1,\infty)} \mu((x,\infty])x^{-1}dx$$

were continuous, the proof would have been complete (clearly T has compact support). However, that's not true. Nevertheless, for every $M > 0$, the map T_M defined by

$$T_M(\mu) := \int_{[1,M]} \mu((x,\infty])x^{-1}dx$$

is continuous. Hence, it suffices to show that for every $\delta > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} P\left[\int_M^\infty \hat{\nu}_n((x,\infty])x^{-1}dx > \delta\right] = 0\,. \tag{5.7}$$

We write

$$
\begin{aligned}
& P\left[\int_M^\infty \hat{\nu}_n((x,\infty])x^{-1}dx > \delta\right] \\
\leq\ & P\left[\int_M^\infty \hat{\nu}_n((x,\infty])x^{-1}dx > \delta, \frac{X_{(k)}}{b(n/k)} \in (1-\eta, 1+\eta)\right] \\
& +P\left[\frac{X_{(k)}}{b(n/k)} \notin (1-\eta, 1+\eta)\right] \\
=\ & A(M,n) + B(n)
\end{aligned}
$$

Note that by Lemma 5.2.2,

$$B(n) \longrightarrow 0\,.$$

Also,

$$
\begin{aligned}
A(M,n) &\le P\left[\int_{M(1-\eta)}^{\infty} \nu_n((x,\infty])x^{-1}dx > \delta\right] \\
&\le \delta^{-1}E\left[\int_{M(1-\eta)}^{\infty} \nu_n((x,\infty])x^{-1}dx\right] \\
&= \delta^{-1}\int_{M(1-\eta)}^{\infty} \frac{n}{k}P[X_1 > b(n/k)x]x^{-1}dx\,.
\end{aligned}
$$

The functions $x \mapsto \frac{n}{k}P[X_1 > b(n/k)x]x^{-1}$ are non-increasing and converging point-wise to $x^{-\alpha-1}$, and hence uniformly on $[M(1-\eta),\infty)$. Thus as $n \longrightarrow \infty$,

$$
\int_{M(1-\eta)}^{\infty} \frac{n}{k}P[X_1 > b(n/k)x]x^{-1}dx \longrightarrow \int_{M(1-\eta)}^{\infty} x^{-\alpha-1}dx = CM^{-\alpha}
$$

for some $C \in (0,\infty)$. This shows

$$
\lim_{M\to\infty}\limsup_{n\to\infty} A(M,n) = 0
$$

and hence (5.7). This completes the proof. $\square$

5.3 How to choose k when the data is truncated?

Since the truncation level M_n is not known, it is desirable to have a sample-based way of deciding on the number of upper order statistics to use in the Hill estimator, particularly when the tails are truncated hard. A natural (in view of the condition (5.3)) choice is to use *a random number* of upper order statistics given by

$$
\hat{k}_n = \left[n\left(\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\big(X_j > \gamma \max_{i=1,\dots,n} X_i\big)\right)^{\beta}\right], \tag{5.8}
$$

where γ and β are user-specified parameters taking values in $(0,1)$ and $[\cdot]$ denotes the integer part. The first main result of this section shows that this choice of the number of upper order statistics leads to a consistent estimator of the reciprocal of the tail exponent in the hard truncation regime.

Theorem 5.3.1. *Suppose that the model* (5.1) *is the hard truncation regime, i.e.,*

$$\lim_{n\to\infty} nP(H > M_n) = \infty\,.$$

Assume that in addition,

$$P(H > M_n)P(L > \epsilon M_n) = o(1/n)\,\textit{for all}\,\epsilon > 0. \tag{5.9}$$

Then, $h(\hat{k}_n, n)$ *consistently estimates* α^{-1}*, where* $\hat{k}_n$ *is as defined in* (5.8).

The proof is using the following lemma.

Lemma 5.3.1. *As* $n \longrightarrow \infty$*,*

$$\frac{\hat{k}_n}{k_n} \xrightarrow{P} 1\,,$$

where

$$k_n := \left[nP(H > \gamma M_n)^\beta\right]\,.$$

Proof. Clearly, all that needs to be shown is

$$\frac{\sum_{j=1}^n \mathbf{1}(X_j > \gamma\hat{M}_n)}{nP(H > \gamma M_n)} \xrightarrow{P} 1 \tag{5.10}$$

where

$$\hat{M}_n := \max_{1\le i\le n} X_i.$$

We shall show that the assumption (5.9) on M_n implies that

$$\frac{\hat{M}_n}{M_n} \xrightarrow{P} 1. \tag{5.11}$$

For this, note that for $0 < x < 1$,

$$\begin{aligned} P\left(M_n^{-1}\max_{1\le i\le n} X_i \le x\right) &\le P(H_j \le xM_n, 1 \le j \le n) \\ &= \{1 - P(H > xM_n)\}^n \\ &\to 0, \end{aligned}$$

the last step following from the assumption that $P(H > M_n) \gg 1/n$. For $x > 1$,

$$P\left(M_n^{-1} \max_{1 \le i \le n} X_i > x\right) \le nP(H > M_n)P(L > (x-1)M_n) \longrightarrow 0\,.$$

This shows (5.11). It's easy to see from the hard truncation assumption that for $0 < \theta_1 < 1$

$$\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \theta_1 M_n\right)}{nP(H > \theta_1 M_n)} \xrightarrow{P} 1$$

and hence for $\theta_2 > 0$,

$$\frac{\sum_{j=1}^n \mathbf{1}\left(j : X_j > \theta_1 M_n\right)}{nP(H > \theta_2 M_n)} \xrightarrow{P} \left(\frac{\theta_1}{\theta_2}\right)^{-\alpha} \tag{5.12}$$

Fix $0 < \epsilon < 1$. Let $0 < \eta < 1$ be such that $(1-\eta)^{-\alpha} < 1 + \epsilon$. Note that

$$\begin{aligned}
& P\left[\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma \hat{M}_n\right)}{nP(H > \gamma M_n)} > 1 + \epsilon\right] \\
\le\ & P\left[\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma \hat{M}_n\right)}{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma(1-\eta) M_n\right)} > 1\right] \\
& + P\left[\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma(1-\eta) M_n\right)}{nP(H > \gamma M_n)} > 1 + \epsilon\right] \\
\le\ & P[(1-\eta)M_n > \hat{M}_n] + P\left[\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma(1-\eta) M_n\right)}{nP(H > \gamma M_n)} > 1 + \epsilon\right]
\end{aligned}$$

By (5.11), $P[(1-\eta)M_n > \hat{M}_n] \longrightarrow 0$. By (5.12),

$$\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma(1-\eta) M_n\right)}{nP(H > \gamma M_n)} \xrightarrow{P} (1-\eta)^{-\alpha} < 1 + \epsilon$$

and hence

$$P\left[\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma(1-\eta) M_n\right)}{nP(H > \gamma M_n)} > 1 + \epsilon\right] \longrightarrow 0.$$

Now let $\eta > 0$ be such that $\gamma(1+\eta) < 1$ and $(1+\eta)^{-\alpha} > 1 - \epsilon$. Note that

$$P\left[\frac{\sum_{j=1}^n \mathbf{1}\left(X_j > \gamma \hat{M}_n\right)}{nP(H > \gamma M_n)} < 1 - \epsilon\right]$$

$$
\begin{aligned}
&\leq P\left[\frac{\sum_{j=1}^{n}\mathbf{1}\left(X_j > \gamma\hat{M}_n\right)}{\sum_{j=1}^{n}\mathbf{1}\left(X_j > \gamma(1+\eta)M_n\right)} < 1\right] \\
&\quad + P\left[\frac{\sum_{j=1}^{n}\mathbf{1}\left(X_j > \gamma(1+\eta)M_n\right)}{nP(H > \gamma M_n)} < 1-\epsilon\right] \\
&\leq P[(1+\eta)M_n < \hat{M}_n] + P\left[\frac{\sum_{j=1}^{n}\mathbf{1}\left(X_j > \gamma(1+\eta)M_n\right)}{nP(H > \gamma M_n)} < 1-\epsilon\right].
\end{aligned}
$$

From here, exactly same arguments as before show that

$$
P\left[\frac{\sum_{j=1}^{n}\mathbf{1}\left(X_j > \gamma\hat{M}_n\right)}{nP(H > \gamma M_n)} < 1-\epsilon\right] \longrightarrow 0
$$

and thus show (5.10), which completes the proof. □

Proof of Theorem 5.3.1. In view of Lemma 5.3.1, it suffices to show that

$$
\frac{1}{k_n}\sum_{i=1}^{\hat{k}_n}\log\frac{X_{(i)}}{X_{(\hat{k}_n)}} \xrightarrow{P} \frac{1}{\alpha}.
$$

Fix $\epsilon > 0$. Fix $0 < \eta < 1/2$ so that $\alpha^{-1}\log\frac{1+\eta}{1-\eta} < \frac{\epsilon}{3}$. We first show that

$$
\begin{aligned}
& P\left[\left|\frac{1}{k_n}\sum_{i=1}^{\hat{k}_n}\log\frac{X_{(i)}}{X_{(\hat{k}_n)}} - \frac{1}{\alpha}\right| > \epsilon\right] \\
\leq\;& P\left[\left|\frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{X_{(i)}}{X_{(k_n)}} - \frac{1}{\alpha}\right| > \frac{\epsilon}{3}\right] \\
&+ P\left[-\log\frac{X_{([k_n(1+\eta)])}}{X_{([k_n(1-\eta)])}} > \frac{\epsilon}{3}\right] \\
&+ P\left[\left|\frac{\hat{k}_n}{k_n} - 1\right| \geq \eta\right]. \qquad (5.13)
\end{aligned}
$$

To see this, suppose that the following are true:

$$
\begin{aligned}
\left|\frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{X_{(i)}}{X_{(k_n)}} - \frac{1}{\alpha}\right| &\leq \frac{\epsilon}{3}, && (5.14)\\
-\log\frac{X_{([k_n(1+\eta)])}}{X_{([k_n(1-\eta)])}} &\leq \frac{\epsilon}{3}, && (5.15)\\
\left|\frac{\hat{k}_n}{k_n} - 1\right| &\leq \eta. && (5.16)
\end{aligned}
$$

By (5.16), it follows that $[(1-\eta)k_n] \le \hat{k}_n \le [(1+\eta)k_n]$. This along with (5.15) shows that

$$\left|\log \frac{X_{(\hat{k}_n)}}{X_{(k_n)}}\right| \le \frac{\epsilon}{3}\,.$$

Invoking (5.14), it follows that

$$\left|\frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{X_{(i)}}{X_{(\hat{k}_n)}} - \frac{1}{\alpha}\right| \le \frac{2\epsilon}{3}\,. \tag{5.17}$$

By (5.15) and (5.16), it also follows that

$$\left|\frac{1}{k_n}\sum_{i=\hat{k}_n\wedge k_n}^{\hat{k}_n\vee k_n} \log \frac{X_{(i)}}{X_{(\hat{k}_n)}}\right| \le 2\eta\frac{\epsilon}{3} \le \frac{\epsilon}{3}\,. \tag{5.18}$$

It follows from (5.17) and (5.18) that

$$\left|\frac{1}{k_n}\sum_{i=1}^{\hat{k}_n} \log \frac{X_{(i)}}{X_{(\hat{k}_n)}} - \frac{1}{\alpha}\right| \le \epsilon\,.$$

This shows (5.13). It follows from Theorem 5.2.1 that

$$P\left[\left|\frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{X_{(i)}}{X_{(k_n+1)}} - \frac{1}{\alpha}\right| > \frac{\epsilon}{3}\right] \longrightarrow 0\,.$$

By Lemma 5.2.2 and the fact that the function $b(\cdot)$, as defined in (5.4), regularly varies with exponent $1/\alpha$, it follows that

$$\frac{X_{([k_n(1+\eta)])}}{X_{([k_n(1-\eta)])}} \xrightarrow{P} \left(\frac{1+\eta}{1-\eta}\right)^{-1/\alpha}\,.$$

and hence the choice of η implies

$$P\left[-\log \frac{X_{([k_n(1+\eta)])}}{X_{([k_n(1-\eta)])}} > \frac{\epsilon}{3}\right] \longrightarrow 0\,.$$

Lemma 5.3.1 implies that

$$P\left[\left|\frac{\hat{k}_n}{k_n} - 1\right| \ge \eta\right] \longrightarrow 0\,.$$

This completes the proof. □

The next result shows that the Hill estimator with the random number of top order statistics is consistent for α^{-1} in the soft truncation regime also.

Theorem 5.3.2. *In the soft truncation regime, $h(\hat{k}_n, n)$ consistently estimates α^{-1}, where $\hat{k}_n$ is as defined in* (5.8).

Proof. Let $\tilde{h}(k, n)$ denote the Hill estimator for the random variables $H_1, \ldots, H_n$ based on the top k order statistics, *i.e.*,

$$\tilde{h}(k,n) := \frac{1}{k}\sum_{i=1}^{k} \log \frac{H_{(i)}}{H_{(k)}} ,$$

where $1 \leq k \leq n$ and $H_{(1)} \geq \ldots \geq H_{(n)}$ are the order statistics from $H_1, \ldots, H_n$. Let β be same as that in the definition of $\hat{k}_n$; see (5.8). By the discussion on page 89 in Resnick (2007), it follows that $\left(\tilde{h}([n^{1-\beta}t], n) : t \geq 1\right)$ converges in probability to the deterministic function that is the constant $1/\alpha$, on $D[1, \infty)$ which is equipped with the topology of uniform convergence on bounded intervals. Since,

$$P\left(\tilde{h}([n^{1-\beta}t], n) \neq h([n^{1-\beta}t], n) \text{ for some } t \geq 1\right) \leq nP(H > M_n) \longrightarrow 0 \, ,$$

it follows that

$$\left(h([n^{1-\beta}t], n) : t \geq 1\right) \xrightarrow{P} \frac{1}{\alpha} \tag{5.19}$$

in $D[1, \infty)$.

Next, we shall investigate the asymptotic behavior of

$$N_n := \sum_{j=1}^{n} \mathbf{1}\big(X_j > \gamma \max_{i=1,\ldots,n} X_i\big) \, .$$

By Proposition 3.21 (page 154) in Resnick (1987) it follows that

$$\sum_{j=1}^{n} \delta_{b_n^{-1} X_j} \Longrightarrow \sum_{j=1}^{\infty} \delta_{\Gamma_j^{-1/\alpha}}$$

on $M_P(0,\infty]$, where

$$b_n := \inf\left\{x > 0 : P(H > x) \le n^{-1}\right\}$$

and $(\Gamma_j : j \ge 1)$ denote the arrivals of an unit rate Poisson Process. An immediate consequence of this is that for $r \ge 1$,

$$\left(b_n^{-1} X_{(i)} : 1 \le i \le r\right) \Longrightarrow \left(\Gamma_i^{-1/\alpha} : 1 \le i \le r\right) .$$

Thus, for $r \ge 1$,

$$\begin{aligned} P\left(N_n = r\right) &= P\left(X_{(r+1)} \le \gamma X_{(1)} < X_{(r)}\right) \\ &\longrightarrow P\left(\Gamma_{r+1}^{-1/\alpha} \le \gamma \Gamma_1^{-1/\alpha} < \Gamma_r^{-1/\alpha}\right) \end{aligned}$$

as $n \longrightarrow \infty$. Thus,

$$N_n \Longrightarrow N := \sum_{j=1}^{\infty} \mathbf{1}\left(\Gamma_j^{-1/\alpha} > \gamma \Gamma_1^{-1/\alpha}\right) .$$

By (5.19), it follows that

$$\left(h([n^{1-\beta}t], n), N_n^{\beta}\right) \Longrightarrow \left(\frac{1}{\alpha}, N^{\beta}\right)$$

in $D[1,\infty) \times \mathbb{N}^{\beta}$, where

$$\mathbb{N}^{\beta} := \{1, 2^{\beta}, 3^{\beta}, \ldots\} .$$

Since the evaluation map from $D[1,\infty) \times \mathbb{N}^{\beta}$ to $\mathbb{R}$ defined by $(x,a) \mapsto x(a)$ is continuous, the continuous mapping theorem completes the proof. □

Once we have shown that the Hill estimator with the random number of upper order statistics is consistent for the reciprocal of the tail exponent, the next natural question is "What is the asymptotic second order behavior? ". In the non-truncated world, it is known that Theorem 1.4.1 holds, for example. It would be nice to know if a similar result would be true in the truncated setting.

We could prove that in the simplest possible case, *i.e.*, when H has a Pareto distribution, L is identically zero and M_n is within a particular range, then the same claim as in Theorem 1.4.1 holds. This is precisely the content of the next result, which is the last main result of this section. We, however, do believe that this can be extended to more general situations and hope to address this in future research.

Theorem 5.3.3. *Let $H, H_1, H_2, \ldots$ be a sequence of i.i.d. Pareto(α) random variables, i.e., have density $\alpha x^{-\alpha-1}$ for $x \geq 1$. Here $\alpha > 0$. Define*

$$X_j := H_j \wedge M_n \,,$$

where M_n is a sequence of positive numbers going to ∞. If the sequence M_n satisfies, in addition, that

$$\lim_{n\to\infty} n M_n^{-\alpha} = \infty$$

and

$$\lim_{n\to\infty} n M_n^{-\alpha(2-\beta)} (\log M_n)^2 = 0 \,, \tag{5.20}$$

then,

$$\sqrt{\hat{k}_n}\left(h(\hat{k}_n, n) - \frac{1}{\alpha}\right) \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right) .$$

Here, $\hat{k}_n$ and β are as in (5.8)*, and for $1 \leq k \leq n$, $h(k,n)$ is the Hill estimator for the random variables $X_1, \ldots, X_n$ based on the top k order statistics, as defined in* (5.2)*.*

The proof is using the following lemma.

Lemma 5.3.2. *Suppose that v_n is a sequence of integers satisfying $1 \ll n M_n^{-\alpha} \ll \sqrt{v_n} \ll \sqrt{n}$ and*

$$\lim_{n\to\infty} \frac{n M_n^{-\alpha}}{\sqrt{v_n}} \log\left\{ M_n \left(\frac{n}{v_n}\right)^{-1/\alpha} \right\} = 0 \,.$$

Let $Y_{(n,1)} \geq \ldots \geq Y_{(n,n)}$ be the order statistics of $Y_{n,1}, \ldots, Y_{n,n}$ that are i.i.d. from the cdf F_n defined as

$$F_n(x) = P(H \leq x | H \leq M_n)\,.$$

Then,

$$\sqrt{v}_n \left(\frac{1}{v_n} \sum_{i=1}^{v_n} \log \frac{Y_{(n,i)}}{Y_{(n,v_n)}} - \frac{1}{\alpha} \right) \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,.$$

Proof. We shall first show that

$$\sqrt{v_n} \left(\frac{1}{v_n} \sum_{i=1}^{n} \delta_{Y_{n,i}/(n/v_n)^{1/\alpha}} \left(y^{-1/\alpha}, \infty\right] - y \right) \Longrightarrow W(y) \tag{5.21}$$

in $D[0,\infty)$, where $D[0,\infty)$ is endowed with the topology of uniform convergence on bounded intervals and W is the standard Brownian Motion on $[0,\infty)$. Let $(\Gamma_i : i \geq 1)$ be the arrivals of a unit rate Poisson process. Define

$$\phi_n(s) := \frac{\Gamma_{n+1}}{v_n} \bar{F}_n \left((n/v_n s)^{1/\alpha} \right)\,,$$

where $\bar{G} := 1 - G$ for any function G. The hypothesis implies that

$$\left(\frac{n}{v_n} \right)^{1/\alpha} \ll M_n$$

and hence for $0 < T < \infty$ and n large enough,

$$\sup_{0 \leq s \leq T} \left| \frac{n}{v_n} \bar{F}_n \left((n/v_n s)^{1/\alpha} \right) - s \right| \leq \frac{1}{1 - M_n^{-\alpha}} \left[T M_n^{-\alpha} + \frac{n M_n^{-\alpha}}{v_n} \right]\,.$$

In view of the hypothesis $n M_n^{-\alpha} \ll \sqrt{v}_n \ll \sqrt{n}$, it follows that

$$\lim_{n\to\infty} \sqrt{v_n} \sup_{0 \leq s \leq T} \left| \frac{n}{v_n} \bar{F}_n \left((n/v_n s)^{1/\alpha} \right) - s \right| = 0\,. \tag{5.22}$$

Also, note that

$$\begin{aligned}
& \sup_{0 \leq s \leq T} \left| \phi_n(s) - \frac{n}{v_n} \bar{F}_n \left((n/v_n s)^{1/\alpha} \right) \right| \\
= & \left| \frac{\Gamma_{n+1}}{n} - 1 \right| \frac{n}{v_n} \bar{F}_n \left((n/v_n T)^{1/\alpha} \right) \\
= & O_p(n^{-1/2}) O(1) \\
= & o_p(v_n^{-1/2})\,.
\end{aligned}$$

This in conjunction with (5.22) shows that

$$\sqrt{v_n}\left(\phi_n(s) - s\right) \longrightarrow 0 \tag{5.23}$$

in probability in $D[0,\infty)$. In particular, this means that ϕ_n converges to the identity in probability. Recall that in $D[0,\infty)$,

$$\sqrt{v_n}\left(\frac{1}{v_n}\sum_{i=1}^{n}\mathbf{1}\left(\Gamma_i \le v_n s\right) - s\right) \Longrightarrow W(s)\,.$$

Hence, it follows by the continuous mapping theorem and Slutsky's theorem that

$$\sqrt{v_n}\left(\frac{1}{v_n}\sum_{i=1}^{n}\mathbf{1}\left(\Gamma_i \le v_n\phi_n(s)\right) - \phi_n(s)\right) \Longrightarrow W(s) \tag{5.24}$$

in $D[0,\infty)$. Let $U_1, U_2, \ldots$ denote i.i.d. standard uniform random variables. Notice that

$$\left(1 - \frac{\Gamma_i}{\Gamma_{n+1}} : 1 \le i \le n\right) \stackrel{d}{=} \left(U_{(n,i)} : 1 \le i \le n\right)\,,$$

where $U_{(n,1)} \ge \ldots \ge U_{(n,n)}$ are the order statistics of $U_1, \ldots, U_n$. The following arguments are similar to those in the proof of Theorem 9.1 in Resnick (2007):

$$\begin{aligned}
\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}\left(y^{-1/\alpha},\infty\right] &= \sum_{i=1}^{n}\mathbf{1}\left(Y_{n,i} > y^{-1/\alpha}(n/v_n)^{1/\alpha}\right)\\
&\stackrel{d}{=} \sum_{i=1}^{n}\mathbf{1}\left(F_n^{\leftarrow}(U_i) > y^{-1/\alpha}(n/v_n)^{1/\alpha}\right)\\
&= \sum_{i=1}^{n}\mathbf{1}\left(F_n^{\leftarrow}(U_{(n,i)}) > y^{-1/\alpha}(n/v_n)^{1/\alpha}\right)\\
&\stackrel{d}{=} \sum_{i=1}^{n}\mathbf{1}\left(F_n^{\leftarrow}\left(1 - \frac{\Gamma_i}{\Gamma_{n+1}}\right) > y^{-1/\alpha}(n/v_n)^{1/\alpha}\right)\\
&= \sum_{i=1}^{n}\mathbf{1}\left(1 - \frac{\Gamma_i}{\Gamma_{n+1}} > F_n\left(y^{-1/\alpha}(n/v_n)^{1/\alpha}\right)\right)\\
&= \sum_{i=1}^{n}\mathbf{1}\left(\frac{\Gamma_i}{\Gamma_{n+1}} < \bar{F}_n\left(y^{-1/\alpha}(n/v_n)^{1/\alpha}\right)\right)\\
&= \sum_{i=1}^{n}\mathbf{1}\left(\Gamma_i < v_n\phi_n(y)\right) \stackrel{a.s.}{=} \sum_{i=1}^{n}\mathbf{1}\left(\Gamma_i \le v_n\phi_n(y)\right)\,.
\end{aligned}$$

This along with (5.23) and (5.24) shows (5.21). An application of Vervaat's lemma shows that

$$\sqrt{v_n}\left(\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}-1\right)\Longrightarrow -W(1) \tag{5.25}$$

jointly with (5.21). This, in particular, shows that

$$\left(\sqrt{v_n}\left\{\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(x,\infty]-x^{-\alpha}\right\},\frac{Y_{(n,v_n)}}{(n/v_n)^{1/\alpha}}\right)$$
$$\Longrightarrow (W(x^{-\alpha}),1)\,,$$

in $D[0,\infty)\times\mathbb{R}$. Since the map from $D[0,\infty)\times\mathbb{R}$ to $\mathbb{R}$ defined by $(x(\cdot),p)\mapsto x(p\cdot)$ is continuous, it follows that

$$\sqrt{v_n}\left\{\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/Y_{(n,v_n)}}(x,\infty]-x^{-\alpha}\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}\right\}$$
$$\Longrightarrow W(x^{-\alpha})\,,$$

in $D[0,\infty)$, jointly with (5.25). From here, we shall proceed as in the proof of Proposition 9.1 in Resnick (2007) to arrive at

$$\left(\sqrt{v}_n\left(\frac{1}{v_n}\sum_{i=1}^{v_n}\log\frac{Y_{(n,i)}}{Y_{(n,v_n)}}-\frac{1}{\alpha}\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}\right),\sqrt{v_n}\left(\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}-1\right)\right)$$
$$\Longrightarrow\left(\int_1^\infty W(x^{-\alpha})\frac{dx}{x},-W(1)\right)\,. \tag{5.26}$$

This implies that

$$\sqrt{v}_n\left(\frac{1}{v_n}\sum_{i=1}^{v_n}\log\frac{Y_{(n,i)}}{Y_{(n,v_n)}}-\frac{1}{\alpha}\right)\Longrightarrow\int_1^\infty W(x^{-\alpha})\frac{dx}{x}-\frac{1}{\alpha}W(1)$$

as desired; hence proving (5.26) suffices. This would be immediate if only the map ψ from $D[0,\infty)$ to $\mathbb{R}$ defined by

$$\psi(x):=\int_1^\infty x(s)\frac{ds}{s}$$

could be applied. This is true because by similar arguments as those in the proof of Theorem 5.2.1, it follows that

$$\psi\left(\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/Y_{(n,v_n)}}(s,\infty]\right)=\frac{1}{v_n}\sum_{i=1}^{v_n}\log\frac{Y_{(n,i)}}{Y_{(n,v_n)}}\,,$$

and that

$$\psi(s^{-\alpha})=\frac{1}{\alpha}\,.$$

Though ψ has bounded support, it is clearly not continuous. Define for $1<T<\infty$, the map ψ_T from $D[0,\infty)$ to $\mathbb{R}$ by

$$\psi_T(x):=\int_1^T x(s)\frac{ds}{s}\,.$$

Clearly ψ_T is a continuous map with compact support and also as $T\longrightarrow\infty$,

$$\psi_T(W(s^{-\alpha}))\Longrightarrow\psi(W(s^{-\alpha}))\,.$$

Thus, to show (5.26), all that remains is to verify that for all $\epsilon>0$,

$$\lim_{T\to\infty}\limsup_{n\to\infty}P\left[\sqrt{v_n}\left|\int_T^\infty\left\{\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/Y_{(n,v_n)}}(s,\infty]-s^{-\alpha}\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}\right\}\frac{ds}{s}\right|>\epsilon\right]=0\,. \tag{5.27}$$

Clearly,

$$\begin{aligned}
&P\left[\sqrt{v_n}\left|\int_T^\infty\left\{\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/Y_{(n,v_n)}}(s,\infty]-s^{-\alpha}\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}\right\}\frac{ds}{s}\right|>\epsilon\right]\\
\le\;&P\left[\sqrt{v_n}\int_T^\infty\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/Y_{(n,v_n)}}(s,\infty]-s^{-\alpha}\frac{n}{v_n}Y_{(n,v_n)}^{-\alpha}\right|\frac{ds}{s}>\epsilon\right]\\
=\;&P\left[\sqrt{v_n}\int_{TY_{(n,v_n)}/(n/v_n)^{1/\alpha}}^\infty\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]-u^{-\alpha}\right|\frac{du}{u}>\epsilon\right]\\
\le\;&P\left[\sqrt{v_n}\int_{T/2}^\infty\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]-u^{-\alpha}\right|\frac{du}{u}>\epsilon\right]\\
&+P\left[Y_{(n,v_n)}/(n/v_n)^{1/\alpha}\le 1/2\right]\,.
\end{aligned}$$

The second term on the right hand side clearly goes to zero, so we forget that. For the first term, note that

$$\begin{aligned}
& P\left[\sqrt{v_n}\int_{T/2}^{\infty}\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]-u^{-\alpha}\right|\frac{du}{u}>\epsilon\right] \\
\leq\ & \frac{\sqrt{v_n}}{\epsilon}E\int_{T/2}^{\infty}\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]-u^{-\alpha}\right|\frac{du}{u} \\
=\ & \frac{\sqrt{v_n}}{\epsilon}\int_{T/2}^{\infty}E\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]-u^{-\alpha}\right|\frac{du}{u} \\
=\ & \frac{\sqrt{v_n}}{\epsilon}\int_{T/2}^{M_n(n/v_n)^{-1/\alpha}}E\left|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]-u^{-\alpha}\right|\frac{du}{u} \\
& +\sqrt{v_n}\int_{M_n(n/v_n)^{-1/\alpha}}^{\infty}u^{-\alpha-1}du \\
\leq\ & \frac{\sqrt{v_n}}{\epsilon}\int_{T/2}^{M_n(n/v_n)^{-1/\alpha}}E\bigg|\frac{1}{v_n}\sum_{i=1}^{n}\delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty] \\
& -\frac{n}{v_n}\bar{F}_n(u(n/v_n)^{1/\alpha})\bigg|\frac{du}{u} \\
& +\frac{\sqrt{v_n}}{\epsilon}\int_{T/2}^{M_n(n/v_n)^{-1/\alpha}}\left|\frac{n}{v_n}\bar{F}_n(u(n/v_n)^{1/\alpha})-u^{-\alpha}\right|\frac{du}{u} \\
& +\sqrt{v_n}\int_{M_n(n/v_n)^{-1/\alpha}}^{\infty}u^{-\alpha-1}du \\
=:\ & I_1+I_2+I_3\,.
\end{aligned}$$

Clearly, by the hypothesis,

$$I_3=O\left(nM_n^{-\alpha}/\sqrt{v_n}\right)=o(1)\,.$$

Some simple algebraic calculations show that

$$\begin{aligned}
I_2 &= \frac{\sqrt{v_n}}{\epsilon}\frac{M_n^{-\alpha}}{1-M_n^{-\alpha}}\int_{T/2}^{M_n(n/v_n)^{-1/\alpha}}\left|u^{-\alpha}-\frac{n}{v_n}\right|\frac{du}{u} \\
&= O\left(M_n^{-\alpha}\sqrt{v_n}+\frac{nM_n^{-\alpha}}{\sqrt{v_n}}\log\left\{M_n\left(\frac{n}{v_n}\right)^{-1/\alpha}\right\}\right) \\
&= o(1)\,.
\end{aligned}$$

Notice that

$$E\left[\frac{1}{v_n}\sum_{i=1}^n \delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]\right] = \frac{n}{v_n}\bar{F}_n(u(n/v_n)^{1/\alpha})\,,$$

and thus, letting C to be a finite positive constant whose value may change from line to line,

$$\begin{aligned} I_1 &\le \frac{\sqrt{v_n}}{\epsilon}\int_{T/2}^\infty \mathrm{Var}\left[\frac{1}{v_n}\sum_{i=1}^n \delta_{Y_{n,i}/(n/v_n)^{1/\alpha}}(u,\infty]\right]^{1/2}\frac{du}{u} \\ &\le C\frac{\sqrt{n}}{\sqrt{v_n}}\int_{T/2}^\infty \left[\bar{F}_n(u(n/v_n)^{1/\alpha})\right]^{1/2}\frac{du}{u} \\ &= C\frac{\sqrt{n}}{\sqrt{v_n}}\int_{T/2}^{M_n(n/v_n)^{-1/\alpha}}\left[\frac{\frac{v_n}{n}u^{-\alpha}-M_n^{-\alpha}}{1-M_n^{-\alpha}}\right]^{1/2}\frac{du}{u} \\ &\le CT^{-\alpha/2}\,. \end{aligned}$$

This shows that

$$\lim_{T\to\infty}\limsup_{n\to\infty} I_1 = 0\,,$$

and thus completes the proof of (5.27), and hence that of (5.26). This completes the proof. □

Proof of Theorem 5.3.3. Define

$$\begin{aligned} U_n &:= \sum_{j=1}^n 1\left(X_j > \gamma M_n\right) \\ \tilde{k}_n &:= \left[n^{1-\beta}U_n^\beta\right] \end{aligned}$$

Since by the assumption $nM_n^{-\alpha} \gg 1$,

$$P(\hat{k}_n \neq \tilde{k}_n) \le P(X_{(1)} \neq M_n) \longrightarrow 0\,,$$

it suffices to show that

$$\tilde{k}_n\left(h(\tilde{k}_n,n)-\frac{1}{\alpha}\right) \Longrightarrow N\left(0,\frac{1}{\alpha^2}\right)\,. \tag{5.28}$$

To this end, we shall show that given any sequence of integers u_n such that

$$\left.\begin{array}{l} k_n := \left[n^{1-\beta} u_n^\beta\right] \gg u_n^2, \\ nM_n^{-\alpha} \ll \sqrt{k_n} \ll \sqrt{n} \\ \text{and } \lim_{n\to\infty} \frac{nM_n^{-\alpha}}{\sqrt{k_n}} \log\left\{M_n\left(\frac{n}{k_n}\right)^{-1/\alpha}\right\} = 0\,, \end{array}\right\} \tag{5.29}$$

$$\left[\tilde{k}_n\left(h(\tilde{k}_n, n) - \frac{1}{\alpha}\right)\middle| U_n = u_n\right] \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,. \tag{5.30}$$

The reason that this suffices for (5.28) is the following. It is easy to see that

$$\frac{U_n}{\gamma^{-\alpha} nM_n^{-\alpha}} \xrightarrow{P} 1$$
$$\frac{\tilde{k}_n}{n\gamma^{-\alpha\beta} M_n^{-\alpha\beta}} \xrightarrow{P} 1\,.$$

Clearly, (5.20) ensures that $\tilde{k}_n \gg U_n^2$, $nM_n^{-\alpha} \ll \sqrt{\tilde{k}_n} \ll \sqrt{n}$ in probability and that

$$\lim_{n\to\infty} \frac{nM_n^{-\alpha}}{\sqrt{\tilde{k}_n}} \log\left(M_n\left(\frac{n}{\tilde{k}_n}\right)^{-1/\alpha}\right) = 0$$

in probability. Recall that any sequence that converges in probability has a subsequence that converges almost surely. Thus, in order to show that proving (5.30) completes the proof of this theorem, we shall show the following. Suppose that if V_n is a sequence of non-negative integer valued random variables such that $V_n \le n$ for all n and

$$\frac{V_n}{nM_n^{-\alpha}} \longrightarrow \gamma^{-\alpha}$$

almost surely, then

$$\left[k_n\left(h(k_n, n) - \frac{1}{\alpha}\right)\middle| V_n = u_n\right] \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,,$$

whenever u_n and k_n are sequences of non-negative integers satisfying (5.29). We shall now show that if this is true, then

$$[n^{1-\beta} V_n^\beta]\left(h([n^{1-\beta} V_n^\beta], n) - \frac{1}{\alpha}\right) \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,. \tag{5.31}$$

Fix $x \in \mathbb{R}$. For $n \geq 1$, define a function f_n on the set of non-negative integers as follows:

$$f_n(v) = \begin{cases} P\left([n^{1-\beta}V_n^\beta]\left(h([n^{1-\beta}V_n^\beta], n) - \frac{1}{\alpha}\right) \leq x \middle| V_n = v\right), & \text{if } P(V_n = v) > 0 \\ 0, & \text{otherwise}\,. \end{cases}$$

Clearly, for $n \geq 1$,

$$P\left([n^{1-\beta}V_n^\beta]\left(h([n^{1-\beta}V_n^\beta], n) - \frac{1}{\alpha}\right) \leq x\right) = Ef_n(V_n)\,.$$

By the assumptions on V_n, it is easy to see that

$$f_n(V_n) \longrightarrow \Phi(\alpha x)$$

almost surely. The bounded convergence theorem implies that (5.31) holds. Thus, proving (5.30) for all sequence u_n satisfying (5.29) completes the proof of this theorem.

To show (5.30), write

$$h(\tilde{k}_n, n) = \frac{1}{\tilde{k}_n}\sum_{i=1}^{U_n}\log\frac{X_{(i)}}{X_{(\tilde{k}_n)}} + \frac{1}{\tilde{k}_n}\sum_{i=U_n+1}^{\tilde{k}_n}\log\frac{X_{(i)}}{X_{(\tilde{k}_n)}} =: \frac{1}{\tilde{k}_n}(S_1 + S_2)$$

Fix any sequence u_n satisfying (5.29) and note that

$$[S_2|U_n = u_n] \stackrel{d}{=} \sum_{i=1}^{k_n-u_n}\log\frac{Y_{(n-u_n,i)}}{Y_{(n-u_n,k_n-u_n)}}\,,$$

where $Y_{(n-u_n,1)} \geq \ldots \geq Y_{(n-u_n,n-u_n)}$ are the order statistics of $(n-u_n)$ i.i.d. random variables generated from the cdf F_n defined as

$$F_n(x) = P(H \leq x | H \leq M_n)\,,$$

and $k_n := [n^{1-\beta}u_n^\beta]$. Since $n \sim n - u_n$ and $k_n \sim k_n - u_n$, it follows that

$$1 \ll (n-u_n)M_n^{-\alpha} \ll \sqrt{k_n - u_n} \ll \sqrt{n - u_n}\,,$$

and

$$\lim_{n\to\infty} \frac{(n-u_n)M_n^{-\alpha}}{\sqrt{n-k_n}} \log\left\{M_n\left(\frac{n-u_n}{k_n-u_n}\right)^{-1/\alpha}\right\} = 0\,.$$

Thus, the hypotheses of Lemma 5.3.2 are satisfied. Hence,

$$\left[\sqrt{k_n}\left(\frac{1}{k_n-u_n}S_2 - \frac{1}{\alpha}\right)\middle| U_n = u_n\right] \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,.$$

This along with the fact that

$$\sqrt{k_n}S_2\left(\frac{1}{k_n-u_n} - \frac{1}{k_n}\right) = \frac{S_2}{k_n-u_n}\frac{u_n}{\sqrt{k_n}} = O_p(1)o(1)$$

shows

$$\left[\sqrt{k_n}\left(\frac{1}{k_n}S_2 - \frac{1}{\alpha}\right)\middle| U_n = u_n\right] \Longrightarrow N\left(0, \frac{1}{\alpha^2}\right)\,. \tag{5.32}$$

It's easy to see that since $1 \leq X_{(i)} \leq M_n$,

$$\begin{aligned} \frac{S_1}{\sqrt{\tilde{k}_n}} &\leq \frac{U_n \log M_n}{\sqrt{\tilde{k}_n}} \\ &= O_p\left(n^{1/2} M_n^{-\alpha(1-\beta/2)} \log M_n\right) \\ &= o_p(1) \end{aligned} \tag{5.33}$$

where the last step follows from (5.20). Clearly, (5.32) and (5.33) show (5.30) and thus complete the proof. □

CHAPTER 6

DATA ANALYSIS

6.1 Introduction

In this chapter we applied the statistical methods of Chapter 4 to two data sets. One data set contains "think times", or delays (in microseconds) between successive request/response exchanges between hosts using a TCP connection. The second data set contains the sizes (in bytes) of objects (files, HTTP responses, email messages, etc.) transferred on TCP connections. Both data sets were acquired by monitoring between 1:30 PM and 2:30 PM on July 24, 2006, the communication links connecting the site of a large commercial enterprize to the Internet. Both data sets exhibit visual evidence of heavy tails, and the Hill estimator confirms that (see below). Our goal is to check if the data sets show statistical evidence of soft or hard truncation of heavy tails.

6.2 Think Times

This data sets contains 2.1×10^7 observations which are plotted on Figure 6.1.

Clearly, the nature of this data set changes over time, and the nature of truncation of heavy tails may potentially change as well. In order to study this effect we have broken the data set into four pieces, with corresponding ranges $\left[0.11 \times 10^7, 0.64 \times 10^7\right]$; $\left[0.8 \times 10^7, 1.6 \times 10^7\right]$; $\left[1.7 \times 10^7, 1.9 \times 10^7\right]$ and $\left[1.95 \times 10^7, 2.1 \times 10^7\right]$. The individual pieces are plotted on Figure 6.2

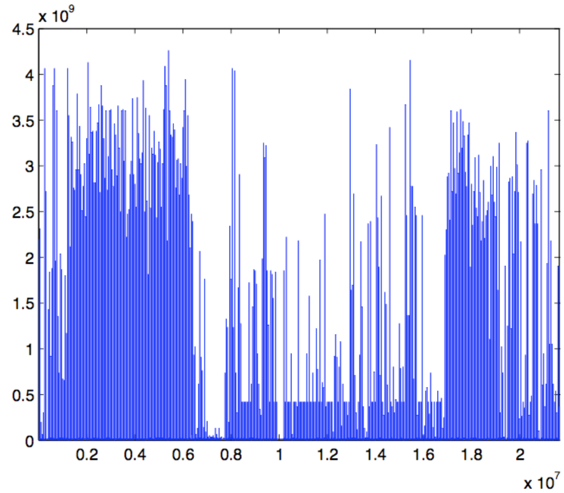

Figure 6.1: Think Times - the entire data set

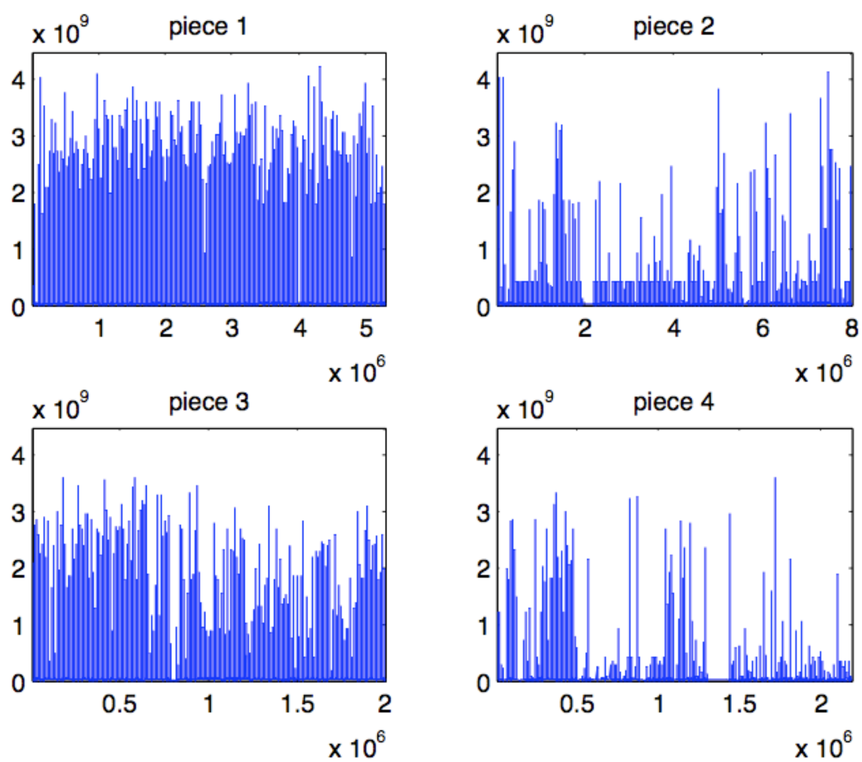

Figure 6.2: Think Times - the different pieces

Table 6.1: Upper bound of α for the different pieces of Think Times

piece	A
1	3.02
2	2.30
3	0.85
4	2.24

The structure of the 4 individual pieces appears to be more stable than that of the entire data sets, and we proceed to analyze each piece separately. To do that, we first ran the Hill estimator with random k given in (5.8) on the first half of each of the 4 pieces. The estimation was conducted using $\beta, \gamma = 0.3, 0.4, 0.5, 0.6, 0.7$ and conservative upper bounds for α were obtained; these are presented in Table 6.1.

We then proceeded to use the second halves of each piece of the Think Times data set to test for soft and hard truncations.

Testing the hypothesis of soft truncation

The test statistic $Z_n(A_1)$ of Section 4.2 was computed for various values of A_1 larger than A. The results are reported in the Table 6.2. Comparing the resulting values of the test statistic with the corresponding quantiles (or their upper bounds) of $Z(A/A_1)$, it is clear that the null hypothesis of soft truncation can be rejected for pieces 1 and 3. For piece 2, there is some evidence against the null hypothesis of hard truncation, while for piece 4 no such evidence exists.

Table 6.2: $Z_n(A_1)$ for the different pieces of Think Times

A/A_1	piece 1	piece 2	piece 3	piece 4
0.5	31.43	5.81	154.05	3.57
0.6	51.59	7.99	205.37	4.72
0.7	77.39	10.74	271.27	6.11
0.8	108.08	14.20	361.74	7.81
0.9	142.78	18.57	491.31	9.91
0.95	161.38	21.16	576.73	11.13

Table 6.3: P-values corresponding to $Z_n(A;\gamma)$ for the different pieces of Think Times

γ	piece 1	piece 2	piece 3	piece 4
0.1	0.85	0.72	0.88	0.33
0.2	0.83	0.98	0.38	0.57
0.3	0.97	0.99	0.79	0.68
0.4	0.94	0.68	0.39	0.43
0.5	0.83	0.63	0.94	0.47
0.6	0.97	0.89	0.83	0.27
0.7	0.91	0.88	0.87	0.40
0.8	0.64	0.85	0.80	0.33
0.9	0.70	0.37	0.85	0.40

Testing the hypothesis of hard truncation

The test statistic $Z_n(A;\gamma)$ of Section 4.3 was computed for various values of γ. The resulting p-values are reported in Table 6.3. Clearly, the hypothesis of hard

Table 6.4: P-values corresponding to $Z_n(A)$ for the different pieces of Think Times

ϵ	piece 1	piece 2	piece 3	piece 4
0.1	1.00	1.00	1.00	1.00
0.2	1.00	1.00	1.00	1.00
0.3	1.00	1.00	1.00	0.91
0.4	1.00	0.73	1.00	0.45

truncation cannot be rejected for any of the four pieces.

Testing a stronger version of the hypothesis of hard truncation

The test statistics $Z_n(A)$ of Section 4.4 was computed and the corresponding p-values calculated for various values of ϵ. These are listed in Table 6.4. It is clear that even the stronger version of the hypothesis of hard truncation cannot be rejected.

6.3 Object Sizes

This data set contains 2.2×10^7 observations. It is plotted in Figure 6.3. It does not appear that the nature of the observations changes with time, so we applied our statistical tests to the entire data set. After running the Hill estimator with random k and parameters β and γ as above, on the first half of the data set, we obtained a conservative upper bound on the value of the tail exponent α; this turned out to be $A = 1.69$. We used the second half of the Object Sizes data set

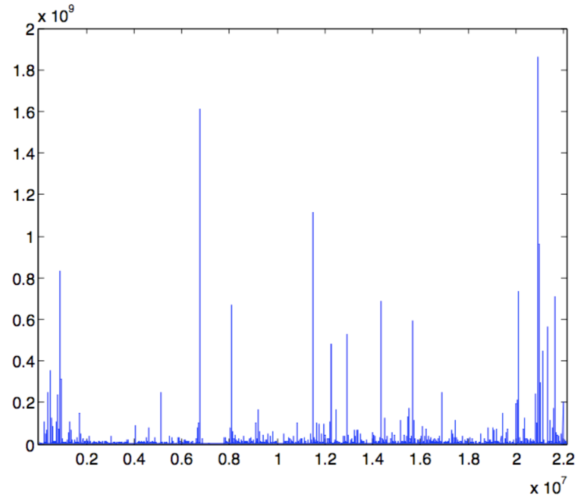

Figure 6.3: Data on Object Sizes

to test for soft and hard truncations.

Testing the hypothesis of soft truncation

We evaluated the test statistic $Z_n(A_1)$ of Section 4.2 for a range of values of A_1 larger than A. The results are reported in Table 6.5. Comparing these with the corresponding quantiles (or their upper bounds) of $Z(A/A_1)$, we see that the hypothesis of soft truncation cannot be rejected.

Testing the hypothesis of hard truncation

We evaluated the test statistic $Z_n(A;\gamma)$ of Section 4.3 for various values of γ, and the obtained p-values are reported in Table 6.6. The null hypothesis of hard truncation cannot be rejected.

Table 6.5: $Z_n(A_1)$ for Object Sizes

A/A_1	$Z_n(A_1)$
0.5	1.75
0.6	2.32
0.7	3.09
0.8	4.10
0.9	5.42
0.95	6.23

Table 6.6: P-values corresponding to $Z_n(A;\gamma)$ for Object Sizes

γ	p-value
0.1	0.50
0.2	0.36
0.3	0.73
0.4	0.77
0.5	0.95
0.6	0.94
0.7	0.94
0.8	0.97
0.9	0.72

Table 6.7: P-values corresponding to $Z_n(A)$ for Object Sizes

ϵ	p-value
0.1	1.00
0.2	0.86
0.3	0.33
0.4	0.08

Testing a stronger version of the hypothesis of hard truncation

We calculated the test statistics $Z_n(A)$ of Section 4.4 for various values of ϵ, and the p-values are given in Table 6.7. The strengthened hypothesis of hard truncation becomes suspicious for $\epsilon = 0.4$, but overall our statistical tests do not produce clear evidence of the level of truncation for the Object Sizes data set.

BIBLIOGRAPHY

Aban, I., Meerschaert, M., and Panorska, A. (2006). Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, 101(473):270–277.

Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York.

Asmussen, S. and Pihlsgard, M. (2005). Performance analysis with truncated heavy-tailed distributions. *Methodology and Computing in Applied Probability*, 7(4):439–457.

Barthelemy, P., Bertolotti, J., and Wiersma, S. (2008). A Lévy flight for light. *Nature*, 453:495–498.

Bartumeus, F., da Luz, M., Vishwanathan, G., and Catalan, J. (2005). Animal search strategies: a quantitative random walk analysis. *Ecology*, 86(11):3078–3087.

Beirlant, J., Guillou, A., Dieckx, G., and Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, 10(3):151–174.

Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439:462–465.

Burrooughs, S. and Tebbens, S. (2001). Upper-truncated power laws in natural systems. *Pure and Applied Geophysics*, 158(4):741–757.

Cline, D. and Hsing, T. (1991). Large deviation probabilities for sums and maxima of random variables with heavy or subexponential tails. Preprint, Texas A&M University.

Corral, A. (2006). Universal earthquake-occurence jumps, correlations with time, and anomalous diffusion. *Physical Review Letters*, 97(17):178501.

de Acosta, A. and Giné, E. (1979). Convergence of moments and related functionals in the general central limit theorem in banach spaces. *Z. Wahr. verw. Geb.*, 48:213–231.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer-Verlag, New York.

Einmahl, J., Fils-Villetard, A., and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli*, 14(1):207–227.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, volume 2. Wiley, New York, 2nd edition.

Gomez, C., Selman, B., Crato, N., and Kautz, H. (2000). Heavy–tailed phenomena in satisfiability and constraint satisfaction problems. *Journal of Automated Reasoning*, 24(1-2):67–100.

Gut, A. (2005). *Probability: A Graduate Course*. Springer.

Heyde, C. (1968). On large deviation probabilities in the case of attraction to a non-normal stable law. *Sankyā Ser. A*, 30:253–258.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174.

Hong, S., Rhee, I., Kim, S., Lee, K., and Chong, S. (2008). Routing performance analysis of human-driven delay tolerant networks using the truncated levy walk model. In *Mobility Models '08: Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models*, pages 25–32, New York. ACM.

Hult, H., Lindskog, F., Mikosch, T., and Samorodnitsky, G. (2005). Functional large deviations for multivariate regularly varying random walks. *Annals of Applied Probability*, 15(4):2651–2680.

Jelenković, P. R. (1999). Network multiplexer with truncated heavy-tailed arrival streams. *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2:625–632.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Verlag.

Maruyama, Y. and Murakami, J. (2003). Truncated Lévy walk of a nanocluster bound weakly to an atomically flat surface: Crossover from superdiffusion to normal diffusion. *Physical Review B*, 67(8):085406.

Microsoft Knowledge Base Article 154997 (2007). Description of the FAT32 file system.

Mikosch, T. (2009). *Non-Life Insurance Mathematics: An Introduction with the Poisson Process*. Springer, Berlin, 2nd edition.

Mikosch, T. and Nagaev, A. V. (1998). Large deviations of heavy-tailed sums with applications to insurance. *Extremes*, 1:81–110.

Nagaev, A. (1969a). Integral limit theorems for large deviations when cramér's condition is not fulfilled i,ii. *Theory of Probability and Its Applications*, 14:51–64 and 193–208.

Nagaev, A. (1969b). Limit theorems for large deviations where cramér's conditions are violated. *Izv. Akad. Nauk UzSSR Ser. Fiz.–Mat. Nauk*, 6:17–22. In Russian.

Nagaev, S. (1979). Large deviations of sums of independent random variables. *Annals of Probability*, 7:745–789.

Prokhorov, Y. (1959). An extremal problem in probability theory. *Theory of Probability and Its Applications*, 4:201–204.

Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.

Resnick, S. (2007). *Heavy-Tail Phenomena : Probabilistic and Statistical Modeling*. Springer, New York.

Rosiński, J. (1990). On series representation of infinitely divisible random vectors. *The Annals of Probability*, 18:405–430.

Rvačeva, E. (1962). On domains of attraction of multi-dimensional distributions. *Selected Translations in Mathematical Statistics and Probability*, 2:183–205. Publisher: IMS-AMS.

Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Scholtz, C. and Contreras, J. (1998). Mechanics of continental rift architecture. *Geology*, 26(11):967–970.

Serrano, M., Flammini, A., and Menczer, F. (2009). Beyond Zipf's law: Modeling the structure of human language. Technical Report.

Varadhan, S. R. S. (1984). *Large Deviations and Applications*. SIAM, Philadelphia.

Zaninetti, L. and Ferraro, M. (2008). On the truncated Pareto distribution with applications. *Central European Journal of Physics*, 6(1):1–6.