Limit laws of weighted power sums of extreme values and Statistical analysis of partition lattices

Ph.D. Thesis

by

LILLIAN ACHOLA OLUOCH

Thesis advisors:

Professor Dr. László Viharos,
Professor Dr. László Zádori

Doctoral School of Mathematics and Computer Science, Bolyai Institute, University of Szeged, Faculty of Science and Informatics

Szeged, 2021

Contents

List of Figures

List of Tables

Preface

This doctoral dissertation is submitted to the Bolyai Institute of the University of Szeged in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy. It contains the results of the following papers [**26, 56, 63, 64**], which were written in collaboration with different co-authors. The content is essentially divided into five chapters addressing both theoretical statistics and applied statistics. The theoretical statistics covers classes of estimators for the tail index and their asymptotic properties. Meanwhile, the applied statistics comprises of application of statistical methods to partition lattices in Algebra. This dissertation does not contain all the scientific papers of the author, but only gives elaborate content related to the two sections of theoretical and applied statistics. The problems that this dissertation intends to address all have their individual histories, therefore separate introductions to each part will be given. In the theoretical section, we address some problems related to the occurrence of rare events.

It appears that we live in an age of disasters like earthquakes, hurricanes, floods, Tsunami, and more recently, the Covid-19 pandemic. The latter has caused devastating global public health impact as well as the noticeable impact on the global economic growth causing much concern to statisticians like the author. These are surprising phenomena that may have no definite rule according to a layman but are happening according to well-defined rules of science. A typical question may be: to what extent can something go wrong? This question can be adequately addressed by observations and estimations of rare occurrences that are considered as extreme events. Extremes are rare or unusual events that occur in real life and are either very small or very large. They are often labelled as outliers and even ignored in classical data analysis due to lack of representativeness and the influence they may have on the measures of central tendency and deviations. Extremes possess heavy tails in real-life situations that no classical distributions predict, but a particular branch of probability theory, the Extreme Value Theory (EVT), offers some in modelling such events. This is a concept that has been established on sound statistical methods that deal with modelling outcomes of extreme events by providing tools for investigating asymptotics of maxima of a sequence of random variables. Estimation of the probability of such events has experienced some difficulties due to the scarcity of data from the rare events. The main objective of estimating this distribution parameter is to enable forecasting the future periods by focusing directly on the tails disregarding the centre of the distribution. This concept forms the theoretical part of this dissertation which will be addressed in chapters 2, 3, and 4.

The applied part of this dissertation involves the practical approach that required a practical application of statistics to lattice theory. It links basic statistics to Algebra. In the friendly atmosphere of Bolyai Institute, my interest in applying mathematical statistics to various fields, ranging from dentistry and epidemiology to zoology, was widely known. This explains that soon I found myself in professor László Zádori's realm, who called my attention to some ongoing research to enumerate some mathematical objects. To find the exact (very large) number $\nu(n)$ of these objects was hopeless both theoretically and in a computer-aided way, but professor László Zádori pointed out that these objects belong to a larger set H with known number of elements. Hence, finding $\nu(n)$ is equivalent to finding the probability that a randomly selected member of H is one of the objects we are interested in. So it was reasonable to take a large sample by computer programs and letting statistics to draw appropriate conclusions. These conclusions are now drawn in Section 3 of Chapter 5. While it is not at all typical to state lattice theoretical statements only with some probabilistic confidence level, including probabilities in information theory is quite standard where the reliability of error-correcting or cryptographic methods are dealt with. But now that one of professor Czédli's recent papers has pointed out that the lattice theoretical objects we deal with have connection with information theory, our goal to take a statistical approach to them becomes less strange.

Later, it was natural to go beyond the original target. Namely, the computer programs we had developed proved to be useful in solving one half of professor László Zádori's 38-year-old problem. These programs also helped in finding the theoretical approach to the other half; see Section 2 of Chapter 5 here. I was privileged to publish my results jointly with professor Gábor Czédli. He put a deep purely lattice theoretical theorem into our joint paper. Of course, this theorem is not in the present dissertation, but Section 2 of my joint paper with Amenah Al-Najafi [**63**] relies heavily on it; see Theorem 4.10 of Section 4 here.

The dissertation is organized as follows: Chapter 2 gives an overview of important literature contributing to both EVT and tail index estimation. In Chapter 3, attention is paid to limit theorems for weighted power sums of extreme values. The results are applied to construct a new class of tail index estimators. Then in Chapter 4, the asymptotic properties of norms of extremal samples are investigated. Here the Gaussian and non-Gaussian limits are obtained. The results are applied to tail index estimation. Finally, Chapter 5 covers application of statistics to partition lattices and some lattice theoretical arguments that "grew out" from this application. In particular, four-element generating sets are presented and a lower bound for the number of four-element generating sets of direct products of two neighbouring partition lattices is established.

CHAPTER 1

Introduction

Normal distributions are usually used in probabilistic modelling phenomena in the real world. One of the importance of this distribution is due to its bell shape. It fits well the empirical data that clusters around the mean value. Normal distributions also possess some desirable mathematical properties.

However, there exist many other phenomena whose distributions deviate from the normal distributions. Some of these include financial assets, fire insurance losses, telecommunications, natural calamities like earthquakes, transmission rates of files and file sizes stored on a server, and environmental science. All these phenomena exhibit data with large values with a high probability, hence they are described as heavy-tailed distributed. In other words, they follow a power-law distribution. The tail index is the shape parameter of these heavy-tailed distributions. The larger the tail index is, the heavier the distributional tail and the more the rare events. Thus estimation of the tail index has attracted much attention in the literature. Moreover, various parametric or semiparametric estimators have been developed from the upper order statistics.

However, one of the drawbacks in the estimation of the tail index is that in most cases, only the information on the maximum value occurring is available for analysis. For example, only the information on a few highest quoted prices is reported to the public for financial data. Whereas in meteorological data, only the highest and lowest temperatures of each day are forecasted. A series of catastrophes have hit the world in recent years. A very recent example is the Covid-19 pandemic in the year 2020. These are indicators that it is crucial nowadays to take also extreme occurrences into account. Although they rarely occur, their consequences are dramatic when they hit unprepared societies. Unfortunately, there are no classical distributions that can predict the tails of these extremes events. Hence, EVT offers an amicable solution to such problems. EVT being a wonderfully much-celebrated theory, it has already been adopted by a wide variety of disciplines. Its first application was to offer solutions to some environmental problems. Later, it rapidly became popular in the finance industry. In recent years, internet traffic and structural reliability are other prime targets for applications of EVT. There exists a very long list of areas in that EVT has played decisive roles in statistical applications. These include but not limited to Structural engineering, Meteorology, Highway traffic as outlined in [**14**], Insurance, and stock markets outlined in [**84, 49**]. While the perfect knowledge of the incidence and occurrence of extreme events is desirable, the application of EVT poses many challenges. The most pertinent of them is the choice of the method of estimation of parameters. Therefore, one of the purposes of this thesis is to focus on estimation of tail index.

Use of adequate and accurate statistical methods is of major importance in all areas of application. The R software (R Development Core Team) has been conveniently used in this dissertation in Chapter 3 and 4 in simulation and analysis of simulated data. This being an open source environment, it incorporates a huge amount of statistical packages and are freely available by the scientific community.

Literature review

It is believed that EVT is originated from an analysis in the field of astronomy. Here, a decision had to be made on whether outliers in the data were essential or not [**57**]. It is confirmed that the oldest mathematical models of extreme value date back to 1925, where Tippet and Fréchet [**83, 39**] investigated the asymptotic behaviour of extreme order statistics. A further step involved dividing the limit distribution into three families of generalized extreme value distribution. The first one was the exponential type distributions wherein the max-domain of attraction of Gumbel's distribution was obtained by [**45**]. The second type was based on the work of Fréchet [**39**]. The last one was identified by Fréchet and was later described in detail by Waloddi Weibull [**90**]. Remarkable advancements in these models have been witnessed. Subsequently, in 1958, Gumbel [**46**] presented the ordered statistics and their exceedance, the distribution of extremes, and their asymptotic distributions. Recently, so much research has covered many areas of EVT encompassing the asymptotic distributions and limit law of extreme value that are the main concerns in this dissertation. Extensive literature has shown that various asymptotic models for extreme values have been developed, see [**16**]. A model that emphasized the availability of several dependent extreme value models was advanced by Galambos [**41**]. Moreover, an exponential model was introduced by Beirlant et al. [**5**] and Feuerverger et al. [**38**] for the extreme order statistics from a Pareto-type distribution. Additionally, a proposal was made by Leadbetter [**60**] on asymptotic distributions of extreme order statistics with relation to Poisson convergence theory. Furthermore, the relationship between the limit distribution and the joint distribution of the sample extremes was established by Weissman [**91**].

Several studies have been advanced to approaches to the estimation of extremal properties and developments on the estimation parameters of extremes in [**40, 52**]. The estimation methods for the number of extreme order statistics that are in the tail was proposed by Danielsson et al. [**27**], of which many applications of stochastic models find vital. Consequently, the class of estimators for the extreme value index has since been derived. These include well-known estimators for the index of a distribution function such as the Hill [**50**], the moment estimator [**32**] and the Pickands estimator [**69**], etc. Aspects, properties, and generalizations of the Hill estimator have been studied and developed in recent researches whereby the asymptotic properties, asymptotic normality, and the volatility of the index have been discussed in [**19, 30, 32, 44, 47**]. It suffices to show that one limitation of the Hill estimator is that it is known to produce poor results in certain situations leading to the Hill horror plots. Different approaches have been advanced to solve this problem by smoothing the Hill estimator [**21, 74**]. Some of the proposed methods depend on a subjective choice of the threshold or the number of extreme value statistics. Thorough consideration is needed since selecting a threshold too high could result in a high variance of the estimator while setting a too low threshold could cause biasedness of the estimator. On the other hand, the Pickands estimator in [**69**], which estimates the distribution parameters by using a simple percentile method, is known to be quite volatile as a function of several order statistics and displays large asymptotic variance and poor efficiency [**93**]. Detailed studies have been performed towards the improvement of the Pickands estimator; see [**13, 76**]. Likewise, in 1989 Dekkers et al. [**32**], using the foundations of the Hill estimator proposed a moment estimator of which required a problematic choice of a number $k = k(n)$ of upper order statistics whereas a few reliable guidelines on this exist [**74**]. Meanwhile, its asymptotic properties were investigated by Beirlant et al.[**4**]. The consistency and asymptotic normality of these estimators have been proved [**30**]. The results indicated no superiority by any estimator as the outcome depended on the distribution parameters in different situations.

General classes of estimators for the tail index of a distribution with a regularly varying upper tail have been developed in various studies. These include the class of kernel estimators by Csörgő et al. [**21**]. The class of universally asymptotically normal weighted doubly logarithmic least-squares estimators was investigated by Viharos [**88**]. Moreover, another class of estimators as scale-invariant functions was represented by Drees [**34**] who introduced applicable methods for constructing estimators having prescribed asymptotic behaviour. More studies have emanated which involves investigating consistency and asymptotic normality of the estimates;

see [**6, 33, 79**]. Norms of extreme values are the targets of another interesting study in EVT. Some in-depth research has also been conducted on the norms of extreme values and their associated limit laws. Central limit theorems were combined with limit theorems by Schlather [**75**] through parameterizing limit theorems. He established the limit laws for the normalized L_p norms. Moreover, the limit distribution of L_p norms of independently and identically distributed (iid) random variables was looked into by Bogachev [**11**] as the sample size n approaches infinity. On the other hand, Biau and Mason [**8**] investigated some new asymptotic properties of the maximum of norms. Limit behaviour of power sum and norms of iid positive samples was suggested by Janßen [**54**] that followed the combination of limit theorems for sums and maxima. Investigation of the limiting distribution of exponential sums was performed by Arous [**2**]. He also explored possible phase transitions due to the growth rate of the parameter n.

CHAPTER 2

Preliminaries

In this chapter we provide the basic definitions and theorems which will be useful as auxiliary facts for the subsequent two chapters. We review some of the statistical background and terminology that will be prevalent in the study of the asymptotic theory of extremal values. This chapter is essentially a reprint of [**2, 29, 53, 72**].

1. Mathematical overview

We introduce the terminologies, review extreme value theory, in particular, we consider the basic concepts of the asymptotic theory of extremes which provides the background for asymptotic statistics.

EVT characterizes the stochastic behaviour of extremes values and rare events. It mainly focuses on the tail of the underlying distribution and therefore essential to adequately test the shape of the tail as it has influence on the estimation of parameters of extremals.

The asymptotic theory provides the necessary and sufficient conditions that ensure the approximation of the probability distribution of the sample maxima.

Let $X, X_1, X_2, \ldots$ be independent identically distributed (iid) random variables with common distribution function $F(x) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$. EVT is concerned with the asymptotic distribution of the suitably centered and normed maxima $M_n := \max(X_1, X_2, \ldots, X_n)$ as $n \to \infty$. The cumulative distribution function (CDF) of M_n is

$$P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x) = P(X_1 \leq x) \cdots P(X_n \leq x) = F^n(x).$$

This distribution function depends on the distribution of X_1, which is usually unknown. Therefore, it is important to consider the asymptotic distribution of M_n as $n \to \infty$.

Suppose there exist sequences of real numbers $a_n > 0$ and b_n and a nondegenerate distribution function G such that

$$P\left(\frac{M_n - b_n}{a_n}\right) = F^n(a_n x + b_n) \to G(x) \tag{2.1}$$

as $n \to \infty$ at every point x of G. Theorem 1.1 below describes the possible limiting distributions in (2.1). These distributions are called extreme value distributions.

The set of distributions F satisfying (2.1) is called the *maximum domain of attraction* of G. The next theorem and the discussion following it show that every extreme value distribution is of one of the three possible types.

THEOREM 1.1 (Fisher and Tippett (1928), Gnedenko (1943)). *The class of extreme value distributions is $G_\gamma(ax+b)$ with $a>0$, $b\in\mathbb{R}$, where*

$$G_\gamma(x) = exp\Big\{-\Big(1+\gamma x\Big)^{-1/\gamma}\Big\}, \quad 1+\gamma x>0, \tag{2.2}$$

with γ real, where for $\gamma = 0$ the right-hand side is interpreted as $\exp(-e^{-x})$.

The parameter γ is called the *extreme value index* that controls the shape of the extreme value distribution.

The parametrization in Theorem 1.1 is due to von Mises (1936) and Jenkinson (1955). An alternative parametrization is the following:

1. For $\gamma > 0$ if we use $G_\gamma((x-1)/\gamma)$ with $\alpha = 1/\gamma$, then we obtain

$$\Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ \exp\{-x^{-\alpha}\}, & x > 0. \end{cases}$$

This class is called the Fréchet class of distributions.

2. For $\gamma < 0$ if we use $G_\gamma(-(1+x)/\gamma)$ with $\alpha = -1/\gamma$, then we obtain

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x < 0 \\ 1, & x \ge 0. \end{cases}$$

This class is called the reverse-Weibull class of distributions.

3. The distribution function G_γ with $\gamma = 0$,

$$G_0(x) = \exp(-e^{-x}),\ x\in\mathbb{R},$$

is the Gumbel distribution function.

2. Maximum domain of attraction

DEFINITION 2.1. A measurable function $f:\mathbb{R}^+\to\mathbb{R}$ that is eventually positive is *regularly varying* at infinity if

$$\lim_{x\to\infty}\frac{f(ux)}{f(x)} = t^\gamma, \quad t>0, \tag{2.3}$$

for some $\gamma\in\mathbb{R}$.

The number γ in (2.3) is called the *index* of regular variation. If $\gamma = 0$ in (2.3), then f is said to be slowly varying.

The next theorem describes the maximum domain of attraction of the extreme value distributions.

THEOREM 2.2 (de Haan and Ferreira [**29**], Theorem 1.2.1). *The distribution function F is in the maximum domain of attraction of the extreme value distribution G_γ if and only if*

(1) for $\gamma > 0 :\ x^* = \sup x : F(x) < 1$ is infinite and

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(x)} = x^{-1/\gamma}$$

for all $x > 0$, i.e. the function $1-F$ is regularly varying at infinity with index $-1/\gamma$;

(2) for $\gamma < 0 :\ x^*$ is finite and

$$\lim_{t\downarrow\ 0} \frac{1-F(x^*-tx)}{1-F(x^*-t)} = x^{-1/\gamma}$$

for all $x > 0$;

(3) for $\gamma = 0 :\ x^*$ can be finite or infinite and

$$\lim_{t\uparrow\ x^*} \frac{1-F(t+xf(t))}{1-F(t)} = e^{-x} \tag{2.4}$$

for all real x, where f is a suitable positive function. If (2.4) holds for some f, then $\int_t^{x^*}(1-F(s))ds < \infty$ for $t < x^*$ and (2.4) holds with

$$f(t) := \frac{\int_t^{x^*}(1-F(s))ds}{(1-F(t))}. \tag{2.5}$$

In this thesis we are concerned with the case $\gamma > 0$. Our main goal is to make inference on the tail of a distribution.

3. Heavy-tailed distribution

Heavy-tailed probability distributions are crucial components of stochastic modelling. A distribution has heavy right tail, if the probability of a huge value is relatively big. Heavy right tails are usually modelled by assuming that the distribution function F has regularly varying upper tail:

$$P(X > x) = 1 - F(x) = x^{-1/\gamma}L(x), \tag{2.6}$$

where L is a slowly varying function at infinity and $\gamma > 0$. In this thesis we assume (2.6) to make inference on the tail of the distribution.

Heavy-tail analysis deals with systems whose behaviour is controlled by large values which periodically impact the system, unlike many other systems whose stability is determined largely by an averaging effect. There has been extensive application of the heavy-tailed distributions in modelling in many areas which include actuarial sciences, economics, risk management and internet traffic. Some scenerios have been outlined by Resnick [**72**] where heavy-tailed analysis have been used. These examples includes: data network study which was performed by the Boston University, the Standard & Poors 500 stock market index and the Danish fire insurance losses. Other examples include magnitudes of earthquakes and floods [**59**] and returns on financial markets [**37**].

4. Classical extreme value index estimators

The extreme value index γ (EVI), measures the heaviness of the tail. The higher values of γ infers heavier right tails. Therefore, estimating γ is crucial in many applications of stochastic models. A number of estimators have been designed for this purpose. These includes but not limited to the Hill estimator, the moment estimator, the generalized Hill estimator, the Pickands estimator and the mixed moment estimator. The best known estimator is the Hill estimator, which was introduced by Hill (1975). Among the aforementioned estimators, we consider the Hill, the moment and the Pickands estimators in this dissertation. In Chapter 3 we compare these estimators with a new class of estimators based on weighted power sum of extreme values.

4.1. The Hill estimator. Let $X_1, X_2, \ldots$ be iid random variables and for each integer $n \geq 1$ let $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics pertaining to the sample $X_1, \ldots, X_n$. The Hill estimator is based on k upper order statistics defined as

$$M_n^{(1)} := \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}, \quad 1 \leq k < n.$$

Consistency of the Hill estimator can be proved if the sequence $k = k_n$ satisfies

$$k \to \infty \quad \text{and} \quad k/n \to 0 \quad \text{as } n \to \infty.$$

4.2. The moment estimator. Define

$$M_n^{(2)} := \frac{1}{k} \sum_{i=0}^{k-1} \big(\log X_{n-i,n} - \log X_{n-k,n} \big)^2.$$

The moment estimator is the following combination of the Hill estimator and the statistic $M_n^{(2)}$:

$$\widehat{\gamma}_n^{(M)} := M_n^{(1)} + 1 - \frac{1}{2}\Big\{1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\Big\}^{-1}.$$

The moment estimator can be used to estimate the shape parameter of any of the three extreme value distributions unlike the Hill estimator.

4.3. The Pickands estimator. The Pickands estimator is based using the order statistics $X_{k,n}$, $X_{2k,n}$, and $X_{4k,n}$. The estimator is

$$\widehat{\gamma}_n^{(P)} := \frac{1}{\log 2} \log \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}, \quad 1 \le k \le \lceil n/4 \rceil,$$

where $\lceil x \rceil$ denotes the integer part of x. The Pickands estimator also can be used to estimate any $\gamma \in \mathbb{R}$, though it suffers from high volatility.

Asymptotic properties of the index estimators were established by several authors under different conditions on the underlying distribution and $k = k_n$ (see the discussion in Chapter 4).

CHAPTER 3

Asymptotic distributions for weighted power sums of extreme values

This chapter is based on a joint paper [**64**] by the author and L. Viharos. Here we considered proving the asymptotic normality for the weighted power sums over the whole heavy-tail model under some constraints on the weights $d_{i,n}$. The results obtained are crucial in the construction of a new class of estimators for the parameter γ.

1. Formulation of the weighted power sums

Let $X, X_1, X_2, \ldots$ be independent random variables with a common distribution function $F(x) = P\{X \leq x\},\ x \in \mathbb{R}$, and for each integer $n \geq 1$ let $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics pertaining to the sample $X_1, \ldots, X_n$. For a constant $\gamma > 0$, let $\mathcal{R}_\gamma$ be the class of all probability distribution functions F such that

$$1 - F(x) = x^{-1/\gamma} L(x), \quad 0 < x < \infty,$$

where L is a function slowly varying at infinity. Without loss of generality we assume that $F(1-) = 0$ for all $F \in \mathcal{R}_\gamma$. If $Q(\cdot)$ denotes the quantile function of F defined as

$$Q(s) = \inf\{x : F(x) \geq s\},\ 0 < s \leq 1, \quad Q(0) = Q(0+),$$

then $F \in \mathcal{R}_\gamma$ if and only if

$$Q(1-s) = s^{-\gamma} \ell(s), \tag{3.1}$$

where ℓ is a slowly varying function at 0. Let k_n be a sequence of integers such that

$$1 \leq k_n < n, \quad k_n \to \infty \quad \text{and} \quad k_n/n \to 0 \quad \text{as } n \to \infty. \tag{3.2}$$

For some constants $d_{i,n}$, $1 \leq i \leq n$, consider the weighted power sums of the extreme values $X_{n-k_n+1,n}, \ldots, X_{n,n}$:

$$S_n(p) := \sum_{i=1}^{k_n} d_{n+1-i,n} \log^p X_{n+1-i,n},$$

where $p > 0$ is a fixed number. Our aim is to study the asymptotic behavior of $S_n(p)$ as $n \to \infty$ whenever $F \in \mathcal{R}_\gamma$.

Csörgő et al. [**20**] found necessary and sufficient conditions for the existence of normalizing and centering constants $A_n > 0$ and C_n such that the sequence

$$\frac{1}{A_n}\left\{\sum_{i=1}^{k_n} X_{n+1-i,n} - C_n\right\}$$

converges in distribution along subsequences of the integers $\{n\}$ to non-degenerate limits and completely described the possible subsequential limiting distributions. Viharos [**85**] generalized this result for linear combinations $\sum_{i=k+1}^{k_n} d_{n+1-i,n} f(X_{n+1-i,n})$ of extreme values, where f is a Borel-measurable function. Assuming $F \in \mathcal{R}_\gamma$ and using the results in [**85**], we will prove asymptotic normality for the properly normalized and centered sequence $S_n(p)$. As an application, we derive a class of asymptotically normal estimators for the parameter γ.

Linear combinations of order statistics are widely studied in the literature. Recently, Barczyk et al. [**3**] obtained limit theorems for L-statistics

$$L_n = \sum_{i=1}^{k_n} c_{i,n} X_{i:k_n},$$

where $k_n \to \infty$ as $n \to \infty$, $c_{i,n}$ are real scores and the order statistics $X_{i:k_n}$ correspond to a possibly non i.i.d. triangular array $(X_{i,n})_{1\le i\le k_n}$ of infinitesimal and row-wise independent random variables with heavy tails. Their approach is related to the extreme order statistics: they give sufficient conditions for the scores $c_{i,n}$ so that only the extreme parts of the L-statistics contribute to the limit law.

We will assume as in [**85**] that the weights $d_{i,n}$ are of the form

$$d_{i,n} = n\int_{(i-1)/n}^{i/n} \bar{L}(t)dt, \quad 1 \le i \le n,$$

for some non-negative continuous function $\bar{L}$ defined on (0,1) which satisfies the following condition:

Condition $\bar{\mathbf{L}}$:

a) There exists a constant $-1/2 < \rho < \infty$ such that $\bar{L}(1-t) = t^\rho \bar{\ell}(t)$ on $(0,1)$ for some function $\bar{\ell}(\cdot)$ slowly varying at 0 and $\bar{\ell}'(t) = t^{-1}\bar{\ell}(t)\varepsilon(t)$ on some $(0,\delta)$ with a continuous function $\varepsilon(\cdot)$ for which $\varepsilon(t) \to 0$ as $t \to 0$.

b) For all $M \ge 1$,

$$\sup_{1/M<y<M}\left|\int_0^y \frac{(\bar{\ell}(u/n) - \bar{\ell}(y/n))u^\rho}{\bar{\ell}(y/n)y^\rho}du\right| \to 0, \quad n \to \infty.$$

Throughout the chapter we use the convention $\int_a^b = \int_{[a,b)}$ when we integrate with respect to a left continuous integrator. Define

$$J(s) = s^{\rho}\bar{\ell}(s), \quad 0 < s < 1,$$

and

$$g(t) = -(\log Q(1-t-))^p,$$

where $Q(1-s-)$ denotes the left-continuous version of the right-continuous function $Q(1-s)$, $0 < s < 1$,

$$K(t) = \int_{1/2}^{t} J(s)dg(s), \quad 0 < t < 1,$$

and

$$\sigma^2(s,t) = \int_s^t \int_s^t (u \wedge v - uv)dK(u)dK(v), \quad 0 \le s \le t \le 1,$$

where $u \wedge v = \min(u,v)$. We introduce the centering sequences

$$\mu_n := -n \int_0^{k_n/n} J(u)g(u)du,$$

and

$$\bar{\mu}_n = -n \int_{1/n}^{k_n/n} J(u)g(u)du - d_{n,n}g\left(\frac{1}{n}\right),$$

while the normalizing sequence will be given by

$$a_n := \begin{cases} \sigma(1/n, k_n/n) & \text{if} \quad \sigma(1/n, k_n/n) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

2. Main results

We state now the main limit theorem of this chapter. Throughout, $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, $\xrightarrow{\mathbb{P}}$ denotes convergence in probability, and limiting and order relations are always meant as $n \to \infty$ if not specified otherwise.

THEOREM 2.1. (i) *Assume that* $F \in \mathcal{R}_\gamma$, (3.2) *holds and suppose that condition* $\bar{\mathbf{L}}$ *is satisfied for the weighs* $d_{i,n}$. *Then*

$$\frac{1}{\sqrt{n}a_n}\left\{\sum_{i=1}^{k_n} d_{n+1-i,n}\log^p X_{n+1-i,n} - \bar{\mu}_n\right\} \xrightarrow{\mathcal{D}} N(0,1). \tag{3.3}$$

(ii) *If in addition to the conditions of* (i) *we have* $(\log n)/k_n^{\varepsilon} \to 0$ *for some* $0 < \varepsilon < \rho + 1/2$, *then* (3.3) *holds with* μ_n *replacing* $\bar{\mu}_n$.

The special case $p = 1$ of Theorem 2.1(i) was stated in Theorem 1.2 of [**86**]. Several estimators exist for the tail index γ among which Hill's estimator is the most classical (see Hill [**50**]). Dekkers et al. [**32**] proposed a moment estimator based on the statistics

$$\frac{1}{k_n}\sum_{i=1}^{k_n}\left(\log\frac{X_{n+1-i,n}}{X_{n-k_n,n}}\right)^j, \quad j = 1, 2. \tag{3.4}$$

The case $j = 1$ yields the Hill estimator. Segers [**77**] investigated more general statistics of the form

$$\frac{1}{k_n}\sum_{i=1}^{k_n} f\left(\frac{X_{n+1-i,n}}{X_{n-k_n,n}}\right), \tag{3.5}$$

for a nice class of functions f, called residual estimators. Segers proved weak consistency and asymptotic normality under general conditions. More recently, Ciuperca and Mercadier [**15**] obtained a class of tail index estimators based on the weighted power sums of the statistics $\big(\log(X_{n+1-i,n}/X_{n-k_n,n})\big)_{1\le i\le k_n}$ and proved limit theorems for the estimators. We use the weighted power sums of the extreme values $(\log X_{n+1-i,n})_{1\le i\le k_n}$ to construct a new class of estimators for γ.

The following proposition describes the asymptotic behavior of the centering and normalizing sequences.

Proposition 2.2. *Assume the conditions of Theorem* 2.1(i). *Then*

$$\sigma(1/n, k_n/n) \sim p\gamma^p\left(\frac{2}{(1+\rho)(1+2\rho)}\right)^{1/2}\left(\frac{k_n}{n}\right)^{\rho+1/2}\left(\log\frac{n}{k_n}\right)^{p-1}\bar{\ell}\left(\frac{k_n}{n}\right) \tag{3.6}$$

and $\mu_n \sim \gamma^p\alpha_n$, *where* $\alpha_n = \frac{k_n}{\rho+1}J\left(\frac{k_n}{n}\right)\left(\log\frac{n}{k_n}\right)^p$ *(*$x_n \sim y_n$ *means that* $x_n/y_n \to 1$*).*

To prove Theorem 2.1 and Proposition 2.2, we need some preparatory results. Recall (2.2). Let G be a distribution function on $\mathbb{R}$. Whenewer G belongs to the maximum domain of attraction of G_γ we write $G \in \Delta(\gamma)$. Set $U(s) := -G^{\leftarrow}(1-s)$, $0 \le s < 1$, where the arrow means the inverse function. From [**20**, equation (1.12)] we know the following statement.

Proposition 2.3. $G \in \Delta(\gamma)$ *if and only if*

$$\lim_{s\downarrow 0}\frac{U(xs)-U(ys)}{U(vs)-U(ws)} = \frac{x^{-\gamma}-y^{-\gamma}}{v^{-\gamma}-w^{-\gamma}},$$

where for $\gamma = 0$ *the limit is understood as* $(\log x - \log y)/(\log v - \log w)$.

Let RV_α^∞ (RV_α^0) denote the class of regularly varying functions at infinity (zero) with index α.

LEMMA 2.4. *Assume the conditions of Theorem 2.1. Then the distribution function* $H(\cdot) := (-K(1-\cdot))^{\leftarrow}$ *satisfies* $H \in \Delta(-\rho)$.

PROOF. A simple calculation yields $K(t) = -\int_{Q(1/2)}^{Q(1-t)} J_1(u)du$, where

$$J_1(u) = pJ(1-F(u))(\log u)^{p-1}u^{-1} \in RV^{\infty}_{-(\rho/\gamma)-1}.$$

If $\rho > 0$ then $K(t) = \int_{Q(1-t)}^{\infty} J_1(u)du + c$, where c is a constant, and by Karamata's theorem (see e.g. [**10**, Theorem 1.5.11]) we obtain

$$K(t) = \frac{\gamma}{\rho}Q(1-t)J_1(Q(1-t))(1+o(1)) + c \quad (t \to 0).$$

Similarly, if $\rho < 0$ then

$$K(t) = \frac{\gamma}{\rho}Q(1-t)J_1(Q(1-t))(1+o(1)) \quad (t \to 0).$$

Theorem 1.5.12 of [**10**] implies that

$$1 - F(Q(1-t)) \sim t \quad (t \to 0). \tag{3.7}$$

Then using (3.7) and $\log Q(1-t) \sim -\gamma \log t$ $(t \to 0)$, we have

$$Q(1-t)J_1(Q(1-t)) \sim p(-\gamma \log t)^{p-1}J(t) \quad (t \to 0).$$

Hence, if $\rho > 0$ then

$$K(t) = t^{\rho}\hat{L}(t)(1+o(1)) + c \quad (t \to 0), \tag{3.8}$$

and if $\rho < 0$ then

$$K(t) = -t^{\rho}\hat{L}(t)(1+o(1)) \quad (t \to 0), \tag{3.9}$$

where

$$\hat{L}(t) = \frac{p\gamma^p}{|\rho|}(-\log t)^{p-1}\bar{\ell}(t) \in RV_0^0. \tag{3.10}$$

Equations (3.8), (3.9) and (3.10) imply that for $\rho \neq 0$,

$$\lim_{s \downarrow 0} \frac{K(xs) - K(ys)}{K(vs) - K(ws)} = \frac{x^{\rho} - y^{\rho}}{v^{\rho} - w^{\rho}}. \tag{3.11}$$

If $\rho = 0$, then for distinct values $0 < x, y < \infty$,

$$K(xs) - K(ys) = \bar{\ell}(\xi)(g(xs) - g(ys)) \tag{3.12}$$

where ξ is between xs and ys. Since $\bar{\ell}$ is slowly varying, we have

$$\bar{\ell}(\xi) \sim \bar{\ell}(s) \quad (s \downarrow 0). \tag{3.13}$$

Moreover, by Lagrange's mean value theorem, with some η between $\log Q(1-(xs)-)$ and $\log Q(1-(ys)-)$,

$$g(xs) - g(ys) = p\eta^{p-1}(\log Q(1-(ys)-) - \log Q(1-(xs)-)). \tag{3.14}$$

Using (3.1) and the fact that $\log Q(1-s-)$ is slowly varying at zero, we have

$$\eta \sim \log Q(1-s-) \sim -\gamma \log s \quad (s \downarrow 0), \tag{3.15}$$

and

$$\log Q(1-(ys)-) - \log Q(1-(xs)-) \to \gamma \log(x/y) \quad (s \downarrow 0). \tag{3.16}$$

By (3.12)-(3.16) it follows that

$$K(xs) - K(ys) \sim p\gamma^p \log(x/y)(-\log s)^{p-1}\bar{\ell}(s) \quad (s \downarrow 0). \tag{3.17}$$

Therefore,

$$\lim_{s\downarrow 0} \frac{K(xs)-K(ys)}{K(vs)-K(ws)} = \frac{\log x - \log y}{\log v - \log w} \tag{3.18}$$

for all distinct $0 < x, y, v, w < \infty$. Equations (3.11), (3.18) and Proposition 2.3 imply the statement of the lemma. □

Choose any sequence of positive constants δ_n such that $n\delta_n < n$ and $n\delta_n \to 0$ as $n \to \infty$. The following two sequences of functions govern the asymptotic behavior of $S_n(p)$:

$$\psi_n(x) = \psi_{n,K}(x) = \begin{cases} \frac{k_n^{1/2}\left\{K\left(\frac{k_n}{n}+x\frac{k_n^{1/2}}{n}\right)-K\left(\frac{k_n}{n}\right)\right\}}{n^{1/2}a_n} & \text{if} \quad -\frac{k_n^{1/2}}{2} \le x \le \frac{k_n^{1/2}}{2}, \\ \psi_n\left(-\frac{k_n^{1/2}}{2}\right) & \text{if} \quad -\infty < x < -\frac{k_n^{1/2}}{2}, \\ \psi_n\left(\frac{k_n^{1/2}}{2}\right) & \text{if} \quad \frac{k_n^{1/2}}{2} < x < \infty, \end{cases}$$

and

$$\varphi_n(y) = \varphi_{n,K}(y) = \begin{cases} \frac{K(y/n)-K(1/n)}{n^{1/2}a_n} & \text{if} \quad 0 < y \le n - n\delta_n, \\ \frac{K(1-\delta_n)-K(1/n)}{n^{1/2}a_n} & \text{if} \quad n - n\delta_n < y < \infty. \end{cases}$$

LEMMA 2.5. *Assume the conditions of Theorem 2.1. Then* $\psi_n(x), \varphi_n(y) \to 0$, $x \in \mathbb{R}$, $y > 0$.

PROOF. The statement is a consequence of Lemmas 2.11 and 2.12 of **[20]** and Lemma 2.4 above. □

PROOF OF PROPOSITION 2.2. If $\rho > -1/2,\ \rho \neq 0$ then by Lemma 2.9 of **[20]** and by Lemma 2.4 above we have

$$\sigma(1/n, k_n/n) \sim \left(\frac{2\rho^2}{(1+\rho)(1+2\rho)}\right)^{1/2} \left(\frac{k_n}{n}\right)^{\rho+1/2} \hat{L}(k_n/n),$$

which is the same as (3.6). If $\rho = 0$, then by (2.29) of **[20]** and by Lemma 2.4 we have $\sigma(1/n, k_n/n) \sim \sigma(0, k_n/n)$, and using Lemma 2.10 of **[20]**, we obtain

$$\lim_{s\downarrow 0} \sqrt{s}(K(\lambda s) - K(s))/\sigma(0,s) = 2^{-1/2} \log \lambda \quad \text{for all } 0 < \lambda < \infty.$$

Then by (3.17),

$$\sigma(0,s) \sim p\gamma^p \sqrt{2s}(-\log s)^{p-1}\bar{\ell}(s) \quad (s \downarrow 0),$$

which implies the statement for $\rho = 0$.

Statement $\mu_n \sim \gamma^p \alpha_n$ follows from the facts $-J(\cdot)g(\cdot) \in RV_\rho^0$, $\log Q(1-s-) \sim -\gamma \log s$ and from Karamata's theorem. □

Proof of Theorem 2.1. The Corollary of [**85**] and Lemma 2.5 imply statement (i). To prove statement (ii) write

$$\bar{\mu}_n = \mu_n - n\int_0^{1/n} J(u)(-g(u))du + d_{n,n}(-g(1/n)) =: \mu_n - r_n^{(1)} + r_n^{(2)}. \tag{3.19}$$

We have to prove that

$$\frac{r_n^{(1)}}{\sqrt{n}a_n} \to 0 \quad \text{and} \quad \frac{r_n^{(2)}}{\sqrt{n}a_n} \to 0. \tag{3.20}$$

By Karamata's theorem, (3.15) and Proposition 2.2, with some constant c we have

$$\begin{aligned}\frac{r_n^{(1)}}{\sqrt{n}a_n} &\sim c\,\frac{n^{-\rho}\bar{\ell}(1/n)(\log n)^p}{\sqrt{n}(k_n/n)^{\rho+1/2}(\log(n/k_n))^{p-1}\bar{\ell}(k_n/n)}\\ &= c\frac{(\log n)^p\bar{\ell}(1/n)}{k_n^{\rho+1/2}(\log(n/k_n))^{p-1}\bar{\ell}(k_n/n)}.\end{aligned}$$

By the Potter bounds ([**10**, Theorem 1.5.6]), for any $A > 1$ and $\delta > 0$, there exist N such that

$$\frac{\bar{\ell}(1/n)}{\bar{\ell}(k_n/n)} \le Ak_n^\delta \quad \text{and} \quad \frac{\log n}{\log(n/k_n)} \le Ak_n^\delta \quad \text{for any} \quad n \ge N.$$

We choose $\delta > 0$ such that $p\delta < \rho - \varepsilon + 1/2$. It follows that with some constant c_1,

$$\frac{r_n^{(1)}}{\sqrt{n}a_n} \le c_1 \frac{\log n}{k_n^{\rho - p\delta + 1/2}} \le c_1 \frac{\log n}{k_n^\varepsilon}$$

if $n \ge N$. A similar upper bound for $r_n^{(2)}/(\sqrt{n}a_n)$ implies (3.20). □

The next corollary describes the asymptotic behavior of the weighted norms $R_n(p) := (S_n(p))^{1/p}$.

COROLLARY 2.6. *Assume the conditions of Theorem* 2.1(ii). *Then*

$$\frac{1}{\gamma\sqrt{2}}\left(\frac{1+2\rho}{1+\rho}\right)^{1/2}\sqrt{k_n}\log\frac{n}{k_n}\left\{\frac{1}{\alpha_n^{1/p}}R_n(p)-\left(\frac{\mu_n}{\alpha_n}\right)^{1/p}\right\}\xrightarrow{\mathcal{D}} N(0,1).$$

By Proposition 2.2 and Corollary 2.6,

$$\widehat{\gamma}_n := \frac{1}{\alpha_n^{1/p}}R_n(p)$$

is an asymptotically normal estimator for γ. This is a generalization of the estimator proposed in [**87**]. Asymptotic normality was proved for the Hill estimator and for the estimators in [**15**] and [**77**] under general conditions but not for every distribution in $\mathcal{R}_\gamma$. However, $\widehat{\gamma}_n$ is asymptotically normal over the whole model $\mathcal{R}_\gamma$.

To investigate the asymptotic bias of the estimator $\widehat{\gamma}_n$, we assume the following conditions:

(B_1) $\quad \sqrt{k_n}\log\frac{n}{k_n}\sup_{0\le u\le k_n/n}\left|\frac{\log\ell(u)}{\log u}\right|\to 0.$

(B_2) $\sqrt{k_n}/\log n\to 0.$

(B_3) $(\log n)/k_n^{\rho+\frac{1}{2}}n\to 0.$

(B_4) $J(s)=s^\rho,\quad 0<s<1.$

Conditions (B_2) and (B_3) imply that $\rho>0$.

PROOF OF COROLLARY 2.6. Using Theorem 2.1 and Proposition 2.2, we obtain

$$\beta_n\left(\frac{S_n(p)}{\alpha_n}-\frac{\mu_n}{\alpha_n}\right)\xrightarrow{\mathcal{D}} N(0,1), \tag{3.21}$$

where

$$\beta_n=\frac{1}{\gamma^p p\sqrt{2}}\left(\frac{1+2\rho}{1+\rho}\right)^{1/2}\sqrt{k_n}\log\frac{n}{k_n}. \tag{3.22}$$

Since $\mu_n/\alpha_n\to\gamma^p$ and $\beta_n\to\infty$, we have $S_n(p)/\alpha_n\xrightarrow{\mathbb{P}}\gamma^p$.

By Lagrange's mean value theorem

$$\left(\frac{S_n(p)}{\alpha_n}\right)^{1/p}-\left(\frac{\mu_n}{\alpha_n}\right)^{1/p}=\frac{1}{p}\xi^{(1/p)-1}\left(\frac{S_n(p)}{\alpha_n}-\frac{\mu_n}{\alpha_n}\right)$$

with some ξ between μ_n/α_n and $S_n(p)/\alpha_n$. Therefore,

$$\beta_n\left(\left(\frac{S_n(p)}{\alpha_n}\right)^{1/p}-\left(\frac{\mu_n}{\alpha_n}\right)^{1/p}\right)\xrightarrow{\mathcal{D}}\frac{1}{p}\gamma^{1-p}N(0,1).$$

$\square$

Corollary 2.7. *Assume the conditions* (B_1)-(B_4), *and the conditions of Theorem* 2.1(i), *and set* $t_n := (\rho+1)\log(n/k_n)$. *Then we have*

(i)

$$\frac{1}{\gamma^p p\sqrt{2}}\left(\frac{1+2\rho}{1+\rho}\right)^{1/2}\sqrt{k_n}\log\frac{n}{k_n}\left\{\frac{S_n(p)}{\alpha_n}-\gamma^p\big(1+pt_n^{-1}\big)\right\}\xrightarrow{\mathcal{D}}N(0,1),$$

(ii)

$$\frac{1}{\gamma\sqrt{2}}\left(\frac{1+2\rho}{1+\rho}\right)^{1/2}\sqrt{k_n}\log\frac{n}{k_n}\big\{\widehat{\gamma}_n-\gamma\big(1+t_n^{-1}\big)\big\}\xrightarrow{\mathcal{D}}N(0,1). \tag{3.23}$$

We show that condition (B_1) is satisfied by the model $\ell(s)=1+b(s)$, where the function b is such that $\sqrt{k_n}\sup_{0\le u\le k_n/n}|b(u)|\to 0$. To prove this, observe that $\sup_{0\le u\le k_n/n}1/|\log u| = 1/\log(n/k_n)$ and hence

$$\begin{aligned}\sqrt{k_n}\log\frac{n}{k_n}\sup_{0\le u\le k_n/n}\left|\frac{\log\ell(u)}{\log u}\right| &\le \sqrt{k_n}\sup_{0\le u\le k_n/n}|\log(1+b(u)|\\ &= \sqrt{k_n}\sup_{0\le u\le k_n/n}|b(s)+O(b^2(s))|\to 0,\end{aligned}$$

if $\sqrt{k_n}\sup_{0\le u\le k_n/n}|b(u)|\to 0$.

In some submodels of (3.1) the Hill estimator can be centered at γ to have normal asymptotic distribution. The strict Pareto model when $\ell\equiv 1$ is the simplest example of these models. This simple model satisfies the conditions of Corollary 2.7. From (3.23) we also see that under these conditions the estimator $\widehat{\gamma}_n$ can not be centered at γ to have asymptotic distribution. However, Corollary 2.7 allows the construction of asymptotic confidence intervals for γ. The estimator $\widehat{\gamma}_n$ is not scale invariant. Accordingly, the slowly varying function $\ell\equiv c$, $c\neq 1$, does not satisfy condition (B_1).

Proof of Corollary 2.7. Proof of (i). To treat $\bar{\mu}_n$, we use the decomposition (3.19). For μ_n we obtain

$$\begin{aligned}\mu_n &= n\int_0^{k_n/n}J(u)(\log u^{-\gamma})^p\left(1+\frac{\log\ell(u)}{\log u^{-\gamma}}\right)^p du\\ &= n\int_0^{k_n/n}J(u)(\log u^{-\gamma})^p du\\ &\quad + n\int_0^{k_n/n}J(u)(\log u^{-\gamma})^p\left[\left(1+\frac{\log\ell(u)}{\log u^{-\gamma}}\right)^p-1\right]du\\ &=: \mu_n^{(1)}+\mu_n^{(2)}.\end{aligned} \tag{3.24}$$

By Karamata's theorem,

$$\int_0^x J(u)(\log u^{-\gamma})^p du \sim \frac{1}{\rho+1} xJ(x)(\log x^{-\gamma})^p \quad \text{as } x \to 0.$$

Therefore, using Condition (B_1), we have

$$\begin{aligned}
&\sqrt{k_n} \log \frac{n}{k_n} \frac{|\mu_n^{(2)}|}{\alpha_n} \\
&\quad \le \frac{n\sqrt{k_n}\log\frac{n}{k_n}}{\alpha_n} \int_0^{k_n/n} J(u)(\log u^{-\gamma})^p du \sup_{0\le u\le k_n/n} \left| \left(1+\frac{\log \ell(u)}{\log u^{-\gamma}}\right)^p - 1 \right| \\
&\quad \sim \gamma^p \sqrt{k_n} \log \frac{n}{k_n} \sup_{0\le u\le k_n/n} \left| \left(1+\frac{\log \ell(u)}{\log u^{-\gamma}}\right)^p - 1 \right|.
\end{aligned}$$

By

$$(1+x)^p = 1 + px + O(x^2) \quad \text{as } x \to 0 \tag{3.25}$$

and condition (B_1) it follows that

$$\sqrt{k_n} \log \frac{n}{k_n} \frac{|\mu_n^{(2)}|}{\alpha_n} \to 0. \tag{3.26}$$

For the first term we obtain

$$\mu_n^{(1)} = \frac{n\gamma^p}{(\rho+1)^{p+1}} \int_{(\rho+1)\log(n/k_n)}^{\infty} t^p e^{-t} dt = \frac{n\gamma^p}{(\rho+1)^{p+1}} \Gamma(p+1, (\rho+1)\log(n/k_n)),$$

where

$$\Gamma(a,x) = \int_x^{\infty} t^{a-1} e^{-t} dt$$

is the incomplete gamma function. It is known that

$$\Gamma(a,x) = x^{a-1} e^{-x} \left(\sum_{j=0}^{n-1} b_j x^{-j} + M_n(x) \right),$$

where $b_j = (a-1)(a-2)\cdots(a-j)$ and

$$M_n(x) = O(x^{-n}) \quad \text{as } x \to \infty \tag{3.27}$$

(see equation (2.02) in [**65**]). Recall the notation $t_n = (\rho+1)\log(n/k_n)$. Then

$$\frac{\mu_n^{(1)}}{\alpha_n} = \gamma^p (1 + pt_n^{-1} + M_2(t_n)). \tag{3.28}$$

For $r_n^{(2)}$ in (3.19) we obtain

$$r_n^{(2)} = \frac{1}{n^\rho(\rho+1)} (\log Q(1-(1/n)-))^p \sim \frac{1}{n^\rho(\rho+1)} (\gamma \log n)^p,$$

implying that

$$\sqrt{k_n}\log\frac{n}{k_n}\frac{|r_n^{(2)}|}{\alpha_n}\sim\gamma^p\frac{(\log n)^p}{k_n^{\rho+\frac{1}{2}}(\log(n/k_n))^{p-1}}. \tag{3.29}$$

Condition (B_2) implies $\log(n/k_n)\sim\log n$. Therefore, by Condition (B_3) we have

$$\sqrt{k_n}\log\frac{n}{k_n}\frac{|r_n^{(2)}|}{\alpha_n}\to 0. \tag{3.30}$$

A similar argument yields that

$$\sqrt{k_n}\log\frac{n}{k_n}\frac{|r_n^{(1)}|}{\alpha_n}\to 0 \tag{3.31}$$

(cf. the proof of Theorem 2.1(ii)). Recall (3.22). Using the decompositions (3.19) and (3.24), equations (3.26), (3.28), (3.30) and (3.31), we obtain

$$\beta_n\left(\frac{S_n(p)}{\alpha_n}-\frac{\bar{\mu}_n}{\alpha_n}\right)=\beta_n\left(\frac{S_n(p)}{\alpha_n}-\gamma^p(1+pt_n^{-1}+M_2(t_n))\right)+o(1). \tag{3.32}$$

Theorem 2.1(i), condition (B_2), (3.27) and (3.32) imply

$$\beta_n\left(\frac{S_n(p)}{\alpha_n}-\gamma^p\left(1+pt_n^{-1}\right)\right)\xrightarrow{\mathcal{D}}N(0,1).$$

This completes the proof of part (i).

Proof of (ii). Using the same argument as in the proof of Corollary 2.6, we have

$$\beta_n\bigg(\left(\frac{S_n(p)}{\alpha_n}\right)^{1/p}-\gamma\left(1+pt_n^{-1}\right)^{1/p}\bigg)\xrightarrow{\mathcal{D}}\frac{1}{p}\gamma^{1-p}N(0,1).$$

Applying (3.25) with $1/p$ replacing p, we obtain

$$\left(1+\frac{p}{t_n}\right)^{1/p}=1+t_n^{-1}+O(t_n^{-2}).$$

Therefore, by condition (B_2)

$$\beta_n\bigg(\left(\frac{S_n(p)}{\alpha_n}\right)^{1/p}-\gamma\big(1+t_n^{-1}\big)\bigg)\xrightarrow{\mathcal{D}}\frac{1}{p}\gamma^{1-p}N(0,1).$$

This completes the proof of part (ii). $\square$

3. Simulation results

In this section we evaluate the performance of the estimator $\widehat{\gamma}_n$ through simulations. In the first simulation study we compare $\widehat{\gamma}_n$ to the Hill, Pickands [**69**] and Moment estimators. Tail index estimators have good performance in the strict Pareto model. However, in practical situations it is very rare when data fit to a simple distribution. For the simulation we use the following model proposed by Hall [**48**]:

$$Q(1-s) = s^{-\gamma} D_1[1 + D_2 s^{\beta}(1+o(1))] \quad \text{as } s \to 0, \tag{3.33}$$

where $D_1 > 0$, $D_2 \neq 0$ and $\beta > 0$ are constants. The Hall model satisfies condition (B_1) if $D_1 = 1$ and $k_n^{\beta+\frac{1}{2}}/n^{\beta} \to 0$.

We repeated the simulations 1000 times and we assumed $n = 1000$ for the sample size and $k_n = 136$ for the sample fraction size. We used $\bar{\ell} \equiv 1$ for the weights $d_{i,n}$. We examined the following two cases of the Hall model:

Case 1: $\beta = 2$, $D_2 = 1$ and $D_1 = 1/\sqrt{e}$.

Case 2: $\beta = 1$, $D_2 = 4/3$ and $D_1 = e^{-2/3}$.

In both cases we assume $o(1) \equiv 0$ in (3.33). Tables 1 and 2 contain the average simulated estimates (mean) and the calculated empirical mean square errors (MSE) for *Case 1*. Using the mean square error as criterion, we see that for $\rho \leq 1$ the performance of $\widehat{\gamma}_n$ generally increases as γ decreases from 2 to 0.5. For $\gamma \geq 1$ the weights improve the performance of $\widehat{\gamma}_n$ significantly ($\rho = 0.5, 1, 2$). For the thin tail pertaining to $\gamma = 0.5$ we also see a trend that the performance of $\widehat{\gamma}_n$ improves as the value of p increases from 1 to 3. The same conclusion holds for $\gamma = 1$ when $\rho = 2$. It can be also seen that $\widehat{\gamma}_n$ with $p = 1, 2, 3$ and appropriate ρ value performs better than the Pickands and the moment estimator. The Pickands estimator has poor performance for $\gamma = 2$. Nonetheless, the Hill and the moment estimator tend to have good estimates.

Tables 3 and 4 contain the simulation results for *Case 2*. This case is farther from the strict Pareto model than *Case 1*. In *Case 2* for $\rho \leq 0.5$ the estimator $\widehat{\gamma}_n$ works slightly better than in the first case. The performance of the Hill estimator is slightly worse in this case, while the other estimators have similar performance compared to the first case.

TABLE 1. Mean in the Hall model for *Case 1*.

mean							
		$\widehat{\gamma}_n$			Hill	Pickands	Moment
ρ	γ	$p=1$	$p=2$	$p=3$			
0	0.5	0.502461	0.5598067	0.6278012	0.4874154	0.5388793	0.4832535
	1	1.252406	1.347012	1.461455	0.9872326	1.021725	0.9745838
	1.5	2.002351	2.136447	2.299039	1.48705	1.52004	1.471576
	2	2.752296	2.926308	3.137432	1.986867	2.022467	1.969981
0.5	0.5	0.4207121	0.4523482	0.4918764	0.4874154	0.5388793	0.4832535
	1	1.088022	1.138332	1.200928	0.9872326	1.021725	0.9745838
	1.5	1.755332	1.826024	1.913608	1.48705	1.52004	1.471576
	2	2.422641	2.514022	2.626971	1.986867	2.022467	1.969981
1	0.5	0.37965551	0.3994002	0.4240878	0.4874154	0.5388793	0.4832535
	1	1.005246	1.03595	1.073641	0.9872326	1.021725	0.9745838
	1.5	1.630837	1.673773	1.726098	1.48705	1.52004	1.471576
	2	2.256427	2.311814	2.379069	1.986867	2.022467	1.969981
2	0.5	0.33886111	0.3486395	0.3606289	0.4874154	0.5388793	0.4832535
	1	0.9227323	0.9375759	0.9552161	0.9872326	1.021725	0.9745838
	1.5	1.506604	1.527265	1.551595	1.48705	1.52004	1.471576
	2	2.090475	2.117078	2.148269	1.986867	2.022467	1.969981

TABLE 2. MSE in the Hall model for *Case 1*.

MSE							
		$\widehat{\gamma}_n$			Hill	Pickands	Moment
ρ	γ	$p=1$	$p=2$	$p=3$			
0	0.5	0.008489717	0.004758226	0.01848487	0.001920372	0.1238975	0.008732585
	1	0.06713682	0.124994	0.2205786	0.007254561	0.1510138	0.01456819
	1.5	0.2601122	0.415274	0.6550551	0.01616043	0.191689	0.02365229
	2	0.579775	0.8761246	1.322759	0.02863798	0.2457045	0.0362088
0.5	0.5	0.006915487	0.002963965	0.0009153005	0.001920372	0.1238975	0.008732585
	1	0.01033168	0.02195434	0.04367552	0.007254561	0.1510138	0.01456819
	1.5	0.07105951	0.1126648	0.1784538	0.01616043	0.191689	0.02365229
	2	0.189099	0.2755773	0.4061787	0.02863798	0.2457045	0.0362088
1	0.5	0.01503467	0.01069469	0.006382895	0.001920372	0.1238975	0.008732585
	1	0.002311005	0.003667682	0.007952766	0.007254561	0.15101382	0.01456819
	1.5	0.02231372	0.03559411	0.05684494	0.01616043	0.191689	0.02365229
	2	0.07504283	0.1068695	0.1538997	0.02863798	0.2457045	0.0362088
2	0.5	0.02645074	0.02340072	0.01992387	0.001920372	0.1238975	0.008732585
	1	0.007996087	0.005951954	0.004100054	0.007254561	0.1510138	0.01456819
	1.5	0.004666964	0.005432634	0.007437678	0.01616043	0.191689	0.02365229
	2	0.01646337	0.02210106	0.03052874	0.02863798	0.2457045	0.0362088

TABLE 3. Mean in the Hall model for *Case 2*.

mean							
		$\widehat{\gamma}_n$			Hill	Pickands	Moment
ρ	γ	$p=1$	$p=2$	$p=3$			
0	0.5	0.4589447	0.5124346	0.5782619	0.4184847	0.609304	0.4811199
	1	1.20889	1.299565	1.411133	0.9183019	1.011627	0.9377913
	1.5	1.958834	2.089031	2.248614	1.418119	1.488072	1.423793
	2	2.708779	2.878913	3.086975	1.917936	1.980745	1.916807
0.5	0.5	0.3846648	0.4124053	0.4486737	0.4184847	0.609304	0.4811199
	1	1.051975	1.09872	1.157892	0.9183019	1.011627	0.9377913
	1.5	1.719284	1.786526	1.870711	1.418119	1.488072	1.423793
	2	2.386594	2.47458	2.58415	1.917936	1.980745	1.916807
1	0.5	0.3484573	0.36485	0.3861637	0.4184847	0.609304	0.4811199
	1	0.9740478	1.001849	1.036418	0.9183019	1.011627	0.9377913
	1.5	1.599638	1.639798	1.689107	1.418119	1.488072	1.423793
	2	2.225229	2.277898	2.34219	1.917936	1.980745	1.916807
2	0.5	0.3135496	0.3210441	0.3304326	0.4184847	0.609304	0.4811199
	1	0.8974208	0.9103876	0.9258765	0.9183019	1.011627	0.9377913
	1.5	1.481292	1.500177	1.522475	1.418119	1.488072	1.423793
	2	2.065163	2.090035	2.119249	1.917936	1.980745	1.916807

TABLE 4. MSE in the Hall model for *Case 2*.

MSE							
		$\widehat{\gamma}_n$			Hill	Pickands	Moment
ρ	γ	$p=1$	$p=2$	$p=3$			
0	0.5	0.002356798	0.001205257	0.008251327	0.0081828	0.1375034	0.00768532
	1	0.04670408	0.09401281	0.1764996	0.01327334	0.1501311	0.01701934
	1.5	0.2177388	0.3566961	0.576717	0.02193555	0.1891187	0.0278058
	2	0.5154611	0.7899285	1.210099	0.03416945	0.2417847	0.04120925
0.5	0.5	0.01375736	0.008212465	0.003358186	0.0081828	0.1375034	0.00768532
	1	0.004917317	0.01222534	0.02793089	0.01327334	0.1501311	0.01701934
	1.5	0.05338891	0.08794855	0.1443359	0.02193555	0.1891187	0.0278058
	2	0.1591721	0.23588	0.3536742	0.03416945	0.2417847	0.04120925
1	0.5	0.02334328	0.0186799	0.01343518	0.0081828	0.1375034	0.00768532
	1	0.002585846	0.00202712	0.003527147	0.01327334	0.1501311	0.01701934
	1.5	0.0145548	0.02439929	0.04097613	0.02193555	0.1891187	0.0278058
	2	0.05925014	0.08613785	0.1266097	0.03416945	0.2417847	0.04120925
2	0.5	0.03507632	0.03235116	0.02909741	0.0081828	0.1375034	0.00768532
	1	0.01217588	0.00972373	0.007239784	0.01327334	0.1501311	0.01701934
	1.5	0.004400968	0.004131617	0.00474087	0.02193555	0.1891187	0.0278058
	2	0.01175158	0.01574687	0.02203485	0.03416945	0.2417847	0.04120925

By Corollary 2.7(ii) we infer that

$$Z_n := \frac{1}{\widehat{\gamma}_n\sqrt{2}}\left(\frac{1+2\rho}{1+\rho}\right)^{1/2}\sqrt{k_n}\log\frac{n}{k_n}\{\widehat{\gamma}_n-\gamma\big(1+t_n^{-1}\big)\}\xrightarrow{\mathcal{D}} N(0,1). \qquad (3.34)$$

Asymptotic confidence intervals for γ can be constructed using either (3.23) or (3.34). In the second simulation study we investigated how fast the distribution result (3.34) kicks in. We simulated the quantity Z_n 5000 times. According to condition (B_2), we used k_n values less than $\log^2 n$. First, we investigated the Fréchet distribution with shape parameter $1/\gamma$ that belongs to the Hall model with parameters $D_1 = 1$, $D_2 = -\gamma/2$ and $\beta = 1$. The simulation was done for $\gamma = 1$, $\rho = 1$, $p = 1$, $n = 900$ and $k_n = 10$. We found empirically that $n = 900$ is the threshold sample size to obtain a good normal approximation in (3.34). Figure 1 contains the histogram with the fitted normal curve and the Q–Q plot of the simulated Z_n quantities with estimated parameters. The mean of the simulated Z_n values is -0.06, the simulated standard deviation is 0.8974. The mean of the simulated $\widehat{\gamma}_n$ values is 1.1116. The bias of the mean is in accordance with the bias term γt_n^{-1} in (3.34). Due to the biased estimator in the leading factor $1/(\widehat{\gamma}_n\sqrt{2})$ of Z_n, the simulated standard deviation of Z_n is smaller than the asymptotic value 1. We performed the chi-square test for normality, and we obtained the p-value 0.2965.

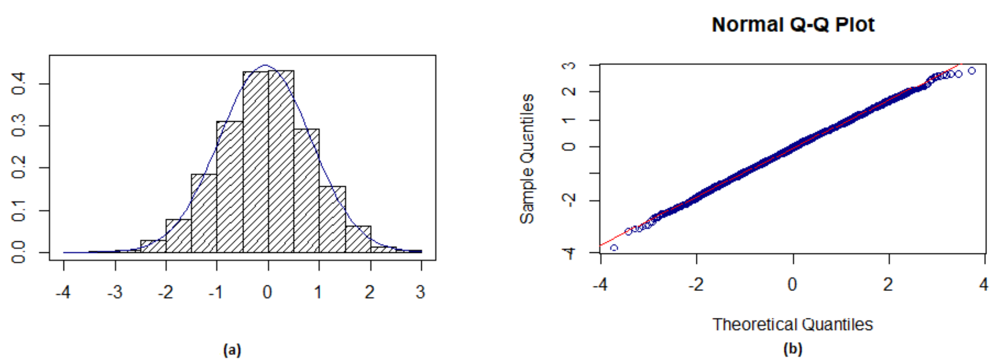

FIGURE 1. Histogram (a) and Q-Q plot (b) for Fréchet Distribution, $n = 900, kn = 10$.

We investigated two more distributions from the Hall model: Case 1: $\gamma = 1$, $D_1 = 1$ and $D_2 = 1/2$, $\beta = 3/4$; Case 2: $\gamma = 2$, $D_1 = 1$ and $D_2 = 1$, $\beta = 1$. We used $\rho = 3$, $p = 2$, $n = 500$ and $k_n = 7$ for Case 1, and $\rho = 1$, $p = 1$, $n = 900$ and $k_n = 10$ for Case 2. These n values are the threshold sample sizes to obtain a good normal approximation in (3.34). We obtained the following numerical results. Case 1: mean of the simulated Z_n values: 0.0013, standard deviation of the Z_n values:

0.9127, mean of the simulated $\widehat{\gamma}_n$ values: 1.0667; Case 2: mean of the simulated Z_n values: -0.0393, standard deviation of the Z_n values: 0.8878, mean of the simulated $\widehat{\gamma}_n$ values: 2.2267. The p-value of the chi-square test for normality is 0.323 for Case 1, and 0.6428 for Case 2. Figures 2 and 3 contain the histograms with the fitted normal curves and the Q–Q plots of the simulated quantities for Case 2 and Case 3, respectively.

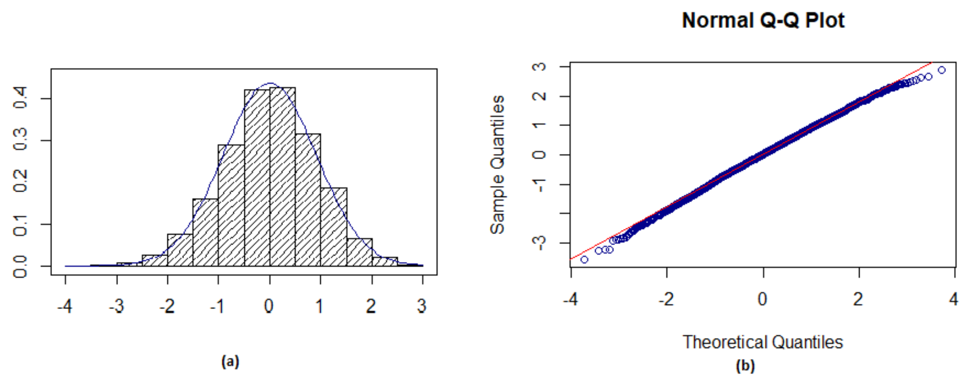

FIGURE 2. Histogram (a) and Q-Q plot (b) for Hall Model Case1, at $n = 500, kn = 7$.

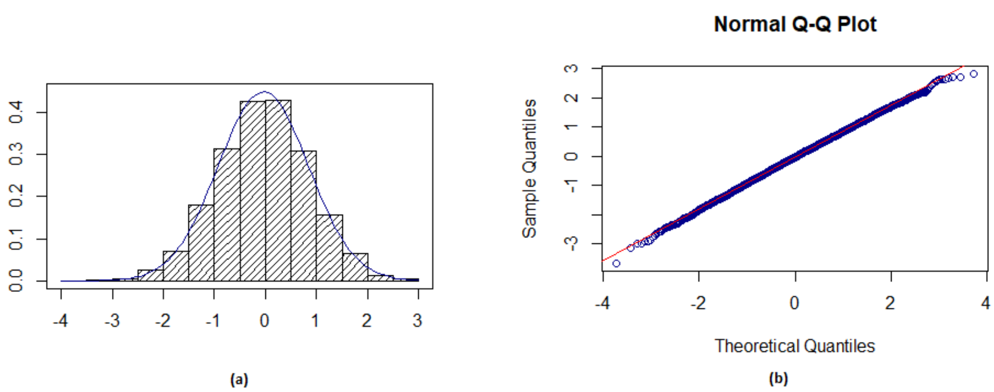

FIGURE 3. Histogram (a) and Q-Q plot (b) for Hall Model Case1, at $n = 900, kn = 10$.

CHAPTER 4

Limit laws for the norms of extremal samples

This chapter is based on the paper [**56**] by P. Kevei, L. Viharos and the author. We considered a class of estimator $\hat{\gamma}(n)$ which is an extension of the Hill estimator. We investigated the asymptotic properties of $\hat{\gamma}(n)$ under conditions of regular varying upper tail. Limit theorems are proved under appropriate assumptions. Gaussian and non-Gaussian (stable) limit are obtained depending on the growth rate of the power sequence p_n. The result is applied to the real data (Danish Fire insurance claim).

1. Introduction

Let $X, X_1, X_2, \ldots$ be independent identically distributed (iid) random variables with common distribution function $F(x) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$. For each $n \geq 1$, let $X_{1,n} \leq \ldots \leq X_{n,n}$ denote the order statistics of the sample $X_1, \ldots, X_n$. Assume that

$$1 - F(x) = x^{-1/\gamma} L(x),$$

where L is a slowly varying function at infinity and $\gamma > 0$. This is equivalent to the condition

$$Q(1-s) = s^{-\gamma} \ell(s), \tag{4.1}$$

where $Q(s) = \inf\{x : F(x) \geq s\}$, $s \in (0,1)$, stands for the quantile function, and ℓ is a slowly varying function at 0. For $p > 0$ introduce the notation

$$S_n(p) = \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\log \frac{X_{n+1-i,n}}{X_{n-k_n,n}} \right)^p. \tag{4.2}$$

The main object of the present paper is the estimate

$$\widehat{\gamma}(n) = \left(\frac{S_n(p)}{\Gamma(p+1)} \right)^{\frac{1}{p}} \tag{4.3}$$

of the tail index, where Γ is the usual gamma function. In what follows we always assume that $1 \leq k_n \leq n$ is a sequence of integers such that $k_n \to \infty$ and $k_n/n \to 0$.

As a special case for $p = 1$ we obtain the well-known Hill estimator of the tail index $\gamma > 0$ introduced by Hill in 1975 [**50**]. For $p = 2$ the estimator was suggested by Dekkers et al. [**32**], where they proved that $S_n(2) \to 2\gamma^2$ a.s. or in probability, depending on the assumptions on k_n, and they proved asymptotic normality of the estimator as well. For general $p > 0$ the properties of the estimator $\widehat{\gamma}(n)$ in (4.3) was investigated by Gomes and Martins [**44**]. Under second-order regular variation assumption they proved weak consistency and asymptotic normality of the estimator $\widehat{\gamma}(n)$. Segers [**77**] considered more general estimators of the form

$$\frac{1}{k_n} \sum_{i=1}^{k_n} f\left(\frac{X_{n+1-i,n}}{X_{n-k_n,n}}\right), \tag{4.4}$$

for a nice class of functions f, called *residual estimators*. Segers proved weak consistency and asymptotic normality under general conditions. More recently, Ciuperca and Mercadier [**15**] investigated weighted version of (4.2). The residual estimator of Segers was further analyzed for special function classes. Paulauskas and Vaičiulis [**66**] considered estimators of the form (4.4) with $f(x) = x^r(\log x)^p$. The classical Hill estimator can be considered as the logarithm of the geometric mean of the variables $X_{n+1-i,n}/X_{n-k,n}$. Based on this interpretation, Brilhante et al. [**12**] introduced the *mean of order p tail index estimator*, Beran et al. [**7**] introduced the *harmonic moment tail index estimator*, while very recently Penalva et al. [**67**] introduced the *Lehmer mean-of-order-p extreme value index estimator*. For a general overview on the generalizations of the Hill estimator we refer to [**67**].

To the best of our knowledge the possibility $p = p_n \to \infty$ in (4.3) was not considered before, which is the main focus of our paper. The estimate $\widehat{\gamma}(n)$ can be considered as $p_n \to \infty$ as the limit law for the norm of the extremal sample. In this direction Schlather [**75**], Bogachev [**11**], and Janßen [**54**] proved limit theorems for norms of iid samples.

In the present paper we investigate the asymptotic properties of $S_n(p_n)$ and $\widehat{\gamma}(n)$ both for $p > 0$ fixed and for $p = p_n \to \infty$. Although the focus of the paper is to obtain asymptotics for large p, in the course we obtain new results for p fixed. In Section 2 in Theorem 2.1 we prove strong consistency of the estimator for p fixed. Strong consistency was only obtained by Dekkers et al. [**32**] for $p = 1$ and $p = 2$, thus our result is new for general p. Asymptotic normality was obtained in several papers for different generalizations of the Hill estimator, see e.g. Gomes and Martins [**44**], Segers [**77**], Paulauskas and Vaičiulis [**66**], and Penalva et al. [**67**] for more general estimators. In all these results second-order regular variation is assumed. In Theorem 2.4 our assumptions on the slowly varying function ℓ are weaker, therefore the asymptotic normality in this generality is new. Our main results are contained in Section 3, where we obtain weak consistency and asymptotic normality when

$p = p_n \to \infty$. Under appropriate assumptions on the power sequence p_n we prove non-Gaussian stable limit theorems. Section 5 contains a small simulation study and data analysis. Here we show that for larger values of p the estimator is not so sensitive to the choice of k_n, which is a critical property in applications. The use of larger p values was already suggested in [**44**] for $p > 0$ fixed. We illustrate this property on the well-known dataset of Danish fire insurance claims, see Resnick [**71**] and Embrechts et al. [**36**, Example 6.2.9].

2. Results for fixed p

In what follows, $U, U_1, U_2, \ldots$ are iid uniform$(0,1)$ random variables, and $U_{1,n} \leq U_{2,n} \leq \ldots \leq U_{n,n}$ stand for the corresponding order statistics. To ease notation we frequently suppress the dependence on n and simply write $k = k_n$. Define $X = Q(1-U)$, $X_i = Q(1-U_i)$ for $i = 1, 2, \ldots$. According to the well-known quantile representation, $X, X_1, X_2, \ldots$ is an iid sequence with common distribution function F, which implies that S_n in (4.2) can be written as

$$S_n(p) = \frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{Q(1-U_{i,n})}{Q(1-U_{k+1,n})}\right)^p \quad \text{for each } n \geq 1, \text{ a.s.} \tag{4.5}$$

First we show strong consistency for $S_n(p)$. Our assumption on the sequence k_n is the same as in Theorem 2.1 in [**32**]. This is not far from the optimal condition $k_n/\log\log n \to \infty$, which was obtained by Deheuvels et al. [**31**] for $p = 1$. In what follows any nonspecified limit is meant as $n \to \infty$.

THEOREM 2.1. *Assume that* (4.1) *holds and* $k_n/n \to 0$, $(\log n)^\delta/k_n \to 0$ *for some* $\delta > 0$. *Then* $S_n(p) \to \gamma^p\Gamma(p+1)$ *a.s., that is for* $p > 0$ *fixed the estimator* $\widehat{\gamma}(n)$ *is strongly consistent.*

Weak consistency holds under weaker assumption on k_n. The following result is a special case of Theorem 2.1 in [**77**], and it follows from representation (4.5) and from the law of large numbers.

THEOREM 2.2. *Assume that* (4.1) *holds, and the sequence* (k_n) *is such that* $k_n \to \infty$, $k_n/n \to 0$. *Then* $S_n(p) \xrightarrow{\mathbb{P}} \gamma^p\Gamma(p+1)$, *that is for* $p > 0$ *fixed the estimator* $\widehat{\gamma}(n)$ *is weakly consistent.*

To prove asymptotic normality we use representation (4.5) where the summands are independent and identically distributed conditioned on $U_{k+1,n}$. Indeed, conditioned on $U_{k+1,n}$

$$(U_{1,n}, \ldots, U_{k,n}) \stackrel{\mathcal{D}}{=} \left(\widetilde{U}_{1,k}U_{k+1,n}, \ldots, \widetilde{U}_{k,k}U_{k+1,n}\right), \tag{4.6}$$

where $\widetilde{U}_1, \widetilde{U}_2, \ldots$ are iid uniform$(0,1)$ random variables, independent of $U_{k+1,n}$, and $\widetilde{U}_{1,k} < \ldots < \widetilde{U}_{k,k}$ stands for the order statistics of $\widetilde{U}_1, \ldots, \widetilde{U}_k$.

To state the result, we need some notation. Introduce the variable for $v \in (0,1)$

$$Y(v) = \log \frac{Q(1-Uv)}{Q(1-v)}, \tag{4.7}$$

where U is uniform$(0,1)$, and $Y(0) = -\gamma \log U$. Note that $Y(v)$ is 'continuous' in v at 0, that is $Y(0) = \lim_{v \downarrow 0} Y(v)$, since for the slowly varying function ℓ in (4.1) we have $\lim_{v\downarrow 0} \ell(vU)/\ell(v) = 1$ a.s. Define

$$m_{p,\gamma}(v) = m_p(v) = \mathbb{E}\left[Y(v)^p\right], \quad \sigma^2_{p,\gamma}(v) = \sigma^2_p(v) = \mathbf{Var}\left(Y(v)^p\right), \tag{4.8}$$

and the corresponding limiting quantities

$$\begin{aligned} m_p &= m_{p,\gamma} = \mathbb{E}\left[(-\gamma \log U)^p\right] = \gamma^p \Gamma(p+1), \\ \sigma_p^2 &= \sigma_{p,\gamma}^2 = \mathbf{Var}((-\gamma \log U)^p) = \gamma^{2p}\left(\Gamma(2p+1) - \Gamma^2(p+1)\right). \end{aligned}$$

Note that these quantities depend on the parameter γ. However, since the value $\gamma > 0$ is fixed, to ease notation we suppress γ.

Central limit theorem with random centering was obtained in Theorem 4.1 in [**77**]. Next, we spell out this result in our case. In the special case $p = 1$ we obtain Theorem 1.6 by Csörgő and Mason [**17**]. The key observation in the proof is representation (4.6). Recall the definition of the centering sequence from (4.8).

THEOREM 2.3. *Assume that* (4.1) *holds, and* $k_n \to \infty$, $k_n/n \to 0$. *Then*

$$\sqrt{k_n}\left(S_n(p_n) - m_p(U_{k+1,n})\right) \xrightarrow{\mathcal{D}} N(0, \sigma_p^2).$$

To obtain asymptotic normality for the estimator, that is, to change the random centering $m_p(U_{k+1,n})$ to m_p, we have to show that

$$\sqrt{k_n}(m_p(U_{k+1,n}) - m_p) \xrightarrow{\mathbb{P}} 0.$$

Since $U_{k+1,n} n/k \to 1$ in probability, this is the same as the deterministic convergence

$$\sqrt{k_n}(m_p(k/n) - m_p) \to 0;$$

see the proof of Theorem 2.4 for the precise version. In the case of the Hill estimator ($p = 1$) Csörgő and Viharos [**18**] obtained optimal conditions under which the random centering $m_p(U_{k+1,n})$ in Theorem 2.3 can be replaced by the deterministic one, $m_p(k/n)$. For general residual estimator this was obtained in Theorem 4.2 in [**77**]. In Theorem 4.5 in [**77**] assuming that the slowly varying function ℓ belongs to the de Haan class Π, conditions were obtained which ensure that the random centering can be replaced by the limit m_p. Our assumptions are weaker, but some second-order conditions are necessary.

Assume that there exist a regularly varying function a and a Borel set $B \subset [0,1]$ of positive measure such that

$$\lim_{v\downarrow 0} \frac{a(v)}{\ell(v)} = 0, \quad \limsup_{v\downarrow 0} \frac{|\ell(uv) - \ell(v)|}{a(v)} < \infty \quad \text{for } u \in B. \tag{4.9}$$

By Theorem 3.1.4 in Bingham et al. [**10**] condition (4.9) implies that the limsup in (4.9) is finite uniformly on any compact set of $(0,1]$. However, in general, uniformity cannot be extended to $[0,1]$.

We emphasize that we do not need exact second-order asymptotics for ℓ, only bounds. In particular, if ℓ belongs to the de Haan class Π (defined at 0) then condition (4.9) holds; see Appendix B in de Haan and Ferreira [**29**], or Chapter 3 in Bingham et al. [**10**]. Therefore, even in the special case $p = 1$, that is, for the Hill estimator, our next result is a generalization of Theorem 3.1 in [**32**]. The asymptotic normality of various generalizations of the Hill estimator are obtained under second-order regular variation for ℓ, see Theorem 4.5 in [**77**], formula (2.7) in [**44**], or Theorem 2 in [**66**]. Our conditions in the next result are weaker.

THEOREM 2.4. *Assume that* (4.9) *holds for* ℓ, *and* k_n *is such that* $k_n \to \infty$, $k_n/n \to 0$, *and*

$$\sqrt{k_n}\frac{a(k_n/n)}{\ell(k_n/n)} \to 0. \tag{4.10}$$

Then, with $\sigma_p^2 = \gamma^{2p}(\Gamma(2p+1) - \Gamma^2(p+1))$,

$$\frac{\sqrt{k_n}}{\sigma_p}\left(S_n(p_n) - \gamma^p\Gamma(p+1)\right) \xrightarrow{\mathcal{D}} N(0,1),$$

and

$$\frac{p\sqrt{k_n}}{\gamma^{1/p-1}\sigma_p}\left(\widehat{\gamma}(n) - \gamma\right) \xrightarrow{\mathcal{D}} N(0,1).$$

We point out that the growth condition (4.10) of the subsequence is the same as in Theorem 4.5 in [**78**] and in the special case $p = 1$ in de Haan [**28**]. However, in [**44**] under the second-order regular variation assumption the asymptotic normality of the estimator was proved under the less restrictive condition

$$\sqrt{k_n}\frac{a(k_n/n)}{\ell(k_n/n)} \to \lambda, \quad \lambda \in \mathbb{R}.$$

For more general estimators the asymptotic normality was proved under the condition above, see Theorem 2 in [**66**], Theorem 2 in [**7**], Theorem 2 in [**12**].

3. Asymptotics for large p

Conditioned on $U_{k+1,n}$ the sum $k_n S_n(p_n)$ in (4.5) is the sum of k_n iid random variables distributed as $Y(U_{k+1,n})$. This allows us to use appropriate uniform version of the results in [**11**] for power sums. These results are spelled out and proved in Section 4.3. As a consequence, we obtain limit theorems with *random centering and norming* for $S_n(p_n)$. In order to change to deterministic centering a precise analysis is needed.

First we need some notation. Let

$$f_v(x) = x^\gamma \ell(v/x), \quad v \in (0,1], \quad f_0(x) = x^\gamma, \quad x > 1.$$

Note that $Y(v)$ is defined for $v \in [0,1)$, while f_v is defined for $v \in [0,1]$. Then f_v is a left-continuous, nondecreasing, regularly varying function at infinity with index γ. Its inverse

$$g_v(y) = \inf\{x \,:\, f_v(x) > y\} = v g_1(y/v^\gamma), \quad v \in (0,1], \quad g_0(y) = y^{1/\gamma},$$

is regularly varying with index $1/\gamma$. Write $f = f_1$ and $g = g_1$. Then, $g(x) = x^{1/\gamma}\widetilde{\ell}(x)$, for a slowly varying function $\widetilde{\ell}$ such that

$$\ell(1/x)^{1/\gamma}\widetilde{\ell}(x^\gamma \ell(1/x)) \sim 1 \quad \text{as } x \to \infty. \tag{4.11}$$

The latter follows from the fact $f(g(x)) \sim g(f(x)) \sim x$. In fact, $\widetilde{\ell}(x)^\gamma$ is the de Bruijn conjugate of $\ell(1/x^{1/\gamma})$, see [**10**, Section 1.7].

Using that $f_v(x) > y$ if and only if $x > g_v(y)$, for $v \in (0,1]$ fixed the tail of $Y(v)$ is

$$\begin{aligned} \mathbb{P}(Y(v) > x) &= \mathbb{P}\left(\log U^{-\gamma}\frac{\ell(Uv)}{\ell(v)} > x\right) \\ &= \mathbb{P}(U^{-\gamma}\ell(Uv) > \ell(v)e^x) \\ &= \mathbb{P}(U^{-1} > g_v(\ell(v)e^x)) \\ &= e^{-x/\gamma}\left[\ell(v)^{1/\gamma}\widetilde{\ell}(v^{-\gamma}\ell(v)e^x)\right]^{-1}, \end{aligned}$$

and for $v = 0$ we have $\mathbb{P}(Y(0) > x) = e^{-x/\gamma}$. Thus, we obtain that the log-tail distribution function

$$h_v(x) := -\log \mathbb{P}(Y(v) > x) = \begin{cases} \frac{x}{\gamma} + \log\left(\ell(v)^{\frac{1}{\gamma}}\widetilde{\ell}(v^{-\gamma}\ell(v)e^x)\right), & v \in (0,1], \\ \frac{x}{\gamma}, & v = 0. \end{cases} \tag{4.12}$$

For any fixed $v \in [0,1]$ we have that $h_v(x) \sim x/\gamma$. In particular, it is regularly varying. For $\zeta > 0$ define η_v as the unique solution to

$$h_v(\eta_v(x)) = \zeta x. \tag{4.13}$$

3.1. Weak laws and Gaussian limit. It is pointed out in [**11**] that the proper rate of the power sequence p_n is $\log k_n$. Let us define the parameter ζ as

$$\zeta = \liminf_{n\to\infty} \frac{\log k_n}{p_n}. \tag{4.14}$$

For $\zeta \le 2$ we need precise assumption on the power sequence, and we assume that

$$k_n \sim e^{\zeta p_n}. \tag{4.15}$$

Therefore, depending on the range of ζ we have different definitions. In the results below we always state which of the two conditions we assume.

For the truncated moments for $v \in [0,1)$ put

$$m_p^1(v) = \mathbb{E}[Y(v)^p I(Y(v) \le \eta_v(p))]$$
$$\sigma_p^1(v) = \left(\mathbb{E}\left[Y(v)^{2p} I(Y(v) \le \eta_v(p))\right]\right)^{1/2}.$$

Recall (4.8) and define the centering and norming functions for $v \in [0,1)$,

$$\widetilde{m}_p(v) = \begin{cases} 0, & \zeta \in (0,1), \\ m_p^1(v), & \zeta = 1, \\ m_p(v), & \zeta \in (1,2), \end{cases} \qquad \widetilde{\sigma}_p(v) = \begin{cases} \sigma_p(v), & \zeta > 2, \\ \sigma_p^1(v), & \zeta = 2. \end{cases} \tag{4.16}$$

To ease notation put $m_p^1 = m_p^1(0)$, $\sigma_p^1 = \sigma_p^1(0)$, $\widetilde{m}_p = \widetilde{m}_p(0)$, and $\widetilde{\sigma}_p = \widetilde{\sigma}_p(0)$.

Weak consistency holds for $\zeta \ge 1$, while asymptotic normality holds for $\zeta \ge 2$. Note that in the borderline cases $\zeta = 1, 2$ the norming is different, and the condition on the subsequence p_n is stronger.

THEOREM 3.1. *Assume that $k_n \to \infty$, $k_n/n \to 0$, and $p_n \to \infty$. If $\zeta > 1$ in* (4.14) *or $\zeta = 1$ in* (4.15) *then*

$$\left(\widetilde{m}_{p_n}(U_{k_n+1,n})\right)^{-1} S_n(p_n) \xrightarrow{\mathbb{P}} 1. \tag{4.17}$$

In both cases $\widehat{\gamma}(n)$ is weakly consistent. Furthermore, if $\zeta > 2$ in (4.14) *or $\zeta = 2$ in* (4.15) *then*

$$\frac{\sqrt{k_n}}{\widetilde{\sigma}_{p_n}(U_{k+1,n})} \left(S_n(p_n) - \widetilde{m}_{p_n}(U_{k+1,n})\right) \xrightarrow{\mathcal{D}} N(0,1), \tag{4.18}$$

and

$$\frac{\sqrt{k_n}\widetilde{m}_{p_n}(U_{k+1,n})}{\widetilde{\sigma}_{p_n}(U_{k+1,n})} p_n \left[\left(\frac{S_n(p_n)}{\widetilde{m}_{p_n}(U_{k+1,n})}\right)^{1/p_n} - 1\right] \xrightarrow{\mathcal{D}} N(0,1). \tag{4.19}$$

Note that both the centering and the norming are random. To change to deterministic values $\widetilde{m}_{p_n}$ and $\widetilde{\sigma}_{p_n}$ further assumptions are needed. We always assume that for the slowly varying function (4.9) holds. For sequences, $k_n \to \infty$, $k_n/n \to 0$, and $p_n \to \infty$ introduce the notation

$$\beta_2 = \limsup_{n\to\infty} -p_n^{-1} \log \frac{a(k_n/n)}{\ell(k_n/n)} \geq \liminf_{n\to\infty} -p_n^{-1} \log \frac{a(k_n/n)}{\ell(k_n/n)} = \beta_1, \tag{4.20}$$

allowing $\beta_1 = \infty$, and let

$$\beta = \begin{cases} \beta_1, & \text{if } \beta_1 \geq 1, \\ \beta_2, & \text{if } 0 < \beta_1 \leq \beta_2 \leq 1 \\ 1, & \text{otherwise.} \end{cases} \tag{4.21}$$

Put $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$. Introduce the notation $H(u) = u - 1 - \log u$, $u > 0$, and for $x \in (0, \infty]$

$$\nu_x = x^{-1} H(2 \vee 2x), \quad \nu_\infty = 2.$$

Then ν is decreasing on $(0, 1]$, and increasing on $[1, \infty)$.

THEOREM 3.2. *Assume that for the slowly varying function ℓ (4.9) holds and $\beta_1 > 0$. If $\zeta > 1$ in (4.14) or $\zeta = 1$ in (4.15) then*

$$(\widetilde{m}_{p_n})^{-1} S_n(p_n) \xrightarrow{\mathbb{P}} 1. \tag{4.22}$$

If $\zeta > 2$ in (4.14) or $\zeta = 2$ in (4.15) then assume additionally that for some $\varepsilon > 0$,

$$\limsup_{n\to\infty} p_n^{-1} \log \left(\sqrt{k_n} \left(\frac{a(k_n/n)}{\ell(k_n/n)} \right)^{(\nu_\beta - \varepsilon) \wedge 1} \right) < \log 2.$$

Then

$$\frac{\sqrt{k_n}}{\widetilde{\sigma}_{p_n}} \left(S_n(p_n) - \widetilde{m}_{p_n} \right) \xrightarrow{\mathcal{D}} N(0, 1), \tag{4.23}$$

and

$$\frac{\sqrt{k_n} \widetilde{m}_{p_n}}{\gamma \widetilde{\sigma}_{p_n}} p_n \left(\widehat{\gamma}(n) - \gamma \right) \xrightarrow{\mathcal{D}} N(0, 1). \tag{4.24}$$

Note that $m_p/\sigma_p \sim 2^{-p}(p\pi)^{1/4}$ as $p \to \infty$.

Under stronger assumptions on the slowly varying function ℓ it is possible to weaken the conditions on k_n and p_n. A stronger condition on ℓ is that the limsup in (4.9) is finite uniformly in $u \in (0, 1]$, that is there exists a regularly varying function a such that

$$\lim_{v \downarrow 0} \frac{a(v)}{\ell(v)} = 0, \quad \limsup_{v \downarrow 0} \sup_{u \in (0,1]} \frac{|\ell(uv) - \ell(v)|}{a(v)} =: K_1 < \infty. \tag{4.25}$$

THEOREM 3.3. *Assume that for the slowly varying function ℓ (4.25) hold. Furthermore, $k_n \to \infty$, $k_n/n \to 0$, and $p_n \to \infty$ such that*

$$p_n \frac{a(k_n/n)}{\ell(k_n/n)} \to 0.$$

If $\zeta > 1$ in (4.14) or $\zeta = 1$ in (4.15) then (4.22) holds. If $\zeta > 2$ in (4.14) or $\zeta = 2$ in (4.15), and

$$\limsup_{n\to\infty} p_n^{-1} \log\left(\sqrt{k_n}\frac{a(k_n/n)}{\ell(k_n/n)}\right) < \log 2$$

then (4.23) and (4.24) hold.

3.2. Non-Gaussian stable limits. Next, we explore the regime $\zeta < 2$. Here we need the precise asymptotic assumption (4.15) on the power sequence p_n. We obtain non-Gaussian limits, where the characteristic exponent of the stable law equals ζ, coming from the growth rate of the power sequence p_n. Therefore, in what follows we use the notation $\zeta = \alpha$.

Let Z_α denote a one-sided α-stable random variable with characteristic function

$$\mathbb{E}e^{\mathrm{i}tZ_\alpha} = \begin{cases} \exp\left\{-\Gamma(1-\alpha)|t|^\alpha e^{-\mathrm{i}\frac{\pi\alpha}{2}\mathrm{sgn}u}\right\}, \\ \exp\left\{\mathrm{i}t(1-a) - \frac{\pi}{2}|t|\left(1 + \mathrm{i}\,\mathrm{sgn}u\,\frac{2}{\pi}\log|t|\right)\right\}, \end{cases}$$

where $a = 0.577\ldots$ stands for the Euler–Mascheroni constant.

THEOREM 3.4. *Assume that $k_n \to \infty$, $k_n/n \to 0$, and $p_n \to \infty$ such that (4.15) holds for some $\zeta = \alpha \in (0,2)$. Then*

$$\frac{k_n}{\eta_{U_{k_n+1,n}}(p_n)^{p_n}}\left(S_n(p_n) - \widetilde{m}_{p_n}(U_{k+1,n})\right) \xrightarrow{\mathcal{D}} Z_\alpha.$$

Moreover, for $\zeta = \alpha \in (0,1)$,

$$p_n\left(\frac{[k_n S_n(p_n)]^{1/p_n}}{\eta_{U_{k_n+1,n}}(p_n)} - 1\right) \xrightarrow{\mathcal{D}} \log Z_\alpha, \tag{4.26}$$

in particular,

$$\widehat{\gamma}(n) \xrightarrow{\mathbb{P}} \gamma\,\alpha e^{1-\alpha}. \tag{4.27}$$

While for $\alpha \in [1,2)$,

$$p_n \frac{k_n \widetilde{m}_{p_n}(U_{k_n+1,n})}{\eta_{U_{k_n+1,n}}(p_n)^{p_n}}\left[\left(\frac{S_n(p_n)}{\widetilde{m}_{p_n}(U_{k_n+1,n})}\right)^{1/p_n} - 1\right] \xrightarrow{\mathcal{D}} Z_\alpha. \tag{4.28}$$

In order to use deterministic norming and centering we need further assumptions on the slowly varying function. Note that $\eta_0(x) = \alpha\gamma x$. Recall (4.20) and (4.21).

THEOREM 3.5. *Assume* (4.15) *and that* (4.9) *holds. Furthermore,* $k_n \to \infty$, $k_n/n \to 0$, *and*

$$\widetilde{\ell}\,(n^\gamma \ell(k/n)) \sim \widetilde{\ell}\,((n/k)^\gamma \ell(k/n)) \tag{4.29}$$

and for $\alpha \in [1,2)$ *assume that*

$$\nu_\beta \beta_1 > \alpha - 1 - \log\alpha = H(\alpha). \tag{4.30}$$

Then for $\alpha \in (0,2)$,

$$\frac{k_n}{(\alpha\gamma p_n)^{p_n}}\left(S_n(p_n) - \widetilde{m}_{p_n}\right) \xrightarrow{\mathcal{D}} Z_\alpha. \tag{4.31}$$

For the estimator $\widehat{\gamma}(n)$ *if* $\alpha \in (0,1)$,

$$\frac{e^{\alpha-1}}{\alpha\gamma}p_n\left[\widehat{\gamma}(n)\left(1+\frac{\log p_n}{2p_n}\right) - \gamma\alpha e^{1-\alpha}\right] \xrightarrow{\mathcal{D}} \log Z_\alpha - \frac{\log 2\pi}{2}. \tag{4.32}$$

while for $\alpha \in (1,2)$,

$$\frac{\sqrt{2\pi}}{\gamma}e^{p_n(\alpha-1-\log\alpha)}p_n^{3/2}\left[\widehat{\gamma}(n) - \gamma\right] \xrightarrow{\mathcal{D}} Z_\alpha, \tag{4.33}$$

and for $\alpha = 1$,

$$\frac{\sqrt{2\pi}}{2\gamma}p_n^{3/2}\left[\widehat{\gamma}(n)\left(1+\frac{\log 2}{p_n}\right) - \gamma\right] \xrightarrow{\mathcal{D}} Z_1. \tag{4.34}$$

Condition (4.29) is rather implicit, since already $\widetilde{\ell}$ is implicit. However, from the proof it will be clear that this is exactly what is needed. In some natural special cases it can be checked. For example, (4.25) implies (4.29). Under some general growth conditions the de Bruijn conjugate, and so $\widetilde{\ell}$ can be determined explicitly, see [**10**, Corollary 2.3.4].

If $\beta = \beta_1$ then $\nu_\beta\beta_1 = H(2 \vee 2\beta) \geq H(2) > H(\alpha)$, that is condition (4.30) is automatic.

Under stronger assumptions on ℓ the result can be simplified.

THEOREM 3.6. *Assume* (4.15) *and that* (4.25) *holds. Furthermore,* $k_n \to \infty$, $k_n/n \to 0$, *and for* $\alpha \in [1,2)$ *assume that* $\beta_1 > H(\alpha)$. *Then* (4.31)*, and depending on the value* α*,* (4.32)*,* (4.33)*, or* (4.34) *hold.*

3.3. Examples. We spell out our results in three special cases. First, we consider the exact Pareto model, when $\ell \equiv 1$. Next, we consider the Hall model, when (4.25) holds. Finally, we consider the nonconstant slowly varying function $\ell(s) = -\log s$.

Example 3.7. The simplest special case is the strict Pareto model, when in (4.1) $\ell \equiv 1$. Then $m_p(v) \equiv m_p = \gamma^p \Gamma(p+1)$, thus the centering and norming do not depend on v. Furthermore, there is no other restriction on the sequence, only $k_n \to \infty$. In fact, $k_n = n$ is possible. Assume that $e^{\zeta p_n} \sim k_n$. Then, a direct consequence of Proposition 10.4 in [**11**] is that depending on the value ζ, (4.24), (4.32), (4.33), or (4.34) hold.

Example 3.8. Assume that the slowly varying function ℓ in (4.1) has the form

$$\ell(u) = c + O(u^\delta) \quad \text{with } \ c > 0, \delta > 0.$$

The asymptotic normality of the Hill estimator was proved for this subclass by Hall [**48**]. Condition (4.25) is satisfied with $a(u) = u^\delta$. By Proposition 4.5, for some $C > 0$

$$|m_{p_n}(u) - m_{p_n}| \le C\Gamma(p_n+1)\gamma^{p_n} u^\delta.$$

Let $p_n = \zeta^{-1} \log k_n$. For $\zeta \ge 2$ assume

$$\limsup_{n\to\infty} \frac{1}{p_n} \log \frac{k_n^{1/2+\delta}}{n^\delta} < \log 2, \tag{4.35}$$

and for $\zeta \in [1,2)$ assume

$$\liminf_{n\to\infty} -\frac{1}{p_n} \log \frac{k_n^\delta}{n^\delta} > H(\zeta). \tag{4.36}$$

Then depending on the value ζ, (4.24), (4.32), (4.33), or (4.34) hold. It is easy to see that both (4.35) and (4.36) are satisfied if $\log k_n = o(\log n)$.

Example 3.9. Finally, let $\ell(s) = -\log s$. Assume that $k_n = (\log n)^d$ for some $d > 0$, and $p_n = \zeta^{-1} \log k_n$. Then simple calculation shows that $\beta = \beta_1 = \beta_2 = \frac{\zeta}{d}$. Furthermore $\widetilde{\ell}(x) = (\gamma/\log x)^{1/\gamma}$, and condition (4.29) holds.

If $\zeta \ge 2$ assume $\zeta/2 - H(2 \vee (2\zeta/d))/\zeta < \log 2$. If $\zeta \in [1,2)$ then condition (4.30) always holds. Then depending on the value ζ, (4.24), (4.32), (4.33), or (4.34) hold.

4. Proofs

4.1. Strong consistency.

Lemma 4.1. *Assume that $k_n/(\log n)^\delta \to \infty$ for some $\delta > 0$, and $k_n/n \to 0$. Then*

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left(-\log \frac{U_{i,n}}{U_{k_n+1,n}} \right)^p \longrightarrow \Gamma(p+1) \quad a.s.$$

PROOF. Let F_n denote the empirical distribution function of the sample $U_1, \ldots, U_n$. Then, integrating by parts, we have

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k}\left(-\log\frac{U_{i,n}}{U_{k+1,n}}\right)^p &= \frac{n}{k}\int_{(0,U_{k,n}]}\left(-\log\frac{u}{U_{k+1,n}}\right)^p \widetilde{d}F_n(u)\\
&= \frac{n}{k}\left[F_n(U_{k,n})\left(-\log\frac{U_{k,n}}{U_{k+1,n}}\right)^p + \int_0^{U_{k,n}} F_n(u)\frac{p}{u}\left(-\log\frac{u}{U_{k+1,n}}\right)^{p-1}\widetilde{d}u\right] \qquad (4.37)\\
&= \left(-\log\frac{U_{k,n}}{U_{k+1,n}}\right)^p + p\frac{n}{k}\int_0^{U_{k,n}/U_{k+1,n}} F_n(U_{k+1,n}s)(-\log s)^{p-1}\frac{1}{s}\widetilde{d}s.
\end{aligned}$$

Theorem 1 by Wellner [**92**] implies that

$$\frac{n}{k}U_{k,n} \to 1 \quad \text{a.s. whenever } \ k_n/\log\log n \to \infty. \qquad (4.38)$$

Thus, the first term on the right-hand side of (4.37) tends to 0 a.s. For the second term

$$\begin{aligned}
&\frac{n}{k}\int_0^{U_{k,n}/U_{k+1,n}} F_n(U_{k+1,n}s)(-\log s)^{p-1}s^{-1}\widetilde{d}s\\
&= \frac{n}{k}U_{k+1,n}\int_0^{U_{k,n}/U_{k+1,n}}(-\log s)^{p-1}\widetilde{d}s\\
&\quad + \frac{n}{k}\int_0^{U_{k,n}/U_{k+1,n}}(F_n(U_{k+1,n}s) - U_{k+1,n}s)(-\log s)^{p-1}s^{-1}\widetilde{d}s\\
&=: I_n + II_n.
\end{aligned}$$

Again by (4.38)

$$I_n \to \int_0^1 (-\log s)^{p-1}\widetilde{d}s = \Gamma(p) \quad \text{a.s.} \qquad (4.39)$$

For the second term, choosing $\nu \in (0, 1/2)$, we have

$$\begin{aligned}
II_n &\sim \int_0^1 \frac{F_n(U_{k+1,n}s) - U_{k+1,n}s}{U_{k+1,n}s}(-\log s)^{p-1}\widetilde{d}s\\
&= \int_0^1 \frac{F_n(U_{k+1,n}s) - U_{k+1,n}s}{(U_{k+1,n}s)^{1/2-\nu}}(-\log s)^{p-1}(U_{k+1,n}s)^{-1/2-\nu}\widetilde{d}s\\
&\le \sup_{u\le U_{k+1,n}}\frac{|F_n(u)-u|}{u^{1/2-\nu}}U_{k+1,n}^{-1/2-\nu}\int_0^1(-\log s)^{p-1}s^{-1/2-\nu}\widetilde{d}s \qquad (4.40)\\
&\le C\left(\frac{\log\log n}{k}\right)^{1/2}\left[\left(\frac{n}{k}\right)^{\nu}\left(\frac{n}{\log\log n}\right)^{1/2}\sup_{u\le 2k/n}\frac{|F_n(u)-u|}{u^{1/2-\nu}}\right],
\end{aligned}$$

where $C > 0$ is a finite constant, not depending on n, k_n. Using Theorem 1(ii) by Einmahl and Mason [**35**] we see that the last term in (4.40) is a.s. bounded, if

$k_n \geq (\log n)^{(1-2\nu)/(2\nu)}$, which holds if ν is close enough to $1/2$. The first term in (4.40) tends to 0. From (4.39), (4.40), and (4.37) the statement follows. $\square$

PROOF OF THEOREM 2.1. By the Potter bounds ([**10**, Theorem 1.5.6]), for any $A > 1$, $\varepsilon > 0$ there exist $x_0 = x_0(A, \varepsilon)$ such that

$$A^{-1}(y/x)^{-\varepsilon} \leq \frac{\ell(x)}{\ell(y)} \leq A(y/x)^{\varepsilon} \quad \text{for any } \ 0 < x \leq y \leq x_0. \tag{4.41}$$

Since $k/n \to 0$, equation (4.38) implies $U_{k+1,n} \to 0$ a.s. Therefore, for n large enough a.s.

$$S_n(p) \leq \frac{1}{k} \sum_{i=1}^{k} \left(-(\gamma + \varepsilon) \log \frac{U_{i,n}}{U_{k+1,n}} + \log A \right)^p. \tag{4.42}$$

First let $p \leq 1$. Using the subadditivity $(a+b)^p \leq a^p + b^p$, $a, b > 0$, by Lemma 4.1 we obtain a.s.

$$\begin{aligned} \limsup_{n\to\infty} S_n(p) &\leq (\gamma+\varepsilon)^p \limsup_{n\to\infty} \frac{1}{k} \sum_{i=1}^{k} \left(-\log \frac{U_{i,n}}{U_{k+1,n}} \right)^p + (\log A)^p \\ &= (\gamma+\varepsilon)^p \Gamma(p+1) + (\log A)^p. \end{aligned}$$

Letting $A \downarrow 1$ and $\varepsilon \downarrow 0$ we have a.s. $\limsup_{n\to\infty} S_n(p) \leq \gamma^p \Gamma(p+1)$.

Next, let $p > 1$. The convexity of the function x^p implies that for any $\varepsilon' > 0$, for $a, b > 0$

$$\begin{aligned} (a+b)^p &\leq (1+\varepsilon')a^p + \left(1 - (1+\varepsilon')^{-1/(p-1)}\right)^{-(p-1)} b^p \\ &=: (1+\varepsilon')a^p + C_{\varepsilon'} b^p. \end{aligned}$$

Therefore, using Lemma 4.1 and (4.42), we obtain a.s.

$$\begin{aligned} \limsup_{n\to\infty} S_n(p) &\leq (\gamma+\varepsilon)^p (1+\varepsilon') \limsup_{n\to\infty} \frac{1}{k} \sum_{i=1}^{k} \left(-\log \frac{U_{i,n}}{U_{k+1,n}} \right)^p + C_{\varepsilon'} (\log A)^p \\ &= (\gamma+\varepsilon)^p (1+\varepsilon') \Gamma(p+1) + C_{\varepsilon'} (\log A)^p. \end{aligned}$$

As $A \downarrow 1$, $\varepsilon \downarrow 0$, $\varepsilon' \downarrow 0$, we have a.s. $\limsup_{n\to\infty} S_n(p) \leq \gamma^p \Gamma(p+1)$.

With the analogous lower bound, the proof is complete. $\square$

4.2. Moment bounds. First we need three simple auxiliary lemmas.

LEMMA 4.2. *For $a \in (0, 1/2)$, $b \in (-1/2, 1/2)$, and $a + b > 0$ we have*

$$|(a+b)^p - a^p| \leq \begin{cases} p|b|, & p \geq 1, \\ 2|b| a^{p-1}, & p \leq 1. \end{cases}$$

PROOF. Simply $(a+b)^p - a^p = bp\xi^{p-1}$, with ξ being between a and $a+b$. If $b > -a/2$ then $\xi \in [a/2, 1]$, thus

$$|(a+b)^p - a^p| \le |b|p\left((a/2)^{p-1} \vee 1\right).$$

If $b < -a/2$ then $\xi \le a$, thus $\xi^{p-1} \le a^{p-1}$ for $p \ge 1$, and

$$|(a+b)^p - a^p| \le |b|pa^{p-1}.$$

While if $b < -a/2$ and $p < 1$

$$\begin{aligned}|(a+b)^p - a^p| &= (a - |b| + |b|)^p - (a-|b|)^p \le |b|^p \\ &= |b||b|^{p-1} \le |b|(a/2)^{p-1}.\end{aligned}$$

$\square$

LEMMA 4.3. *For $x \ge p > 0$ we have*

$$\int_x^\infty e^{-y} y^p \widetilde{d}y \le x^{p+1} e^{-x} (x-p)^{-1}.$$

PROOF. Simple calculation gives that

$$\begin{aligned}\int_x^\infty e^{-y} y^p \widetilde{d}y &= x^{p+1} e^{-x} \int_1^\infty e^{-x(u-1)+p\log u} \widetilde{d}u \\ &= x^{p+1} e^{-x} \int_1^\infty e^{-(x-p)(u-1)-p(u-1-\log u)} \widetilde{d}u \\ &\le x^{p+1} e^{-x} \int_1^\infty e^{-(x-p)(u-1)} \widetilde{d}u \\ &= x^{p+1} e^{-x} (x-p)^{-1}.\end{aligned}$$

$\square$

LEMMA 4.4. *For $\zeta = 1$ as $p \to \infty$ for the truncated moments we have*

$$m_p^1 \sim \left(\frac{\gamma p}{e}\right)^p \frac{\sqrt{p\pi}}{\sqrt{2}}, \quad \textit{and} \quad \sigma_p^1 \sim \left(\frac{2\gamma p}{e}\right)^p (p\pi)^{1/4}.$$

PROOF. Since $\eta_0(p) = \zeta\gamma p$, by definition

$$m_p^1 = \gamma^p \int_0^p y^p e^{-y} \widetilde{d}y \quad \text{and} \quad (\sigma_p^1)^2 = \gamma^{2p} \int_0^{2p} y^{2p} e^{-y} \widetilde{d}y.$$

We have

$$\int_0^p y^p e^{-y} \widetilde{d}y = p^{p+1} e^{-p} \int_0^1 e^{-p(x-1-\log x)} \widetilde{d}x.$$

The exponent is negative and $x - 1 - \log x \sim (x-1)^2/2$ as $x \uparrow 1$. Thus

$$\int_0^1 e^{-p(x-1-\log x)} \widetilde{d}x \sim \sqrt{\pi/(2p)}.$$

$\square$

Proposition 4.5. *Assume* (4.25) *and that*

$$\lim_{v\downarrow 0} p_v \frac{a(v)}{\ell(v)} = 0. \tag{4.43}$$

Then there exists $v_0 > 0$ such that for all $v \in (0, v_0)$

$$|m_{p_v}(v) - m_{p_v}| \leq 2K_1 \frac{a(v)}{\ell(v)} \gamma^{p_v - 1} \Gamma(p_v + 1).$$

Proof. To ease notation put

$$\eta(u,v) = \left(-\gamma \log u + \log \frac{\ell(uv)}{\ell(v)}\right)^p - (-\gamma \log u)^p. \tag{4.44}$$

We have by (4.1)

$$\begin{aligned} m_p(v) - m_p &= \mathbb{E}\left[\left(\log \frac{Q(1-Uv)}{Q(1-v)}\right)^p - (-\gamma \log U)^p\right] \\ &= \mathbb{E}\left[\left(-\gamma \log U + \log \frac{\ell(Uv)}{\ell(v)}\right)^p - (-\gamma \log U)^p\right] \\ &= \int_0^1 \eta(u,v) \widetilde{d}u =: I_1(\delta) + I_2(\delta), \end{aligned}$$

where I_1, I_2 are the integrals on $(0, 1-\delta)$, $(1-\delta, 1)$, with $\delta \in (0, 1/2)$.

First we deal with the integral on $(0, 1-\delta)$. By (4.41), for any $\varepsilon > 0$, $A > 1$, there is $v_0 > 0$ such that for $v \leq v_0$, $u \in (0,1)$,

$$A^{-1} u^{\varepsilon} \leq \frac{\ell(uv)}{\ell(v)} \leq A u^{-\varepsilon}, \tag{4.45}$$

implying that uniformly on $u \in (0, 1-\delta]$,

$$\frac{\log \frac{\ell(uv)}{\ell(v)}}{-\log u} \to 0 \quad \text{as } v \downarrow 0. \tag{4.46}$$

Writing

$$\frac{\ell(uv) - \ell(v)}{\ell(v)} = \frac{a(v)}{\ell(v)} \frac{\ell(uv) - \ell(v)}{a(v)},$$

by (4.25) we see that the first factor tends to 0 and the second factor is bounded. Therefore, uniformly in $u \in [0,1]$,

$$\log \frac{\ell(uv)}{\ell(v)} \sim \frac{a(v)}{\ell(v)} \frac{\ell(uv) - \ell(v)}{a(v)} \quad \text{as } v \downarrow 0. \tag{4.47}$$

By (4.46) and (4.47), if (4.43) holds then, uniformly on $u \in [0, 1-\delta]$,

$$\left(1 + \frac{\log \frac{\ell(uv)}{\ell(v)}}{-\gamma \log u}\right)^p - 1 \sim p\,(-\gamma \log u)^{-1} \frac{a(v)}{\ell(v)} \frac{\ell(uv) - \ell(v)}{a(v)}. \tag{4.48}$$

Thus,

$$I_1(\delta) \le p\frac{a(v)}{\ell(v)}\,\frac{3}{2}K_1\gamma^{p-1}\int_0^{1-\delta}(-\log u)^{p-1}\,\widetilde{d}u. \tag{4.49}$$

Next, we turn to I_2. Note that (4.47) holds, but (4.46) does not, because $\log u$ can be small. Choosing $\delta > 0$ small enough we can achieve that $-\gamma\log(1-\delta) \in (0, 1/2)$ and by (4.47) also that $\log \ell(uv)/\ell(v) \in (-1/2, 1/2)$ for v small and $u \in [1-\delta, 1]$. Therefore, we can apply Lemma 4.2 with $a = -\gamma \log u$ and $b = \log(\ell(uv)/\ell(v))$ together with (4.47) and (4.25), and we obtain for $p \le 1$ that

$$\begin{aligned}|\eta(u,v)| &\le 2\left|\log\frac{\ell(uv)}{\ell(v)}\right|(-\gamma\log u)^{p-1}\\ &\le \frac{a(v)}{\ell(v)}2K_1(-\gamma\log u)^{p-1}.\end{aligned}$$

While, for $p \ge 1$

$$|\eta(u,v)| \le p\left|\log\frac{\ell(uv)}{\ell(v)}\right| \le p\frac{a(v)}{\ell(v)}K_1.$$

Summarizing,

$$I_2(\delta) \le \begin{cases}\frac{a(v)}{\ell(v)}2K_1\gamma^{p-1}\int_{1-\delta}^1(-\log u)^{p-1}\widetilde{d}u, & p \le 1,\\ p\frac{a(v)}{\ell(v)}K_1\delta, & p \ge 1.\end{cases} \tag{4.50}$$

The bounds (4.49) and (4.50) imply the statement. □

PROPOSITION 4.6. *Assume* (4.9) *and let*

$$\beta_2 := \limsup_{v\downarrow 0}\frac{-\log\frac{a(v)}{\ell(v)}}{p_v} \ge \liminf_{v\downarrow 0}\frac{-\log\frac{a(v)}{\ell(v)}}{p_v} := \beta_1, \tag{4.51}$$

allowing $\beta_1 = \infty$. *Assume either* $\beta_1 \ge 1$ *or* $\beta_2 \le 1$, *and define* β *as in* (4.21). *Then for any* $\varepsilon > 0$ *there exists a* $K > 0$ *such that for* v *small enough*

$$|m_{p_v}(v) - m_{p_v}| \le K\left(\frac{a(v)}{\ell(v)}\right)^{(\nu_\beta-\varepsilon)\wedge 1}(\gamma+\varepsilon)^{p_v}\,\Gamma(p_v+1).$$

Note that if $p > 0$ is fixed then $\beta = \infty$ and we obtain the same bound as in Proposition 4.5.

PROOF. The difference compared to the previous proof is that (4.25) does not hold uniformly in $[0, 1]$, which implies that the integral of $\eta(u, v)$ in (4.44) on the interval $[0, \delta]$ has to be treated differently.

By Theorem 3.1.4 in [**10**] (translating the results from infinity to zero, by defining $\overline{\ell}(x) = \ell(x^{-1})$, $\overline{a}(x) = a(x^{-1})$)

$$\limsup_{v\downarrow 0} \sup_{u\in[\delta,1]} \frac{|\ell(uv) - \ell(v)|}{a(v)} =: K_1(\delta) < \infty.$$

This implies that the bound (4.50) on $[1-\delta, 1]$ remains true and on $[\delta, 1-\delta]$ as in (4.49) we have

$$\int_\delta^{1-\delta} \eta(u,v)\widetilde{d}u \le p\frac{a(v)}{\ell(v)}\frac{3}{2}K_1\gamma^{p-1}\int_\delta^{1-\delta} (-\log u)^{p-1}\,\widetilde{d}u. \tag{4.52}$$

Recall (4.44) and let

$$J_1 = \int_0^{b(v)} \eta(u,v)\widetilde{d}u, \quad J_2 = \int_{b(v)}^{\delta} \eta(u,v)\widetilde{d}u, \tag{4.53}$$

where

$$b(v) = \left(\frac{a(v)}{\ell(v)}\right)^2 \wedge e^{-2p}. \tag{4.54}$$

By Theorem 3.1.4 in [**10**] for any $\varepsilon > 0$ there is $v_0(\varepsilon) > 0$ and $K_2(\varepsilon) > 0$ such that

$$\frac{|\ell(uv) - \ell(v)|}{a(v)} \le K_2(\varepsilon)u^{-\varepsilon} \quad \text{for all } u \le 1, v \le v_0(\varepsilon). \tag{4.55}$$

By (4.54) and (4.51) for $\varepsilon_1 > 0$ small enough

$$p\frac{a(v)}{\ell(v)}b(v)^{-\varepsilon_1} \to 0. \tag{4.56}$$

Using (4.55), for $u \ge b(v)$

$$\frac{|\ell(uv) - \ell(v)|}{\ell(v)} \le K_2(\varepsilon_1)\frac{a(v)}{\ell(v)}u^{-\varepsilon_1} \le K_2(\varepsilon_1)\frac{a(v)}{\ell(v)}b(v)^{-\varepsilon_1} \to 0,$$

therefore

$$\left|\log\frac{\ell(uv)}{\ell(v)}\right| \sim \frac{|\ell(uv) - \ell(v)|}{\ell(v)} \le K_2(\varepsilon_1)\frac{a(v)}{\ell(v)}u^{-\varepsilon_1}.$$

By (4.56) for $u \in [b(v), \delta]$ the asymptotic equality in (4.48) holds, thus for J_2 in (4.53)

$$\begin{aligned} J_2 &\sim \int_{b(v)}^{\delta} (-\gamma\log u)^p p(-\gamma\log u)^{-1}\frac{a(v)}{\ell(v)}\frac{\ell(uv) - \ell(v)}{a(v)}\widetilde{d}u \\ &\le p\frac{a(v)}{\ell(v)}K_2(\varepsilon_1)\int_{b(v)}^{\delta} (-\gamma\log u)^{p-1}u^{-\varepsilon_1}\,\widetilde{d}u \\ &\le p\frac{a(v)}{\ell(v)}K_2(\varepsilon_1)(1-\varepsilon_1)^{-p}\gamma^{p-1}\Gamma(p), \end{aligned} \tag{4.57}$$

where at the last inequality we used that

$$\int_0^1 (-\log u)^{p-1} u^{-\varepsilon_1}\, \widetilde{d}u = \int_0^\infty y^{p-1} e^{-(1-\varepsilon_1)y} \widetilde{d}y$$
$$= (1-\varepsilon_1)^{-p}\, \Gamma(p).$$

On $(0, b(v))$ using (4.45), $b(v) \to 0$, Lemma 4.3, and that $-\log b(v) - p \geq (-\log b(v))/2$ we obtain for v small enough

$$\begin{aligned} J_1 &\leq 2 \int_0^{b(v)} (-(\gamma+\varepsilon)\log u + \log A)^p\, \widetilde{d}u \\ &\leq 2(\gamma + 2\varepsilon)^p \int_0^{b(v)} (-\log u)^p\, \widetilde{d}u \\ &= 2(\gamma+2\varepsilon)^p \int_{-\log b(v)}^\infty y^p e^{-y} \widetilde{d}y \\ &\leq 2(\gamma+2\varepsilon)^p (-\log b(v))^{p+1} e^{\log b(v)} (-\log b(v) - p)^{-1} \\ &\leq 4(\gamma+2\varepsilon)^p\, (-\log b(v))^p\, b(v). \end{aligned} \tag{4.58}$$

Note that for $\log x > p$

$$\frac{(\log x)^p}{x} \frac{e^p}{p^p} = \exp\left\{ -p \left(\frac{\log x}{p} - 1 - \log \frac{\log x}{p} \right) \right\}$$
$$= \exp\left\{ -pH\left(\frac{\log x}{p} \right) \right\}.$$

Thus with $x = b(v)^{-1}$

$$\left(\frac{e}{p}\right)^p (-\log b(v))^p\, b(v) = \exp\left\{ -pH\left(2 \vee \frac{-2\log(a(v)/\ell(v))}{p} \right) \right\}$$
$$= \left(\frac{a(v)}{\ell(v)} \right)^{\frac{p}{-\log(a(v)/\ell(v))} H\left(2 \vee \frac{-2\log(a(v)/\ell(v))}{p}\right)}.$$

The function $\nu_x = x^{-1}\, H(2 \vee 2x)$ is strictly decreasing on $(0,1]$, and strictly increasing on $[1,\infty)$ attaining its unique minimum at 1. Continuing (4.58) for any $\varepsilon_2 > 0$ for v small enough

$$J_1 \leq \frac{4}{\sqrt{p\pi}} (\gamma+2\varepsilon)^p\, \Gamma(p+1) \left(\frac{a(v)}{\ell(v)} \right)^{\nu_\beta - \varepsilon_2}.$$

Combining with (4.57), (4.52), and (4.50) the result follows. □

As an easy consequence of the moment bounds we show that the random centering and norming can be substituted with the deterministic one.

PROOF OF THEOREM 2.4. The theorem is an immediate consequence of Theorem 2.3 and Proposition 4.6. Indeed, by Proposition 4.6

$$\sqrt{k}\,|m_p(U_{k+1,n}) - m_p| \le c\sqrt{k}\frac{a(U_{k+1,n})}{\ell(U_{k+1,n})} = \sqrt{k}\frac{a(k/n)}{\ell(k/n)}\frac{a(U_{k+1,n})}{a(k/n)}\frac{\ell(k/n)}{\ell(U_{k+1,n})}.$$

By the assumption $\sqrt{k}a(k/n)/\ell(k/n) \to 0$, while the last two factors tends to 1, since a and ℓ are regularly varying and $U_{k+1,n} \sim k/n$.

The central limit theorem for $\widehat{\gamma}(n)$ follows from the previous result using the delta method, see Agresti [**1**, Section 14.1]. □

4.3. Limit results for power sums. In this section we assume that $p = p_n$ tends to infinity at a certain rate. We prove the analogues of Bogachev's result [**11**, Section 2] for the random variables $Y(v)$ uniformly in v. As the log-tail distribution function h_v in (4.12) is regularly varying, for each $v \in [0,1)$ fixed all the following results are consequences of Bogachev's results. However, the main difficulty in our setup is the additional parameter v, in which we need some kind of uniformity. We apply these results to prove limit theorems for the $S_n(p_n)$ and $\widehat{\gamma}(n)$.

Recall (4.7). Let $Y(v), Y_1(v), Y_2(v), \dots$ be iid random variables, and put

$$Z_n(p,v) = \sum_{i=1}^{n} Y_i(v)^p.$$

First we determine the asymptotic behavior of the moments as $p \to \infty$.

LEMMA 4.7. *For any $\varepsilon > 0$ there is a $p_0 > 0$ such that for $v \in [0,1)$, $p > p_0$*

$$(\gamma - \varepsilon)^p\,\Gamma(p+1) \le m_p(v) \le (\gamma + \varepsilon)^p\,\Gamma(p+1). \tag{4.59}$$

In particular, as $p \to \infty$ uniformly in v

$$\frac{\log m_p(v)}{p} - \log p \to \log\gamma\, - 1.$$

PROOF. First note that if X is a nonnegative random variable for which $\mathbb{P}(X > x) > 0$ for any x then for any $K > 0$

$$\mathbb{E}X^p \sim \mathbb{E}X^p I(X > K) \quad \text{as } p \to \infty.$$

This implies that for any $\varepsilon > 0$ and $a > 0$ there exist $p_0 = p_0(\varepsilon, a)$ such that for $p > p_0$

$$(1-\varepsilon)^p\,\mathbb{E}(X+a)^p \le \mathbb{E}X^p \le (1+\varepsilon)^p\,\mathbb{E}(X-a)^p. \tag{4.60}$$

Using the Potter bounds (see (4.45)) and (4.60), for any $\varepsilon > 0$ there exists $A > 1$ and $p_0 > 0$ such that for $v \in [0,1]$, $p > p_0$

$$\begin{aligned} m_p(v) &= \mathbb{E}\left(\log\left(U^{-\gamma}\frac{\ell(Uv)}{\ell(v)}\right)\right)^p \\ &\le \mathbb{E}\left(\log\left(U^{-(\gamma+\varepsilon)}A\right)\right)^p \\ &\le (\gamma+\varepsilon)^p\mathbb{E}\left(\log U^{-1} + \frac{\log A}{\gamma+\varepsilon}\right)^p \\ &\le ((1+\varepsilon)(\gamma+\varepsilon))^p\Gamma(p+1). \end{aligned}$$

Together with an analogous lower bound, (4.59) follows. The second part simply follows from Stirling's formula. $\square$

If $V_n(v)$ is a sequence of random variables indexed by $v \in [0,1)$, then $V_n(v)$ *converges in distribution uniformly in* v *to a random variable* W, if for each continuity point x of the distribution function of W

$$\lim_{n\to\infty}\sup_{v\in[0,1)}|\mathbb{P}(V_n(v) \le x) - \mathbb{P}(W \le x)| = 0.$$

Similarly, $V_n(v)$ *converges in probability uniformly in* v *to a random variable* W, if for each $\varepsilon > 0$

$$\lim_{n\to\infty}\sup_{v\in[0,1)}\mathbb{P}(|V_n(v) - W| > \varepsilon) = 0.$$

For the sequence $p = p_n$ let

$$\liminf_{n\to\infty}\frac{\log n}{p_n} = \zeta \ge 0; \tag{4.61}$$

for $\zeta \le 2$ we need the stronger assumption

$$n \sim e^{\zeta p_n}. \tag{4.62}$$

To obtain a weak law of large numbers we need that $\zeta > 1$.

Proposition 4.8. *If* $\zeta > 1$ *in* (4.61) *or* $\zeta = 1$ *in* (4.62) *then uniformly for* $v \in [0,1)$ *as* $p_n \to \infty$

$$\frac{Z_n(p_n,v) - n\widetilde{m}_{p_n}(v)}{n\widetilde{m}_{p_n}(v)} \xrightarrow{\mathbb{P}} 0.$$

Proof. Let $\zeta > 1$. We follow the proof of Theorem 2.1 in **[11]**. Fix $\varepsilon > 0$, and let $r \in [1,2]$. Using the Markov inequality, the Marcinkiewicz–Zygmund inequality (see e.g. [**68**, 2.6.18]), and the subadditivity we have with some $c_r > 0$

$$\begin{aligned}
&\mathbb{P}\left(\frac{|Z_n(p,v) - nm_p(v)|}{nm_p(v)} > \varepsilon\right) \\
&\le (\varepsilon n m_p(v))^{-r}\mathbb{E}|Z_n(p,v) - nm_p(v)|^r \\
&\le c_r(\varepsilon n m_p(v))^{-r}\mathbb{E}\left(\sum_{i=1}^n (Y_i(v)^p - m_p(v))^2\right)^{r/2} \\
&\le c_r(\varepsilon n m_p(v))^{-r} n\mathbb{E}|Y(v)^p - m_p(v)|^r \\
&\le c_r\varepsilon^{-r}n^{1-r}\frac{m_{rp}(v)}{m_p(v)^r}.
\end{aligned} \tag{4.63}$$

By Lemma 4.7 for any $\varepsilon_1 > 0$ we can choose $p_0 > 0$ such that for $v \in [0,1)$ and $p > p_0$

$$\frac{m_{rp}(v)}{m_p(v)^r} \le \frac{(\gamma+\varepsilon_1)^{rp}\Gamma(rp+1)}{(\gamma-\varepsilon_1)^{rp}\Gamma(p+1)^r} \le (1+\varepsilon_2)^{rp}\frac{\Gamma(rp+1)}{\Gamma(p+1)^r},$$

with $\varepsilon_2 = 2\varepsilon_2/(\gamma - \varepsilon_2)$. Thus, by the Stirling formula

$$\begin{aligned}
\limsup_{p\to\infty}\frac{1}{p}\log\frac{m_{rp}(v)}{n^{r-1}m_p(v)^r} &\le r\log(1+\varepsilon_2) + r\log r - (r-1)\liminf_{p\to\infty}\frac{\log n}{p} \\
&\le r\log(1+\varepsilon_2) + r\log r - (r-1)\zeta.
\end{aligned} \tag{4.64}$$

As $\zeta > 1$ we can choose $r \in [1,2]$ such that $r\log r - (r-1)\zeta < 0$. Then choosing ε_1 small enough we see that the right-hand side in (4.64) is negative, implying that the right-hand side in (4.63) tends to 0.

For $\zeta = 1$ the result is a consequence of Proposition 4.12. We only need that $n\widetilde{m}_p(v)/\eta_v(p)^p \to \infty$, which follows from (4.70) in Lemma 4.11 with $r = \zeta = 1$. $\square$

For the central limit theorem we need further restriction on p_n.

PROPOSITION 4.9. *If $\zeta > 2$ in (4.61) or $\zeta = 2$ in (4.62) then uniformly on $[0,1)$*

$$\frac{Z_n(p_n,v) - n\widetilde{m}_{p_n}(v)}{\sqrt{n}\widetilde{\sigma}_{p_n}(v)} \xrightarrow{\mathcal{D}} N(0,1).$$

PROOF. Let $\zeta > 2$. By Lyapunov's theorem (see e.g. Theorem 27.3 in Billingsley [**9**]) it is enough to show that for some $\delta > 0$ uniformly in v

$$\frac{n}{(\sqrt{n}\sigma_p(v))^{2+\delta}}\mathbb{E}|Y(v)^p - m_p(v)|^{2+\delta} \to 0$$

as $n \to \infty$. By Lemma 4.7 $\sigma_p(v) \sim \sqrt{m_{2p}(v)}$ as $p \to \infty$. Thus we have to show that

$$\frac{m_{p(2+\delta)}(v)}{n^{\delta/2}m_{2p}(v)^{1+\delta/2}} \to 0.$$

As in the proof of Proposition 4.8

$$\limsup_{p\to\infty} \frac{1}{p} \log \frac{m_{p(2+\delta)}(v)}{n^{\delta/2} m_{2p}(v)^{1+\delta/2}} \leq -\frac{\delta}{2}\zeta + \log(1+\varepsilon) + (2+\delta)\log(1+\delta/2).$$

We have to choose $\delta > 0$ such that

$$\frac{2}{\delta}(2+\delta)\log\left(1+\frac{\delta}{2}\right) < \zeta.$$

This is possible for $\zeta > 2$.

For $\zeta = 2$ we defer the proof after Proposition 4.12. □

In the range $\zeta \in (0,2)$ we need (4.62), the finer assumption on the sequence p_n. For the error term in h_v in (4.12)

$$\begin{aligned} h_v(x) - h_0(x) &= \log\Big(\ell(v)^{1/\gamma}\widetilde{\ell}(v^{-\gamma}\ell(v)e^x)\Big) \\ &= \log \ell(v)^{1/\gamma}\widetilde{\ell}(v^{-\gamma}\ell(v)) + \log\frac{\widetilde{\ell}(v^{-\gamma}\ell(v)e^x)}{\widetilde{\ell}(v^{-\gamma}\ell(v))}. \end{aligned} \tag{4.65}$$

By the inverse relation (4.11) the first term is small for v small, while the second term can be bounded using the Potter bounds, thus for any $\varepsilon > 0$ there exist $x_0 > 0$ such that for $x > x_0$

$$\left|\log\Big(\ell(v)^{1/\gamma}\widetilde{\ell}(v^{-\gamma}\ell(v)e^x)\Big)\right| \leq \varepsilon x,$$

implying that for $x > x_0$

$$|h_v(x) - x/\gamma| \leq \varepsilon x.$$

Also, for η_v in (4.13) there exist $x_0 > 0$ such that for $x > x_0$

$$|\eta_v(x) - \gamma\zeta x| \leq \varepsilon x. \tag{4.66}$$

Using these bounds, we can prove the uniform version of Lemma 5.4 in [**11**].

Lemma 4.10. *For any $K > 0$*

$$\lim_{p\to\infty} \sup_{v\in[0,1], x\in[K^{-1},K]} \left|h_v(\eta_v(p)) - h_v(\eta_v(p)x^{1/p}) + \zeta\log x\right| = 0.$$

Proof. We have by (4.12)

$$h_v(\eta_v(p)) - h_v(\eta_v(p)x^{1/p}) = \frac{\eta_v(p)}{\gamma}(1 - x^{1/p}) + \log\frac{\widetilde{\ell}(v^{-\gamma}\ell(v)e^{\eta_v(p)})}{\widetilde{\ell}(v^{-\gamma}\ell(v)e^{\eta_v(p)x^{1/p}})}.$$

Using (4.66) and that $1 - x^{1/p} \sim -p^{-1}\log x + O(p^{-2})$, we see that the first term tends to $-\zeta\log x$. This further implies, using also the uniform convergence theorem that the second term above tends to 0, proving the statement.

We also note that the argument above shows that the uniform convergence theorem for the regularly varying h_v holds uniformly in $v \in [0,1]$. □

Once we have the uniform convergence in v, and the uniform moment bound (4.59), the proofs of Lemma 6.1, 6.2, and 6.3 in [**11**] go through. We omit the proof.

LEMMA 4.11. *For any $r > 0$ and $\tau > 0$ uniformly in $v \in [0,1)$*

$$\begin{aligned}&\lim_{p\to\infty} \frac{e^{\zeta p}}{\eta_v(p)^{pr}} \mathbb{E}\left[Y(v)^{rp}\left(I(Y(v) \le \eta_v(p)\tau^{1/p}) - I(Y(v) \le \eta_v(p))\right)\right] \\ &= \begin{cases} \frac{\zeta}{r-\zeta}(\tau^{r-\zeta} - 1), & r \neq \zeta, \\ \zeta \log \tau, & r = \zeta. \end{cases}\end{aligned} \tag{4.67}$$

For any $\tau > 0$ and $r > \zeta$

$$\lim_{p\to\infty} \frac{e^{\zeta p}}{\eta_v(p)^{pr}} \mathbb{E}\left[Y(v)^{rp} I(Y(v) \le \eta_v(p)\tau^{1/p})\right] = \frac{\zeta}{r-\zeta}\tau^{r-\zeta}, \tag{4.68}$$

while for $\tau > 0$ and $r < \zeta$

$$\lim_{p\to\infty} \frac{e^{\zeta p}}{\eta_v(p)^{pr}} \mathbb{E}\left[Y(v)^{rp} I(Y(v) > \eta_v(p)\tau^{1/p})\right] = \frac{\zeta}{\zeta - r}\tau^{r-\zeta}. \tag{4.69}$$

For $r = \zeta$

$$\lim_{p\to\infty} \frac{e^{\zeta p}}{\eta_v(p)^{\zeta p}} \mathbb{E}\left[Y(v)^{\zeta p} I(Y(v) \le \eta_v(p))\right] = \infty. \tag{4.70}$$

Recall the notation (4.16). Again, if $\zeta \le 2$ then ζ equals the characteristic exponent of the limiting stable law. Therefore, we use the notation $\zeta = \alpha$.

PROPOSITION 4.12. *Assume that* (4.62) *holds with $\zeta = \alpha \in (0,2)$. Then as $n \to \infty$, uniformly in $v \in [0,1)$*

$$\frac{1}{\eta_v(p_n)^{p_n}} \left[Z_n(p_n, v) - n\widetilde{m}_{p_n}(v)\right] \xrightarrow{\mathcal{D}} Z_\alpha.$$

PROOF. We use the classical criteria for convergence of sums of independent random variables, see Theorem 25.1 in Gnedenko and Kolmogorov [**42**].

First, by (4.12), (4.13), and Lemma 4.10, uniformly in $v \in [0,1)$

$$n\mathbb{P}(Y(v)^p > \eta_v(p)^p x) = ne^{-h_v(\eta_v(p))} e^{h_v(\eta_v(p)) - h_v(\eta_v(p)x^{1/p})} \to x^{-\alpha}. \tag{4.71}$$

Next, applying Lemma 4.11 with $r = 2$, uniformly in $v \in [0,1)$

$$\lim_{\tau\downarrow 0} \limsup_{n\to\infty} \frac{n}{\eta_v(p)^{2p}} \mathbb{E}\left[Y(v)^{2p} I(Y(v) \le \eta_v(p)\tau^{1/p})\right] = 0.$$

Therefore, we already have that the normed sum converges with an appropriate centering, and the limit is a one-sided α-stable law. To see that the centering is correct note that

$$\lim_{p\to\infty}\left(\frac{n}{\eta_v(p)^p}\mathbb{E}\left[Y(v)^p I\left(Y(v)\le \eta_v(p)\tau^{1/p}\right)\right]-\frac{n\widetilde{m}_p(v)}{\eta_v(p)^p}\right)$$
$$=\begin{cases}\frac{\alpha}{1-\alpha}\tau^{1-\alpha}, & \alpha\neq 1,\\ \log\tau, & \alpha=1.\end{cases}$$

Indeed, this follows from (4.68) for $\alpha<1$, from (4.69) for $\alpha>1$, and from (4.67) for $\alpha=1$. □

We end this section with the proof of the central limit theorem in the borderline case $\alpha=2$.

Proof of Proposition 4.9 for $\alpha=2$. Here we use again the classical criteria [**42**, Theorem 25.1], but specified to the Gaussian law.

Using (4.70) with $\alpha=r=2$ we obtain that uniformly in $v\in[0,1)$

$$\sigma_p^1(v)e^p/\eta_v(p)^p\to\infty. \tag{4.72}$$

Thus, for any $x>0$ fixed and $\tau>0$ large, for n large enough

$$n\mathbb{P}(Y(v)^p>\sqrt{n}\sigma_p^1(v)x)\le n\mathbb{P}(Y(v)^p>\eta_v(p)^p\tau),$$

which by (4.71) converges to τ^{-2}. Thus, for any $x>0$

$$n\mathbb{P}(Y(v)^p>\sqrt{n}\sigma_p^1(v)x)\to 0. \tag{4.73}$$

For the truncated variance

$$\begin{aligned}&\frac{n}{(\sigma_p^1(v)\sqrt{n})^2}\mathbb{E}\left[Y(v)^{2p}I\left(Y(v)\le(\sigma_p^1(v)\sqrt{n}\tau)^{1/p}\right)\right]\\&=1+\frac{\mathbb{E}\left[Y(v)^{2p}I\left(\eta_v(p)\le Y(v)\le(\sigma_p^1(v)\sqrt{n}\tau)^{1/p}\right)\right]}{(\sigma_p^1(v))^2}.\end{aligned} \tag{4.74}$$

For the second term for $\delta\in(0,1)$ by (4.69) with $\alpha=2$, $r=2-\delta$

$$\begin{aligned}&\frac{\mathbb{E}\left[Y(v)^{2p}I\left(\eta_v(p)\le Y(v)\le(\sigma_p^1(v)\sqrt{n}\tau)^{1/p}\right)\right]}{(\sigma_p^1(v))^2}\\&\le\frac{(\sigma_p^1(v)\sqrt{n}\tau)^\delta}{(\sigma_p^1(v))^2}\mathbb{E}\left[Y(v)^{(2-\delta)p}I(\eta_v(p)\le Y(v))\right]\\&\sim\frac{2}{\delta}\left(\frac{\sigma_p^1(v)e^p}{\eta_v(p)^p}\right)^{-(2-\delta)}\frac{n^{\delta/2}}{e^{p\delta}},\end{aligned}$$

which tends to 0 by (4.72). Furthermore, by (4.72)

$$\frac{\left[\mathbb{E}\left(Y(v)^pI(Y(v)\le(\sigma_p^1(v)\sqrt{n}\tau)^{1/p})\right)\right]^2}{\mathbb{E}\left(Y(v)^{2p}I(Y(v)\le(\sigma_p^1(v)\sqrt{n}\tau)^{1/p})\right)}\le\frac{(m_p(v))^2}{(\sigma_p^1(v))^2}.$$

Using Lemma 4.7 and (4.66) it is simple to show that the latter quantity tends to 0 uniformly in $v \in [0,1)$. Thus, from (4.74) we obtain for any $\tau > 0$

$$\begin{aligned}\lim_{n\to\infty} \frac{n}{\left(\sqrt{n}\sigma_p^1(v)\right)^2}\Big\{&\mathbb{E}\left[Y(v)^{2p} I\left(Y(v) \le (\sigma_p^1(v)\sqrt{n}\tau)^{1/p}\right)\right] \\ &- \left(\mathbb{E}\left[Y(v)^p I\left(Y(v) \le (\sigma_p^1(v)\sqrt{n}\tau)^{1/p}\right)\right]\right)^2\Big\} = 1.\end{aligned} \tag{4.75}$$

Finally, by (4.69) with $\alpha = 2$, $r = 1$

$$\begin{aligned}&\frac{n}{\sigma_p^1(v)\sqrt{n}}\mathbb{E}\left(Y(v)^p I(Y(v) > (\sigma_p^1(v)\sqrt{n}\tau)^{1/p})\right) \\ &\le \frac{\sqrt{n}}{\sigma_p^1(v)}\mathbb{E}\left(Y(v)^p I(Y(v) > \eta_p(v)\tau^{1/p})\right) \sim \frac{\sqrt{n}}{\sigma_p^1(v)}\frac{\eta_v(p)^p}{e^{2p}}\tau^{-1}\end{aligned}$$

which tends to 0 by (4.72). Together with (4.73) and (4.75) this implies the statement. □

4.4. Proofs for asymptotics for large p.

Proof of Theorem 3.1. The limiting relations (4.17) and (4.18) are immediate consequences of Propositions 4.8 and 4.9. Indeed, for the law of large numbers by representation (4.5)

$$\begin{aligned}&\mathbb{P}\left(|S_n(p_n)/m_{p_n}(U_{k+1,n}) - 1| > \varepsilon\right) \\ &= \int_0^1 \mathbb{P}\left(|Z_n(p_n, v)/(n m_{p_n}(v)) - 1| > \varepsilon\right) d\widetilde{\mathbb{P}}(U_{k+1,n} \le v),\end{aligned}$$

which tends to 0, since the integrand tends to 0 uniformly. The proof of (4.18) is similar.

The weak consistency follows from Lemma 4.7 and (4.17).

The CLT (4.19) follows from Lemma 9.1 in [**11**] and Theorem 3.1. To apply Lemma 9.1 in [**11**] we only need to show that

$$\frac{\sqrt{k_n}\, m_{p_n}(U_{k+1,n})}{\sigma_{p_n}(U_{k+1,n})} \to \infty.$$

This follows easily from Lemma 4.7 as for $\varepsilon > 0$ small enough

$$\liminf_{n\to\infty} p_n^{-1} \log \frac{\sqrt{k_n}\, m_{p_n}(U_{k+1,n})}{\sigma_{p_n}(U_{k+1,n})} \ge \frac{\zeta}{2} - \log 2 - \log(1+\varepsilon) > 0.$$

□

Proof of Theorem 3.2. First note that $U_{k+1,n} n/k \to 1$ in probability, and since a and ℓ are regularly varying functions $U_{k+1,n}$ can be changed to k/n.

For the first result we have to show that $m_p(k/n)/m_p \to 1$. This follows from Proposition 4.6. Indeed, for any $\varepsilon > 0$

$$\left| \frac{m_p(k/n)}{m_p} - 1 \right| \le K(1+\varepsilon)^p \left(\frac{a(k/n)}{\ell(k/n)} \right)^{\nu_\beta - \varepsilon} .$$

Taking logarithm and dividing by p we see that the right-hand side above is negative for $\varepsilon > 0$ small enough.

For the central limit theorem, $\sigma_p(k/n)/\sigma_p \to 1$ follows again from Proposition 4.6, thus $\sigma_p(U_{k_n+1,n})/\sigma_p \to 1$ also follows as above. To change the centering, using again Proposition 4.6

$$\begin{aligned} \frac{\sqrt{k}}{\sigma_{p_n}} |m_p(k/n) - m_p| &= \frac{m_p \sqrt{k}}{\sigma_p} \frac{|m_p(k/n) - m_p|}{m_p} \\ &\le c\sqrt{k}(1+\varepsilon)^p \frac{\Gamma(p+1)}{\sqrt{\Gamma(2p+1)}} \left(\frac{a(k/n)}{\ell(k/n)} \right)^{\widetilde{\nu}}, \end{aligned} \tag{4.76}$$

with $\widetilde{\nu} = 1 \wedge (\nu_\beta - \varepsilon)$. Taking logarithm, dividing by p, and using the Stirling formula

$$\begin{aligned} &\limsup_{p\to\infty} p^{-1} \log \left[\sqrt{k}(1+\varepsilon)^p \, \frac{\Gamma(p+1)}{\sqrt{\Gamma(2p+1)}} \left(\frac{a(k/n)}{\ell(k/n)} \right)^{\widetilde{\nu}} \right] \\ &\le \log(1+\varepsilon) - \log 2 + \limsup_{p\to\infty} p^{-1} \log \left[\sqrt{k} \left(\frac{a(k/n)}{\ell(k/n)} \right)^{\widetilde{\nu}} \right]. \end{aligned}$$

Since $\varepsilon > 0$ in (4.76) is as small as we wish, the result follows.

Now (4.24) follows from (4.19) using Bogachev's transfer lemma, as above. □

Proof of Theorem 3.3. The proof goes as the previous proof, but we use Proposition 4.5. □

Proof of Theorem 3.4. The first result follows from Proposition 4.12. Combining with Bogachev's transfer lemma we obtain (4.26) and (4.28). To use the transfer lemma for $\alpha \in [1, 2)$ we have to check that

$$\frac{k_n \widetilde{m}_{p_n}(U_{k+1,n})}{\eta_{U_{k+1,n}}(p_n)^{p_n}} \to \infty.$$

For $\alpha > 1$ by Lemma 4.7 and (4.66) the left-hand side above is at least

$$\frac{e^{\alpha p_n}(\gamma - \varepsilon)^{p_n} \Gamma(p_n + 1)}{((\alpha\gamma - \varepsilon)p_n)^{p_n}} \ge \left(\frac{e^\alpha}{\alpha} \right)^{p_n} (1-\varepsilon)^{p_n},$$

which tends to ∞ for $\varepsilon > 0$ small enough. For $\alpha = 1$ the result follows from (4.70) in Proposition 4.11 with $r = \alpha = 1$.

To see (4.27), note that as $p_n \to \infty$,

$$\left(\frac{S_n(p_n)}{\Gamma(p_n+1)}\right)^{1/p_n} (k_n\Gamma(p_n+1))^{1/p_n} \frac{1}{\eta_{U_{k+1,n}}(p_n)} = \frac{(k_n S_n(p_n))^{1/p_n}}{\eta_{U_{k+1,n}}(p_n)} \to 1.$$

Thus (4.27) follows from the asymptotics

$$\frac{(k_n\Gamma(p_n+1))^{1/p_n}}{\eta_{U_{k+1,n}}(p_n)} \to \frac{e^{\alpha-1}}{\alpha\gamma}.$$

□

Proof of Theorem 3.5. First we show that we can change to deterministic normalization, i.e.

$$\lim_{n\to\infty}\left(\frac{\eta_{U_{k+1,n}}(p_n)}{\alpha\gamma p_n}\right)^{p_n} = 1. \tag{4.77}$$

We have $|\eta_v(x) - \alpha\gamma x| = \gamma|h_v(\eta_v(x)) - h_0(\eta_v(x))|$ by the definition of η_v in (4.44). Therefore,

$$\left|\frac{\eta_v(p_n)}{\alpha\gamma p_n} - 1\right| \le \frac{1}{\alpha p_n}\left|h_v(\eta_v(p_n)) - h_0(\eta_v(p_n))\right|,$$

from which we see that (4.77) follows if we show the convergence

$$\lim_{n\to\infty}\left[h_{U_{k+1,n}}(\eta_{U_{k+1,n}}(p_n)) - h_0(\eta_{U_{k+1,n}}(p_n))\right] = 0.$$

This holds, since the first term on the right-hand side of (4.65) tends to 0 as $v = U_{k+1,n} \to 0$ by (4.11). Changing $U_{k+1,n}$ to k/n, with $v = k/n$ and $x = \eta_v(p_n) \sim \gamma\log k$, we see that the second term tends to 0 by assumption (4.29). Thus (4.31) holds for $\alpha < 1$.

For $\alpha \in [1,2)$ we need to handle the centering as well. For $\alpha > 1$ by Proposition 4.6

$$\frac{k_n|m_{p_n} - m_{p_n}(v)|}{(\alpha\gamma p_n)^{p_n}} \le \frac{2K_1}{\gamma}(1+\varepsilon)^p \frac{e^{\alpha p_n}}{(\alpha p_n)^{p_n}}\Gamma(p_n+1)\left(\frac{a(v)}{\ell(v)}\right)^{\widetilde{\nu}},$$

with $\widetilde{\nu} = 1 \wedge (\nu_\beta - \varepsilon)$. As before we can substitute $U_{k+1,n}$ to k/n. Thus

$$\begin{aligned}
&\limsup_{n\to\infty} p_n^{-1}\log\frac{k_n|m_{p_n} - m_{p_n}(k/n)|}{(\alpha\gamma p_n)^{p_n}} \\
&\le \log(1+\varepsilon) + \alpha - 1 - \log\alpha + \widetilde{\nu}\limsup_{n\to\infty} p_n^{-1}\log\left(\frac{a(k/n)}{\ell(k/n)}\right) \\
&= \log(1+\varepsilon) + H(\alpha) - \widetilde{\nu}\beta_1,
\end{aligned}$$

which is negative for $\varepsilon > 0$ small, under our assumptions. Thus (4.31) follows.

To prove (4.32), write

$$p_n\left(\frac{[k_n S_n(p_n)]^{1/p_n}}{\alpha\gamma p_n} - 1\right) = p_n\left(\widehat{\gamma}(n)\frac{(k_n\Gamma(p_n+1))^{1/p_n}}{\alpha\gamma p_n} - 1\right).$$

Simple calculation shows that

$$\frac{k_n^{1/p_n}}{\alpha\gamma p_n}\Gamma(p_n+1)^{1/p_n} = \frac{e^{\alpha-1}}{\alpha\gamma}\left(1+\frac{\log(2\pi p_n)}{2p_n}+o(1/p_n)\right).$$

Thus (4.32) and (4.33) follows from Bogachev's transfer lemma. For (4.34) we use the asymptotics of the truncated moments in Lemma 4.4. □

Proof of Theorem 3.6. The proof goes as the previous proof, but we use Proposition 4.5. □

5. Simulation study

The purpose of this small simulation study is to show that understanding the behavior of $\widehat{\gamma}(n)$ for large values of p is not only a mathematical challenge. The use of larger p values sometimes is beneficial in practical situations, which was already pointed out by Gomes and Martins [**44**]. However, we do not intend to provide neither a theoretical nor a practical comparison of the various tail index estimators. For a comprehensive simulation study, as well as for a practical criteria for the choice of k and p, we refer to [**44**].

Note that for $p = 1$ we obtain the usual Hill estimator. In Theorem 5.1 Segers [**77**] proved the optimality of the Hill estimator among residual estimators. We also see from (4.24) that the asymptotic variance increases with p. However, in practical situation higher p values turns out to be useful.

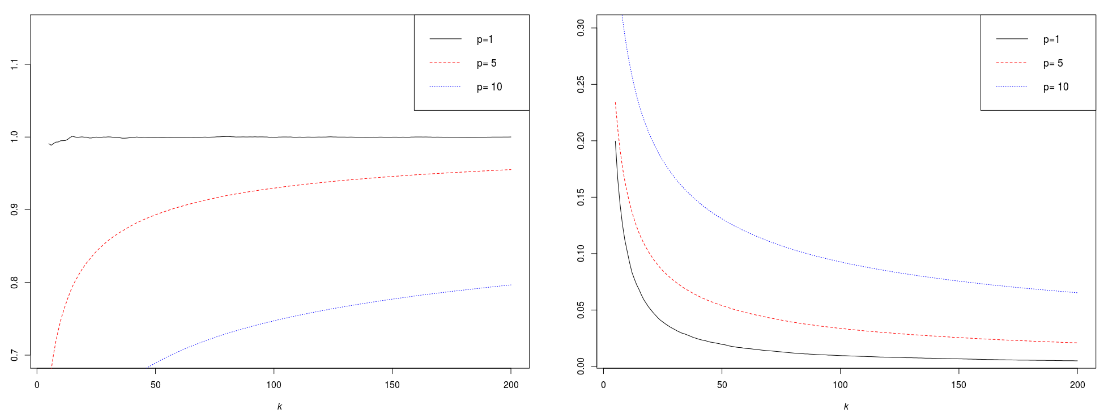

Figure 1. Mean (left) and MSE (right) in the strict Pareto model with $\gamma = 1$.

In the simulations below $n = 1000$ and we repeated the simulations 5000 times. In all the figures the mean and mean squared error (MSE) are calculated for different values of k_n. We plotted the estimators as a function of k in the range $[5, 200]$.

For $k \geq 200$ the estimators do not change much, and we have to estimate from a negligible portion, that is $k_n/n \to 0$.

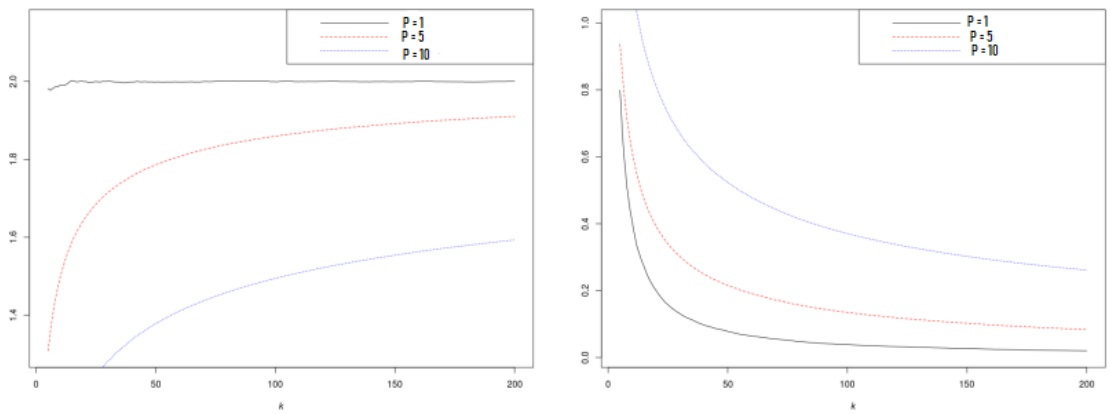

FIGURE 2. Mean (left) and MSE (right) in the strict Pareto model with $\gamma = 2$.

In Figures 1 and 2 we see that the Hill estimator is the best in the strict Pareto model. In this case $Q(1-s) = s^{-\gamma}$. For $p = 10$ we also see that the estimator is not consistent, as $\zeta = (\log k)/10 \ll 1$. In fact we see the graph $\gamma \cdot \zeta e^{1-\zeta} = \gamma k^{-1/10} \log k\, e/10$. Note that $e^5 \approx 150$, so loosely speaking the estimator for $p = 5$ is weakly consistent only for $k \geq 150$, while $e^{10} \approx 22,000$, so asymptotic normality starts to hold for $k \geq 22,000$. Therefore, for $k \leq 200$ smaller p values should be used. We chose larger values to illustrate better the difference. We also note that for large data sets we may use larger p values.

However, in practice it is very unusual to encounter data which fit to a nice distribution everywhere. It is more common that the large values fit to a Pareto-type

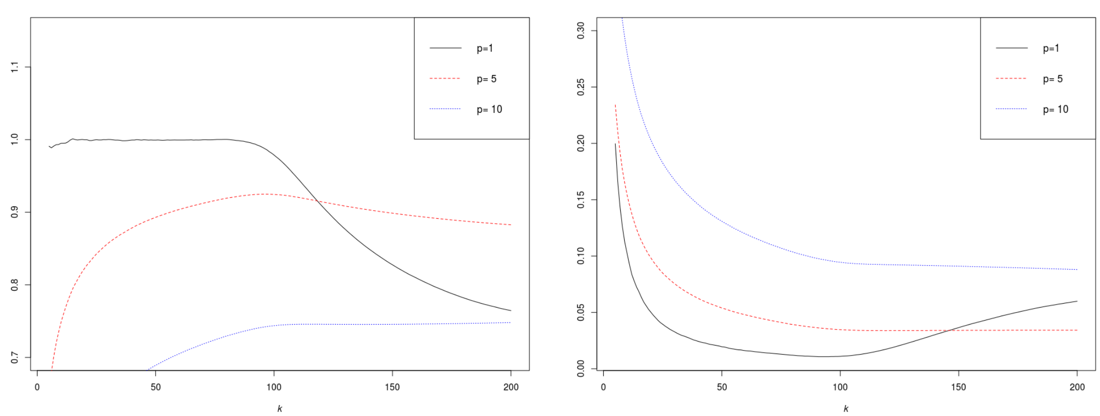

FIGURE 3. Mean and MSE for a sample with quantile function (4.78) with $\gamma = 1$.

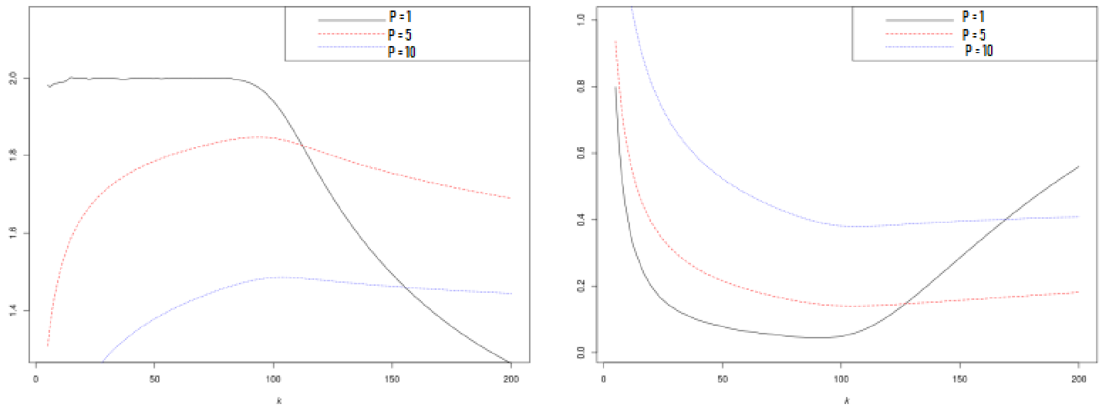

FIGURE 4. Mean and MSE for a sample with quantile function (4.78) with $\gamma = 2$.

distribution, while the smaller values behave as a light-tailed distribution. Consider the quantile function

$$Q(1-s) = \begin{cases} s^{-\gamma}, & \text{if } s \leq 0.1, \\ \frac{10^{\gamma}}{\log 10} \log s^{-1}, & \text{if } s \geq 0.1, \end{cases} \tag{4.78}$$

which is a mixture of an exponential quantile and a strict Pareto quantile. The parameter of the exponential is chosen such that Q is continuous. Figures 3 and 4 contain the simulation results for $\gamma = 1$ and $\gamma = 2$. In this simple model we already see the advantage of larger p values. Note that the Hill estimator is very sensitive to the change of k_n for those values where the quantile function changes. Indeed, for $k_n \leq 100$ we basically have a sample from a strict Pareto distribution, and for those values the Hill estimator is the best. For $k_n = 200$ we already see the exponential part of the sample, and the Hill estimator changes drastically (for $\gamma = 1$ from 0.98 to 0.76), while for $p = 5$ the change is not as large (from 0.92 to 0.88).

Next, we further add a nonconstant slowly varying function to the quantile. A logarithmic factor in the tail of the random variable cannot be detected in practice, but it makes significantly more difficult to determine the underlying index of regular variation. We modify the construction in (4.78) and consider the quantile function

$$Q(1-s) = \begin{cases} s^{-\gamma}(\log s^{-1})^3, & \text{if } s \leq 0.1, \\ 10^{\gamma}(\log 10)^2 \log s^{-1}, & \text{if } s \geq 0.1. \end{cases} \tag{4.79}$$

Note again that the function is continuous. We see from the simulation results in Figures 5 and 6 that in this setup the estimators with larger p values work much better than the Hill estimator. These estimators are not so sensitive for the change

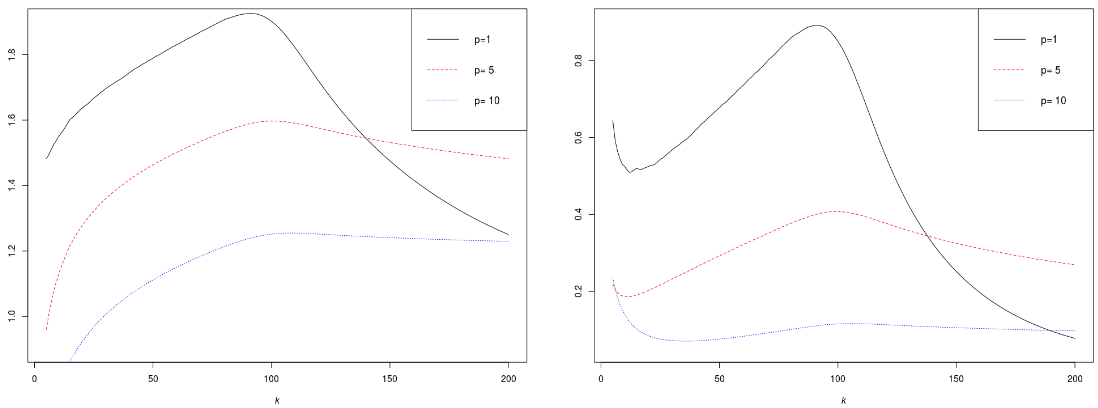

FIGURE 5. Mean and MSE for a sample with quantile function (4.79) with $\gamma = 1$.

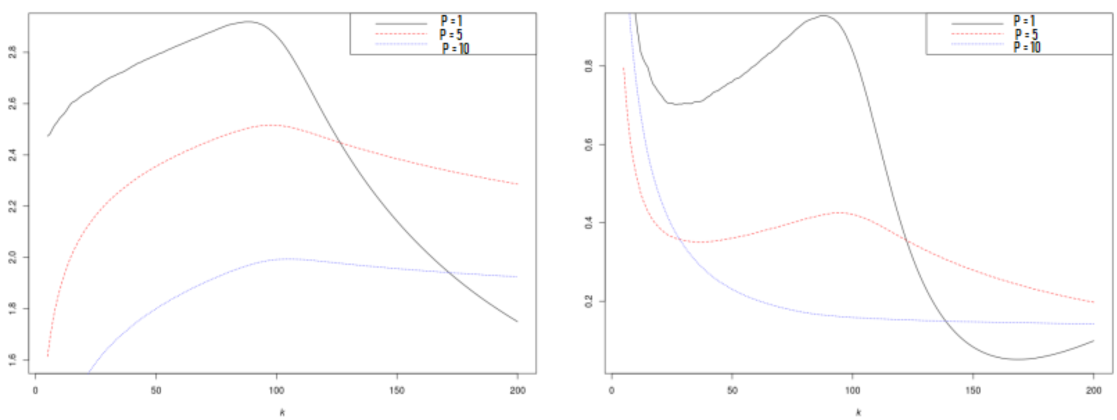

FIGURE 6. Mean and MSE for a sample with quantile function (4.79) with $\gamma = 2$.

in the nature of the quantile function. We also see that heavier tails are in favor of larger p values.

It was pointed out in [**44**] that in various models under second-order regular variation for a wide range of p values (usually $p \in (1, 5]$) the estimator $\widehat{\gamma}(n)$ with p fixed is more efficient than the Hill estimator. The variance of the estimator has a unique minimum at $p = 1$ (the Hill estimator), but the bias decreases in p, which is the decisive factor in some models, see Figures 3 and 4 in [**44**].

We also apply the estimator with different p values to real data. We chose the data set of Danish fire insurance losses, which consists of 2167 fire losses in millions of Danish Kroner. The data set is included in the R package evir, and was analyzed in [**71**] and in [**36**, Example 6.2.9]. In Figure 7 we plotted the estimate for $1/\gamma$,

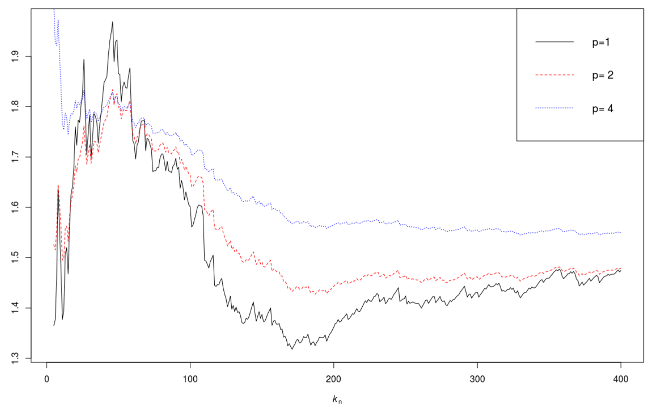

FIGURE 7. Hill type plots of $\hat{\gamma}(n)^{-1}$ for the Danish fire insurance claim with different p values.

i.e. we plotted $1/\hat{\gamma}(n)$ against k_n, to obtain the Hill plot in **[71]** for $p = 1$. Resnick **[71]** used various techniques to obtain smoother plot. In our setting larger p values naturally produce smoother plots.

CHAPTER 5

A statistical approach to partition lattices with some theoretical "by-products"

Outline and source. This chapter entails investigating four-element generating sets of a partition lattice and establishing a lower bound for the number of four-element generating sets of direct products of two neighbouring partition lattices. The chapter is divided into four sections, which are taken from [**26**] and [**63**]. Section 1 is introductory and combines the corresponding parts of [**26**] and [**63**]. Sections 2 and 3 are reprints from [**26**]. Section 4 is taken from [**63**] but more details are given here. Namely, Remark 4.2 and, to prove it, Cases 4.3–4.8 have been added to the original publication. In connection with this, Lemma 4.1 has become a separate statement. Also, there is a notational change in Section 4: $\nu(n)$ from [**63**] has been changed to $\nu_2(n)$ in order to avoid confusion with Sections 2 and 3, which inherit the meaning of $\nu(n)$ from [**26**].

1. Introduction to partition lattices and their generating sets

H. Strietz proved in 1975 that the minimum size of a generating set of the partition lattice $\mathrm{Part}(n)$ on the n-element set ($n \geq 4$) equals 4; see [**81, 82**]. This classical result forms the foundation for this chapter. In [**94**], Strietz's results have been echoed by L. Zádori (1983), who gave a new elegant proof confirming the outcome. Based on his approach, several studies have indeed emerged henceforth concerning
four-element generating sets of partition lattices. In particular, the Strietz–Zádori result was extended to infinite partition lattices by Czédli [**22, 23, 24**].

The papers and results mentioned above were devoted to the *existence* of four-element generating sets. The study of the *number* of small generating sets of partition lattices have started recently with Czédli [**25**] and Czédli and Oluoch [**26**]. In particular, [**26**] gives a lower bound for the number $\nu(n)$ of four-element generating sets of $\mathrm{Part}(n)$ as well as a statistical approach to $\nu(n)$ for small values of n. Also, in [**26**], we have recently proved that certain direct products of partition lattices are also 4-generated. In particular, some direct powers of $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$

is four-generated for $n \geq 7$; this fact will turn out to be important in the present chapter.

When dealing with $\nu(n)$ in [**26**], it caused some difficulty that $\mathrm{Part}(n)$ has very many elements and a complicated structure; for example, each finite lattice is embeddable in $\mathrm{Part}(n)$ for some n by a classical result of Pudlák and Tůma [**70**]. This explains that, instead of determining the exact or the asymptotic value of $\nu(n)$, we could only give a lower bound for $\nu(n)$. Note that this lower bound is better than what would trivially follow from Zádori [**94**] with $n!$ automorphisms taken into account, but we knew that this lower bound was far from being sharp. The reason is that we proved the validity of the lower bound by presenting four-element generating sets obtained by a *special* construction that goes back to the technique of Zádori [**94**], but there is no hope to construct *all* four-element generating sets. Since the lower bound for $\nu(n)$ in [**26**] is not sharp, we have developed some computer programs for investigating $\nu(n)$ for some small values of n. The results obtained in this way are analyzed using a computer-assisted statistical approach to $\nu(n)$ for small values of n. This analyzis constitute Section 3 here.

Once the above-mentioned computer program was available, we could use it to solve an old problem of Zádori [**94**]; this solution is presented here in Section 2. (To harmonize with [**26**], this section precedes Section 3 in the dissertation.)

In Section 4, based on [**63**], we give a lower bound on the number $\nu_2(n)$ of 4-element generating sets of the direct product $\mathrm{Part}(n)\times\mathrm{Part}(n+1)$ for $n \geq 7$ using the main theorem of [**26**]. Again, like in case of $\mathrm{Part}(n)$ and $\nu(n)$, the complicated structure of $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$ prevents us from determining the exact value of $\nu_2(n)$. Hence, we only present a lower bound for $\nu_2(n)$ for large n. Note that for $n \in \{1, 2, 3, 4, 5\}$, some computer programs have been developed and used for the investigation of $\nu_2(n)$ in [**63**] but these investigations are not included in the dissertation.

2. A solution of Zádori's problem on $(1+1+2)$-generation of small partition lattices

We know from Zádori [**94**] that, for $n \geq 7$, the lattice $\mathrm{Part}(n)$ of all partitions of the n-element set 1,...,n has a so-called $(1+1+2)$-*generating set*, that is, a four-element generating set of which two elements (and only two elements) are comparable. The question whether $\mathrm{Part}(5)$ and $\mathrm{Part}(6)$ have $(1+1+2)$-generating sets was left open in Zádori [**94**]. The purpose of this section is to prove the following two statements, which solve Zádori's problem.

PROPOSITION 2.1. *The partition lattice* $\mathrm{Part}(6)$ *has a* $(1+1+2)$-*generating set.*

PROPOSITION 2.2. *Every four-element generating set of* Part(5) *is an antichain. Hence,* Part(5) *has no* $(1+1+2)$*-generating set.*

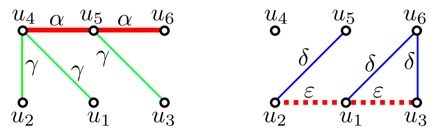

FIGURE 1. With $\beta := \alpha + \varepsilon$, the set $\{\alpha, \beta, \gamma, \delta\}$ is a $(1+1+2)$-generating set of Equ(6).

As usual, associated with a partition U of A, we define an *equivalence relation* π_U of A as the collection of all pairs $(x, y) \in A^2$ such that x and y belong to the same block of U. As it is well known, the equivalence relations and the partitions of A mutually determine each other, and $\pi_U \leq \pi_V$ if and only if $U \leq V$. (Here $\pi_U \leq \pi_V$ means that $\pi_U \subseteq \pi_V$ as sets of pairs of elements of A.) Hence, the *lattice* Equ(A) *of all equivalence relations* of A (in short, the *equivalence lattice* of A) is isomorphic to Part(A). In what follows, we do not make a sharp distinction between a partition and the corresponding equivalence relation; no matter which of them is given, we can use the other one without warning. Typically, we speak of *partitions lattices* in the main statements but we prefer to speak of *equivalence lattices* in the proofs.

CONVENTION 2.3. We are going to define our equivalence relations and the corresponding partitions by (undirected edge-coloured) graphs; multiple edges are allowed. On it's vertex set $\{u_1, ..., u_6\}$, the graph on the left of Figure 1 defines $\alpha \in$ Equ(A) in the following way: deleting all edges but the α-colored ones, the components of the remaining graph are the blocks of the partition associated with α. In other words, $\langle x, y \rangle \in \alpha$ if and only if there is an α-coloured path from vertex x to vertex y in the graph, that is, a path (of possibly zero length) all of whose edges are α-colored. The equivalences γ, δ, and ε are defined analogously.

NOTATION 2.4. Following Czédli [**25**], we adopt the following notation. Assume that A is a base set and we are interested in its partitions or, equivalently, in its equivalence relations. For elements $u_1, \dots, u_k$ of A, the partition of A with block $\{u_1, \dots, u_k\}$ such that all the other blocks are singletons will be denoted by

$$[\![u_1, \dots u_k]\!]^{\mathrm{e}}.$$

Usually but not always, the elements $u_1, \dots, u_k$ are assumed to be pairwise distinct. Note that $[\![u_1, u_1]\!]^{\mathrm{e}}$ is Δ, the least equivalence relation of A, that is, the zero element of Equ(A). For $\kappa, \lambda \in$ Equ(A), the *meet* and the *join* of κ and λ, denoted by $\kappa\lambda$ (or

$\kappa \cdot \lambda$) and $\kappa + \lambda$, are the intersection and the transitive hull of the union of κ and λ, respectively. The usual precedence rules apply; for example, $xy + xz$ stands for $(x \wedge y) \vee (x \wedge z)$. *Lattice terms* are composed from variables and join and meet operation signs in the usual way; for example, $f(x_1, x_2, x_3, x_4) = x_1(x_3 + x_4) + (x_1 + x_3)x_4$ is a quaternary lattice term. Given a lattice L and $u_1, \dots, u_k \in L$, the *sublattice generated* by $\{u_1, \dots, u_k\}$ is denoted and defined by

$$[u_1, \dots, u_k]_{\mathrm{lat}} := \{f(u_1, \dots, u_k) : u_1, \dots, u_k \in L,\ f \text{ is a lattice term}\}. \tag{5.1}$$

Our arguments will often use the following technical lemma from Zádori [**94**], which has been used also in Czédli [**22, 23, 24**] and in some other papers like Kulin [**58**]. Note that the proof of this lemma is straightforward.

LEMMA 2.5 ("Circle Principle"). *Let $d_0, d_1, \dots, d_{t-1}$ be pairwise distinct elements of a set A. Then, for any $0 \le i < j \le t-1$ and in the lattice* $\mathrm{Equ}(A)$,

$$\begin{aligned} [\![d_i, d_j]\!]^{\mathrm{e}} = \big([\![d_i, d_{i+1}]\!]^{\mathrm{e}} + [\![d_{i+1}, d_{i+2}]\!]^{\mathrm{e}} \cdots + [\![d_{j-1}, d_j]\!]^{\mathrm{e}}\big) \cdot \big([\![d_j, d_{j+1}]\!]^{\mathrm{e}} \\ + \cdots + [\![d_{t-2}, d_{t-1}]\!]^{\mathrm{e}} + [\![d_{t-1}, d_0]\!]^{\mathrm{e}} + [\![d_0, d_1]\!]^{\mathrm{e}} + \cdots + [\![d_{i-1}, d_i]\!]^{\mathrm{e}}\big). \end{aligned} \tag{5.2}$$

Consequently, $[\![d_i, d_j]\!]^{\mathrm{e}} \in [\, [\![d_0, d_1]\!]^{\mathrm{e}}, [\![d_1, d_2]\!]^{\mathrm{e}}, \dots, [\![d_{t-2}, d_{t-1}]\!]^{\mathrm{e}}, [\![d_{t-1}, d_0]\!]^{\mathrm{e}}\,]_{\mathrm{lat}}$.

For later reference, note the following. If all the joinands (formally, the summands) in (5.2) are substitution values of appropriate quaternary terms, then so is $[\![d_i, d_j]\!]^{\mathrm{e}}$ of a longer quaternary term, which is defined according to (5.2) and

$$\text{which we denote by } \widehat{e}_{d_i, d_j}. \tag{5.3}$$

EXAMPLE 2.6. In $A := \{a_0, a_1, a_2, a_3, a_4, a_5\}$, we obtain $[\![a_1, a_3]\!]^{\mathrm{e}}$ after computing the joins $[\![a_1, a_2]\!]^{\mathrm{e}} + [\![a_2, a_3]\!]^{\mathrm{e}}$ and $[\![a_3, a_4]\!]^{\mathrm{e}} + [\![a_4, a_5]\!]^{\mathrm{e}} + [\![a_5, a_0]\!]^{\mathrm{e}} + [\![a_0, a_1]\!]^{\mathrm{e}}$ and finally, taking their meet.

Armed with our conventions and notations, the proofs in this section runs as follows.

PROOF OF PROPOSITION 2.1. Let $A := \{u_1, u_2, \dots, u_6\}$. Figure 1, according to Convention 2.3, indicates that we consider the following equivalences of A

$$\begin{aligned} \alpha &= [\![u_4, u_5, u_6]\!]^{\mathrm{e}}, & \gamma &= [\![u_1, u_2, u_4]\!]^{\mathrm{e}} + [\![u_3, u_5]\!]^{\mathrm{e}} \\ \varepsilon &= [\![u_1, u_2, u_3]\!]^{\mathrm{e}}, & \delta &= [\![u_1, u_3, u_6]\!]^{\mathrm{e}} + [\![u_2, u_5]\!]^{\mathrm{e}}, \end{aligned} \tag{5.4}$$

and let $\beta := \alpha + \varepsilon$. Since $\alpha < \beta$, the set $X := \{\alpha, \beta, \gamma, \delta\}$ is of order type $1 + 1 + 2$. Let S be the sublattice generated by X; we are going to show that $X = \mathrm{Equ}(A)$.

Observe that

$$\begin{aligned}
&[\![u_2,u_1]\!]^{\mathrm{e}} = \beta\gamma \in S, && [\![u_1,u_3]\!]^{\mathrm{e}} = \beta\delta \in S,\\
&[\![u_5,u_4]\!]^{\mathrm{e}} = \alpha(\gamma + [\![u_1,u_3]\!]^{\mathrm{e}}) \in S, && [\![u_6,u_5]\!]^{\mathrm{e}} = \alpha(\delta + [\![u_2,u_1]\!]^{\mathrm{e}}) \in S,\\
&[\![u_4,u_2]\!]^{\mathrm{e}} = \gamma(\delta + [\![u_5,u_4]\!]^{\mathrm{e}}) \in S, \text{ and} && [\![u_3,u_6]\!]^{\mathrm{e}} = \delta(\gamma + [\![u_6,u_5]\!]^{\mathrm{e}}) \in S.
\end{aligned}$$

Hence, Lemma 2.5 applies to the circle $\langle u_1, u_3, u_6, u_5, u_4, u_2\rangle$, and we obtain that all atoms of $\mathrm{Part}(A)$ are in S. But $\mathrm{Part}(A)$ is an *atomistic lattice*, that is, each of its elements is the join of some atoms; this completes the proof of Proposition 2.1. □

PROOF OF PROPOSITION 2.2. Unfortunately, we have no elegant proof. However, we have computer programs on the websites [1] that list all four-element generating sets of Part(5); there are exactly 5305 such sets. And we have another program that checks if these 5305 sets are antichains. The application of this program completes the proof. □

3. Computer-assisted results and statistical analysis of the number of four-element generating sets of a partition lattice

3.1. Estimating confidence intervals. In this section, we use some well-known facts of statistics; see, for example, Hodges and Lehmann [**51**, page 255], Lefebvre [**61**, Chapter 6.2], and mainly Mendenhall, Beaver and Beaver [**62**].

Following Zádori [**94**] and several papers developing his construction further, the letter n denotes the size of the base set of a partition lattice. Hence, we denote the size of a statistical sample by $N \in \mathbb{N}^+$ even if n would fit into the traditions of statistics better. Also, the confidence level is usually denoted by $1-\alpha$, but many earlier papers use α to denote one of the generators of $\mathrm{Part}(n)$. Hence, the confidence level will be denoted by $1-\alpha_{\mathrm{conf}}$.

Note that an event of a binomial model corresponds to an indicator variable, which is a random variable with two possible values, 0 and 1; this allows us to simplify what follows below. Note also that for a large N, $\sqrt{N/(N-1)}$ is very close to 1; for example, it is 1.000020000 (up to nine digits) for $N = 25000$. Hence, replacing $N-1$ by N in (3.1) as some sources of information do, the error would be neglectable (and smaller than what rounding can cause).

Assume that an experiment has only two possible outcomes: "success" with probability p and "failure" with probability $q := 1-p$ but none of p and q is known. In order to obtain some information on p, a random *sample* is taken, that is, the

[1] http://www.math.u-szeged.hu/~czedli/ and http://www.math.u-szeged.hu/~oluoch/

experiment is repeated N times independently. Let s denote the number of those experiments that ended up with "success". Then, of course,

$$\text{we estimate } p \text{ by } \widehat{p} := s/N, \tag{5.5}$$

but we would also like to know how much we can rely on this estimation. Therefore, let $\widehat{q} := 1 - \widehat{p}$, pick a "confidence level" $1 - \alpha_{\text{conf}} \in (0,1) \subset \mathbb{R}$, we let

$$\widehat{\sigma} := \sqrt{\frac{\widehat{p} \cdot \widehat{q}}{N-1}}, \tag{5.6}$$

and determine the positive real number $z(\alpha_{\text{conf}})$ from the equation

$$1 - \alpha_{\text{conf}} = \int_{-z(\alpha_{\text{conf}})}^{z(\alpha_{\text{conf}})} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx. \tag{5.7}$$

Note that the function to be integrated in (5.7) is the density function of the *standard normal distribution* and the $z(\alpha_{\text{conf}})$ for many typical values of α_{conf} are given in practically all books on statistics. In this paper, to maintain five-digit accuracy, we used the values given in Table 1. Finally, we define the

$$\textit{confidence interval } I(\alpha_{\text{conf}}) \text{ to be } [\,\widehat{p} - z(\alpha_{\text{conf}})\widehat{\sigma},\, \widehat{p} + z(\alpha_{\text{conf}})\widehat{\sigma}\,]. \tag{5.8}$$

Let us emphasize that while p is a concrete real number, the confidence interval is *random*, because it depends on a randomly chosen sample. Taking another N-element sample with the same N (that is, repeating the experiments N times again), (with very high probability in general) a different confidence interval is obtained.

We cannot claim that the confidence interval $I(\alpha_{\text{conf}})$ surely contains the unknown probability p. Furthermore, as it has been pointed out by an anonymous referee, it may even happen that $I(\alpha_{\text{conf}})$ contains p only with *very little* probability. For example, if $p = 10^{-100}$ and $N = 2$, then a random N-element sample yields that $\widehat{p} = 0$ and $p \notin I(\alpha_{\text{conf}}) = [0, 0]$ with probability $q^2 = 1 - 2 \cdot 10^{-100} + 10^{-200} \approx 1$. However, the Moivre-Laplace theorem, which is a particular case of the central limit theorem, implies that whenever $p \notin \{0, 1\}$, then

$$\text{the probability of } p \in I(\alpha_{\text{conf}}) \text{ tends to } 1 - \alpha_{\text{conf}} \text{ as } N \to \infty. \tag{5.9}$$

$\alpha_{\rm conf}$	0.100	0.050	0.010	0.001
$1-\alpha_{\rm conf}$	0.900	0.950	0.990	0.999
$z(\alpha_{\rm conf})$	1.64485	1.95996	2.57583	3.29053

TABLE 1. $z(\alpha_{\rm conf})$ for some confidence levels $1-\alpha_{\rm conf}$; taken from `https://mathworld.wolfram.com/ConfidenceInterval.html`

Next, assume that G is a given subset of a large finite set F such that $\emptyset \neq G \neq F$. (For example, and this is what is going to happen soon, F can be the set of all four-element subsets of $\mathrm{Part}(n)$, for $n \in \{4,5,\dots,9\}$, and G can be $\{H \in F : [H]_{\rm lat} = \mathrm{Part}(n)\}$.) We know the size $|F|$ of F but that of G is usually unknown for us. We would like to obtain some information on $|G|$; this task is equivalent to getting information on the proportion $p := |G|/|F|$. With respect to uniform distribution, p is the probability that a randomly selected element of F belongs G. Take an N-element sample, that is, select N members of F independently, and declare "success" if a randomly chosen member belongs to G. With a fixed confidence level $1-\alpha_{\rm conf}$, compute the confidence interval $I(\alpha_{\rm conf})$ according to (5.7) and (5.8). Then we conclude from (5.8) and (5.9) that

$$\left.\begin{array}{l}\text{with approximate probability } 1-\alpha_{\rm conf}\text{, the } N\text{-element}\\ \text{sample has been chosen so that } (\widehat{p}-z(\alpha_{\rm conf})\widehat{\sigma})\cdot|F| \leq\\ |G| \leq (\widehat{p}+z(\alpha_{\rm conf})\widehat{\sigma})\cdot|F|.\end{array}\right\} \tag{5.10}$$

3.2. Computer programs. Two disjoint sets of computer programs were developed and all data to be reported in (this) Section 3 were achieved by these programs. Furthermore, a sufficient amount of these data, including $\nu(4) = 50$ and $\nu(5) = 5\,305$ from Table 3, were achieved independently by different persons (namely, by both authors of **[26]**, with different programs, different attitudes to computer programming, and different computers. This fact gives us a lot of confidence in our programs and the results obtained by them even if some results that needed too much performance from our computers and programs were achieved only by one of the above-mentioned two settings.

The first setting includes some programs written in Bloodshed *Dev-Pascal* v1.9.2 (Freepascal) under Windows 10 and also in *Maple* V. Release 5 (1997); these programs are available from professor G. Czédli's website[2]. The strategy is to represent a partition by a lexicographically ordered list of its blocks separated by

[2]http://www.math.u-szeged.hu/~czedli/

zeros. For example, for $n = 8$, the partition $\{\{4, 6\}, \{1, 5, 3, 7\}, \{2, 8\}\}$ is represented by the vector

$$\langle 1, 3, 5, 7, 0, 2, 8, 0, 4, 6, 0, -1, -1, -1, -1, -1, -1\rangle \tag{5.11}$$

where -1 means that there is no more block. This representation is unique, which allows us to order any set of partitions lexicographically. The benefit of this lexicographic ordering is that it takes only $O(t)$ step to decide if a given partition belongs to a t-element set of partitions; this task occurs many times when computing sublattices generated by four partitions. The collection of all partitions of $\{1, 2, \dots, n\}$ was computed recursively by a Maple program. This collection was saved into a txt file. After inputting this file, the hard job of generating sublattices was done by Pascal programs, which took care of efficiency in some ways. Since all the partitions were input by these Pascal programs into an array, a random partition was selected by selecting its index as a random number of the given range. Note that Maple was also used to do some computations, trivial for Maple, to obtain some numbers occurring in this section; see, for example, the numbers in Table 2.

The second set consist of programs written in *R* (64 bit, version 3.6.3) and *Python* 3.8 which works well with any operating system. Implementing the urn model of Stam [**80**] in R, random members of $\mathrm{Part}(n)$ were selected in a sophisticated way as follows; note that this method does not require that $\mathrm{Part}(n)$ be stored in the computer. The method consists of two steps. First, choose a positive integer u according to the probability distribution

$$P(u = j) = \frac{j^n}{e \cdot j! \cdot \mathrm{Bell}(n)}, \qquad j \in \mathbb{N}^+, \tag{5.12}$$

where e is the well-known constant $\lim_{j\to\infty}(1 + 1/j)^j$. It is pointed out in Stam [**80**] that $\sum_{j=1}^{\infty} P(u = j) = 1$, so (5.12) is indeed a probability distribution. Fortunately, the series in (5.12) converges very fast and $\sum_{j=1}^{2n} P(u = j)$ is very close to 1. Hence, though the program chose u from $\{1, 2, \dots, 2n\}$ rather than from $\mathbb{N}^+$, the error is neglectable. In the next step, put the numbers 1, 2, $\dots$, n into u urns at random, according to the uniform distribution on the set of urns and independently from each other. Finally, the contents of the nonempty urns constitute a random partition that we were looking for.

By the main result of Stam [**80**], the algorithm just described yields a uniform distribution on the set $\mathrm{Part}(n)$ of all partitions (apart from the above-mentioned neglectable error caused by using $\sum_{j=1}^{2n}$ instead $\sum_{j=1}^{\infty}$).

The implementation of partitions involved importing *sympy* (external library with functions for computing *Partitions*) into Python. Additionally, *itertool* which is a built-in function in Python was necessary for computing various combinations for the four partitions of which meets and joins were evaluated. The combination of these two functions was crucial for the computation of the sublattices generated by the four partitions.

The lion's share of the computation was done on a desktop computer with AMD Ryzen 7 2700X Eight-Core Processor 3.70 GHz. With the speed 3.70 GHz, the total amount of "pure computation time" was a bit more than two and a half weeks; see Tables 3 and 4. By "pure computation time" we mean that a single copy of the program was running without being disturbed by other programs and without letting the computer go to an idle state. The whole computation took more than a month because of several breaks when the computer was idle or it was turned off or it was used for other purposes.

3.3. Data obtained by computer programs. The results obtained by computers are given in Tables 3 and 4. In particular, Table 3 gives the number $\nu(n)$ of four-element generating sets for $n \in \{4, 5, 6\}$. Clearly, $\nu(7)$ cannot be determined by our programs and computers, although this task might be possible with thousands or millions of similar computers working jointly for a few years or so. (But this is just a first impression not supported by real analysis.)

Table 4 shows what we have obtained from random samples. For a given n, let

$$p = p(n) := \nu(n) \cdot \binom{\mathrm{Bell}(n)}{4}^{-1}; \tag{5.13}$$

this is the exact theoretical probability that a random four-element subset of $\mathrm{Part}(n)$ generates $\mathrm{Part}(n)$. In accordance with (5.5), $\widehat{p} = \widehat{p}(n)$ is s/N; the table contains $100\widehat{p}$ up to five digits. The least and the largest endpoints of the confidence interval $I(\alpha_{\mathrm{conf}})$ are denoted by $(1-\alpha_{\mathrm{conf}})_*$ and $(1-\alpha_{\mathrm{conf}})^*$, respectively. For example, with this notation, $[0.999_*,\ 0.999^*]$ is the confidence interval $I(1-0.999)$, and this interval contains p with approximate probability 0.999.

In order to enlighten the meaning of Table 4 even more, consider, say, the entries of its last two rows in the column for $n = 7$. Taking (5.10) and Table 2 also into account, we obtain that

$$p(7) \in [0.0157753,\ 0.0159877] \text{ with approximate probability 0.999 and} \tag{5.14}$$

$$\nu(7) \in [3.86180 \cdot 10^8,\ 3.91381 \cdot 10^8] \text{ with approximate probability 0.999.} \tag{5.15}$$

However, let us note the following. From rigorous mathematical point of view, not even $\nu(7) \geq 2$ is proved by Table 4. Indeed, it is theoretically possible (although very unlikely) that all the 238 223 generating sets the program found after fifteen million experiments are the same. The right interpretation of (5.14) and (5.15) is that if very many lattice theorists and statisticians pick random samples of the same size (that is, they pick fifteen million four-elements subsets of $\mathrm{Part}(n)$ and count the generating sets among them), then these samples give many different intervals but approximately 99.9 percent of these colleagues find intervals that happen to contain $p(7)$ and $\nu(7)$, respectively.

Based on experience with generating sets of $\mathrm{Equ}(n)$ and Table 4, we risk formulating the following conjecture.

CONJECTURE 3.1. The set $\{p(n) : 4 \le n \in \mathbb{N}^+\}$ has a positive lower bound.

It would be too early to formulate the rest of our feelings as a conjecture, so we formulate them in an open problem as follows.

PROBLEM 3.2. Is it true that

$$p(6) > p(7) \le p(8) \le p(9) \le p(10) \le p(11) \le \dots$$

and, for all $4 \le n \in \mathbb{N}^+$, $p(n) \ge 3/200$? Note that we already know from Table 3 that $p(4) > p(5) > p(6) > 3/200$.

Since $100 \cdot p(6) = 1.613014768$ is exactly known and it is sufficiently "far away" from the confidence interval $I(0.001) = [1.57753, 1.59877]$ for the unknown $100 \cdot p(7)$, our confidence in $p(6) > p(7)$ is even more than approximately 0.999. However, we are not really confident in, say, $p(8) \le p(9)$ even if this inequality is more likely to hold than to fail.

n	7	8	9
$\binom{\mathrm{Bell}(n)}{4}$	24 480 029 875	12 222 513 708 615	8 330 299 023 110 190

TABLE 2. The number of four-element subsets of $\mathrm{Part}(n)$ for $n \in \{7, 8, 9\}$

n	4	5	6
$\binom{\mathrm{Bell}(n)}{4}$	1 365	270 725	68 685 050
$\nu(n)$	50	5 305	1 107 900
%, i.e., $100p(n)$	3.663003663	1.959553052	1.613014768
computer time	0.11sec	68 sec	38 hours

TABLE 3. The (exact) number $\nu(n)$ of the four-element generating sets of $\mathrm{Equ}(n)$ for $n \in \{4, 5, 6\}$

n	4	5	6	7	8	9
$100p(n)$	3.6630037	1.9595531	1.613014768			
N	10 000 000	10 000 000	10 000 000	15 000 000	500 000	25 000
time	8 minutes	27 min	3h+33min	102 hours	95 h	166 h
s	367 221	196 243	161 768	238 223	8 244	438
$100\widehat{p}(n)$	3.67221	1.96243	1.61768	1.58815	1.64880	1.75200
0.900_*	3.66243	1.95522	1.61112	1.58284	1.61918	1.61551
0.900^*	3.68199	1.96964	1.62424	1.59346	1.67842	1.88849
0.950_*	3.66055	1.95383	1.60986	1.58183	1.61350	1.58936
0.950^*	3.68387	1.97103	1.62550	1.59448	1.68410	1.91464
0.990_*	3.65689	1.95113	1.60740	1.57984	1.60241	1.53826
0.990^*	3.68753	1.97373	1.62796	1.59647	1.69519	1.96574
0.999_*	3.65264	1.94800	1.60455	1.57753	1.58954	1.47896
0.999^*	3.69178	1.97686	1.63081	1.59877	1.70806	2.02504

TABLE 4. Statistics with N experiments that yielded s many 4-element generating sets of $\mathrm{Part}(n)$ for $n \in \{4, \dots, 9\}$

4. A lower bound for the number of 4-element generating sets of direct products of two neighbouring partition lattices

Recall that $\nu_2(n)$ is the number of the 4-element generating sets of $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$. This section is taken from **[63]** but, like in **[63]**, we follow closely the approach presented in **[26]**, where Theorem 4.4 states that certain direct products of direct powers of partitions lattices are still 4-generated. In particular, for any integer $5 \leq n$, the direct product $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$ is four-generated, i.e., 1 is a lower bound for $\nu_2(n)$; of course, a much larger lower bound is presented here for large values of n. It is worth noting that Czédli **[25]** has shown that the study of small generating sets of partitions lattices and their direct products have connection with information theory.

For elements $x_1, \ldots, x_4$ of a lattice L, we say that $\langle x_1, \ldots, x_4 \rangle$ is a *generating quadruple* and $\{x_1, \ldots, x_4\}$ is a *generating set* of L if the smallest sublattice of L containing each of $x_1, \ldots, x_4$ is L itself or, equivalently, $[[x_1, x_2, x_3, x_4]]_{lat}$ defined in (5.1) is L.

LEMMA 4.1. *With* $t^* = \binom{n-6}{(n-5)/2}$, *if* $n \geq 7$ *and* n *is odd, then the lattice* $\mathrm{Part}(n)^{t^*} \times \mathrm{Part}(n+1)^{t^*}$ *is 4-generated.*

The exponent t^* given above is not the best (=largest) possible value; simply because the exponent supplied by Theorem 4.4 of **[26]** is not the best either; the reason has been mentioned in the Introduction part of this chapter. However,we also prove the following.

REMARK 4.2. The exponent t^* in Lemma 4.1 is the best exponent that one can extract from (4.7)–(4.9) and Theorem 4.4 of **[26]**.

PROOF OF LEMMA 4.1. The exponent for the direct product of two neighbouring partition lattices provided by **[26**, Theorem 4.4] depends on the following parameters: d, i, the parity on n, and the boolean parameter "φ or τ", to be defined below. The *upper* and *lower integer parts* of a real number x will be denoted by $\lceil x \rceil$ and $\lfloor x \rfloor$, respectively; for example, $\lceil \sqrt{2}\, \rceil = 2$ and $\lfloor \sqrt{2} \rfloor = 1$. For a convenient way to reference **[26]**, (4.7) , (4.8), and (4.9) of **[26]** motivate us to define

$$\varphi(u,k) := \begin{cases} \binom{k-1}{u-1}, & \text{if} \quad u,k \in \mathbb{N}^+ \quad \text{and} \quad u-1 \leq \lceil (k-1)/2 \rceil, \\ \binom{k-1}{\lceil (k-1)/2 \rceil}, & \text{if} \quad u,k \in \mathbb{N}^+ \quad \text{and} \quad u-1 > \lceil (k-1)/2 \rceil \end{cases} \tag{5.16}$$

and

$$\tau(u,k) := \left\{ \left\langle \binom{k-1}{j-1}, \binom{k-1}{j} \right\rangle : j \in \{1, 2, \ldots, \min(u,k) - 1\} \right\}. \tag{5.17}$$

The boolean parameter "φ or τ" can either return the "choose φ and use (5.16) above and (4.7)–(4.8) from **[26]**", or the "choose τ and use (5.17) above and (4.9) from **[26]**" value. It can be observed that there are different cases depending on "φ or τ" and whether n is odd or even. In this way, we can establish different cases given some specific conditions. Note in advance, for all the forthcoming cases, that (with the notation from (4.19) of **[26]**), we are going to have $n_i' = n$ and $n_i'' = n+1$ for n odd while $n_i'' = n$ and $n_{i+1}' = n+1$ for n even.

CASE 4.3. Here we assume that n is odd, $i = 1$ (see Theorem 4.4 of [**26**] for the meaning of i), and we go after the upper line of (4.15) of [**26**], that is, we use φ from (5.16) rather than τ from (5.17). Note that the role of our φ is to combine (4.7) and (4.8) of [**26**]. So we use φ as a lower estimate of sba(-,-) occurring in (4.15) of [**26**]. Define

$$w = w_i := \binom{(m_i+3)/2}{2} = \binom{(n-1)/2}{2}, \tag{5.18}$$

where $m_i = m_1 = n-4$ by (4.19) of [**26**]. Hence $\varphi(d, m_1 - 1) = \varphi(d, n-5) = \binom{n-6}{d-1}$, for small d or $\binom{n-6}{\lceil (n-6)/2 \rceil}$, for big values of d. Looking for the largest number in the $(n-6)th$ row of Pascal's triangle, d is selected such that $d - 1 = \lceil (n-6)/2 \rceil$. Temporarily, let $s := \binom{n-6}{\lceil (n-6)/2 \rceil}$. Then $q := s - p$ where p is a parameter. If p is increased then it implies that $w \cdot q$ decreases, thus the minimum $\min(p, w \cdot q)$ takes its maximum when $p = w \cdot q$. Hence after substituting $s - q$ for p, the equality $s - q = w \cdot q$ turns into $q = s - w \cdot q$. Thus

$$q = s - w \cdot q \implies (w+1)q = s \implies q = \frac{s}{w+1} \implies w \cdot q = \frac{w}{w+1} \cdot s.$$

It should be noted that $\frac{s}{w+1}$ need not be an integer, however, either $q = \lfloor \frac{s}{w+1} \rfloor$ (the lower integer part) or $q = \lceil \frac{s}{w+1} \rceil$ (the upper integer part) is the best. This results in two possibilities:

$$t^* = w \cdot \left\lfloor \frac{s}{w+1} \right\rfloor = \binom{(n-1)/2}{2} \cdot \left\lfloor \binom{n-6}{\lceil \frac{n-6}{2} \rceil} \cdot \left(\binom{(n-1)/2}{2} + 1 \right)^{-1} \right\rfloor \tag{5.19}$$

and

$$t^* = p = s - q = \binom{n-6}{\lceil \frac{n-6}{2} \rceil} - \left\lceil \binom{n-6}{\lceil \frac{n-6}{2} \rceil} \cdot \left(\binom{(n-1)/2}{2} + 1 \right)^{-1} \right\rceil. \tag{5.20}$$

CASE 4.4. Consider n is odd and $i = 2$, we follow (5.16). Then $m_i = m_2 = n-4$, and $w = w_i := \binom{(m_i+3)/2}{2}$. Hence, using the upper line of (4.16) of [**26**],

$$p + q = \varphi(m_2 - d, m_2 - d - 1) = \varphi(m - 4 - d, n - 5 - d) = \binom{n-6-d}{(n-6-d)/2} =: s$$

so $p + q =: s$ is on the vertical axis of symmetry of Pascal's triangle and becomes larger when d is reduced. But how small can d be? The stipulation in [**26**] is that d is odd and $2 = i \le d+2$. This allows us to let $d = 1$. If $2 = i \le d+1$ and we let $d = 1$, then $s = \binom{n-6-1}{(n-6-1)/2}$ but this would be a smaller s compared to the previous and would yield a smaller value of t^* than those in (5.19) and (5.20). Hence, this subcase is dropped since the target is the largest t^*.

CASE 4.5. Now we assume that n is odd and $2 < i \le d+1$ and we follow (5.16), then we still have that $m_i = n-4$ and $w = w_i := \binom{(n-1)/2}{2}$. By (4.17 of [**26**]),

$$p+q = \varphi(d+3-i, m_i-d-1) = \varphi(d+3-i, n-5-d) \le \binom{n-6-d}{(n-6-d)/2} =: s. \quad (5.21)$$

Since $d \ge 1$, $n-6-d < n-6$, the s in (5.21) yields a smaller t^* than (5.19) and (5.20), hence this subcase can also be disregarded.

CASE 4.6. In this Particular case, we consider when n is odd, $i = 1$, we use τ from (5.17), $w = w_i = \binom{(n-1)/2}{2}$ and $m = m_i = n-4$. Then by (4.15) of [**26**],

$$\begin{aligned} \langle p,q\rangle &\in \tau(d, m-1) = \tau(d, n-5) \\ &= \left\{ \left\langle \binom{n-6}{j-1}, \binom{n-6}{j} \right\rangle : j \in \{1, 2, \ldots, \min(n-5, d) - 1\} \right\}. \end{aligned} \quad (5.22)$$

If d increases, this set gets larger hence results in more pairs $\langle p, q\rangle$, of which among them we get a pair that gives a larger $t^* = \min(p, w \cdot q)$. So we let d be as large as possible conditioned on $d \le m_1 = m = n-4$, so $d = n-4$, then we have that $\min(n-5, d) - 1 = n-6$. Then, according to (5.22), (p, q) denotes a pair of two consecutive numbers of the $(n-6)th$ row of Pascal's triangle. We aim at maximizing $\min(p, w \cdot q)$. Since p and q are the two neighbouring elements in the middle of Pascal's triangle, then $p = q = \binom{n-6}{\lceil \frac{n-6}{2} \rceil} = \min(p, q)$ which results in a

$$t^* = \binom{n-6}{\lceil \frac{n-6}{2} \rceil}, \quad (5.23)$$

which is larger that (5.19) and (5.20). (This is the largest t^* found so far.)

CASE 4.7. In this case we still let $w = \binom{(n-1)/2}{2}$ and $m = m_2 = n-4$, when n is odd and we follow τ in (5.17) while $i = 2$, then using (4.16)of [**26**] we get that

$$\begin{aligned} \langle p, q\rangle &\in \tau(m-d, m-d-1) = \tau(n-4-d, n-5-d) \\ &= \left\{ \left\langle \binom{n-6-d}{j-1}, \binom{n-6-d}{j} \right\rangle : j \in \{1, 2, \ldots, n-6-d\} \right\}. \end{aligned}$$

Since $d \ge 1$, the upper number of each of the binomial coefficients above is smaller thus result in a t^* smaller than t^* in (5.23), hence this case is disregarded.

CASE 4.8. Meanwhile, we also consider $2 < i \le d+2$, and n is odd, $w = \binom{(n-1)/2}{2}$ and $m = m_i = n-4$, we follow τ in (5.17) and (4.18) of [**26**]. Then

$$\begin{aligned} \langle p, q\rangle &\in \tau(d+3-i, m-d-1) = \tau(d+3-i, n-5-d) \\ &= \left\{ \binom{n-6-d}{j-1} \binom{n-6-d}{j} : j \in \{1, 2, \ldots, \min(d-4-i, n-6-d)\} \right\}. \end{aligned}$$

This case also results in a t^* smaller than t^* in (5.23), hence it is also disregarded.

These results indicate that the largest t^* when n is odd, under various conditions is the t^* in (5.23). This completes the proof of Remark 4.2. $\square$

REMARK 4.9. Assume that n is even, $n \geq 7$ and choose the parameters in Theorem 4.4 of **[26]** as follows. For brevity in inline formulas, $\mathrm{binc}(x, y)$ will denote the binomial coefficient with upper and lower parameters x and y, respectively. Let $d = 1$, $i = 1$, $n = n_1''$, $m_1 = n - 5$, $m_2 = n - 3$, $n_2' = n + 1 = m_2 + 4$. Then $w_1 = (m_1 + 3)(m_1 + 1)/8 = (n - 2)(n - 4)/8$ by (4.14) of **[26]**, $q_1 = \mathrm{sba}(1, m_1 - 1) \geq \mathrm{binc}(m_1 - 2, 0) = 1$ by (4.7) and (4.15) of **[26]**, and $p_2 \geq \mathrm{sba}(m_2 - d, m_2 - d - 1) = \mathrm{sba}(n - 4, n - 5) \geq \mathrm{binc}(n - 6, (n - 6)/2)$ by (4.8) and (4.16) of **[26]**. Since (4.20) and Remark 4.3 of **[26]** allow $\min(w_1 q_1, p_2)$ and the computation above shows that $w_1 q_1 \geq (n - 2)(n - 4)/8$ and $p_2 \geq \mathrm{binc}(n - 6, n/2 - 3)$, it follows that with $t = t_n$ defined in (5.26), (5.27) also holds for n even. That is, in other words,

$$\underbrace{\mathrm{Part}(n) \times \cdots \times \mathrm{Part}(n)}_{t \text{ times}} \times \underbrace{\mathrm{Part}(n+1) \times \cdots \times \mathrm{Part}(n+1)}_{t \text{ times}} \tag{5.24}$$

is 4-generated for all $n \geq 7$. Although Theorem 4.4 of **[26]** would allow a larger $t_n = t$ for a *large* even n, we have no explicit formula for this larger t_n and the t_n given in (5.25) and in the lower line of (5.26) is the only possibility that Theorem 4.4 of **[26]** yields for $n = 8$. Thus we use the following exponent when n is even:

$$t^* = \min\left((n-2)(n-4)/8, \binom{n-6}{n/2-3} \right). \tag{5.25}$$

Note that the magnitude of t^* for n even is considerably less than that of t^* for n odd, because Theorem 4.4 of **[26]** performs worse for n even than for n odd; the reason is hidden in the construction that proves Theorem 4.4 in **[26]**.

Now we are in the position to state the main result of this section.

THEOREM 4.10. *Let $n \geq 7$ be an integer number and define*

$$t_n := \begin{cases} \dbinom{n-6}{(n-5)/2}, & \text{if } n \text{ is odd, and} \\[2ex] \min\left((n-2)(n-4)/8, \dbinom{n-6}{n/2-3} \right), & \text{if } n \text{ is even.} \end{cases} \tag{5.26}$$

Then $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$ *has at least* $t_n^2 \cdot n! \cdot (n+1)!/2$ *many* 4*-element generating sets.*

Proof. To ease our forthcoming notation, we will write t instead of t_n. By Lemma 4.1 and Remark 4.9,

$$\mathrm{Part}(n)^t \times \mathrm{Part}(n+1)^t \text{ is 4-generated.} \tag{5.27}$$

First, we are dealing with generating quadruples of a special kind. Let us fix a quadruple $\langle \vec\alpha, \vec\beta, \vec\gamma, \vec\delta \rangle$ such that $\{\vec\alpha, \vec\beta, \vec\gamma, \vec\delta\}$ generates the direct product (5.24). With more details, this quadruple consists of

$$\begin{aligned}
\vec\alpha &= \langle \alpha'_1, \quad \alpha'_2, \quad ..., \quad \alpha'_t, \quad \alpha''_{t+1}, \quad \alpha''_{t+2}, \quad ..., \quad \alpha''_{2t}\rangle,\\
\vec\beta &= \langle \beta'_1, \quad \beta'_2, \quad ..., \quad \beta'_t, \quad \beta''_{t+1}, \quad \beta''_{t+2}, \quad ..., \quad \beta''_{2t}\rangle,\\
\vec\gamma &= \langle \gamma'_1, \quad \gamma'_2, \quad ..., \quad \gamma'_t, \quad \gamma''_{t+1}, \quad \gamma''_{t+2}, \quad ..., \quad \gamma''_{2t}\rangle,\\
\vec\delta &= \langle \delta'_1, \quad \delta'_2, \quad ..., \quad \delta'_t, \quad \delta''_{t+1}, \quad \delta''_{t+1}, \quad ..., \quad \delta''_{2t}\rangle.
\end{aligned} \tag{5.28}$$

We also need the "columns" of (5.28), which we write in row vectors as follows:

$$\vec g^{(1)} = \langle \alpha'_1, \beta'_1, \gamma'_1, \delta'_1\rangle, \qquad \dots, \qquad \vec g^{(t)} = \langle \alpha'_t, \beta'_t, \gamma'_t, \delta'_t\rangle, \tag{5.29}$$

$$\vec h^{(t+1)} = \langle \alpha''_{t+1}, \beta''_{t+1}, \gamma''_{t+1}, \delta''_{t+1}\rangle, \quad \dots, \quad \vec h^{(2t)} = \langle \alpha''_{2t}, \beta''_{2t}, \gamma''_{2t}, \delta''_{2t}\rangle. \tag{5.30}$$

It would not be too hard to observe that the quadruples in (5.29) are pairwise different and the same holds for (5.30), but actually we are going to prove even more. But first, we need to fix some notation. The set of all permutations of $\{1, 2, \dots, n\}$ will be denoted by S_n; the meaning of S_{n+1} is analogous. Each $\pi \in S_n$ induces an automorphism $\widehat\pi$ of $\mathrm{Part}(n)$ in the natural way. That is, for $\varepsilon \in \mathrm{Part}(n)$, a pair $\langle i, j\rangle$ is collapsed by ε if and only if $\langle \pi(i), \pi(j)\rangle$ is collapsed by $\widehat\pi(\varepsilon)$. Let $\widehat\pi^*$ denote the componentwise action of $\widehat\pi$ on quadruples. In particular, $\widehat\pi^*(g^{(i)})$ is $\langle \widehat\pi(\alpha'_i), \widehat\pi(\beta'_i), \widehat\pi(\gamma'_i), \widehat\pi(\delta'_i)\rangle$ by the definition of $\widehat\pi^*$. Note that $\widehat\pi^*$ is an automorphism of the direct power $\mathrm{Part}(n)^4$. We claim that

$$\left.\begin{array}{l}\text{for any } i, i' \in \{1, \dots, t\} \text{ and } \pi_1, \pi_2 \in S_n, \text{ if}\\ \langle i, \pi_1\rangle \neq \langle i', \pi_2\rangle, \text{ then } \widehat\pi_1^*(g^{(i)}) \neq \widehat\pi_2^*(g^{(i')}),\\ \text{and}\end{array}\right\} \tag{5.31}$$

$$\left.\begin{array}{l}\text{for any } j, j' \in \{t+1, \dots, 2t\} \text{ and } \sigma_1, \sigma_2 \in\\ S_{n+1}, \text{ if } \langle j, \sigma_1\rangle \neq \langle j', \sigma_2\rangle, \text{ then } \widehat\sigma_1^*(h^{(j)}) \neq\\ \widehat\sigma_2^*(h^{(j')}).\end{array}\right\} \tag{5.32}$$

It suffices to deal with (5.31), because the argument for (5.32) is similar. Suppose that (5.31) fails and pick $i, i' \in \{1, \dots, t\}$ and $\pi_1, \pi_2 \in S_n$ such that

$$\langle i, \pi_1\rangle \neq \langle i', \pi_2\rangle \tag{5.33}$$

but $\widehat\pi_1^*(g^{(i)}) = \widehat\pi_2^*(g^{(i')})$. This equality means that

$$\langle \widehat\pi_1(\alpha'_i), \widehat\pi_1(\beta'_i), \widehat\pi_1(\gamma'_i), \widehat\pi_1(\delta'_i)\rangle = \langle \widehat\pi_2(\alpha'_{i'}), \widehat\pi_2(\beta'_{i'}), \widehat\pi_2(\gamma'_{i'}), \widehat\pi_2(\delta'_{i'})\rangle. \tag{5.34}$$

We let $\pi := \pi_2^{-1} \circ \pi_1$; note that we compose permutations from right to left, that is, $(\pi_2^{-1} \circ \pi_1)(x) = \pi_2^{-1}(\pi_1(x))$. Note also that $\widehat{\pi} = \widehat{\pi}_2^{-1} \circ \widehat{\pi}_1$. Hence, (5.34) yields that

$$\widehat{\pi}(\alpha'_i) = (\widehat{\pi}_2^{-1} \circ \widehat{\pi}_1)(\alpha'_i)) = \widehat{\pi}_2^{-1}(\widehat{\pi}_1(\alpha'_i)) = \widehat{\pi}_2^{-1}(\widehat{\pi}_2(\alpha'_{i'})) = (\widehat{\pi}_2^{-1} \circ \widehat{\pi}_2)(\alpha'_{i'}) = \alpha'_{i'}.$$

Similarly for the rest of components. So

$$\widehat{\pi}^*(\vec{g}^{(i)}) = \langle \widehat{\pi}(\alpha'_i), \widehat{\pi}(\beta'_i), \widehat{\pi}(\gamma'_i), \widehat{\pi}(\delta'_i) \rangle = \langle \alpha'_{i'}, \beta'_{i'}, \gamma'_{i'}, \delta'_{i'} \rangle = \vec{g}^{(i')}. \tag{5.35}$$

Now let f be a quaternary lattice term. Using that $\widehat{\pi}$ is a lattice automorphism and thus it commutes with f, let us compute:

$$\widehat{\pi}(f(\alpha_i, \beta_i, \gamma_i, \delta_i)) = f(\widehat{\pi}(\alpha_i), \widehat{\pi}(\beta_i), \widehat{\pi}(\gamma_i), \widehat{\pi}(\delta_i))) \overset{(5.35)}{=} f(\alpha'_{i'}, \beta'_{i'}, \gamma'_{i'}, \delta'_{i'}). \tag{5.36}$$

Since $\{\vec{\alpha}, \vec{\beta}, \vec{\gamma}, \vec{\delta}\}$ generates the direct product (5.24), for each

$$\vec{\mu} = (\mu'_1, \mu'_2, \dots, \mu'_i, \dots, \mu'_{i'}, \dots, \mu'_t, \mu''_{t+1}, \dots, \mu''_j, \dots, \mu''_{2t}) \tag{5.37}$$

of the direct product (5.24), there is a quaternary lattice term f such that $\vec{\mu}$ is of the form

$$\begin{aligned} &\vec{\mu} = f\bigl(\vec{\alpha},\ \vec{\beta},\ \vec{\gamma},\ \vec{\delta}\bigr) = \\ &\quad \langle \dots, \underbrace{f(\alpha'_i, \beta'_i, \gamma'_i, \delta'_i)}_{\mu'_i}, \dots, \underbrace{f(\alpha'_{i'}, \beta'_{i'}, \gamma'_{i'}, \delta'_{i'})}_{\mu'_{i'}}, \dots, \underbrace{f(\alpha''_j, \beta''_j, \gamma''_j, \delta''_j)}_{\mu''_j}, \dots \rangle \end{aligned} \tag{5.38}$$

where $j \in \{t+1, \dots, 2t\}$. (Note that j and μ''_j will only be needed later, not here.) Combining (5.36), (5.37) and (5.38), it follows that

$$\widehat{\pi}(\mu'_i) = \widehat{\pi}(f(\alpha'_i, \beta'_i, \gamma'_i, \delta'_i)) = f(\alpha'_{i'}, \beta'_{i'}, \gamma'_{i'}, \delta'_{i'}) = \mu'_{i'}. \tag{5.39}$$

Now if $\pi_1 = \pi_2$, then π and $\widehat{\pi}$ are the identity permutations and (5.39) turns into $\mu'_i = \mu'_{i'}$. But this is a contradiction since $i \neq i'$ by (5.33) and so the fact that $\vec{\mu}$ in (5.37) is an arbitrary $(2t)$-tuple of (5.24) allows us to choose μ'_i and $\mu'_{i'}$ such that $\mu'_i \neq \mu'_{i'}$. Thus $\pi_1 \neq \pi_2$ and the automorphism $\widehat{\pi}$ is not identity map of $\mathrm{Part}(n)$. However, in the arbitrary $(2t)$-tuple (5.37), we can pick $\mu'_i \in \mathrm{Part}(n)$ arbitrarily, and we can let $\mu'_{i'} := \mu'_i$ regardless if $i' = i$ or $i' \neq i$. With this choice of $\mu'_{i'}$, we obtain from (5.39) that $\widehat{\pi}(\mu'_i) = \mu'_i$ for all $\mu'_i \in \mathrm{Part}(n)$, which contradicts the fact that now $\widehat{\pi}$ is not the identity map. The argument proving (5.31) is complete. Then, as we have already mentioned, (5.32) is also true.

Next, we claim that, for all $i \in \{1, \dots, t\}$ and $j \in \{t+1, \dots, 2t\}$,

$$\{\langle \alpha'_i, \alpha''_j \rangle, \langle \beta'_i, \beta''_j \rangle, \langle \gamma'_i, \gamma''_j \rangle, \langle \delta'_i, \delta''_j \rangle\} \text{ generates } \mathrm{Part}(n) \times \mathrm{Part}(n+1). \tag{5.40}$$

Let $\langle \mu'_i, \mu''_j \rangle$ be an arbitrary element of $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$. We can extend the pair $\langle \mu'_i, \mu''_j \rangle$ to a $(2t)$-component vector $\vec{\mu}$ as in (5.37). As (5.38), shows, $\langle \mu'_i, \mu''_j \rangle$ is of the form

$$\begin{aligned}\langle \mu'_i, \mu''_j\rangle &= \langle f(\alpha'_i, \beta'_i, \gamma'_i, \delta'_i), f(\alpha''_j, \beta''_j, \gamma''_j, \delta'_j)\rangle\\ &= f(\langle \alpha'_i, \alpha''_j\rangle, \langle \beta'_i, \beta''_j\rangle, \langle \gamma'_i, \gamma''_j\rangle, \langle \delta'_i, \delta''_j\rangle).\end{aligned} \tag{5.41}$$

with some quaternary lattice term f. Hence, $\langle \mu'_i, \mu''_j\rangle$ belongs to the sublattice generated by $\{\langle \alpha'_i, \alpha''_j\rangle, \langle \beta'_i, \beta''_j\rangle, \langle \gamma'_i, \gamma''_j\rangle, \langle \delta'_i, \delta''_j\rangle\}$, proving (5.40).

Next, based on (5.40), we state even more than (5.40). Namely, we state that

$$\begin{gathered}\text{for every } i \in \{1,\dots,t\}, \text{ for every } j \in \{t+1,\dots,2t\}, \text{ and for arbitrary permutations } \pi \in S_n\\ \text{and } \sigma \in S_{n+1},\\ \langle\langle \widehat{\pi}(\alpha'_i), \widehat{\sigma}(\alpha''_j)\rangle, \langle \widehat{\pi}(\beta'_i), \widehat{\sigma}(\beta''_j)\rangle, \langle \widehat{\pi}(\gamma'_i), \widehat{\sigma}(\gamma''_j\rangle, \langle \widehat{\pi}(\delta'_i), \sigma(\widehat{\delta''_j})\rangle\rangle\\ \text{is a generating quadruple of } \operatorname{Part}(n)\times\operatorname{Part}(n+1).\end{gathered} \tag{5.42}$$

Clearly, the map κ: $\operatorname{Part}(n) \times \operatorname{Part}(n+1) \to \operatorname{Part}(n) \times \operatorname{Part}(n+1)$, defined by $\langle \mu'_i, \mu''_j\rangle \mapsto \langle \widehat{\pi}(\mu'_i), \widehat{\sigma}(\mu''_j)\rangle$, is bijective. Since lattice operations are computed component-wise and since both $\widehat{\pi}$ and $\widehat{\sigma}$ are automorphisms, it follows that κ is an automorphism of the direct product $\operatorname{Part}(n) \times \operatorname{Part}(n+1)$. Therefore, the element-wise κ-image of a generating set is again a generating set and (5.40) implies (5.42).

Next, we count how many generating quadruples occur in (5.42). Each of the parameters i and j can be chosen in t ways. Hence, the pair of subscripts $\langle i, j\rangle$ can be chosen in t^2 ways. There are $n! = |S_n|$ ways to chose the parameter π and, similarly, $(n+1)!$ ways to pick a permutation σ. Therefore,

$$\text{there are } t^2 \cdot n! \cdot (n+1)! \text{ ways to chose a quadruple } \langle i, j, \pi, \sigma\rangle \text{ with components occurring in (5.42).} \tag{5.43}$$

We need to show that whenever a meaningful quadruple $\langle i', j', \pi', \sigma'\rangle$ of parameters is different from the quadruple occurring in (5.43) then, for the corresponding generating quadruple of $\operatorname{Part}(n) \times \operatorname{Part}(n+1)$,

$$\begin{gathered}\langle\langle \widehat{\pi}(\alpha'_i), \widehat{\sigma}(\alpha''_j)\rangle, \langle \widehat{\pi}(\beta'_i), \widehat{\sigma}(\beta''_j)\rangle, \langle \widehat{\pi}(\gamma'_i), \widehat{\sigma}(\gamma''_j)\rangle, \langle \widehat{\pi}(\delta'_i), \widehat{\sigma}(\delta''_j)\rangle\rangle \neq\\ \langle\langle \widehat{\pi}'(\alpha'_{i'}), \widehat{\sigma}'(\alpha''_{j'})\rangle, \langle \widehat{\pi}'(\beta'_{i'}), \widehat{\sigma}'(\beta''_{j'})\rangle, \langle \widehat{\pi}'(\gamma'_{i'}), \widehat{\sigma}'(\gamma''_{j'})\rangle, \langle \widehat{\pi}'(\delta'_{i'}), \widehat{\sigma}'(\delta''_{j'})\rangle\rangle.\end{gathered} \tag{5.44}$$

Here $\widehat{\pi}'$ denotes $\widehat{\pi'}$ and similarly for $\widehat{\sigma}'$, of course. So assume that $\langle i, j, \pi, \sigma\rangle \neq \langle i', j', \pi', \sigma'\rangle$. Then $\langle i, \pi\rangle \neq \langle i', \pi'\rangle$ or $\langle j, \sigma\rangle \neq \langle j', \sigma'\rangle$. Since the first t components of (5.24) and the last t components play a similar role, we can assume that $\langle i, \pi\rangle \neq \langle i', \pi'\rangle$. Then, applying (5.31) with $\langle \pi, \pi'\rangle$ playing the role of $\langle \pi_1, \pi_2\rangle$ and taking

(5.29) account, we obtain that

$$\begin{aligned}\langle\widehat{\pi}(\alpha'_i),\widehat{\pi}(\beta'_i),\widehat{\pi}(\gamma'_i),\widehat{\pi}(\delta'_i)\rangle &= \widehat{\pi}^*(\vec{g}^{(i)}) \neq \widehat{\pi}'^*(\vec{g}^{(i')})\\ &= \langle\widehat{\pi}'(\alpha'_{i'}),\widehat{\pi}'(\beta'_{i'}),\widehat{\pi}'(\gamma'_{i'}),\widehat{\pi}'(\delta'_{i'})\rangle.\end{aligned} \tag{5.45}$$

Thinking of the first components of the pairs occurring in (5.44), we obtain that (5.45) implies (5.44). This shows the validity of (5.44). Now, (5.42), (5.43) and (5.44) together imply that

$$\begin{aligned}&\text{the number of generating quadruples}\\ &\text{we have considered is } t^2\cdot n!\cdot(n+1)!\,.\end{aligned} \tag{5.46}$$

Next, consider a generating quadruple

$$\langle\langle\widehat{\pi}(\alpha'_i),\widehat{\sigma}(\alpha''_j)\rangle,\langle\widehat{\pi}(\beta'_i),\widehat{\sigma}(\beta''_j)\rangle,\langle\widehat{\pi}(\gamma'_i),\widehat{\sigma}(\gamma''_j\rangle,\langle\widehat{\pi}(\delta'_i),\sigma(\widehat{\delta''_j}\rangle)\rangle \tag{5.47}$$

from (5.42). It determines a generating set

$$\{\langle\widehat{\pi}(\alpha'_i),\widehat{\sigma}(\alpha''_j)\rangle,\langle\widehat{\pi}(\beta'_i),\widehat{\sigma}(\beta''_j)\rangle,\langle\widehat{\pi}(\gamma'_i),\widehat{\sigma}(\gamma''_j\rangle,\langle\widehat{\pi}(\delta'_i),\sigma(\widehat{\delta''_j}\rangle)\}. \tag{5.48}$$

Using the same technique with quaternary lattice terms as in the neighbourhood of (5.38), it is straightforward to see that the first components of the pairs in (5.48) generate Part(n). We know from Zádori [**94**] that Part(n) cannot be generated with fewer than four elements. Hence, there are four different first components in (5.48), implying that (5.48) is a 4-element set, so a 4-element generating set.

Assume that a generating quadruple

$$\langle\langle\widehat{\pi}'(\alpha'_{i'}),\widehat{\sigma}'(\alpha''_{j'})\rangle,\langle\widehat{\pi}'(\beta'_{i'}),\widehat{\sigma}'(\beta''_{j'})\rangle,\langle\widehat{\pi}'(\gamma'_{i'}),\widehat{\sigma}'(\gamma''_{j'})\rangle,\langle\widehat{\pi}'(\delta'_{i'}),\widehat{\sigma}'(\delta''_{j'})\rangle\rangle \tag{5.49}$$

different from (5.47) gives the same generating set (5.48) as (5.47). In the worst case, there could be $4! = 24$ different generating quadruples giving the same set (5.48); if this was the case then the denominator in the theorem would be 24 rather than 2. But in [**26**], α_i and δ_i were constructed in a way that each of them has some specific property that distinguish it from the rest of the four partitions. These specific properties are explicitly described in page 422 and (the beginning of) page 423 in [**26**]. We do not give the exact details of these properties here; we only mention that for a large odd n, α_i is the only partition out of α_i, β_i, γ_i and δ_i that has an $(n{+}1)/2$-element block and has exactly two blocks. The specific properties described in [**26**] are clearly preserved by automorphisms. This implies that $\widehat{\pi}(\alpha'_i) = \widehat{\pi}'(\alpha'_{i'})$ and $\widehat{\pi}(\delta'_i) = \widehat{\pi}'(\delta'_{i'})$. Although we did not characterize β_i and γ_i by individual properties among the four partitions constructed in [**26**], we did characterize the *set* $\{\beta_i,\gamma_i\}$ by such a property in pages 422–423 of [**26**]; this property again is preserved by automorphisms. Hence, $\{\widehat{\pi}(\beta'_i),\widehat{\pi}(\gamma'_i)\}$ is necessarily the same as the set $\{\widehat{\pi}'(\beta'_{i'}),\widehat{\pi}'(\gamma'_{i'})\}$. This implies that there are only at most two ways to choose the quadruple (5.49): either it is the same as (5.47), or we get it from (5.47) by interchanging the middle two pairs of partitions. Now we are in the position to conclude that the number of 4-element generating sets is at least half of the number of generating quadruples given in (5.46). This completes the proof of Theorem 4.10. $\square$

Bibliography

[1] A. Agresti: Categorical data analysis. Wiley Series in Probability and Statistics. Wiley-Interscience *John Wiley and Sons*, New York, second edition, 2002.

[2] G. B. Arous, L. V. Bogachev, and S.A Molchanov: Limit theorems for sums of random exponentials. *Probab. Theory Relat. Fields.*, 132:579–612, 2005.
(DOI) 10.1007/s00440-004-0406-3

[3] A. Barczyk, A. Janssen, and M. Pauly: The asymptotics of Lstatistics for non i.i.d. variables with heavy tails. *Probab. Math. Statist.*, 31(2):285–299, 2011.

[4] J. Beirlant, A. Guillou, G. Dierckx, et al.: Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* 10:151–174, 2007.

[5] J. Beirlant, G.Dierckx, Y. Goegebeur, and G.Matthys: Tail index estimation and an exponential regression model. *Extremes* 2:177–200, 1999.

[6] J. Beirlant, P. Vynckier, and J.L. Teugels: Excess functions and estimation of the extreme value index. *Bernoulli*, 2:293–318, 1996.

[7] J. Beran, D. Schell, and M. Stehlík: The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Statist. Math.*, 66(1):193–220, 2014.

[8] G. Biau, and D.M. Mason: High-Dimensional p-Norms.*arXiv: Statistics Theory*, 21–40, 2015. arXiv:1311.0587v1

[9] P. Billingsley: Probability and measure: Wiley Series in Probability and Mathematical Statistics. *John Wiley and Sons*, Inc., New York, third edition, 1995.

[10] N. H. Bingham, C. M. Goldie, and J. L. Teugels: Regular variation, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[11] L. V. Bogachev: Limit laws for norms of IID samples with Weibull tails. *J. Theor. Probab.*, 19:849–873, 2006.

[12] M. F. Brilhante, M. I. Gomes, and D. Pestana: A simple generalisation of the Hill estimator. *Comput. Statist. Data Anal.*, 57:518–535, 2013.

[13] L. Canto e Castro, and S. Dias: Generalized Pickands estimators for the tail index parameter and max-semistability. *Extremes* 14:429–449, 2011.

[14] E. Castillo: Extreme Value Theory in Engineering. *Academic Press*, 1988.

[15] G. Ciuperca and C. Mercadier: Semi-parametric estimation for heavy-tailed distributions. *Extremes*, 13(1):55–87, 2010.

[16] S. Coles: Classical Extreme Value Theory and Models. In: An Introduction to Statistical Modeling of Extreme Values. *Springer Series in Statistics. Springer*, 19–29, 2001.

[17] S. Csörgő and D. M. Mason: Central limit theorems for sums of extreme values. *Math. Proc. Cambridge Philos. Soc.*, 98(3):547–558, 1985.

[18] S. Csörgő and L. Viharos: On the asymptotic normality of Hill's estimator. *Math. Proc. Cambridge Philos. Soc.*, 118(2):375–382, 1995.

[19] S. Csörgő, and L. Viharos: Asymptotic normality of least-squares estimators of tail indices. *Bernoulli* 3:351–370, 1997.

[20] S. Csörgő, E. Haeusler, and D. M. Mason: The asymptotic distribution of extreme sums. *Ann. Probab.*, 19(2):783–811, 1991.

[21] S. Csörgő, P. Deheuvels, and D. Mason: Kernel estimates of the tail index of a distribution. *Ann. Stat.*, 13:1050–1077, 1985.

[22] G. Czédli: Four-generated large equivalence lattices. *Acta Sci. Math. (Szeged)*, 62:47–69, 1996.

[23] G. Czédli: Lattice generation of small equivalences of a countable set. Order 13:11–16, 1996.

[24] G. Czédli: (1+1+2)-generated equivalence lattices. *J. Algebra*, 221:439–462, 1999.

[25] G. Czédli: Four-generated direct powers of partition lattices and authentication. *Publicationes Mathematicae (Debrecen)*, accepted; `https://arxiv.org/pdf/2004.14509`

[26] G. Czédli and Lillian Oluoch: Four-element generating sets of partition lattices and their direct products. *Acta Sci. Math. (Szeged)*, 86:405–448, 2020.

[27] J. L. Danielsson, L. De Haan, L. Peng, and C.G.de Vries: Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.* 76:226–248, 2001.

[28] L. De Haan: Extreme value statistics. In Extreme value theory and applications. *Kluwer Acad. Publ.*, Dordrecht, 1:93–122, 1994.

[29] L. De Haan and A. Ferreira: Extreme value theory. *Springer Series in Operations Research and Financial Engineering. Springer*, New York, 2006.

[30] L. De Haan, and S. Resnick: On asymptotic normality of the hill estimator, Communications in statistics. *Stochastic models*, 14:849–866, 1998.

[31] P. Deheuvels, E. Haeusler, and D. M. Mason: Almost sure convergence of the Hill estimator. *Math. Proc. Cambridge Philos. Soc.*, 104(2):371–381, 1988.

[32] A. L. M. Dekkers, J. H. J Einmahl, and L. De Haan: A Moment Estimator for the Index of an Extreme-Value Distribution. *Ann. Statist.* 17:1833—1855, 1989.

[33] A. L. M. Dekkers and L. De Haan: On the estimation of the extreme value index and large quantile estimation, Ann. Statist., 17, 1795-1832 by Mixture Modelling. *Extremes* 6:25–47, 1989.

[34] H. Drees: A general class of estimators of the extreme value index. *J. Stat. Planing and Inference:*, 66:95–112, 1998.

[35] J. H. J. Einmahl and D. M. Mason: Laws of the iterated logarithm in the tails for weighted uniform empirical processes. *Ann. Probab.*, 16(1):126–141, 1988.

[36] P. Embrechts, C. Klüppelberg, and T. Mikosch: Modelling extremal events. *Applications of Mathematics (New York)*, Springer-Verlag, Berlin, 1997.

[37] P. Embrechts, S. Resnick, and G. Samorodnitsky: Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal*, 3(2):30–41, 1999. DOI: 10.1080/10920277.1999.10595797 .

[38] A. Feuerverger, and P. Hall: Estimating a tail exponent by modeling departure from a Pareto distribution, *Ann. Statist.* 27:760–781, 1999.

[39] M. Fréchet: Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon.*, Math. 6:93–116, 1927.

[40] J. Galambos, J. Lechner, and E. Simiu: Multivariate threshold methods. *Extreme Value Theory and Applications*, 225–248, 1994.

[41] J. Galambos: Asymptotics; Stable Laws for Extremes; Tail Properties. In: de Oliveira J.T. (eds) *Statistical Extremes and Applications.* NATO ASI Series. Springer, 1984.

[42] B. V. Gnedenko and A. N. Kolmogorov: Limit distributions for sums of independent random variables. *Addison-Wesley Publishing Company*, Inc., Cambridge, Mass., 1954.

[43] M. I. Gomes, and A. Guillou: An Extreme Value Theory and Statistics of Univariate Extremes: A Review. *International Statistical Review, Wiley* , 83 (2):263–292, 2015.

[44] M.I. Gomes, and M.J. Martins: Generalizations of the Hill estimator; asymptotic versus finite sample behaviour. *Journal of statistical planning and inference*, 93:161–180, 2001.

[45] E.J. Gumbel: Les valeurs extrêmes des distributions statistiques *Annales de l'Institut Henri Poincaré* 5 (2):115–158, 1935.

[46] E. J. Gumbel: Statistics of extremes. Columbia University Press. New York, 1958.

[47] E. Haeusler, and J.L. Teugels: On asymptotic normality of Hill's estimator for the exponent of regular variation, *Ann. Statist.* 13:743–746, 1985.

[48] P. Hall: On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, 44(1):37–42, 1982.

[49] R. Herrera, and B. Schipp: Self exciting extreme value models for stock market crushes. *Statistical inference, econometric analysis and matrix algebra*, 2008.

[50] B.M. Hill: A simple general approach to inference about the tail of a distribution. *Ann. Stat*, 3:1163–1174, 1975.

[51] J. L. Hodges Jr, and E. L. Lehmann: Basic concepts of probability and statistics. Second edition Holden-Day, San Francisco, California-London-Amsterdam 1970, xiii+441 pp.

[52] J. R. M. Hosking, J. R. Wallis, and E. F. Wood: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27:251–261, 1985.

[53] M. Jacob , C. Neves , V.D. Greetham: Extreme Value Theory. In: Forecasting and Assessing Risk of Individual Electricity Peaks. *Mathematics of Planet Earth.* Springer, Cham. 2020.

[54] A. Janßen: Limit laws for power sums and norms of i.i.d. samples. *Probab. Theory Relat. Fields*, 146:515–533, 2010.

[55] A. F. Jenkinson: The frequency distribution of annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteorol. Soc.*, 81:158–171, 1955.

[56] P. Kevei, L. Oluoch, and L. Viharos: Limit laws for the norms of extremal samples. Submitted to *Journal of Statistical Planning and Inference*, http://arxiv.org/abs/2004.12736

[57] S. Kotz and S. Nadarajah: Extreme Value Distribution: Theory and Applications. *World Scientific*, 2000.

[58] J. Kulin: Quasiorder lattices are five-generated. *Discuss. Math. Gen. Algebra Appl.*, 36:59–70, 2016.

[59] J. L. Latchman, F. D. O. Morgan and W. P. Aspinall: Temporal changes in the cumulative piecewise gradient of a variant of the Gutenberg–Richter relationship, and the imminence of extreme events. *Earth-Science Reviews* , 87(3-4):94–112, 2008.

[60] M. R. Leadbetter, G. Lindgren and H. Rootzén: Asymptotic Distributions of Extremes. In: Extremes and Related Properties of Random Sequences and Processes : *Springer Series in Statistics.* Springer, New York, 1983.

[61] M. Lefebvre: Applied Probability and Statistics. Springer, New York, 2006. xviii+356 pp.

[62] W. Mendenhall, R. J. Beaver, and B. M. Beaver: Introduction to Probability and Statistics. 14th (student) edition, BROOKS/COLE CENGAGE Learning, ISBN-13: 978-1-133-10375-2, ISBN-10: 1-133-10375-8, 2013.

[63] L. Oluoch and A. Al-Najafi: Lower bound for the number of 4-element generating sets of direct products of two neighboring partition lattices. *Discussiones Mathematicae — General Algebra and Applications*, 2021. Accepted.

[64] L. Oluoch and L. Viharos: Asymptotic distributions for weighted power sums of extreme values. *Acta Sci. Math. (Szeged)*, 2020. Accepted.

[65] F. W. J. Olver: Asymptotics and special functions. *AKP Classics. A K Peters, Ltd.*, Wellesley, MA, 1997. Reprint of the 1974 original [Academic Press, New York; MR0435697 (55 # 8655)].

[66] V. Paulauskas and M. Vaičiulis : A class of new tail index estimators. *Ann. Inst. Statist. Math.*, 69(2):461–487, 2017.

[67] H. Penalva, M. I. Gomes, F. Caerio, and M. M. Neves: A couple of non reduced bias generalized means in extreme value theory: An asymptotic comparison. *REVSTAT Stat. J.*, 18(3):281–298, 2020.

[68] V. V. Petrov: Limit theorems of probability theory, volume 4 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1995.

[69] J. Pickands III.: Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics.*, 3 (1):119–131, 1975.

[70] P. Pudlák and J. Tůma: Every finite lattice can be embedded in a finite partition lattice. *Algebra Universalis*, 10:74–95, 1980.

[71] S. Resnick: Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin: The Journal of the IAA*, 27(1):139–151, 1997.

[72] S. Resnick: Heavy tail phenomena probabilistic and Statistical modelling. *Springer Series on Operations research and Financial Engineering*, 2007.

[73] S. Resnick, C. Staăricaă: Smoothing the Moment Estimator of the Extreme Value Parameter. *Extremes*, 1:263–293 , 1999.

[74] S. Resnick, and C. Staăricaă: Smoothing the Hill estimator. *Adv. in Appl. Prob.*, 29:271–293, 1997.

[75] M. Schlather: Limit Distributions Of Norms Of Vectors Of Positive I.I.D. Random Variables. *Ann. Probab*,29(2):862–881, 2001.

[76] J. Segers: Generalized Pickands estimators for the extreme value index. *Journal of Statistical Planning*, 2005.

[77] J. Segers: Residual estimators. *J. Statist. Plann. Inference*, 98(1-2):15–27, 2001.

[78] J. Segers: Abelian and Tauberian theorems on the bias of the Hill estimator. *Scand. J. Statist.*, 29(3):461–483, 2002.

[79] R. L. Smith: Estimating Tails of Probability Distributions. *Ann. Statist.*, 15(3):1174—1207, 1987.

[80] A. J. Stam: Generation of a random partition of a finite set by an urn model. *J. Combin. Theory Ser. A*, 35:231–240, 1983.

[81] H. Strietz: Finite partition lattices are four-generated. In: *Proc. Lattice Th. Conf. Ulm*, 257–259, 1975.

[82] H. Strietz: Über Erzeugendenmengen endlicher Partitionenverbände. (German) *Studia Sci. Math. Hungar.*, 12:1–17, 1977.

[83] L. C. H. Tippett: On the extreme individuals and the range of samples taken from a normal population. *Biometrika* 17:364-387, 1925.

[84] C. Tudor: Implication of extreme values theory for stock market proceedings of international academic conference. *International institute of social and economic science*, 2018.

[85] L. Viharos: Asymptotic distributions of linear combinations of extreme values. *Acta Sci. Math. (Szeged)*, 58(1-4):211–231, 1993.

[86] L. Viharos: Limit theorems for linear combinations of extreme values with applications to inference about the tail of a distribution. *Acta Sci. Math. (Szeged)*, 60(3-4):761–777, 1995.

[87] L. Viharos: Estimators of the exponent of regular variation with universally normal asymptotic distributions. *Math. Methods Statist.*, 6(3):375–384, 1997.

[88] L. Viharos: Weighted least squares estimators of tail indices. *Probab. Math. Statist.*, 19(2):249–265, 1999.

[89] R. von Mises: La distribution de la plus grande de n valeurs. Rev. Math. Union Interbalcanique 1, 141-160 (1936) Reproduced in: Selected Papers of Richard von Mises, Amer. *Math. Soc.*, Vol. 2:271–294, 1964.

[90] W. Weibull: A statistical distribution function of wide applicability. *ASME Journal of Applied Mechanics,* 18 (3):293–297, 1951.

[91] I. Weissman: Statistical Estimation in Extreme Value Theory. In de Oliveira J.T. (eds) Statistical Extremes and Applications. NATO ASI Series. *Springer*, 131, 1998.

[92] J. A. Wellner: Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete*, 45(1):73–88, 1978.

[93] S.Yun: On a generalized Pickands estimator of the extreme value index. *Journal of Statistical Planning and Inference*, 102:389–409, 2002.

[94] L. Zádori: *Generation of finite partition lattices.* In: Lectures in universal algebra (Proc. Colloq. Szeged, 1983), Colloq. Math. Soc. János Bolyai **43**, North-Holland, 573–586.

Acknowledgments

First and foremost, I wish to give thanks to God Almighty, for His showers of blessings, and unfailing love, and for filling me with patience and endurance throughout my research work to successful completion.

This work would not have been possible without the financial support from the Stipendium Hungaricum Scholarship. I am greatly indebted to The Tempus Foundation for noting the great potential in me and provided the Dissertation Scholarship besides the Stipendium Hungaricum Scholarship.

It is with immense gratitude that I acknowledge the support and help of my esteemed supervisors Prof. László Viharos and Prof. László Zádori for their invaluable supervision, continuous support, patience, motivation, and tutelage during the course of my Ph.D. study. Their immense knowledge and experience have encouraged me in all the time of my academic research and daily life. I cannot find words to express my gratitude to Prof. László Stachó and Prof. Gábor Czédli for their mentorship, insightful comments and suggestions, and for their unwavering support and belief in me. I consider it an honor to work with Prof. Péter Kevei and Prof. Dénes Attila. Their treasured support and joint co-authorship were really influential in shaping this dissertation.

My gratitude extends to the Faculty of Science and Informatics, Bolyai Institute for the opportunity to undertake my studies at the Department of Stochastics, University of Szeged. Special thanks to Csilla Kertész for her kind help and support that has made my study and life in Szeged wonderful. I am indebted to my many colleagues in the Mathematics department for their support and encouragement. I cherish time spent together in the office, and in social settings.

It gives me great pleasure in acknowledging the support and help of Mr. Oginga S. B. Otieno for immensely improving my programming skills by imparting his exemplary computer programming skills.

Nobody has been more important to me in the pursuit of this research than the members of my family. I will forever remain grateful to my parents whose love, prayers, care, sacrifices for educating me, and guidance are with me in whatever I pursue. My deepest gratitude to my loving husband Benn, your encouragement when times got tough and rough is much appreciated and duly noted. My heartfelt thanks to my children Billy, Eunice, and Adrian for allowing me time away from you to research and write and for providing unending inspiration.

Finally, my thanks go to all the people who have in one way or the other supported me to complete the research work directly or indirectly.

Summary

This dissertation contains research on both theoretical and applied statistics. The aim of the theoretical part is to focus on the estimation of the tail index by proposing a new class of estimators for γ with asymptotic properties. The applied part addresses the practical approach that required a practical application of statistics to lattice theory in Algebra. This approach tends to solve Prof. László Zádori's 38-year-old problem on Lattice theory. This dissertation is based on the papers [**64**],[**56**], [**26**] and [**63**].

In chapter 2, the basic definitions and theorems that are used as auxiliary facts for the subsequent chapters 3 and 4 are given. Some statistical background and terminology on extreme value theory are reviewed. Some important properties of asymptotic normality (asymptotic theory of extremes) are also outlined. In this section a description of Heavy-tailed probability distributions as crucial components of stochastic modeling were given. We also describe some Classical extreme value index estimators in particular the Hill, Pickand, and the Moments estimators.

In chapter 3, we pay much attention to the asymptotic normality for the power sums, the consistency, and limit theorems under appropriate assumptions. This chapter is based on a joint paper [**64**]. Weighted power sums were formulated by considering the necessary and sufficient conditions for the existence of normalizing and centering constants. The asymptotic normality for the power sums over the whole heavy-tail model under some constraints on the weights $d_{i,n}$ was proved. Theorems based on Viharos [**86**] and Ciuperca and Mercadier [**15**] were crucial in developing a statistic based on the weighted γ. Using a corollary that describes the asymptotic behavior of the weighted norm $R_n(p)$, an estimator $\hat{\gamma}_n$ was obtained. It's asymptotic normality and asymptotic bias were proved under certain conditions. The results indicated that $\hat{\gamma}_n$ is asymptotically normal over the whole model $\mathbf{R}_\gamma$ while Hill estimator and other known estimators were asymptotically normal but not for every distribution in $\mathbf{R}_\gamma$.

The performance of the estimator was evaluated through simulation study where the performance of $\hat{\gamma}_n$ was compared to the Hill, the Pickand, and the Moment estimators. The Tail index estimators showed good performance in the strict Pareto model which is a simple distribution. For simulation of a more complex distribution

, two cases from Hall model proposed by [**48**] were considered. Using the mean square error as criterion, it was evident that for $\rho \leq 1$ the performance of $\hat{\gamma}_n$ generally increases as γ decreases from 2 to 0.5. For $\gamma \geq 1$ the weights improve the performance of $\hat{\gamma}_n$ significantly ($\rho = 0.5, 1, 2$). It was also observed that $\hat{\gamma}_n$ with $p = 1, 2, 3$ and appropriate ρ value performs better than the Pickands and the moment estimator. In case 2 for $\rho \leq 0.5$ the estimator $\hat{\gamma}_n$ works slightly better than in case 1.

In another simulation study, the rate of convergence to a normal distribution was established. Additionally, the empirical threshold sample sizes for Fréchet distribution, Hall model case 1 and case 2 were determined as $k_n = 900, 500, 900$ respectively. The graphical representation for these thresholds are visualized in terms of histograms and Q-Q plots in which a good fit to a standard normal distribution was observed.

In chapter 4, we relied on the research paper [**56**] in which we proposed another class of estimators for estimating the tail index. The asymptotic properties of $\hat{\gamma}_n$ under conditions of the regular varying upper tail were investigated. The consistency and limit theorem was also proved under appropriate assumptions. The possibility of having a class of estimator $\hat{\gamma}_n$ for $p = p_n \to \infty$ was considered. Special cases for this estimator had been considered by Hall for $p = 1$ and Dekker et al. for $p = 2$. The main focus for this chapter was to consider the estimate $\hat{\gamma}_n$ as $p_n \to \infty$ as the limit law for the norm of the extremal sample.

The asymptotic properties of the class of estimators both for $p > 0$ fixed and for $p_n \to \infty$ were investigated and interestingly, in the course new results for p fixed was also obtained. In this case, the random variables were assumed to be uniformly distributed and hence the normalized and centered sequence was rewritten in another form given in (4.5). In Theorem 2.1, strong consistency of the estimator for p fixed was established conditioned on $k_n/n \to 0$, $(\log n)^\delta/k_n \to 0$ for some $\delta > 0$.

Weak consistency holds under weaker assumption on k_n ($k_n \to \infty$, $k_n/n \to 0$), which follows from representation (4.5) and from the law of large numbers. In Theorem 2.4 assumptions on the slowly varying function ℓ are weaker, hence giving a new dimension to the asymptotic normality in this case. Therefore, Theorem 2.2 indicates weak consistency of the estimator for p fixed. Proving asymptotic normality, the representation (4.5) was used, where the summands are independent and identically distributed conditioned on $U_{k+1,n}$ as shown in (4.6). The asymptotic normality of various generalizations of the Hill estimator is obtained under second-order regular variation for ℓ but we obtain the result from weaker conditions as given in Theorem (2.4). It is necessary to point out that the growth condition (4.10) of the subsequence is the same as in Theorem 4.5 in [**78**].

The main results are contained in Section 3, of this chapter shows results of weak consistency and asymptotic normality when $p_n \to \infty$. The limit theorems with *random centering and norming* for $S_n(p_n)$ were obtained. The results of weak laws and Gaussian limits that were obtained are summarized in Theorem 3.1 in which conditions for weak consistency are outlined. Assuming that conditions for the slowly varying function in (4.9) holds then, Theorem 3.2 summarizes the weak consistency and asymptotic normality for $S_n(p)$. It was evident that weak consistency holds for $\zeta \geq 1$ while asymptotic normality holds for $\zeta \geq 2$. Under appropriate and precise asymptotic assumptions on the power sequence p_n, non-Gaussian stable limit theorems were also proved for $\zeta \geq 2$. The characteristic exponent of the stable law was found to equal ζ, coming from the growth rate of the power sequence p_n. The results for the non-Gaussian stable limits are given in Theorem 3.5.

A simulation study was performed to investigate the behavior of $\widehat{\gamma}(n)$ for large values of p. The estimators were plotted as a function of k in the range $[5, 200]$. For $k \geq 200$, the estimators did not exhibit much change. The quantile function which is a mixture of an exponential and a strict Pareto quantile was constructed such that the parameter of the exponential Q was continuous. The results indicated that for $k_n = 200$, the exponential part of the sample and the Hill estimator changed drastically (for $\gamma = 1$ from 0.98 to 0.76), while for $p = 5$ the change was not as large (from 0.92 to 0.88). From the simulation results in Figures 5 and 6 the setup of the estimators with larger p values worked much better than the Hill estimator. These estimators were not so sensitive to the change in the nature of the quantile function and it is also worth noting that the heavier tails favored larger p values. On asymptotic bias, it was clear that the bias decreased in p which is a critical property in application in some models. This property was illustrated on the well-known dataset of Danish fire insurance claims for different p. It was observed that larger p values naturally produce smoother plots.

In chapter 5, we present the result of the two research papers [**26**] and [**63**] where we investigated four-element generating sets of a partition lattice and established a lower bound for the number of four-element generating sets of direct products of two neighbouring partition lattices. The study of the number of small generating sets of partition lattices finds its base on the recent research papers by Czédli [**25**]. We proved two statements, which solve Zádori's problem on the whether Part(5) and Part(6) have $(1 + 1 + 2)$-generating sets. We also proved that certain direct products of partition lattices are also 4-generated. In particular, some direct powers of $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$ is four-generated for $n \geq 7$. We gave lower bound for the number $\nu(n)$ of four-element generating sets of $\mathrm{Part}(n)$ as well as a statistical approach to $\nu(n)$ for small values of n. The results obtained in this way are analyzed using a computer-assisted statistical approach to $\nu(n)$ for small values of n. A lower bound for the number of 4-element generating sets of direct products of two neighbouring partition lattices was also investigated. The results indicates that the largest t^* when n is odd, under various conditions is the t^* in (5.23) while when n is even is obtained by t^* in (5.25). Other results shows that for arbitrary permutations $\pi \in S_n$ and $\sigma \in S_{n+1}$, $\langle\langle\widehat{\pi}(\alpha'_i), \widehat{\sigma}(\alpha''_j)\rangle, \langle\widehat{\pi}(\beta'_i), \widehat{\sigma}(\beta''_j)\rangle, \langle\widehat{\pi}(\gamma'_i), \widehat{\sigma}(\gamma''_j\rangle, \langle\widehat{\pi}(\delta'_i), \sigma(\widehat{\delta}''_j\rangle)\rangle$ is a generating quadruple of $\mathrm{Part}(n) \times \mathrm{Part}(n+1)$, with $t^2 \cdot n! \cdot (n+1)!$ ways to choose a quadruple $\langle i, j, \pi, \sigma\rangle$.

Journal publications

[1] Gábor Czédli and **Lillian Oluoch**, (2020), Four-element generating sets of partition lattices and their direct products. *Acta Sci. Math. (Szeged)* **86**, 405–448.

[2] Dénes, A., Ibrahim, M.A., **Oluoch, L.** et al.,(2019), Impact of weather seasonality and sexual transmission on the spread of Zika fever. *Sci Rep* 9,17055, https://doi.org/10.1038/s41598-019-53062-z.

[3] **Lillian Oluoch**, László Viharos,(2020), Asymptotic distributions for weighted power sums of extreme values. *Acta Sci. Math. (Szeged)*, accepted.

[4] Peter Kevei, **Lillian Oluoch**, László Viharos,(2020), Limit laws for the norms of extremal samples, submitted.

[5] **Lillian Oluoch** and Amenah Al-Najafi,(2021), Lower bound for the number of 4-element generating sets of direct products of two neighboring partition lattices. *Discussiones Mathematicae — General Algebra and Applications*, accepted.

Declaration of Authorship

I, OLUOCH LILLIAN ACHOLA , declare that this thesis titled, *'Limit laws of weighted power sums of extreme values and Statistical analysis of partition lattices'* and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exceptionn of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed................................

Date.................................