Estimation problems for distributions with heavy tails

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Zhaozhi Fan
aus Shandong, VR China

Göttingen 2001

D 7
Referent: Prof. Dr. Manfred Denker
Korreferent: Prof. Dr. Edgar Brunner
Tag der mündlichen Prüfung: 02.05.2001

Contents

Chapter 1

Introduction

Statistical analysis and stochastic modelling are usually based on special assumptions on the distribution of the parent population. The most widely accepted assumption is normality.
This is based on the fact that any phenomenon, which is influenced by many independent factors, none of these being dominant, is approximately normally distributed. Moreover, normal distributions are simple to deal with and completely determined by their means and variances. Furthermore, the class of normal distributions is closed under convolution.

Contrary to this common model, in many situations of daily life one observes dominant single factors leading to the conclusion that normality is not an appropriate model. Another situation when normality is not an adequate description of a data set arise with observations far away from the average. Here normal distribution is inadequate to describe the experiment, unless one assumes a high variance. Stock returns, for example, were originally supposed to follow a normal distribution or simple mixtures of normal distributions (see [49]). However, empirical studies show so many outliers such that normal distribution cannot be used to describe the data set (see Teichmöler [47]). It is also known that the fluctuation of stock returns cannot be explained by a white noise. The underlying distribution must possess a much fatter tail than a normal distribution.
Similar situations appear in the fields of telecommunications (see [44]), hydrology (see [50]), biology and sociology (see [25] and [51]).

In general, a distribution with fat tail will be called a heavy-tailed distribution.

Among the examples of heavy-tailed distributions are Pareto distributions, student distributions and the family of α-stable distributions with $0 < \alpha < 2$.

Particularly interesting heavy-tailed distributions are those of polynomial form, i.e.* as $x \to \infty$,

$$P(X > x) \sim C_1 x^{-\alpha} + C_2 x^{-\alpha_2} + \cdots + C_n x^{-\alpha_n},$$

$$P(X < -x) \sim D_1 x^{-\beta} + D_2 x^{-\beta_2} + \cdots + D_m x^{-\beta_m},$$

where

$$0 < \alpha < \alpha_2 < \cdots < \alpha_n,\ 0 < \beta < \beta_2 < \cdots < \beta_m,$$

$$C_i, D_j \in R,\ i = 1, \ldots, n,\ j = 1, \ldots, m,\ C_1 > 0,\ D_1 > 0,\ n, m \in \mathcal{N}.$$

Because of their simplicity, they are of particular interest to the statistician. In this thesis,a more general class is considered where

$$P(X > x) \sim x^{-\alpha} L_1(x),$$
$$P(X < -x) \sim x^{-\beta} L_2(x),$$

as $x \to \infty$. Here $L_1(\cdot)$, $L_2(\cdot)$ are slowly varying functions† and α and β are positive real numbers.

The numbers α and β are called the tail indices. They characterize the thickness of the left and right tails, respectively.

We shall mainly be interested in the case where the two tails have the same index, i.e. $\alpha = \beta$. Furthermore, we shall restrict ourselves to the case $L_1(\cdot) = C \cdot L_2(\cdot)$, $C \in R^+$.

If the index α takes values in the interval (0,2], then the associated distribution lies in the domain of attraction of a stable law (see Feller[13], Ibragimov & Linnik[16]).

For an α-stable distribution, $\alpha \in (0, 2)$‡, the tail probabilities are of the form

$$P(X > x) \sim C_1 x^{-\alpha},$$
$$P(X < -x) \sim C_2 x^{-\alpha},$$

as $x \to \infty$, where $C_1 + C_2 > 0$.

Thus, α-stable distributions are a special type of heavy-tailed distributions with tail index α, where $\alpha \in (0, 2)$.

Mandelbrot (see [32, 33, 34, 35]) and Fama (see [10]) suggested to use stable distributions with tail index $1 < \alpha < 2$ to model stock prices; however, stable distributed samples are only rarely observed in practice. It is more suitable to model the observations with distributions in the domain of attraction of some stable law.

*For the explanation of the symbol $\sim$ I refer to page 17.

†See section 3.1 for details.

‡The case $\alpha = 2$ corresponds to the normal distribution, which is not heavy tailed.

Tail indices persist under convolutions and stable distributions appear as limit laws of properly normalized means. For this reason parameter estimation of Tail indices has been an important field of research for quite a long time.

Fama and Roll (see [11, 12]) started in 1968 to develop a method to estimate the index of a stable distribution.

Some years later, Press (see [41]) found another approach to estimate the parameter based on the empirical characteristic function of a stable distribution.[‡].

Yet another important approach is due to Zolotarev (see [52])in 1986. It uses moment properties of the logarithm of the absolute value of a stable random variable and the sign of the observations[§].

In the more general case of a heavy-tailed distribution the estimation of the tail index is more complicated due to the slowly varying function L.
In 1975 Hill (see [23]) investigated the special case of a heavy-tailed distribution with Pareto-type tail. He constructed a simple estimator for the tail index by maximizing the conditional Likelihood function. In fact, Hill's estimator is a plotting procedure. It depends on the number k of largest observations chosen to calculate the estimate[¶], where k is, of course, less than the sample size n.

In the same year Pickands (see [40]) constructed a different estimator for the same type of distributions. In contrast to Hill, he selected four of the observations to estimate the unknown parameter, namely the kth, the $(2k)$th, the $(3k)$th and the $(4k)$th order statistics. Here $4k$ is less than n. Consistency of the estimator can be obtained by choosing k appropriately.

De Haan and Resnick (see [21]) considered heavy-tailed distributions with regularly varying tails. Using the properties of the regularly varying function, they constructed a very simple estimator[‖].

The estimator of Csörgö, Deheuvels and Mason (see [9]), also constructed for heavy-tailed distributions with regularly varying tail, has a very general form. Here a suitably chosen kernel on the real line plays a very important role[**].

In fact, the estimators of Hill and De Haan & Resnick are special cases of this

[‡]See page 9 and chapter 5.
[§]See page 11 and chapter 5.
[¶]For details I refer to section 2.2.1.
[‖]See section 2.2.2.
[**]See section 2.2.

estimator, corresponding to special kernels†, respectively.

In 1981 Teugels (see [48]) studied the case where the attracting distribution is asymmetric stable. He ordered the sample in a special way: the observations were rearranged into the order with ascending absolute values. Selecting k of the so ordered statistics, $k < n$, he obtained a sequence of estimates.

Recently, De Haan and Pereira (see [20]) developed a further estimator based on the order statistics. In contrast to the approaches mentioned above, they took into account the conditions under which the distribution is attracted to a stable one††. They used all but the k largest order statistics of the absolute values of the observations to calculate the estimate. By choosing an adequate k one obtains consistency of the estimator. Moreover, under certain assumptions on the slowly varying function involved in the distribution tail, the estimator is asymptotically normally distributed.

One disadvantage of all these estimators lies in the fact that they are susceptible to departures from the assumption that the distribution has Pareto- type tail.

If one confines the attention to data that are so far out in the tail that the Pareto assumption is valid, then the effective sample size can be small. But in this case the estimator of the tail index may have relatively large variance (see Feuerverger and Hall [14]).

Finally, we want to mention the estimator of Meerschaert and Scheffler (see [37]), which is robust to the assumption on the parent distribution. In their setting the distribution of the parent population lies in the domain of normal attraction of an α-stable law. In this case the distribution of the centralized sum of squares of the observations converges to an $\alpha/2$-stable distribution. The construction of the estimator is based on this fact‡‡.

For further estimation methods, we refer to Holt and Crow (see [26]), Mittnick and Rachev (see [38]), and the bibliography of Zolotarev (see [52]).

As indicated above, none of the described estimators is perfect. From different point of views, each one has some disadvantages. Indeed, they can hardly be used simply and effectively in practice. The best convergence rate, for example, is given by $(\log n)^{-1}$, unless one puts some other restrictions on the slowly varying function L§. For this reason the application of heavy tailed distributions is often avoided in practical situations.

---

†See section 2.2.

††See Feller [13] page 312, theorem 1.

‡‡See section 2.2

§See chapter 3 for details.

From a statistical viewpoint there is a need for simple, easy-to-calculate estimates. This is particularly true for the "shape" parameter, or the tail index, α (see Hall [24]). In this thesis we construct some estimators which serve the practical needs in a better way. The estimators have a simpler structure and perform better in simulations. They use not only some, but all the observations. Most importantly, the new estimators are robust with respect to assumptions on the distribution of the parent population, i.e. each of the estimators performs very well no matter whether the distribution of the parent population is stable or has some other heavy tail. Moreover, the convergence rate of our estimators, is better than $(\log n)^{-1}$[§].

In the following we want to describe the construction of the estimators briefly.

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. random variables with strictly stable distribution. The sum-preserving property of stable laws says

$$\frac{X_1 + X_2 + \cdots + X_n}{n^{1/\alpha}} \stackrel{d}{=} X_1,$$

and this leads to

$$\log n \left[\frac{\log |\sum_{i=1}^{n} X_i|}{\log n} - \frac{1}{\alpha} \right] \stackrel{d}{=} \log |X_1|.$$

Let

$$\widehat{\alpha_n} = \frac{\log n}{\log |\sum_{i=1}^{n} X_i|}, \tag{1.1}$$

then $\widehat{\alpha_n} \to \alpha$ in probability as $n \to \infty$.

It will be shown that this estimation method performs quite similar in the case of more general heavy-tailed parent distributions[*].

Let us first consider the case when the sample size n is large ($n \geq 1000$). The main idea for the construction of a new type of estimator of α is the following. Partition the sample $X_1, X_2, ..., X_n$ into k sub-samples $X_{j1}, X_{j2}, ..., X_{jm}, j = 1, 2, ..., k$, each having sample size m, $m = [n/k]$. Every sub-sample provides an estimator $\widehat{\alpha_j} = \frac{\log m}{\log |\sum_{i=1}^{m} X_{ji}|}$, $j = 1, 2, \dots, k$. Let $\widehat{\alpha_s}(k)$ denote the average of these estimators:

$$\widehat{\alpha_s}(k) = \frac{1}{k} \sum_{j=1}^{k} \widehat{\alpha_j}.$$

Since the convergence of $\widehat{\alpha_n}$ to α is rather slow, the bias and variance of $\widehat{\alpha_n}$ is relatively

[§]See chapter 3 for details.

[*]See chapter 3 for details.

close to those of $\widehat{\alpha_j}, j = 1, 2, \ldots, k$, even if n is large. This suggests the idea to replace $\widehat{\alpha_n}$ by the average $\widehat{\alpha_s}(k)$.
Obviously, the variances satisfy the equation

$$Var(\widehat{\alpha_s}(k)) = \frac{1}{k} \cdot Var(\widehat{\alpha_j}), j = 1, 2, \ldots, k.$$

Hence, the variance of the new estimator $\widehat{\alpha_s}(k)$ tends to zero when k tends to infinity. Another advantage of the estimator $\widehat{\alpha_s}(k)$ is the fact that it is asymptotically normally distributed, as $k \to \infty$.

If the sample size is small ($n << 1000$), we cannot proceed in the same manner since the bias of the estimator $\widehat{\alpha_n}$, defined in (1.1), might be large if the size of the sub-samples is small. But the theory of U-statistics provides the possibility to use a generalized average. The estimator with U-statistics structure has the form

$$\widehat{\alpha_U} = \binom{n}{m}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_m \le n} h(X_{i_1}, X_{i_2}, \ldots, X_{i_m})$$

where h is some kernel function and $m < n$. In view of the form of $\widehat{\alpha_n}$ we define the kernel h as follows:

$$h(x_1, x_2, \ldots, x_m) = \frac{\log m}{\log |\sum_{i=1}^m x_i|}.$$

Note that in case of a large sample size, $\widehat{\alpha_U}$ can be used to estimate α, too. Using the theory of U-statistics its asymptotic normality is easily established[†].

In general, the calculation of a U-statistic is elaborate, since we need to average $\binom{n}{m}$ terms. However, because of the dependence of the involved terms, we can omit some summands in $\widehat{\alpha_U}$ without unduly inflating the variance. This gave us the idea of the incomplete U-statistics. The sub-sample estimator $\widehat{\alpha_s}(k)$ is a special type of it.

The effort to simulate the U-statistic $\widehat{\alpha_U}$ led to the following re-sampling procedure. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. observations from a population, whose distribution is attracted to some α-stable law; let $Y_1, Y_2, \ldots, Y_n$ denote independent Bernoulli random variables with parameter p, independent of $X_1, X_2, \ldots, X_n$. For every observation $y_1, y_2, \ldots, y_n$ of $Y_1, Y_2, \ldots, Y_n$, the following expression provides an estimator of α:

$$\widehat{\alpha_{r1}} = \frac{\log \sum_{i=1}^n y_i}{\log |\sum_{i=1}^n X_i y_i|}.$$

[†]See chapter 3.

If $\sum_{i=1}^{n} y_i = L$, we have chosen L of the n observations $X_1, X_2, \cdots, X_n$ for the summation, $\sum_{i=1}^{n} X_i y_i = \sum_{l=1}^{L} X_{i_l}$ say, and

$$\widehat{\alpha_{r1}} = \frac{\log L}{\log |\sum_{l=1}^{L} X_{i_l}|}$$

has the same form as $\widehat{\alpha_n}$[‡] with $n = L$.
Repeating this procedure k times, we obtain k estimates $\widehat{\alpha_{rj}}, j = 1, 2, \ldots, k$. The average

$$\widehat{\alpha_{Rs}} = \frac{1}{k} \sum_{j=1}^{k} \widehat{\alpha_{rj}}$$

is the so called re-sampling estimator.
By neglecting the extremal estimates from $\widehat{\alpha_{rj}}, j = 1, 2, \ldots, k$, we obtain a robust estimator.

Again, assume that $X_1, X_2, \cdots, X_n$ is a sequence of i.i.d. random variables, and let the parent distribution be α_0-stable ($\alpha_0 \in (0, 2)$ is the unknown parameter). We construct an estimator for α_0 by minimizing the Kolmogorov distance between the empirical distribution function $\widehat{F_n}$, corresponding to the given sample, and a class $\mathcal{F}$ of distributions indexed by $\alpha \in (0, 2)$, which is defined as follows: First, consider the random variables

$$\{\frac{X_i + X_j}{2^{1/\alpha}}; i, j = 1, 2, \ldots, n\}$$

for all $\alpha \in (0, 2)$; then, for every α, let $\widetilde{F_n^\alpha}$ denote their empirical distribution function. Finally, $\mathcal{F}$ is defined by

$$\mathcal{F} = \{\widetilde{F_n^\alpha} | \alpha \in (0, 2)\}.$$

The sum-preserving property of stable random variables guarantees, that there is only one value of $\alpha \in (0, 2)$, for which the corresponding Kolmogorov distance is minimal. This value of α, denoted by $\widehat{\alpha_{md}}$, is the minimum distance (MD) estimate of the unknown parameter α_0. The MD estimator can also be explained as a U-statistic. Hence it shares their properties[†].

The organisation of the thesis is as follows: In chapter 2, we review some of the most competitive existing estimators for the tail index. This is done for both cases when the parent distribution is stable and when it is a heavy-tailed distribution.

[‡] See page 5
[†] See chapter 4.

In chapter 3, we propose some new estimators for the tail index of heavy-tailed distributions. A re-sampling estimator, an estimator with U-statistic structure and a sub-sample estimator are developed. Large deviation probabilities of the estimators are studied.

In chapter 4, we construct an estimator for the stable exponent. The construction is based on the Minimum-Distance consideration.

In chapter 5, we analyse the results of the simulations. Some comparisons with other estimators are given. The simulations show in the case of a large sample size,that our sub-sample estimator is similar to that of Press or Zolotarev in the sense that all of them are unbiased and have very small standard deviations. When the sample size is small, our re-sampling estimator for the stable index performs even better than that of Press and Zolotarev. The new estimators perform better than those of Hill and Resnick & de Haan when estimating the tail index. Finally, the estimator $\widehat{\alpha_n}$ is compared with the re-sampling estimator in the case of a small sample size. The simulation result shows that the latter one has a much smaller bias and a smaller standard deviation.

Chapter 2

Overview

In this chapter we briefly review some of the most popular estimators in the case of a stable parent population, Pareto type population and more general heavy-tailed populations. We refer to the corresponding references for details. We shall compare their performance with our new estimators (in the appropriate setting of the parent distribution) by simulations. The results will be displayed in chapter 5.

2.1 Estimators of stable exponents

We begin with an introduction to estimators of stable exponents.

Starting with the work of Fama and Roll[11, 12], the estimation theory for the stable index has drawn some attention. Popular estimators are nowadays the methods of Press(1975), Zolotarev(1986), and the procedures based on point process(see Marohn[36] and the references there) or on regression of tail probabilities against the logarithm of the corresponding arguments (see Samorodnitsky and Taqqu[45], chapter 1), or on conditions of convergence towards a stable distribution (see De Haan and Pereira[20]). We begin with the first two.

2.1.1 The estimator of Press

The density of stable distributions have closed analytical form only in rare situations. If α takes values of $2,\ 1,\ 1/2$, the corresponding distributions are respectively normal, Cauchy and Lévy distribution. However, their characteristic functions are known explicitely in an analytical form. The log characteristic function for a random variable X following a stable law is of form:

$$\log(\phi(t)) = i\mu t - |t\sigma|^{\alpha}[1 + i\beta \mathrm{sign(t)}\omega(\mathrm{t},\alpha)] \tag{2.1}$$

where

$$\omega(t,\alpha) = \begin{cases} \tan(\pi\alpha/2), & \text{for} \quad \alpha \neq 1, \\ \frac{2}{\pi}\log(|t|), & \text{for} \quad \alpha = 1. \end{cases}$$

If the parameter μ in equation (2.1) equals 0, the underlying stable distribution is strict, and if $\beta = 0$, the stable distribution is symmetric.

Consider the empirical characteristic function as an estimate of the theoretical one. Clearly, this is an unbiased and consistent estimator. It is possible to calculate estimators of the unknown parameters from the analytical form of the characteristic function. Since the empirical characteristic function is consistent the derived estimators are easily seen to be consistent as well. This kind of estimator is due to Press (see [41]).

Let n independent observations $X_1, X_2, \ldots, X_n$ be given.

Denote the empirical characteristic function of the parent distribution by

$$\widehat{\phi}(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(itX_j).$$

$\widehat{\phi}(t)$ is computable for all values of t. And for every given value of t, the terms included in the summation of $\widehat{\phi}(t)$ are i.i.d. and bounded above by 1. By the law of large number it is a consistent estimator of $\phi(t)$.

Different ways can be used to get estimators of the corresponding parameters. An analytical estimator which involves minimal computation is a version of the method of moments based on the empirical characteristic function.

We begin with the simplest situation:

$$|\phi(t)| = \exp(-\gamma|t|^{\alpha}).$$

For two different non-zero values of t, say t_1, t_2, we get two equations

$$\begin{aligned} \gamma|t_1|^{\alpha} &= -\log(|\phi(t_1)|), \\ \gamma|t_2|^{\alpha} &= -\log(|\phi(t_2)|). \end{aligned}$$

Solving these two equations and replacing $\phi(t)$ by its estimated values, gives the following estimator of α:

$$\widehat{\alpha_p} = \frac{\log\left(\left|\frac{\log\left(\left|\widehat{\phi(t_1)}\right|\right)}{\log\left(\left|\widehat{\phi(t_2)}\right|\right)}\right|\right)}{\log\left(\left|t_1/t_2\right|\right)}$$

The asymptotic normality of $\widehat{\alpha_p}$ has been proved (see Press[41]).

2.1.2 Estimator of Zolotarev

The estimator of Zolotarev is a kind of moment estimator derived from specially transformed original data.

α-stable random variables have finite moments only of order smaller than α. Hence the traditional moment method can not be used in this situation. However, the logarithm of the absolute values of the stable random variables have finite moments of any order. In the case of strict stable distribution, one has simple analytical representations of their moments (see Zolotarev[52]). Together with the moments of the sign variable of the stable variable, one can get estimates of the unknown parameters by the method of moments. The estimates of the parameters are due to Zolotarev (see [52]).

Suppose X is a random variable with an α-stable distribution. Define two random variables U and V in the following way:

$$U = \mathrm{sign(X)},\ \mathrm{V} = \mathrm{log(|X|)}.$$

When the stable distribution is strict, we have the following relation‡:

$$\frac{1}{\alpha^2} = \frac{6}{\pi^2} \cdot \mathrm{Var(V)} - \frac{3}{2}\mathrm{Var(U)} + 1$$

Let $X_1, X_2, \dots, X_n$ be i.i.d. samples from some population, which has a strict α-stable distribution. We can construct two i.i.d. sequences $U_1, U_2, \dots, U_n$ and $V_1, V_2, \dots, V_n$. Denote by S_v^2 and S_u^2 the corresponding sample variances,

$$\begin{aligned} S_v^2 &= \tfrac{1}{n-1}\textstyle\sum_{i=1}^n (V_i - \overline{V})^2, \\ S_u^2 &= \tfrac{1}{n-1}\textstyle\sum_{i=1}^n (U_i - \overline{U})^2, \end{aligned}$$

where $\overline{U}$ and $\overline{V}$ are the corresponding sample means.

Now one obtains the estimator of $\frac{1}{\alpha^2}$:

$$\widehat{\alpha_z^{-2}} = \frac{6}{\pi^2} \cdot S_v^2 - \frac{3}{2} \cdot S_v^2 + 1.$$

As an estimator of α^2, $\widehat{\alpha^2}$ is unbiased and consistent (see [52],chapter 4).

2.2 Estimators of the tail index of heavy-tailed distributions

Now we come to the case of the heavy-tailed population.
The estimators of the tail index are mainly based on two approaches: the maximum

‡See Zolotarev [52], chapter 4

likelihood and the moment properties of the corresponding order statistics. Some examples are Hill(1975), Pickands(1975), de Haan & Resnick(1980), Teugels(1981), Csörgö, Deheuvels & Mason(1985), Hosking, Wallis & Wood(1985).

Most of them are using the largest order statistics. Hill's estimator is mostly used when the distribution tail is of Pareto type.

2.2.1 Hill's estimator

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. positive random variables with a distribution F of Pareto type with index α, that is

$$1 - F(x) \sim C \cdot x^{-\alpha}, \text{ as } \mathrm{x} \longrightarrow \infty,$$

and let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ denote the order statistic.
In the special case of $F(x) = Cx^{\alpha}$ when $x \geq D$ for some $D > 0$, one can easily calculate the maximum likelihood estimator of α. This is called Hill's estimator. It is formally defined by

$$\widehat{\alpha_H^{-1}} = \frac{1}{k} \sum_{i=1}^{k} \log(X_{(i)}) - \log(X_{(k+1)}). \tag{2.2}$$

When k, as a function of n, is chosen to satisfy the following conditions:

$$k(n) \to \infty, \quad k(n) = o(n), \quad \text{as } \mathrm{n} \to \infty,$$

Hill's estimator is consistent. The asymptotic normality of the estimator has been shown under suitable assumptions (see [22], [43] and the references there).
However, as pointed out in [43], Hill's estimator fails to give an accurate estimate of the tail index, in particular for stable distributions as a special kind of heavy-tailed distribution with Pareto type tails.[§]. It is another disadvantage that it uses only a part of the observations to estimate the unknown parameter.

2.2.2 Other estimators

The following estimators are based on similar assumptions on the parent distribution.
Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. positive random variables with a heavy-tailed distribution,let us say,

$$1 - F(x) \sim C \cdot x^{-\alpha}, x \longrightarrow \infty.$$

Let $Z_1 \geq Z_2 \geq \cdots \geq Z_n$ denote the associated order statistics in descending order.

[§]see [43], fig.4

Pickands estimator

Under the assumptions we described in this section, Pickands constructed an estimator of α^{-1} as follows,

$$\widehat{\alpha^{-1}_{picands}} \;=\; (\log(2))^{-1}\log(\frac{Z_m - Z_{2m}}{Z_{2m} - Z_{4m}}).$$

When m is appropriately selected, the estimator is consistent¶.

De Haan & Resnick estimator

Based on the same assumptions as above, De Haan and Resnick developed the following estimator:

$$\widehat{\alpha^{-1}_{DR}} \;=\; (\log(k))^{-1}\{\log(Z_1) \;-\; \log(Z_k)\}.$$

For the consistency and asymptotic normality of the estimator, the parameter k should be appropriately chosen. In fact, k should satisfy the following conditions:

$$k \;\to\; \infty,\;\; k/n \;\to\; 0, \text{ as } n \to \infty.$$

Csörgö, Deheuvels & Mason estimator

Csörgö, Deheuvels & Mason proposed an estimator when the upper tail of the distribution is regularly varying with order $-\alpha$†. The estimator has the form:

$$a_n \;=\; a_{n,\lambda} \;=\; (\sum_{j=1}^{n} K(\frac{j}{n\lambda})\{\log^+(Z_j) \;-\; \log^+(Z_{j+1})\})^{-1}(\int_0^{1/\lambda} K(v)dv),$$

or the following discrete form,

$$\widetilde{a_n} \;=\; \widetilde{a_{n,\lambda}} \;=\; (\sum_{j=1}^{n} K(\frac{j}{n\lambda})\{\log^+(Z_j) \;-\; \log^+(Z_{j+1})\})^{-1}(\sum_{j=1}^{n} \frac{1}{n\lambda}K(\frac{j}{n\lambda})),$$

¶See [40] for details.

†See [19] for details.

where $K(\cdot)$ is a kernel and λ is the bandwidth[‡].
When $K(u) = 1_{\{0<u<1\}}$ or $K(u) = u^{-1}1_{\{0<u<1\}}$ and $\lambda = k/n$, $\widetilde{a_n}$ is the estimator of Hill or De Haan & Resnick, respectively.

Teugels' estimator

In fact, Teugels' estimator is used for the parameters included in the distribution attracted to a stable law. The parameters α and p, which are to be estimated, satisfy the following relations:

$$\begin{aligned} 1 - F(x) &\sim px^{-\alpha}L(x) \\ F(-x) &\sim qx^{-\alpha}L(x), \end{aligned}$$

when $x \to \infty$.

Suppose now that $X_1, X_2, \ldots X_n$ are i.i.d. observations from the population F, and $\tilde{X}_1, \tilde{X}_2, \ldots \tilde{X}_n$ are reordered as follows:

$$|\tilde{X}_1| \le |\tilde{X}_2| \le \cdots \le |\tilde{X}_n|$$

With a special choice of k_n,

$$\widehat{\gamma} = \frac{T_{n,k_n}}{k_n}$$

is an asymptotically unbiased estimator of

$$\frac{\alpha}{1-\alpha}(p-q) := \gamma,$$

where $k_n \to \infty$ and $\mathrm{k_n = o(n)}$ as $\mathrm{n} \to \infty$ and where

$$T_{k_n} = (\sum_{i=1}^{n-k_n} \tilde{X}_i - (n-k_n)\nu)/\tilde{X}_{n-k_n+1}$$

with

$$\nu = \begin{cases} 0, & \text{if } 0 < \alpha < 1 \\ \mu := E(X), & \text{if } 1 < \alpha < 2 \end{cases}$$

The estimator of Meerschaert and Scheffler

[‡]A kernel is a nonnegative non-increasing function $k(u)$ defined on R^+ and satisfies $\int_0^\infty k(u)du = 1$. The bandwidth is the reciprocal of the length of the interval, on which one takes an average.

At the end of this chapter, we introduce the estimator of Meerschaert and Scheffler. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with distribution in the domain of normal attraction of an α-stable law. Then the distribution of the sum of the centred squares

$$S_n = \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, converges to some $\alpha/2$-stable law when $n \to \infty$. Based on this fact, the following estimator is constructed (see [37]):

$$\widehat{\gamma_n} = \frac{\log_+ \sum_{i=1}^{n} (X_i - \bar{X})^2}{2 \log n}$$

where $\log_+(x) = \max\{\log x, 0\}$.

In theorem 1 of [37] it is shown that the estimator is consistent, asymptotically unbiased and L^2-convergent.

Chapter 3

Estimating the stable index

In the previous chapter, we have introduced some known estimation methods for the tail index of distributions with heavy tails. Because of the difficulty in treating the involved slowly varying function, and also because the distribution functions have no closed analytical form, they use either only the largest order statistics, or they are limited to use the empirical characteristic functions. Hence their applications are limited to the situations of their own assumptions.

In this chapter we construct simpler estimators which are robust with respect to the assumptions on the parent distributions. This is the most important advantage of our new estimators over their competitors.

Starting from the sum-preserving property of stable random variables, we construct a consistent estimator for the stable index. Since the estimator only satisfies a large deviation principle of polynomial order, its convergence rate is very slow. The variance of the estimator can be relatively large even when the sample size is large. We develop an estimator with U-statistics structure to improve the variance.
The calculation of the U-statistic estimator using a large sample is quite cumbersome because the asymptotic variance has to be estimated.
In order to simplify the calculation, we suggest to use incomplete U-statistics. A resampling procedure is proposed – this is in fact an incomplete U-statistic with random order[†].

Let $X_1, X_2, ..., X_n$ be i.i.d. observations from population X with distribution F, which is attracted to an α-stable distribution, denoted by $F \in DA(\alpha)$.

[†]See section 3.2 for details.

3.1 Construction of the Estimator

Any α-stable distribution belongs to its own domain of attraction. Hence it has a heavy tail. The corresponding tail index, called the stable index, is denoted by α. We start with the situation that F is a strictly α-stable distribution[†].

By the property of stable distributions we know that

$$S_n = \sum_{i=1}^{n} X_i \stackrel{\mathrm{d}}{=} n^{\frac{1}{\alpha}} X.$$

It follows that

$$\widehat{\alpha_n^{-1}} = \frac{\log(|\sum_{i=1}^{n} X_i|)}{\log(n)} = \frac{\log(|S_n|)}{\log(n)} \tag{3.1}$$

is a reasonable choice for an estimator of the parameter α^{-1}. It is log-stable distributed. As to the more general situation of stable attraction, we have a similar relation in the sense of weak convergence. A slowly varying function is added to the calculation. According to the property of slowly varying functions, the same form of estimation can still be used.

Suppose now $X_1, X_2, ..., X_n$ are i.i.d. copies of X with distribution F attracted to a strict α-stable law, such that

$$x^{\alpha}(1 - F(x)) \sim {}^{*}pL(x),$$

$$x^{\alpha}F(-x) \sim qL(x),$$

as $x \to \infty$ where $p, q \geq 0, p + q = 1$, and $L(x)$ is a slowly varying function with Karamata's representation

$$L(x) = c(x)\exp\left(\int_B^x \frac{\varepsilon(t)}{t} dt\right),$$

$c(t) \to c, \varepsilon(t) \to 0$ as $t \to \infty$; $B > 1$ and c is some constant.

The normalizing constants a_n of the convergence of the partial sum to a stable law can be chosen to satisfy the following equation,

$$nL(a_n) = a_n^{\alpha} \cdot \frac{\Gamma(2-\alpha)}{|1-\alpha|} \cdot |\cos\frac{\pi\alpha}{2}|^{\ddagger},$$

and we will always assume this choice.

[†] See page 9 for detail.

[*] We use "$\sim$" to denote the equivalence of two limiting process in the sense: $A_n \sim B_n$ iff $\lim_{n\to\infty} A_n/B_n = 1$ through the text.

[‡] The righthand side equals $a_n^{\alpha} \cdot \frac{\pi}{2}$, when $\alpha = 1$.

According to theorem 2.1.1 of [16], a_n must be of the form

$$a_n = n^{\frac{1}{\alpha}} h(n),$$

and $h(n)$ is also slowly varying in the sense of Karamata. That is we have

$$\frac{\sum_{i=1}^n X_i}{n^{1/\alpha} h(n)} \Rightarrow^{\dagger} S,$$

where S is some stable distributed random variable.

Theorem 3.1.1. Let $\widehat{\alpha_n^{-1}}$ be defined by (3.1). For any given $\epsilon > 0$, we have the following large deviation principle:

$$\lim_{n\to\infty} \frac{\log P\{\widehat{\alpha_n^{-1}} - \alpha^{-1} < -\epsilon\}}{\log n} = -\epsilon;$$

$$\lim_{n\to\infty} \frac{\log P\{\widehat{\alpha_n^{-1}} - \alpha^{-1} > \epsilon\}}{\log n} = -\alpha\epsilon.$$

Remark: The proof of the theorem is a straight forward calculation of the corresponding large deviation probabilities using a lemma of Heyde [17] for the second equation and the Fourier inversion formula for the first one.

Lemma 3.1.2. Suppose S_n, a_n, X, X_n are defined as above and $b_n \to \infty$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \frac{P\{|S_n| > a_n b_n\}}{nP\{|X| > a_n b_n\}} = 1.$$

or equivalently,

$$\lim_{n\to\infty} \frac{P\{|S_n| > a_n b_n\}}{P\{\max_{1\le k\le n} |X_k| > a_n b_n\}} = 1.$$

The proof of theorem 3.1.1:

Proof. For any given $\epsilon > 0$, as $n \to \infty$,

$$\begin{aligned} P\{\widehat{\alpha_n^{-1}} - \alpha^{-1} > \epsilon\} &= P\{\widehat{\alpha_n^{-1}} - \alpha^{-1} - \frac{\log(h(n))}{\log(n)} > \epsilon - \frac{\log(h(n))}{\log(n)}\} \\ &= P\{\frac{|S_n|}{a_n} > n^{\epsilon}/h(n)\} \end{aligned}$$

[†] Through out the text,"$\Rightarrow$" means convergence in distribution.

$$
\begin{array}{cl}
\overset{lemma3.1.2}{\sim} & nP\{\frac{|X|}{a_n} > n^{\epsilon}/h(n)\} \\
= & nP\{|X| > a_n n^{\epsilon}/h(n)\} \\
= & nP\{|X| > n^{1/\alpha+\epsilon}\} \\
\sim & n \cdot (n^{1/\alpha+\epsilon})^{-\alpha} L(n^{1/\alpha+\epsilon}) \\
= & n^{-\alpha\epsilon} L(n^{1/\alpha+\epsilon}),
\end{array}
$$

where L is slowly varying at infinity. Taking the logarithm of the last term and dividing it by $\log n$, we obtain the second equation.

The first part of the theorem can be proved as follows:

$$
\begin{array}{cl}
P & \{\widehat{\alpha_n^{-1}} - \alpha^{-1} < -\epsilon\} = P\{\frac{|S_n|}{n^{\frac{1}{\alpha}}} < n^{-\epsilon}\} \\
=^{*} & \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp(i \cdot n^{-\epsilon}t) - \exp(-i \cdot n^{-\epsilon}t)}{it} f_n(t)dt \\
= & \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{2\sin(n^{-\epsilon}t)}{t} f_n(t)dt \\
= & \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{2\sin(n^{-\epsilon}t)}{t} [f(\frac{t}{n^{1/\alpha}})]^n dt \\
\sim^{\dagger} & \frac{2}{\pi} n^{-\epsilon} \int_0^{\infty} \exp(-c_0|t|^{\alpha} L(n^{1/\alpha}) \cdot (1 - i\beta \mathrm{sign}(t)\tan(\frac{\pi\alpha}{2}))(1+o(1)))dt \\
\sim & \frac{2}{\pi} n^{-\epsilon} \int_0^{\infty} \exp(-c_0 t^{\alpha} L(n^{1/\alpha})(1+o(1))) \cos(c_1 t^{\alpha} L(n^{1/\alpha})(1+o(1)))dt \\
\sim & n^{-\epsilon} c_0^{-1/\alpha} L(n^{1/\alpha})^{-1/\alpha} \frac{2}{\pi\alpha} \int_0^{\infty} \exp(-u) u^{\frac{1}{\alpha}-1} \cos(c_1 u) du \\
= & n^{-\epsilon} c_0^{-1/\alpha} L(n^{1/\alpha})^{-1/\alpha} \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha} \cdot \left(1 + \beta^2 \tan^2(\pi\alpha/2)\right)^{-\frac{1}{2\alpha}} \\
\cdot & \cos\left(\frac{1}{\alpha}\tan^{-1}(|\beta|\tan(\pi\alpha/2))\right) \\
\overset{\mathrm{def}}{=} & C n^{-\epsilon} c_0^{-1/\alpha} L(n^{1/\alpha})^{-1/\alpha} \\
\leq & \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha} n^{-\epsilon} c_0^{-1/\alpha} L(n^{1/\alpha})^{-1/\alpha},
\end{array}
$$

---

*The inversion formula is used here. $f_n(t)$ is the characteristic function of $S_n/n^{1/\alpha}$.

†If a distribution F is attracted to an α-stable law, then its characteristic function is of the form: $f(\frac{t}{n^{1/\alpha}}) = \exp\{-c_0(|t|/n^{1/\alpha})^{\alpha} L(n^{1/\alpha}/|t|) \cdot (1 - i\beta \mathrm{sign}(t)\tan(\frac{\pi\alpha}{2})) + o((|t|/n^{1/\alpha} L(n^{1/\alpha}/|t|))^{-1})\}$, if $\alpha \neq 1$ and F is centred (see [16] theorem 2.6.5). If $\alpha = 1$, the imaginary part of the characteristic is different (see [1] theorem 2).

where $c_0 = \frac{\Gamma(2-\alpha)}{|\alpha-1|}|\cos\frac{\pi\alpha}{2}|$, $c_1 = |\beta|\tan(\frac{\pi\alpha}{2})$, and

$$C = \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha} \cdot \left(1+\beta^2\tan^2(\pi\alpha/2)\right)^{-\frac{1}{2\alpha}} \cdot \cos\left(\frac{1}{\alpha}\tan^{-1}(|\beta|\tan(\pi\alpha/2))\right).$$

If $\beta = 0$, then

$$C = \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha}$$

□

Note that for small sample sizes the slowly varying term in the previous proof can be relatively large, since we have only a polynomial large deviation principle. Hence we use the following large deviation probabilities to construct test to some given hypotheses:

Let $L(x)$ be a slowly varying function such that in Karamata's representation the function ε satisfies

$$\varepsilon(t) = o(\frac{1}{\log(t)}), (t \to \infty).$$

In this case for any $\epsilon > 0$, as $n \to \infty$,

$$\begin{aligned}
P\{\widehat{\alpha_n^{-1}} - \alpha^{-1} - \frac{\log(h(n))}{\log(n)} > \epsilon\} &= P\{\frac{|S_n|}{a_n} > \exp(\epsilon\log(n))\} \\
&\overset{lemma 3.1.2}{\sim} nP\{\frac{|X|}{a_n} > \exp(\epsilon\log(n))\} \\
&= nP\{|X| > a_n\exp(\epsilon\log(n))\} \\
&\sim \frac{nL(a_n n^\epsilon)}{a_n^\alpha n^{\alpha\epsilon}} \\
&= n^{-\alpha\epsilon}\frac{nL(a_n n^\epsilon)}{nL(a_n)} \\
&= n^{-\alpha\epsilon}\frac{nL(n^{\frac{1}{\alpha}+\epsilon}h(n))}{nL(n^{\frac{1}{\alpha}}h(n))} \\
&\overset{*}{\sim} n^{-\alpha\epsilon}.
\end{aligned}$$

Here the relation (*) holds, because

$$\frac{L(n^{\frac{1}{\alpha}+\epsilon}h(n))}{L(n^{\frac{1}{\alpha}}h(n))} \sim \exp(\int_{n^{\frac{1}{\alpha}}h(n)}^{L(n^{\frac{1}{\alpha}+\epsilon}h(n))}\frac{\varepsilon(t)}{t}dt)$$

$$
\begin{aligned}
&= \exp(\int_{\frac{1}{\alpha}+\frac{\log(h(n))}{\log(n)}}^{\frac{1}{\alpha}+\epsilon+\frac{\log(h(n))}{\log(n)}} \log(n)\varepsilon(n^z)dz) \\
&(\quad set \quad t = n^z) \\
&\sim \exp(\log(n)\int_{\frac{1}{\alpha}}^{\frac{1}{\alpha}+\epsilon} \varepsilon(n^z)dz) \\
&= \exp(\epsilon\varepsilon(n^{\frac{1}{\alpha}}+\epsilon\delta)\log(n)) \\
&(\quad \delta \in [0,1]) \\
&\to \exp(0) = 1.
\end{aligned}
$$

The other probability can be calculated as follows:

$$
\begin{aligned}
P \quad & \{\widehat{\alpha_n^{-1}} - \alpha^{-1} - \frac{\log(h(n))}{\log(n)} < -\epsilon\} = P\{\frac{|S_n|}{n^{\frac{1}{\alpha}}h(n)} < n^{-\epsilon}\} \\
= \quad & \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin(n^{-\epsilon}t)}{t}[f(|t|/n^{1/\alpha}h(n))]^n dt \\
\sim \quad & \frac{1}{\pi}n^{-\epsilon}\int_{-\infty}^{\infty} \exp(-c_0|t|^\alpha \frac{L(n^{1/\alpha}h(n))}{h^\alpha(n)}(1-i\beta sign(t)\tan(\frac{\pi\alpha}{2}))(1+o(1)))dt \\
\sim \quad & \frac{1}{\pi}n^{-\epsilon}\int_{-\infty}^{\infty} \exp(-|t|^\alpha(1-i\beta \text{ sign}(t)\tan(\frac{\pi\alpha}{2})))dt \\
= \quad & n^{-\epsilon}\frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha}\cdot\left(1+\beta^2\tan^2(\pi\alpha/2)\right)^{-\frac{1}{2\alpha}} \\
\cdot \quad & \cos\left(\frac{1}{\alpha}\tan^{-1}(|\beta|\tan(\pi\alpha/2))\right) \\
:= \quad & Cn^{-\epsilon} \\
\leq \quad & \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha}n^{-\epsilon}.
\end{aligned}
$$

where $c_0 = \frac{\Gamma(2-\alpha)}{|\alpha-1|}|\cos\frac{\pi\alpha}{2}|$, and

$$
C = \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha}\cdot\left(1+\beta^2\tan^2(\pi\alpha/2)\right)^{-\frac{1}{2\alpha}}\cdot\cos\left(\frac{1}{\alpha}\tan^{-1}(|\beta|\tan(\pi\alpha/2))\right).
$$

And if $\beta = 0$, then

$$
C = \frac{2\Gamma(\frac{1}{\alpha})}{\pi\alpha}
$$

The large deviation probabilities can be used to construct test procedure for hypotheses on α. Hence it is of importance for the practical purposes.

For example, suppose that the following hypothesis is to be tested:

$$
H_0 : \alpha = \alpha_0,\ H_1 : \alpha \neq \alpha_0.
$$

Theorem 3.1 provides the following test with the confidence level A:

Reject H_0, if $\widehat{\alpha_n^{-1}} > \alpha_0^{-1} + \frac{\log h(n)}{\log n} + \epsilon_+$, or $\widehat{\alpha_n^{-1}} < \alpha_0^{-1} + \frac{\log h(n)}{\log n} - \epsilon_-$,

where ϵ_+ satisfies $n^{-\alpha_0 \epsilon_+} \leq A/2$ and ϵ_- satisfies $\frac{2\Gamma(\frac{1}{\alpha_0})}{\pi \alpha_0} n^{-\epsilon_-} \approx A/2$; otherwise we say there is no significant differences between α_0 and the stable index of the the observed population. Note that the rejection region depends partly on the slowly varying function. In the situation where the slowly varying function is unknown and the sample size is large, we can approximately omit it. But at this time, we should take a smaller rejection region.

3.2 Estimating the stable index using estimator with U-statistic structure

3.2.1 U-statistic as an estimator

It is well known that U-statistics have usually better statistical properties when used to estimate an unknown parameter. Based on the statistic we have used to estimate the index of an attracting law, we can construct the following estimator of α^{-1} with U-statistic structure.

Suppose now, that $X_1, X_2, ..., X_n$ are i.i.d. copies from X with a distribution F being attracted to an α-stable law S. Let $m < n$, and denote by h the function

$$h(x_1, x_2, ..., x_m) = \frac{\log |\sum_{i=1}^m x_i|}{\log m}. \tag{3.2}$$

h is considered as the kernel of a U-statistic defined by

$$\widehat{\alpha_U^{-1}} = U_n(h) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} h(X_{i_1}, X_{i_2}, ..., X_{i_m}). \tag{3.3}$$

This is an estimator of α^{-1}.

By the discussion above we know that when m is large enough,

$$Eh(X_1, X_2, ..., X_m)^k < \infty \tag{3.4}$$

for every $k \in \mathcal{N}$. In fact, if $m \to \infty$, we have (see [37], theorem 2)

$$E(h(X_1, X_2, ..., X_m) - \frac{1}{\alpha}) \to 0, \tag{3.5}$$

$$E(h(X_1, X_2, ..., X_m) - \frac{1}{\alpha})^2 \to 0 \tag{3.6}$$

that is, the expectation of h tends to α^{-1} and the mean square error tends to 0.

The distribution of $\widehat{\alpha_U^{-1}}$ here is difficult to obtain, its asymptotic variance also. A very good calculation method to overcome this difficulty is jackknifing the variance([2]). Let U_n^0 denote the U-statistic based on the observations $X_1, X_2, ..., X_n$ and U_{n-1}^i the U-statistic based on all observations except X_i, that is $X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_n$. Let $\overline{U_{n-1}}$ denote the average of the U_{n-1}^i, $i = 1, 2, ..., n$. It is known that

$$S_n^2 := (n-1)^{-1} \sum_{i=1}^{n} \left(U_{n-1}^i - \overline{U_{n-1}}\right)^2 \xrightarrow{p} m^2\zeta_1, \tag{3.7}$$

where $\zeta_1 = Var(E(h(X_1, X_2, \cdots, X_m)|X_1) - Eh(X_1, X_2, \cdots, X_m))$

In this way we obtain the consistency of $\widehat{\alpha_U^{-1}}$ as an estimator of α^{-1} and also its asymptotic normality :

Theorem 3.2.1. Suppose $X_1, X_2, ..., X_n$ are i.i.d. observations with distribution F, which is attracted to some α-stable law, and let $\widehat{\alpha_U^{-1}}$ be defined by (3.3). As an estimator of α^{-1}, $\widehat{\alpha_U^{-1}}$ is consistent, when $m \to \infty$, $m < n$. In addition, if m is fixed, we have also

$$\sqrt{n} \cdot S_n^{-1} \left(\widehat{\alpha_U^{-1}} - E\widehat{\alpha_U^{-1}}\right) \Rightarrow N(0,1), \text{ as } n \to \infty. \tag{3.8}$$

Proof. The consistency of $\widehat{\alpha_U^{-1}}$ follows directly from the a.s.convergence of U-statistics and the consistency of $\widehat{\alpha^{-1}}$.

We have obviously, $\zeta_m < \infty$, so by the asymptotic normality of U-statistics, it holds, as $n \to \infty$, m is fixed,

$$\sqrt{n} \left(U_n(h) - E\widehat{\alpha_U^{-1}}\right) \Rightarrow N(0, m^2\zeta_1),$$

that is

$$\frac{\sqrt{n}}{m\zeta_1^{\frac{1}{2}}} \cdot \left(U_n(h) - E\widehat{\alpha_U^{-1}}\right) \Rightarrow N(0,1). \tag{3.9}$$

Moreover, by (3.7)

$$\sqrt{n} \left(S_n^{-1} - \frac{1}{m\zeta_1^{\frac{1}{2}}}\right) \cdot \left(U_n(h) - E\widehat{\alpha_U^{-1}}\right) \sim -\frac{(S_n^2 - m^2\zeta_1)}{2m^2\zeta_1} \cdot \frac{\sqrt{n}}{m\zeta_1^{\frac{1}{2}}} \cdot \left(U_n(h) - E\widehat{\alpha_U^{-1}}\right),$$

hence from (3.7) and (3.9) we get the result. ∎

Note that the conclusion in the above theorem can also be described in the following way: $X_1, X_2, \ldots, X_n$ are i.i.d. copies of X in the domain of attraction of stable law, then we have

$$\sqrt{n} \cdot S_n^{-1} \left(\widehat{\alpha_U^{-1}} - \frac{1}{\alpha} + c_m \right) \stackrel{d}{\Rightarrow} N(0,1), \text{ as } n \to \infty. \tag{3.10}$$

where $c_m \to 0$ as $m \to \infty$. That is $\widehat{\alpha_U^{-1}}$ is an asymptotically normal estimator of α^{-1} with very small bias when the sample size is large.

Note The asymptotic normality of $\widehat{\alpha_U^{-1}}$ holds only when m is fixed or $m = o(n^{1/2})$; Otherwise its asymptotic limit can be some Wiener-Ito stochastic integral with respect to some Wiener measure or Poisson measure. It is a U-statistic with infinite degree m. The convergence is seriously affected by the growth rate of m (see Rempala and Gupta [42]). In fact, we have the following theorem:

Theorem 3.2.2. Suppose $X_1, X_2, ..., X_n$ are i.i.d. observations with the same distribution F, which is attracted to an α-stable law , and let $\widehat{\alpha_U^{-1}}$ be defined by (3.3). If $m \to \infty$ and $m = o(n^{1/2})$, as $n \to \infty$, then

$$\sqrt{n} \cdot S_n^{-1} \left(\widehat{\alpha_U^{-1}} - E\widehat{\alpha_U^{-1}} \right) \Rightarrow N(0,1), \text{ as } n \to \infty. \tag{3.11}$$

Proof. Let
$g_c(X_1, X_2, ..., X_c) = E(h(X_1, X_2, ..., X_m)|X_1, X_2, ..., X_c) - Eh(X_1, X_2, ..., X_m)$, $c = 1, 2, ..., m$. Hoeffding's decomposition of the U-statistic $\widehat{\alpha_U^{-1}}$ has the form

$$\widehat{\alpha_U^{-1}} - Eh(X_1, X_2, ..., X_m) = \frac{m}{n} \sum_{i=1}^{n} g_1(X_i) + R_n.$$

The main part of the proof is to verify that the variance of the remainder term R_n is of smaller order compared with the first sum.

As we know, if X has a distribution in the domain of attraction of an α-stable law, then $X^2 \text{sign}(X)$ has a distribution attracted to an $\alpha/2$-stable law. So that without loss of generality, we can suppose that the stable index α is less than 1. The kernel of the U-statistic can be written in the form:

$$\begin{aligned} h(X_1, X_2, ..., X_m) = & \frac{\log(m-1)}{\alpha \log m} + \frac{\log l(m-1)}{\log m} + \frac{\log |Y_{m-1}|}{\log m} \\ & + \frac{\log |1 + \frac{X_1}{Y_{m-1}(m-1)^{1/\alpha} l(m-1)})|}{\log m}, \end{aligned}$$

where $Y_{m-1} = \frac{\sum_{i=2}^{m} Xi}{(m-1)^{1/\alpha} l(m-1)}$ is independent with X_1.

Hence

$$
\begin{aligned}
\zeta_1 =& Var(g_1(X_1)) \\
=& \frac{Var(E(\log|1+\frac{X_1}{Y_{m-1}(m-1)^{1/\alpha}l(m-1)}|\,|X_1))}{(\log m)^2} \\
:=& \frac{VarZ}{(\log m)^2}.
\end{aligned}
$$

While

$$EZ \approx \frac{1}{\alpha m},$$

using the following inequality we have,

$$
\begin{aligned}
EZ^2 \geq & m^2|EZ|^2 P(|Z| \geq m|EZ|) \\
\approx & \frac{1}{\alpha^2}P(|Z| \geq \frac{1}{\alpha}) \\
= & \frac{1}{\alpha^2}P(\frac{X_1}{Y_{m-1}} \geq (m-1)^{1/\alpha}l(m-1)(e^{1/\alpha}-1)) \\
+ & \frac{1}{\alpha^2}P(\frac{X_1}{Y_{m-1}} \leq -(m-1)^{1/\alpha}l(m-1)(1+e^{1/\alpha})) \\
+ & \frac{1}{\alpha^2}P(\frac{X_1}{Y_{m-1}} \leq \ -(m-1)^{1/\alpha}l(m-1)(1-e^{-1/\alpha})) \\
- & \frac{1}{\alpha^2}P(\frac{X_1}{Y_{m-1}} \leq \ -(m-1)^{1/\alpha}l(m-1)(1+e^{-1/\alpha})) \\
:= & I_1 + I_2 + I_3 - I_4.
\end{aligned}
$$

Now we calculate I_1, the other terms are in the same order as I_1.

$$
\begin{aligned}
\alpha^2 I_1 := & P(\frac{X_1}{Y_{m-1}} \geq \ A_m) \\
= & \int_0^{+\infty} P(X_1 \geq A_m s) f_{Y_{m-1}}(s)ds + \int_{-\infty}^0 P(X_1 \leq A_m s) f_{Y_{m-1}}(s)ds \\
\geq & \int_\epsilon^{+\infty} P(X_1 \geq A_m s) f_{Y_{m-1}}(s)ds + \int_{-\infty}^{-\epsilon} P(X_1 \leq A_m s) f_{Y_{m-1}}(s)ds \\
\sim & \frac{c_1}{m-1}\int_\epsilon^{+\infty} s^{-\alpha} f(s)ds + \frac{c_2}{m-1}\int_{-\infty}^{-\epsilon} |s|^{-\alpha} f(s)ds \\
\sim^{\dagger} & \frac{1}{m}\cdot d_1,
\end{aligned}
$$

where $A_m = (m-1)^{1/\alpha}l(m-1)\cdot(e^{1/\alpha}-1)$, $Y_{m-1} = \frac{\sum_{i=2}^m X_i}{(m-1)^{1/\alpha}l(m-1)}$, l is slowly

† In this approximation we have used the theorem of uniforme convergence of regularly varying functions with negative index ([5] pp23, theorem 1.5.2) and the property of the normalizing constant of the partial sum. The convergence of the last step is because that s^{-1} is bounded on integration intervals. The integrals are finite because $\alpha < 1$.

varying at infinity, $f_{Y_{m-1}}$ is the density of Y_{m-1}, f is the density of a stable distribution, c_1, c_2, d_1 are positive constants.

$$\begin{aligned}\zeta_1 &= \frac{Var(Z)}{(\log m)^2} \\ &\approx \frac{d}{m(\log m)^2},\end{aligned}$$

where d is some positive constant.

Furthermore, following the method of [37] theorem 2, we have

$$\zeta_m \approx \frac{Var(\log|Y|)}{(\log m)^2},$$

where Y is an α-stable random variable, and by the theory of U-statistics we have

$$\zeta_1 \leq \frac{\zeta_2}{2} \leq \frac{\zeta_3}{3} \leq \ldots \leq \frac{\zeta_m}{m},$$

whence we obtain that

$$\begin{aligned}Var(R_n) =& \sum_{c=2}^{m} \binom{m}{c}^2 \binom{n}{c}^{-1} \zeta_c \\ =& \frac{m^2}{n} Var(g_1(X_1)) \cdot O(m^2/n).\end{aligned}$$

The condition $m = o(n^{1/2})$ and the Jackknifing estimate of $m^2\zeta_1$ complete the proof. $\square$

Note: Using U-statistics to estimate the stable index does not reduce the bias of the estimator, but the variance becomes smaller. Our estimator is improved in the sense that it possesses smaller mean square error (MSE).

3.2.2 Estimating the stable index using incomplete U-statistics

The estimator of a stable index with U-statistics structure has very nice properties. However, if the sample size n or the degree m is large, the calculation of $\widehat{\alpha_U^{-1}}$ can be quite elaborate, since it involves averaging $\binom{n}{m}$ terms. Because of the dependence of the involved summands, we can omit some summands of $\widehat{\alpha_U^{-1}}$ without unduly inflating the variance. We are thus led to consider a special form of incomplete U-statistics.

Let the assumptions be the same as in the previous section.
Suppose that the sample size has the form $n = km$, otherwise we delete some terms. The i.i.d. observations $X_1, X_2, ..., X_n$ are partitioned into k independent sub-samples, each having a sample size m. From each of the sub-samples we get one estimator of the stable index, $\widehat{\alpha_{mj}^{-1}}$, $j = 1, 2, ..., k$, and the average of these estimators is our incomplete U-statistic

$$\widehat{\alpha_{U0}^{-1}} = \frac{1}{k}\sum_{j=1}^{k}\widehat{\alpha_{mj}^{-1}}.$$

In this case, the variance of $\widehat{\alpha_{U0}^{-1}}$ is

$$Var(\widehat{\alpha_{U0}^{-1}}) = \frac{1}{k}Var(\widehat{\alpha_{mj}^{-1}}),\ j = 1, 2, ..., k.$$

Comparing it with the variance that of the complete U-statistic, we get its asymptotic relative efficiency (ARE) as:

$$ARE = \lim_{n\to\infty} Var(\widehat{\alpha_U^{-1}})/Var(\widehat{\alpha_{U0}^{-1}}) = m\zeta_1/\zeta_m,$$

for given m. By lemma 1.1.4 of [30] we have

$$m\zeta_1/\zeta_m < 1.$$

So the ARE depends on the ratio ζ_1/ζ_m and may be close to 1.

The asymptotic normality of the estimator can be obtained directly from the properties of U-statistics. For practical use, we have also the following result:

Theorem 3.2.3. Let $\widehat{\alpha_{U0}^{-1}}$ and $\widehat{\alpha_{mj}^{-1}}$ $\ j = 1, 2, ..., k$ be defined as above, and

$$s_k^2 = \frac{1}{k-1}\sum_{i=1}^{k}\left(\widehat{\alpha_{mj}^{-1}} - \widehat{\alpha_{U0}^{-1}}\right)^2.$$

Then if $n \to \infty$, and m is fixed, we have the following normal approximation:

$$\sqrt{k}\frac{\widehat{\alpha_{U0}^{-1}} - E\widehat{\alpha_{U0}^{-1}}}{s_k} \Rightarrow N(0, 1). \tag{3.12}$$

Proof. The proof of the theorem is direct. From the partition of the sample we have that the sub-estimators $\widehat{\alpha_{mj}^{-1}}$, $\ j = 1, 2, ..., k$ are independent and identically distributed random variables, and $\widehat{\alpha_{U0}^{-1}}$ is exactly the sample mean.

The expectations of $\widehat{\alpha_{U0}^{-1}}$ and $\widehat{\alpha_{mj}^{-1}}$, $\ j = 1, 2, ..., k$ are all equal to each other. The result follows now from routine calculations. □

Note that the exact distribution of $\widehat{\alpha_{U0}^{-1}}$ is almost impossible to obtain, so we cannot construct the exact confidence interval of α^{-1}. But from the above theorem we can asymptotically construct tests for hypotheses on α^{-1} by constructing asymptotic confidence intervals.

Our incomplete U-statistic is obtained by choosing k disjoint subsets of the sample successively. We can use other methods to select subsamples. For example, it is possible to choose the subsets at random with replacement. This is indeed a type of bootstrap. Suppose we have selected L subsets, or subsamples say, each having sample size m, at random with replacement. Then the estimator with incomplete U-statistic structure has the variance

$$Var(\widehat{\alpha_{U0}^{-1}}) = \zeta_m/L + (1 - L^{-1})Var(\widehat{\alpha_U^{-1}}).$$

Hence, repeating the resampling procedure infinitely often, we obtain a variance asymptotically very close to that of the complete U-statistic.

The random resampling procedure without replacement has similar properties. Furthermore, the resampling procedure can also be designed to get smallest variance for a given subsample size m.

From the discussion above we know, that the degree m of the underlying U-statistic can be freely chosen in the range of the sample size. But for different m, the corresponding U-statistics have different variances. Hence it is necessary to optimize the choice of m.

We discuss the optimization problem of our estimator $\widehat{\alpha_{U0}^{-1}}$ with Incomplete U-statistic structure.

We start with the simplest case where the parent distribution is stable. According to the properties of stable random variables, we know that

$$\begin{aligned}
\widehat{\alpha_{U0}^{-1}} - \alpha^{-1} =& \frac{1}{k}\sum_{j=1}^{k}(\widehat{\alpha_{mj}^{-1}} - \alpha^{-1}) \\
\stackrel{\mathrm{d}}{=}& \frac{1}{k}\sum_{j=1}^{k}\frac{\log|X_j|}{\log m}.
\end{aligned}$$

Its mean square error can be calculated as:

$$\begin{aligned}
MSE(\widehat{\alpha_{U0}^{-1}}) =& E\left(\frac{1}{k}\sum_{j=1}^{k}\frac{\log|X_j|}{\log m}\right)^2 \\
=& \frac{1}{k}\frac{E(\log|X_1|)^2}{(\log m)^2} + \frac{k-1}{k}\frac{(E\log|X_1|)^2}{(\log m)^2}
\end{aligned}$$

$$=^{\dagger} \frac{\tau^2}{(\log m)^2} + \frac{1}{k(\log m)^2} \frac{\pi^2}{12} (\frac{2}{\alpha^2} + 1),$$

where $X_1, X_2, ..., X_n$ are i.i.d. random variables with strictly stable distribution, and $\tau = E \log |X_1|$. In the case of a standard stable distribution i.e. the scale parameter $\sigma = 1$, then $\tau = C(1/\alpha - 1)$, here $C = 0.5772$ is the Euler constant.

Now we can start the optimization procedure with a given initial value of α_0^{-1}. The optimal value of m is that minimizing the mean square error of $\widehat{\alpha_{U0}^{-1}}$. This can be carried out by iteration.

In case that the parent distribution is attracted to a stable law, we have to consider the slowly varying function in addition. We suggest a similar procedure as in the case of stable distribution, but we need further information about the slowly varying function.

3.2.3 Estimating the stable index by resampling

The estimator of α^{-1} with the structure of U-statistic can be treated as a special kind of resampling from the known observations. In the following we give a randomized resampling procedure.

Suppose that $X_1, X_2, ..., X_n$ are i.i.d. observations from some population attracted to an α-stable law, and let $Y_1, Y_2, ..., Y_n$ be i.i.d. copies of a Bernoulli distributed parent random variable Y with parameter p, independent of X. Define a new random sequence of $Z_i, i = 1, 2, ..., n$ by

$$Z_i = X_i \cdot Y_i,$$

then, as $z \to \infty$,

$$\begin{aligned} P\{|Z_i| > z\} &= P\{|X_i| \cdot Y_i > z\} \\ &= P\{|X_i| > z\} \cdot p \\ &\sim p \cdot z^{-\alpha} L(z). \end{aligned}$$

Hence we have the normalizing constant

$$a_n = (np)^{\frac{1}{\alpha}} l(n),$$

for the convergence of the partial sum of $Z_i, i = 1, 2, ..., n$, where $l(n)$ is a slowly varying function.

Considering the attraction property for the re-sampled observations, we have

$$\frac{\log |\sum_{i=1}^n Z_i|}{\log np} \to \frac{1}{\alpha},$$

†See Zolotarev [52] chapter 4. We assume further the underlying stable distribution is symmetric.

as $n \to \infty$.

Furthermore, by the strong law of large numbers, we have also

$$\frac{\sum_{i=1}^{n} Y_i}{n} \xrightarrow{a.s.} p.$$

Summarizing we obtain the following theorem:

Theorem 3.2.4. Suppose $X_i, i = 1, 2, ..., n$, are i.i.d. copies of $X \sim F$, which is attracted to some α-stable law and let $Y_i, i = 1, 2, ..., n$ be observations from a Bernoulli distribution with parameter p. X and Y are assumed to be independent. Then

$$\widehat{\alpha_{r1}^{-1}} = \frac{\log |\sum_{i=1}^{n} X_i \cdot Y_i|}{\log \sum_{i=1}^{n} Y_i}$$

is a consistent estimator of α^{-1}.

Suppose further that we have k i.i.d. Bernoulli samples of size n, $Y_{ij}, i = 1, \dots, n,\ j = 1, \dots, k$. From every Bernoulli sample we get a resample from that in the domain of stable attraction. Then we have k estimates of α^{-1} from the re-sampled observations. The average of the estimates is our re-sampled estimator:

$$\widehat{\alpha_R^{-1}} = \frac{1}{k} \sum_{j=1}^{k} \widehat{\alpha_{rj}^{-1}},$$

where

$$\widehat{\alpha_{rj}^{-1}} = \frac{\log |\sum_{i=1}^{n} X_i \cdot Y_{ij}|}{\log \sum_{i=1}^{n} Y_{ij}}$$

The above theorem is derived from $\widehat{\alpha_n^{-1}}$ by a minor change. Note however, that it opens another possibility to do simulations for U-statistic in a simpler way. We can also omit the upper and lower extremal values of the repeated estimates of this procedure to get a robust estimate of the stable index. Using this method, we get a good estimate of the unknown parameter in a simple way which may be close to the complete U-statistic. This method is simpler than that of the incomplete U-statistic, which takes re-samples at random without replacement.

$\widehat{\alpha_R^{-1}}$ is a consistent estimator of α^{-1} and also has the property of asymptotic normality.

3.3 Estimating stable index in the case of dependence

Now we consider dependent sequences of random variables.

Theorem 3.3.1. Let $(X_n, n \geq 1)$ be a strictly stationary sequence of random variables such that,

$$P\{|X_1| > x\} = x^{-\alpha}L(x) \quad (x > 0),$$

where $0 < \alpha < 2$ and $L(x)$ is a slowly varying function with Karamata's representation such that

$$\varepsilon(t) = o(\frac{1}{\log(t)}).$$

Define b_n by the following relation,

$$\lim_{n\to\infty} nP\{|X_1| > b_n\} = 1.$$

Suppose that $(X_n, n \geq 1)$ is exponentially ψ-mixing ($i.e.$ $\psi(n) \leq K\eta^n$, $n = 1, 2, ...$ for some $K > 0$, and $0 < \eta < 1$) and such that

$$\psi(1) < \infty.$$

Then, for all $x_n = O(n^d)$, where $d \in R^+$, as $n \to \infty$,

$$x_n^\alpha |P(\frac{S_n}{b_n} \in x_n A) - nP(\frac{X_1}{b_n} \in x_n A)| \to 0,$$

for all $A \in \mathcal{B}$, where $\mathcal{B}$ is Borel σ-algebra, $(0 \bar{\in} \bar{A})$.

In particular,

$$nP(\frac{X_1}{b_n} \in A) \to \nu(A)$$

implies

$$x_n^\alpha P(\frac{S_n}{b_n} \in x_n A) \to \nu(A)$$

as $n \to \infty$.

Proof. Without loss of generality, we can suppose that $A = (1, \infty)$,
then we have

$$\begin{aligned}
& x_n^\alpha \sum_{1\leq i<j\leq n} P(X_i > b_n x_n, X_j > b_n x_n) \\
\leq \;& x_n^\alpha \sum_{1\leq i\leq n} (n-i)(K\eta^i + 1)P(X_1 > b_n x_n)^2 \\
\leq \;& x_n^\alpha \frac{n(n-1)}{2}(1+K)(x_n^{-\alpha} b_n^{-\alpha} L(x_n b_n))^2 \\
= \;& x_n^{-\alpha}(1+K)\frac{n(n+1)}{2n^2}(\frac{L(x_n b_n)}{L(b_n)})^2
\end{aligned}$$

$$\sim \quad \frac{1+K}{2} x_n^{-\alpha} \to 0,$$

where

$$\begin{aligned}
\frac{L(x_n b_n)}{L(b_n)} &= \exp(\int_{b_n}^{x_n b_n} \frac{\varepsilon(t)}{t} dt) \\
&= \exp(\int_{\frac{1}{\alpha}\log(n)+\log(h(n))}^{\frac{1}{\alpha}\log(n)+\log(h(n))+\log(x_n)} \varepsilon(\exp(z))dz) \\
&\sim \exp(\int_{\frac{1}{\alpha}}^{\frac{1}{\alpha}+\frac{\log(x_n)}{\log(n)}} \varepsilon(\exp(x\log(n)))\log(n)dx) \\
&= \exp(\log(x_n)\varepsilon(n^\delta)) \\
&(\quad \frac{1}{\alpha} < \delta < \frac{1}{\alpha} + \frac{\log(x_n)}{\log(n)}) \\
&= \exp(\log(x_n) o(\frac{1}{\delta\log(n)})) \\
&= \exp(0(\frac{\log(x_n)}{\delta\log(n)})) = 1
\end{aligned}$$

Repeating the proof in [4], we have,

$$\begin{aligned}
& x_n^\alpha |P(\frac{S_n}{b_n} > x_n) - nP(\frac{X_1}{b_n} > x_n)| \\
\leq \quad & x_n^\alpha \sum_{1\leq i<j\leq n} P(X_i > b_n x_n, X_j > b_n x_n) \\
\to \quad & 0.
\end{aligned}$$

□

From theorem 3.3.1 we have the following corollary:

Corollary 3.3.2. Under the conditions of theorem 3.3.1, we have that the sequence $\frac{S_n}{b_n}, n \geq 1$ converges weakly to a strictly α-stable law μ, whose Lévy measure has density

$$f(x) = |x|^{-1-\alpha}(c\mathbb{1}_{\{x>0\}} + (1-c)\mathbb{1}_{\{x<0\}}),$$

and where $b_n = n^{\frac{1}{\alpha}} h(n)$.

Similar to the independent case, we have the following theorem.

Theorem 3.3.3. Under the assumptions of theorem 3.3.1,

$$\widehat{\alpha_n^{-1}} = \frac{\log(|\sum_{i=1}^{n} X_i|)}{\log(n)}$$

is a consistent estimator of α^{-1}.

Proof. The proof follows directly from theorem 3.3.1 and some simple calculations. □

Note that the hypotheses testing can be constructed in the similar way to the independent case.

3.4 Estimating the stable index using the maximum

It is well known that the statistical properties of the sum of a sample from some law attracted to a stable distribution is seriously influenced by its maximum. From Heyde's lemma 3.1.2 we know that we can do some inferences based on the maximum.

Suppose that $X_1, X_2, ..., X_n, ...$, are i.i.d. random variables having distribution function F, and F belongs to the domain of attraction of an α-stable law. Define ,

$$M_n = \max_{1 \le k \le n} \{|X_k|\},$$

then we have the following results:

Theorem 3.4.1. Let $X_1,\ X_2,\ \cdots,\ X_n$ be i.i.d. random variables with distribution F, which is attracted to an α-stable law. Then

$$\widehat{\alpha_m^{-1}} = \frac{\log(M_n)}{\log(n)}$$

is a consistent estimator of α^{-1}.

The large deviation probabilities of $\widehat{\alpha_m^{-1}}$ are as follows,

Theorem 3.4.2. Under the same conditions as theorem 3.4.1, for any given $\epsilon > 0$,

$$P\{\widehat{\alpha_m^{-1}} - \alpha^{-1} - \frac{\log h(n)}{\log n} < -\epsilon\} \sim \exp(-n^{\alpha\epsilon}), \quad (n \to \infty).$$

Furthermore, if in Karamata's representation of $L(x)$

$$\varepsilon(x) = o(\frac{1}{\log(x)}), \quad (x \to \infty),$$

then,

$$P\{\widehat{\alpha_m^{-1}} - \alpha^{-1} - \frac{\log h(n)}{\log n} > \epsilon\} \sim n^{-\alpha\epsilon}, \quad (n \to \infty).$$

Proof. The first part of the theorem follows from the uniform convergence of the extremal distribution in finite intervals.
The second part follows immediately from lemma 3.1.2 and the proof of theorem 3.1.1. □

In fact, we have also the approximate distribution of $\widehat{\alpha_m^{-1}}$:

Theorem 3.4.3. If in Karamata's representation of the slowly varying function $L(\cdot)$ holds that

$$\varepsilon(t) = o(\frac{1}{\log(t)}) \quad (t \to \infty),$$

then for any given y, we have the following approximation,

$$P\left\{\widehat{\alpha_m^{-1}} - \alpha^{-1} - \frac{\log h(n)}{\log n} < y\right\} \sim \exp\left\{-n^{-\alpha y}\right\}, \tag{3.13}$$

as $n \to \infty$.

Proof. For any given $y < 0$, we have,by the uniform convergence of extremal distribution in a finite interval,

$$\begin{aligned} P\left\{\widehat{\alpha_m^{-1}} - \alpha^{-1} - \frac{\log h(n)}{\log n} < y\right\} =& P\left\{\widehat{\alpha_m^{-1}} - \alpha^{-1} < y + \frac{\log(h(n))}{\log(n)}\right\} \\ =& P\left\{\frac{M_n}{a_n} < n^y\right\} \\ \sim& \exp\left\{-n^{-\alpha y}\right\}. \end{aligned}$$

as $n \to \infty$, where $h(n)$ is a slowly varying function defined by the normalizing constant $a_n = n^{1/\alpha}h(n)$.

By lemma 3.1.2 and the proof of Theorem 3.1.1, we have, for $y > 0$,

$$\begin{aligned} P\left\{\widehat{\alpha_m^{-1}} - \alpha^{-1} - \frac{\log h(n)}{\log n} < y\right\} =& P\left\{\widehat{\alpha_m^{-1}} - \alpha^{-1} < y + \frac{\log(h(n))}{\log(n)}\right\} \\ =& P\left\{\frac{M_n}{a_n} < n^y\right\} \\ \sim& 1 - n^{-\alpha y} \\ \sim& \exp\left\{-n^{-\alpha y}\right\}. \end{aligned}$$

□

From the above theorem we can also construct a test of the hypothesis in section 3.1.

3.4.1 Pareto distribution as an example

In the case of a Pareto distribution with parameter $\alpha > 0$, we know from the theory of extreme values, that the suitable normalized maximum of the observations has an extremal asymptotic distribution.

Suppose now the observations $X_1, X_2, \cdots, X_n$ are i.i.d. copies of a Pareto distribution with parameters $\alpha > 0$ and $\sigma \in R^+$, that is the distribution has the form:

$$F(x) = 1 - \left(\frac{\sigma}{x}\right)^{\alpha}, \, x > \sigma,$$

then from Pfeifer[39] we have that, for $x > 0$,

$$P\left\{\frac{M_n}{n^{1/\alpha}} < x\right\} \longrightarrow \exp(-x^{-\alpha}), \; (n \to \infty)$$

where$M_n = \max\{X_1, X_2, \cdots, X_n\}$, is the maximum or the largest order statistics of $X_1, X_2, \cdots, X_n$.

From here we have in fact,

$$\begin{aligned} &P\left\{\widehat{\alpha_m^{-1}} - \alpha^{-1} < y\right\} \\ =&P\left\{\frac{M_n}{n^{1/\alpha}} < n^y\right\} \\ \longrightarrow& \exp\{-n^{-\alpha y}\}. \end{aligned}$$

So we know now that the corresponding estimator of the Pareto index is consistent.

3.5 Methods to improve the estimator of α

As we already mentioned above, the maximum of the observations seriously influence the statistical properties of the sample sum. The amplitude of the maximum usually does not perform very regularly. This means that sometimes we will see large bias when we are estimating the parameters of interest using $\widehat{\alpha_n^{-1}}$.

On the other hand, the convergence rate of our estimator is very slow. The scale parameter will provide great bias when it is not equal to 1 even when the sample size is very large.

Consequently, an improvement of the estimator of α is necessary, from the consideration of both minimizing the variance and obtaining robustness.

In the following we consider some ways to improve the estimator.

The estimator as Incomplete U-statistic

Let X be strictly distributed stable random variable with scale parameter σ and index α. From Zolotarev(1986)[52] (chapter 4, pp218-220) we know that

$$E\log|X| = \log\sigma + C\cdot(\frac{1}{\alpha} - 1),$$

where $C = 0.577216$ is the Euler constant.

If we have i.i.d. sample $X_1,\ X_2,\ \cdots,\ X_n$ from the parent population X, then an unbiased estimator of the right hand side of the above equation is

$$\overbrace{\log\sigma + C\cdot(\frac{1}{\alpha} - 1)} = \frac{1}{n}\sum_{i=1}^{n}\log|X_i|$$

Suppose now we have $N = kn$ observations from some strictly stable distribution with index α and scale parameter σ. The observations are partitioned into the following way,

$$X_{11}, X_{12}, ..., X_{1n}, X_{21}, X_{22}, ..., X_{2n}, ...X_{k1}, X_{k2}, ..., X_{kn}$$

From every sub-sample $X_{i1}, X_{i2}, ..., X_{in}$, we can get an estimator of α^{-1}, say $\widehat{\alpha_{ni}^{-1}}$, $i = 1, 2, ..., k$, and we have

$$\log n(\widehat{\alpha_{ni}^{-1}} - \frac{1}{\alpha}) \stackrel{d}{=} \log|S_i|,$$

where S_i, $i = 1, 2, \ldots, nk$ are random variables with the same distribution as that of X.

Taking the average of the k estimators $\widehat{\alpha_i^{-1}}, i = 1, 2, ..., k$, we have

$$\widehat{\alpha_{U0}^{-1}} = \frac{1}{k}\sum_{i=1}^{k}\widehat{\alpha_i^{-1}} \tag{3.14}$$

and

$$\log n(\widehat{\alpha_{U0}^{-1}} - \frac{1}{\alpha}) \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\log|S_i|.$$

If k is large enough, the right hand side of the above equation tends to the expectation of $\log|S_i|$ and can be estimated from the sample mean of $\log|X_1|, \log|X_2|, \cdots, \log|X_n|$. Now we get our unbiased estimator of $1/\alpha$ as follows:

$$\widehat{\alpha_{U0s}^{-1}} = \widehat{\alpha_{U0}^{-1}} - \frac{1}{kn \cdot \log n} \sum_{i=1}^{k} \sum_{j=1}^{n} \log|X_{ij}|. \tag{3.15}$$

Theorem 3.5.1. Suppose $X_1, X_2, ..., X_N$ are i.i.d. observations from an strictly α-stable distribution, $N = kn$. Let $\widehat{\alpha_{U0s}^{-1}}$ be defined by (3.15), then we have

$$E\widehat{\alpha_{U0s}^{-1}} = \frac{1}{\alpha}.$$

If both k and n go to infinity, we have asymptotic normality of $\widehat{\alpha_{us}^{-1}}$ which is unbiased and has an improved variance.

The scale parameter can be estimated as follows:

$$\widehat{\log\sigma} = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} \log|X_{ij}| - C \cdot (\frac{1}{\widehat{\alpha_{us}^{-1}}} - 1).$$

Note that $\widehat{\log\sigma}$ is an asymptotically unbiased estimator of $\log\sigma$.

In the case of other heavy-tailed distributions, we can also use the estimator $\widehat{\alpha_{U0s}^{-1}}$ as an approximation. The simulation shows that it will overestimate the real parameter by about 10%.

From the above partition of the sample, we get the estimator of the corresponding parameter α^{-1}, both from the sample sum and from the maximum of the observations. The two kinds of estimators are usually not equal to each other but the consistency of the two estimators we have already proved. Furthermore the two estimators are linearly dependent. This gives us a very important hint to improve the estimator of the parameter. One way to do this is simply taking the average of both of them and get a new estimator as follows:

$$\widehat{\alpha_{mu}^{-1}} = \frac{1}{2}\left(\widehat{\alpha_{U0}^{-1}} + \widehat{\alpha_{m}^{-1}}\right). \tag{3.16}$$

where $\widehat{\alpha_{U0}^{-1}}$ is defined as above and $\widehat{\alpha_{m}^{-1}}$ is the average of the sub-estimators based on the sub-maximum of the partitioned sample:

$$\widehat{\alpha_{m}^{-1}} = \frac{1}{k} \sum_{i=1}^{k} \widehat{\alpha_{mi}^{-1}} \tag{3.17}$$

$$:= \frac{1}{k} \sum_{i=1}^{k} \frac{\log(M_n^i)}{\log(n)}. \tag{3.18}$$

Here $M_n^i = \max\{X_{i1}, X_{i2}, \ldots, X_{in}\}$

The asymptotic normality of $\widehat{\alpha_{mu}^{-1}}$ will also hold. Both of the two estimators have finite variances, very small absolute deviations when the sample size m is large enough. The average of the two estimators has smaller variance. The simulation results in chapter 5 will show this.

In fact, we can prove that the joint limit of (S_n, M_n) is (U, V), where U is some α-stable random variable and V has an extremal distribution by considering the hybrid form of characteristic and distribution function in the following way:

$$\phi_n(t, v) := E[\exp(itS_n) \cdot \mathbb{1}\{M_n \leq v\}].$$

The so called hybrid function has a limit under suitable normalising of the arguments S_n and M_n (see [7] for detail.). So the consistency of the averaged estimator $\widehat{\alpha_{mu}}$ holds.

3.6 The case of non-strict population

In the above sections we have discussed the estimating method of the stable index α and the corresponding properties of the estimators. All what has been said so far assumed that the corresponding parent population is strictly distributed. If the strictness assumption does not hold, then the estimator can usually not be used in the case when $\alpha > 1$. If we want the estimator still to be applicable, we should first of all, make the observations strictly distributed. In what follows we have some suggestions how to transform the non-strictly distributed observations into a symmetric one.

Suppose $X_1, X_2, \cdots, X_n$ be i.i.d. copies of some population X attracted to a stable law, X is not necessarily strictly distributed.

Let now $Y_1, Y_2, \cdots, Y_n$ be i.i.d. random variables with the distribution

$$P\{Y_i = 1\} = P\{Y_i = -1\} = \frac{1}{2},$$

and define new random variables by

$$Z_i = X_i \cdot Y_i,\ i = 1, 2, \cdots, n.$$

Then we have, for any given $x > 0$,

$$\begin{aligned} P\{Z_i < x\} = & \ \tfrac{1}{2} \cdot [P\{X < x\} + P\{X > -x\}] \, , \\ P\{Z_i > -x\} = & \ \tfrac{1}{2} \cdot [P\{X < x\} + P\{X > -x\}] \end{aligned}$$

that is the random variables $Z_i, \ i = 1, 2, \cdots n$ are symmetric, and of course also strictly distributed.

Furthermore, if the attracting stable law has index α, we have, as $x \to \infty$,

$$\begin{aligned} P\{Z_i > x\} = & \ \tfrac{1}{2}[P\{X > x\} + P\{X < -x\}] \sim \tfrac{1}{2}x^{-\alpha}L(x), \\ P\{Z_i < -x\} = & \ \tfrac{1}{2}[P\{X > x\} + P\{X < -x\}] \sim \tfrac{1}{2}x^{-\alpha}L(x). \end{aligned}$$

Note that the procedure can also be repeated, and an average of the estimators is available.

Another way to make our estimators available under non-strict assumption is to make the exponent smaller, less than 1, say. Then the attracting property holds without adding a corresponding constant term, that is the convergence is the same as in the strict case. In fact, if $X \sim F$, and F is in the domain of attraction of α-stable law, let $Y = |X|^2 \cdot \mathrm{sign(X)}$, then Y is also in the domain of attraction of stable law, but the index of the attracting stable distribution will be changed to $\alpha/2$. From the assumption that $\alpha < 2$ we know, $\alpha/2 < 1$. Then we have the corresponding estimator of α as follows:

$$\widehat{\alpha_q^{-1}} = \frac{\log|\sum_{i=1}^n Y_i|}{2\log n}.$$

The simulation result shows that even when the parent population is strict, the above adjustment gives also better estimates.

There is still another methods to make a distribution strict: Suppose $X_1, X_2, ..., X_n$, where $n = 2m$, are i.i.d. observations from a distribution F. Then the random variables $Y_1, Y_2, ..., Y_m$, where $Y_i = X_{2i-1} - X_{2i}, i = 1, 2, ..., m$ are strictly distributed.

Now all the estimators based on the properties of attraction we have used before can be applied.

Chapter 4

Estimation of stable index with Minimum-Distance

In this chapter, we construct an estimator for the stable index when the parent distributions are in a special class of heavy-tailed distributions – strictly stable distributions.

The sum-preserving property led us to the construction of new samples and, furthermore, their empirical distribution functions. We construct an estimator for the stable index by minimizing the Kolmogorov distance of distribution functions. The MD estimator can also be explained as a U-statistic. So it possesses "double" good properties.

4.1 The construction of the estimator

When the involved stable distribution F is supposed to be strict, we know from the definition of stable distribution that, if $X_1,\, X_2$ are independent random variables with the same distribution F, then the sum of them is also stably distributed, that is,

$$\frac{X_1 \,+\, X_2}{2^{1/\alpha}} \overset{\mathrm{d}}{=} X_1$$

where $\overset{\mathrm{d}}{=}$ means equivalence in distribution.

Suppose now $X_1,\, X_2,\, \cdots,\, X_n$ are i.i.d. copies of a strictly distributed stable random variable X with exponent α_0.

Let h denote the kernel

$$h(x_1,\, x_2) \,=\, \mathbb{1}_{\left(x_1+x_2 \,\le\, 2^{1/\alpha}\cdot t\right)} \,-\, \frac{1}{2}\left(\mathbb{1}_{(x_1 \le t)} + \mathbb{1}_{(x_2 \le t)}\right),$$

where $t \in R$ and $0 \,<\, \alpha \,<\, 2$,

and let a U-statistic with kernel h be defined by

$$U_n(\alpha, t) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j).$$

The properties of U-statistics guarantee that $U_n(\alpha, t)$ is a strongly consistent estimator of

$$E(h(X_1, X_2)) = P(X \le 2^{\frac{1}{\alpha} - \frac{1}{\alpha_0}} \cdot t) - P(X \le t).$$

Let

$$A(\alpha) = \sup_{t \in R} |Eh(X_1, X_2)|.$$

We know that $A(\alpha)$ is monotone in $\frac{1}{\alpha} - \frac{1}{\alpha_0}$ and attains its minimum only when $\alpha = \alpha_0$.

Now we have the estimator of α based on the above analysis:

$$\widehat{\alpha_{md}} = \arg \min_{\alpha \in [0,2]} \sup_{t \in R} |U_n(\alpha, t)|. \tag{4.1}$$

Theorem 4.1.1. Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. copies of X with strictly α_0-stable distribution. Let $\widehat{\alpha_{md}}$ be defined as in 4.1.Then as an estimator of α_0, $\widehat{\alpha_{md}}$ is consistent.

Proof. Fix α and t. Since the kernel of the U-statistic is bounded, that is,

$$|h(x_1, x_2)| \le 1,$$

it follows from the strong consistency of U-statistics that, as $n \to \infty$

$$U_n(\alpha, t) \longrightarrow E(h(X_1, X_2)) \quad a.s.$$

If the exponent of X is equal to α_0, then

$$Eh(X_1, X_2) = P(X \le 2^{\frac{1}{\alpha} - \frac{1}{\alpha_0}} \cdot t) - P(X \le t),$$

especially we have for all $t \in R$

$$U_n(\alpha_0, t) \longrightarrow 0 \quad a.s.$$

as $n \to \infty$.

The convergence in the above situation is in fact uniformly in t, that is, we have also, as $n \to \infty$

$$\sup_{t \in R} |U_n(\alpha_0, t)| \to 0 \quad a.s.$$

Now we have, for any given $\epsilon > 0$, if n is large enough,

$$\sup_{t \in R} |U_n(\alpha_0, t)| < \epsilon, \quad a.s.$$

so that, for $n \to \infty$,

$$0 < \sup_{t\in R} |U_n(\alpha_0, t)| - \min_{\alpha\in[0,2]} \sup_{t\in R} |U_n(\alpha, t)| \leq \epsilon \quad a.s.$$

that is

$$\sup_{t\in R} |U_n(\alpha_0, t)| \leq \min_{\alpha\in[0,2]} \sup_{t\in R} |U_n(\alpha, t)| + \epsilon \quad a.s..$$

On the other hand, if $\alpha \neq \alpha_0$, we have, for some $t_0 \neq 0$,

$$|P(X \leq 2^{\frac{1}{\alpha}-\frac{1}{\alpha_0}} \cdot t_0) - P(X \leq t_0)| = \delta > 0$$

hence, for n large enough,

$$|U_n(\alpha, t_0)| \geq |P(X \leq 2^{\frac{1}{\alpha}-\frac{1}{\alpha_0}} \cdot t_0) - P(X \leq t_0)| - \frac{\delta}{2} > \frac{\delta}{2} \quad a.s.$$

and

$$\sup_{t\in R} |U_n(\alpha, t)| \geq |P(X \leq 2^{\frac{1}{\alpha}-\frac{1}{\alpha_0}} \cdot t_0) - P(X \leq t_0)| - \frac{\delta}{2} > \frac{\delta}{2} \quad a.s.$$

Let $\epsilon = \delta/4$, we have

$$\begin{aligned} \sup_{t\in R} |U_n(\alpha, t)| &> \min_{\alpha\in[0,2]} \sup_{t\in R} |U_n(\alpha, t)| + \frac{\delta}{2} - \frac{\delta}{4} \\ &= \min_{\alpha\in[0,2]} \sup_{t\in R} |U_n(\alpha, t)| + \frac{\delta}{4}, \end{aligned}$$

that is: α_0 is almost sure the unique asymptotic solution to

$$\sup_{t\in R} |U_n(\alpha, t)| = \min_{\alpha\in[0,2]} \sup_{t\in R} |U_n(\alpha, t)|.$$

Finally it follows that

$$\widehat{\alpha_{md}} \longrightarrow \alpha_0, \quad a.s.$$

□

4.2 Another look at the estimator

The estimator $\widehat{\alpha_{md}}$ is in fact the minimum-distance estimator with respect to a special distance – Kolmogorov-distance:

$$d_k(F, G) = \sup_{t\in R} |F(t) - G(t)|,$$

where F, G are two distribution functions.

Suppose now that $X_1,\, X_2,\, \cdots,\, X_n$ are i.i.d. copies of the parent population X, with distribution F and $F \in \mathcal{F}$, a class of distribution functions.

Let

$$G_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(X_i \le x)}$$

be the empirical distribution function, then the corresponding minimum-distance estimator of F based on Kolmogorov-distance is the function $F_0 \in \mathcal{F}$, such that

$$d_k(F_0,\, G_n) = \min_{F \in \mathcal{F}} d_k(F,\, G_n).$$

The consistency of the minimum-distance estimator under some conditions holds as in the following proposition of Parr and Schucany[53]:

Proposition 4.2.1. Let T_n be the asymptotic minimum-distance estimator based on the empirical distribution function G_n with respect to Kolmogorov-distance

$$d_k(F,\, G) = \sup_{x \in R} |F(x) - G(x)|.$$

If (i) there is a point $\theta_0 \in \Omega$, the parameter space, such that

$$\inf_{\theta \in \Omega} d_k(F_\theta,\, G) = d_k(F_{\theta_0},\, G),$$

(ii) $\lim_{l \to \infty} d_k(F_{\theta_l},\, G) = d_k(F_{\theta_0},\, G)$ implies $\lim_{l \to \infty} \theta_l = \theta_0$,

then

$$P(\lim_{n \to \infty} T_n = \theta_0) = 1$$

In our assumptions here, the sample comes from some stable population F with exponent $\alpha_0 \in [0, 2]$.

The sum-preserving property of the stable distributions guarantees that,

$$\frac{X_1 + X_2}{2^{\frac{1}{\alpha_0}}} \stackrel{\mathrm{d}}{=} X_1,$$

hence for $\alpha_0 \in [0, 2]$, we have also a corresponding empirical distribution function of U-statistic structure:

$$\widetilde{F_n(\alpha,\, t)} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbb{1}_{\left(\frac{X_1 + X_2}{2^{\frac{1}{\alpha_0}}} \le t\right)}.$$

If α_0 is the exponent, then

$$E\mathbb{1}_{\left(\frac{X_1 + X_2}{2^{\frac{1}{\alpha_0}}} \le t\right)} = P(X \le t) = F(t),$$

and the boundedness of the kernel guarantees the strong consistency of the empirical distribution function $\widetilde{F_n(\alpha_0,\,\cdot)}$ with respect to $F(\cdot)$.

Now we consider the class of distributions

$$\mathcal{F} = \{\widetilde{F_n(\alpha,\,\cdot)} \,:\, \alpha \in [0,2]\}$$

A distribution or equivalently a value of $\alpha_0 \in [0,2]$ is to be chosen by minimizing the Kolmogorov-distance between $\widetilde{F_n(\alpha,\,\cdot)}$ and the empirical distribution function

$$G_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(X_i \le \cdot)}.$$

That is we are choosing the value of $\widehat{\alpha_{md}}$, such that

$$\min_{\alpha\in[0,2]} d_k(\widetilde{F_n(\alpha,\cdot)},\, G_n(\cdot)) \;=\; d_k(\widetilde{F_n(\widehat{\alpha_{md}},\cdot)},\, G_n(\cdot))$$

From the properties of minimum-distance estimators, we know that our estimator is consistent. And also as an M-estimator, it is robust.

Chapter 5

Simulations

In this chapter we compare our estimators to the most commonly used ones by simulation. Our estimators are simulated in the following two cases:

(1) the parent distribution is stable;

(2) the parent distribution is attracted to some α-stable law.

For different kind of parent distributions we have simulated our estimators with small sample size and large sample size.

5.1 Simulation of the estimators for stable index

5.1.1 The simulation results about $\widehat{\alpha_n^{-1}}$

We begin our simulation with the comparison of our estimator

$$\widehat{\alpha_n^{-1}} = \frac{\log(|\sum_{i=1}^{n} X_i|)}{\log(n)},$$

with that of Press and Zolotarev. The first parent population is chosen to have a Cauchy distribution, i.e. the stable index of the distribution is equal to 1. The sample size of the simulated sample is 1000. 1000 repetitions are performed. In the estimator of Press, we choose two different values of t's, 0.1 and 0.9 say, to calculate the estimate.

The random observations are easily taken by using SAS random number generating-programme. The simulation result shows, that all the three estimators are unbiased and all have very small standard deviations:

estimating α when Cauchy, n = 1000, 1000 times

ESTIMATOR	MEAN	STD	MAX	MIN
$\widehat{\alpha_n}$	0.99	0.067	1.256	0.795
AL_Z	1.00	0.040	1.15	0.895
AL_P	1.00	0.039	1.13	0.863

Under the same assumption, we have simulated our resampling estimator 100 times. We take the sample size $n = 500$ and 2000 times re-sampled, the corresponding resampling rates are $p1 = \exp(-1), \quad p2 = \exp(-2)$. This time we have chosen 1 and 2 as the values of t in the estimator of Press. The simulation results are shown as follows:

estimating the index of Cauchy, n = 500, 100 times

ESTIMATOR	MEAN	STD	MAX	MIN
$\widehat{\alpha_n}$	1.011079	0.221422	1.685062	0.615358
$\widehat{\alpha}_{r1}$	0.992299	0.160639	1.454322	0.713796
$\widehat{\alpha}_{r2}$	0.991106	0.115437	1.250524	0.757033
AL_p	1.002055	0.212363	1.79101	0.504879
AL_z	1.016161	0.073563	1.204795	0.876207

From the results above we know, that the estimator of Zolotarev performs best in this situation. And our resampling estimator with resampling rate $r2$ performs almost as good as that of Zolotarev.
The simulation results of the estimator of Press depend seriously on the choice of the corresponding t values, therefor a regression procedure has been proposed in [31].

We continue with the simulation results of the estimator $\widehat{\alpha_{U0}}$ with incomplete U-statistic structure. The parent populations are stable laws† but not necessarily Cauchy distribution. The sample size is $20 \cdot 50 = 1000$. We have partitioned the sample into 20 groups, or subsamples say, each with a sample size of 50. 1000 repetitions are taken. The simulated mean, standard deviation, the maximal and minimal observation and the 95%-,5%-quantiles are in the table.

The simulation results show that the estimator is asymptotically unbiased and the simulated standard deviations are very small. The simulated 90% confidence intervals of the parameters have very short lengths.

†The simulation of the stable random variables are based on the method of Chambers, Mallows and Stuck [6], i.e.

$$S(\alpha,\beta) = \frac{\sin\alpha(\Phi - \Phi_0)}{(\cos\Phi)^{1/\alpha}} \cdot \left(\frac{\cos(\Phi - \alpha(\Phi - \Phi_0))}{W}\right)^{(1-\alpha)/\alpha}, \alpha \neq 1,$$

where W is standard exponential variable, Φ is uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$, $\Phi_0 = -\frac{\pi}{2}\beta(k(\alpha)/\alpha)$, $k(\alpha) = 1 - |1 - \alpha|$, β is the skewness of the stable distribution. We take $\beta = 0$ here for simplicity.

simulation–stable–1000 times

α	MEAN	STD	MAXL	MINL	95%	5%
0.1	0.0885	0.005645	0.110968	0.070367	0.097555	0.079778
0.2	0.179975	0.011743	0.233847	0.142098	0.199641	161013
0.3	0.272814	0.018714	0.339677	0.222868	0.30479	0.243001
0.4	0.368623	0.025317	0.452184	0.299404	0.410056	0.327046
0.5	0.468771	0.035003	0.616874	0.375852	0.528203	0.412902
0.6	0.565932	0.041946	0.719763	0.453906	0.636796	0.500755
0.7	0.674596	0.052877	0.873042	0.540581	0.770397	0.592311
0.8	0.780657	0.060605	1.036442	0.606317	0.886737	0.686527
0.9	0.885335	0.073696	1.206059	0.689409	1.010124	0.770722
1.0	0.986729	0.0797	1.338058	0.754691	1.122484	0.859136
1.1	1.096306	0.091665	1.446826	0.800295	1.256774	0.954191
1.2	1.200678	0.104251	1.761665	0.921855	1.384139	1.003652
1.3	1.305742	0.12012	1.892169	1.002769	1.523353	1.131205
1.4	1.406406	0.129193	2.060279	1.088249	1.641126	1.220774
1.5	1.512264	0.134448	2.115736	1.212528	1.761875	1.322811
1.6	1.602115	0.155149	2.473736	1.267983	1.886896	1.399116
1.7	1.700103	0.16121	2.53578	1.371294	1.985334	1.475616
1.8	1.783365	0.17714	2.633028	1.43042	2.079229	1.543659
1.9	1.864068	0.185169	2.814284	1.488579	2.193833	1.60923

Next we present the comparison simulation of our estimator $\widehat{\alpha_{U0}}$ with that of Press and Zolotarev. The estimator of Press performs better when $\alpha > 1$ and that of Zolotarev is better when $\alpha < 1$. Our estimator $\widehat{\alpha_{U0}}$ performs almost as good as the best of Press and Zolotarev estimator all the time. Note that the estimator of Press is difficult to be simulated 1000 times when the parameter $\alpha < 0.4$. The computer refused to calculate the cosine or sine transformation of the observed data. The estimator of Zolotarev has also a problem. We have observed many samples from some stable distribution, where the associated estimates of $1/\alpha^2$ take negative values. In the following two tables, MEAN$_{U0}$ and STD$_{U0}$ denote the simulated mean values and the standard deviations of our estimator $\widehat{\alpha_{U0}}$, MAX$_{U0}$ and MIN$_{U0}$ the maximums and the minimums. The sub-index -z and -p represent that of Zolotarev estimator and Press estimator. The sample size here is $20 \cdot 50 = 1000$

comparison simulation–stable–n = 1000, 1000 times

α	MEAN$_{U0}$	MEANz	MEANp	STD$_{U0}$	STDz	STDp
0.4	0.370	0.401	0.400	0.0243	0.0137	0.0354
0.5	0.468	0.501	0.501	0.0327	0.0171	0.0369
0.6	0.565	0.599	0.600	0.0420	0.0218	0.0376
0.7	0.674	0.701	0.703	0.0539	0.0266	0.0414
0.8	0.785	0.803	0.802	0.0664	0.0324	0.0438
0.9	0.887	0.904	0.900	0.0723	0.0402	0.0455
1.0	0.991	1.003	1.000	0.0847	0.0498	0.0505
1.1	1.094	1.102	1.103	0.0865	0.0547	0.0516
1.2	1.203	1.207	1.203	0.1010	0.0679	0.0570
1.3	1.310	1.309	1.303	0.1126	0.0811	0.0562
1.4	1.405	1.415	1.407	0.1197	0.0971	0.0619
1.5	1.505	1.507	1.505	0.1302	0.1188	0.0645
1.6	1.61	1.61	1.605	0.1550	0.1337	0.0646
1.7	1.69	1.71	1.705	0.1568	0.1647	0.0648
1.8	1.79	1.82	1.80	0.1825	0.1860	0.05859
1.9	1.88	1.94	1.90	0.2165	0.2221	0.0500

comparison simulation–stable–1000 times

α	MAX$_{U0}$	MAXz	MAXp	MIN$_{U0}$	MINz	MINp
0.4	0.44	0.45	0.54	0.30	0.36	0.29
0.5	0.58	0.56	0.61	0.37	0.44	0.37
0.6	0.70	0.67	0.74	0.45	0.54	0.49
0.7	0.86	0.81	0.86	0.53	0.61	0.57
0.8	1.08	0.93	0.93	0.64	0.69	0.68
0.9	1.11	1.05	1.05	0.71	0.79	0.78
1.0	1.30	1.21	1.18	0.76	0.87	0.87
1.1	1.43	1.29	1.27	0.84	0.93	0.95
1.2	1.61	1.44	1.42	0.94	0.97	1.05
1.3	1.70	1.66	1.47	1.06	1.07	1.14
1.4	1.87	1.79	1.62	1.13	1.17	1.23
1.5	2.01	1.95	1.71	1.20	1.19	1.31
1.6	2.24	2.15	1.78	1.31	1.28	1.41
1.7	2.32	2.46	1.90	1.36	1.27	1.50
1.8	2.65	2.66	1.95	1.4	1.32	1.59
1.9	2.80	3.08	2.01	1.47	1.42	1.75

In the case of **small sample size**, $n = 200$, for example, we have simulated our resampling estimator 200 times with stable parent distributions. Our resampling procedure gives very good estimates of the unknown parameters. The estimators of Press or Zolotarev perform not so well.

When α takes values greater than 1, the estimator of Zolotarev performs not well, but our resampling estimator and Press estimator perform very well in this situation.

When α takes values less than 1, all three estimators perform well, but the Press estimator performs relatively worse.

In the following there are some results of the above simulation. Here MEAN,STD,R,Q31 represent the simulated mean, standard deviation, range and the interquartile-range of the estimators respectively, the sub-index -r, -z, -p denote our resampling estimator, the estimator of Zolotarev and that of Press, accordingly.

comparison simulation–stable–200 times

α	MEANr	MEANz	MEANp	STDr	STDz	STDp
0.3	0.277025	0.300399	0.288717	0.032808	0.022099	0.067261
0.6	0.570816	0.600956	0.600964	0.067654	0.049142	0.08977
0.9	0.903337	0.923222	0.916601	0.115214	0.086165	0.100267
1.2	1.201662	1.213172	1.20449	0.149892	0.171192	0.134035
1.5	1.55455	1.631239	1.537978	0.182088	0.330738	0.144458
1.8	1.876307	2.023097	1.819125	0.161744	0.730564	0.107914

comparison simulation–stable–200 times

α	Rr	Rz	Rp	Q31r	Q31z	Q31p
0.3	0.166538	0.119223	0.351355	0.047731	0.032222	0.089129
0.6	0.373284	0.238977	0.465285	0.080948	0.067191	0.11831
0.9	0.588226	0.470496	0.561045	0.160312	0.110152	0.141955
1.2	0.805052	1.256783	0.7073	0.180267	0.181793	0.1694
1.5	0.998092	2.913216	0.80398	0.251428	0.357218	0.187832
1.8	0.804096	6.916831	0.572065	0.214085	0.529533	0.146426

5.2 The simulation of estimators of the tail index

Now we come to the simulation of the estimators for the tail index of heavy-tailed distributions.

The simulation of the random variables, whose distributions have heavy tails, is based on the following theorem. We have generated all our observations in the same way.

Theorem 5.2.1. Let X be a random variable with probability density $f(x), \quad x \in R$, and satisfying the following condition:

$$\lim_{x \to 0} f(x) = m > 0.$$

Then

$$Y = sign(X) \cdot |X|^{-\frac{1}{\alpha}}$$

has a distribution with heavy tails of tail index α, i.e.

$$y^\alpha P\{|Y| > y\} \sim L(y), \quad (y \to \infty),$$

where $L(y)$ is slowly varying at infinity.

Proof. For any real $y > 0$,

$$\begin{aligned} y^\alpha P\{|Y| > y\} &= y^\alpha P\{|X|^{-\frac{1}{\alpha}} > y\} \\ &= y^\alpha P\{|X| < y^{-\alpha}\} \\ &= y^\alpha \int_{-y^{-\alpha}}^{y^{-\alpha}} f(t)dt \\ &= \int_{-1}^{1} f(zy^{-\alpha})dz \widehat{=} L(y) \end{aligned}$$

Now for any given $\lambda > 0$,

$$\begin{aligned} \frac{L(\lambda y)}{L(y)} &= \frac{\int_{-1}^{1} f(z\lambda^{-\alpha}y^{-\alpha})dz}{\int_{-1}^{1} f(zy^{-\alpha})dz} \\ &\to \frac{\int_{-1}^{1} \lim_{t\to 0} f(t)dz}{\int_{-1}^{1} \lim_{t\to 0} f(t)dz} = 1 \\ &(\quad y \to \infty) \end{aligned}$$

It follows that $L(y)$ is slowly varying and the distribution of Y has heavy tails of index α.

□

Based on the above theorem, we have simulated our estimator $\widehat{\alpha_{U0}}$ of tail index when the parent distributions have heavy tails. Here the tail indices are taken to be equal to 0.1 to 1.9 with step size 0.1.

For each value of α, we have taken a sample of size $30 \cdot 100 = 3000$. 1000 times are simulated.

The simulation results show that our estimator $\widehat{\alpha_{U0}}$ underestimates the real parameters a little, about 5% lower than what we are estimating, however the standard deviations are very small. All the 1000 estimates are distributed very near to the theoretical value we want to estimate. The simulated ranges of the estimator are very narrow. See the following table for detail.

simulation result–DA(α)–1000 times

α	MEAN	STD	MAXL	MINL
0.1	0.092987	0.004282	0.109353	0.080635
0.2	0.1857	0.008701	0.215748	0.161254
0.3	0.27933	0.01327	0.325539	0.242901
0.4	0.372936	0.017568	0.429575	0.319699
0.5	0.466254	0.022651	0.539905	0.40122
0.6	0.560768	0.027235	0.660525	0.488642
0.7	0.655925	0.031885	0.767183	0.556924
0.8	0.752233	0.038655	0.869054	0.632564
0.9	0.847303	0.040948	1.022709	0.727041
1.0	0.942731	0.04766	1.114957	0.804083
1.1	1.040359	0.053285	1.233191	0.892092
1.2	1.136058	0.056685	1.305414	0.959664
1.3	1.232885	0.065645	1.497309	1.038785
1.4	1.32609	0.07605	1.579052	1.157234
1.5	1.425414	0.080079	1.734725	1.203831
1.6	1.525933	0.086198	1.845454	1.287219
1.7	1.622746	0.090743	1.93784	1.378038
1.8	1.718552	0.101254	2.097628	1.451296
1.9	1.810259	0.10826	2.156351	1.542873

5.2.1 The simulation results of $\widehat{\alpha_{mu}^{-1}}$

As we have pointed out in chapter 3, when we use our estimator with incomplete U-statistic structure, we get at the same time another estimator from the maxima of every subsamples. The two estimators are different but both of them are consistent estimators of the tail index. Hence their average should perform better than both of them.
Suppose now the observations are partitioned into the following way:

$$X_{11}, X_{12}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}, \dots, X_{k1}, \dots, X_{kn_k}.$$

From every subsample we get two different estimators of the corresponding parameter α^{-1} based on respectively the sum and the maximum of the subsamples. The average of the two estimators is denoted by $\widehat{\alpha_{mu}^{-1}}$

Here are simulation results for $\alpha = 0.1$ to 1.9 with sample size $30 \cdot 100 = 3000$. 1000 repetitions are computed.

From the simulation results we know that the performance of the averaged estimator $\widehat{\alpha_{mu}^{-1}}$ is perfect both in the sense of unbiasedness and standard deviation. The ranges of the simulated estimates are very short.

simulation results for $\widehat{\alpha_{mu}}$ n=3000, 1000 times

α	MEAN	STD	MAX	MIN
0.1	0.100129	0.004868	0.11461	0.084937
0.2	0.200443	0.009037	0.233376	0.167738
0.3	0.300843	0.013788	0.350241	0.257719
0.4	0.403175	0.018217	0.462523	0.347983
0.5	0.502868	0.023505	0.574583	0.427065
0.6	0.605573	0.029425	0.694585	0.506735
0.7	0.707793	0.033489	0.807223	0.596969
0.8	0.810774	0.038715	0.942016	0.695136
0.9	0.912324	0.045318	1.069898	0.773246
1.0	1.012038	0.049916	1.185196	0.848996
1.1	1.115918	0.053939	1.329899	0.959299
1.2	1.213177	0.060468	1.433545	1.012969
1.3	1.314458	0.068438	1.567453	1.088237
1.4	1.415756	0.069709	1.705125	1.223999
1.5	1.50704	0.076623	1.771287	1.29424
1.6	1.60049	0.083238	1.943089	1.397052
1.7	1.687451	0.089191	2.033835	1.42573
1.8	1.782468	0.096933	2.173826	1.506238
1.9	1.867337	0.098592	2.144365	1.589966

Next consider the simulation results of the resampling procedure. The sample size here is 400. We have compared the unimproved and the resampling estimator of α. The resampling rate is 0.1, 2000 re-samples and 100 repetitions are taken.

The simulation results show that the estimator without resampling adjustment behaves very badly when the sample size is small. But when the resampling procedure is used, we get a great improvement both in the sense of unbiasedness and standard deviation.

Let $\widehat{\alpha_{rmu}}$ denote the estimator $\widehat{\alpha_{mu}}$ when it is used in the resampling procedure. In the following table, the MEAN and STD denote the means and the standard deviations of the estimator $\widehat{\alpha_n}$. MEANr and STDr denote those of the estimator re-sampled.

The simulation results tell us that it is very dangerous to use the unadjusted estimator $\widehat{\alpha_n}$ directly, especially in the case when α is larger than 1. However, the re-sampled estimator performs very well. The unbiasedness of the estimator holds and the simulated standard deviations are very small.

simulation results for $\widehat{\alpha_{rmu}}$ †

α	MEAN	MEANr	STD	STDr
0.1	0.096865	0.099792	0.01913	0.00915
0.2	0.20038	0.204956	0.034985	0.018902
0.3	0.288847	0.301298	0.055859	0.030822
0.4	0.399371	0.408202	0.082475	0.038405
0.5	0.507288	0.508608	0.108042	0.041682
0.6	0.631442	0.610819	0.262979	0.059225
0.7	0.703653	0.721994	0.124382	0.070203
0.8	0.811178	0.809378	0.209398	0.071019
0.9	0.915695	0.917442	0.35802	0.090973
1.0	1.021822	1.011007	0.396262	0.105381
1.1	1.002088	1.134574	1.06149	0.104662
1.2	1.400864	1.261483	1.090945	0.111339
1.3	1.362876	1.319596	1.336016	0.114629
1.4	1.576241	1.415312	0.928443	0.121628
1.5	1.526868	1.540436	0.446798	0.143157
1.6	2.360023	1.637475	8.937643	0.142141
1.7	2.305255	1.692761	2.809973	0.152969
1.8	1.634679	1.792639	1.760745	0.147203
1.9	1.682808	1.860281	1.508459	0.169249

5.3 Comparison with the most popular estimators

The most important advantages of our estimator are that we have used all of the observations and the formula is very simple. The following simulated comparisons show that our estimator performs very well also. In the simulation study, our averaged estimator are simulated together with the sub-sampling procedure. Our estimator is unbiased and has always the smallest standard deviations comparing with that of Hill's estimator and Resnick & De Haan's estimator. Furthermore, our estimator has the smallest maximal estimate and largest minimal of the three estimators.

Here are the simulations of a comparison with Hill's estimator and that of Resnick & de Haan, where the parent distributions are in the domain of attraction of stable distribution and the sample size is $20 \cdot 50 = 1000$. 1000 repetitions are computed.

† Here the sample size is $n = 400$, resampling rate $r = 0.1$, 2000 times re-sampled, 100 times simulated.

Comparison simulation results –with Hill's and R & D's estimators

α	MEAN	MEANh	MEANrd	STD	STDh	STDrd
0.1	0.100138	0.100113	0.094118	0.006377	0.010028	0.021495
0.2	0.200255	0.200443	0.187498	0.013125	0.020616	0.043228
0.3	0.300065	0.30113	0.280856	0.019625	0.030194	0.063631
0.4	0.400841	0.402636	0.376279	0.026961	0.041024	0.085248
0.5	0.500449	0.49651	0.468862	0.034278	0.050213	0.103935
0.6	0.602717	0.600002	0.563899	0.040459	0.059673	0.128889
0.7	0.703386	0.703505	0.655011	0.046491	0.068876	0.143837
0.8	0.804661	0.799965	0.756091	0.053867	0.078057	0.17507
0.9	0.904469	0.899727	0.835005	0.058458	0.086739	0.191711
1.0	1.007727	1.002484	0.942006	0.068427	0.102838	0.215082
1.1	1.105628	1.100189	1.018997	0.072097	0.108952	0.237571
1.2	1.210411	1.199682	1.118505	0.082714	0.121091	0.264387
1.3	1.320271	1.305397	1.227301	0.087078	0.130423	0.280187
1.4	1.424354	1.403946	1.322749	0.096013	0.141552	0.298518
1.5	1.520911	1.504556	1.40251	0.103351	0.155608	0.387344
1.6	1.623432	1.603543	1.487837	0.113431	0.160963	0.39942
1.7	1.723349	1.697324	1.60345	0.116472	0.16677	0.424284
1.8	1.821861	1.797191	1.671671	0.126096	0.17903	0.446097
1.9	1.927302	1.900913	1.775463	0.132651	0.191674	0.468612
2.0	2.027878	1.999309	1.869232	0.143364	0.201738	0.439047
N	2.06267	3.461711	5.078234	0.197853	0.312094	0.612247

In the above table, MEAN, STD denote the simulated mean values and standard deviations of our estimator and that with the sub-index -h or -rd represent that of Hill's or Resnick and De Haan's estimators. N denotes the standard normal distribution.

The simulation results in the above table show that, our estimator performs always better than that of Hill's estimator and Resnick & De Haan's estimator. In fact, the variance of Hill's estimator is about 2 times that of ours. And the estimator of Resnick and De Haan preforms worse.

One more thing to note: When the three estimators are used to estimate the index of a normal distribution, only our estimator performs very well. The estimates of the other two estimators are totally misleading.

Comparison simulation results –with Hill's and R & D's estimators

α	MAX	MAXh	MAXrd	MIN	MINh	MINrd
0.1	0.120088	0.135445	0.182661	0.083024	0.073885	0.040132
0.2	0.243224	0.288085	0.353281	0.15307	0.141522	0.075422
0.3	0.363572	0.420337	0.547133	0.238139	0.20962	0.098449
0.4	0.50867	0.532018	0.70949	0.33076	0.300739	0.128819
0.5	0.620336	0.685364	0.921321	0.399494	0.354632	0.165548
0.6	0.745266	0.844892	1.106638	0.488072	0.444192	0.255988
0.7	0.851254	0.967654	1.224282	0.565276	0.47959	0.28343
0.8	0.977676	1.150082	1.369684	0.618817	0.60449	0.26943
0.9	1.107294	1.257708	1.550806	0.710638	0.670727	0.337919
1.0	1.275338	1.440594	1.910853	0.805363	0.748811	0.373364
1.1	1.351287	1.500437	1.863937	0.878313	0.832397	0.398581
1.2	1.475956	1.674282	2.194764	0.974076	0.833509	0.414142
1.3	1.629395	1.784487	2.321554	1.03133	0.982383	0.511172
1.4	1.767652	1.919196	2.714056	1.050087	1.021057	0.579257
1.5	1.938837	2.042981	3.451483	1.226818	1.06383	0.444437
1.6	2.019092	2.126089	3.192805	1.259138	1.197484	0.625627
1.7	2.070122	2.340255	3.18834	1.360355	1.264621	0.629478
1.8	2.349523	2.414025	3.330987	1.416648	1.335004	0.681997
1.9	2.364064	2.54558	3.554365	1.519032	1.384768	0.74339
2.0	2.47429	2.86183	3.467732	1.557095	1.486861	0.634264
N	3.224972	4.62248	7.685469	1.574036	2.595376	3.383567

In the above table, MAX and MIN denote the simulated maxima and minima of our estimator. Q1 and Q2 in the following table represent the 75%- and 25%-quantiles. The sub-index -h and -rd have the same meaning as in the the first of the three tables.

Comparison simulation results –with Hill's and R & D's estimators

α	Q3	Q3h	Q3rd	Q1	Q1h	Q1rd
0.1	0.104363	0.106163	0.108204	0.095656	0.093032	0.079869
0.2	0.208384	0.213763	0.214676	0.191320	0.185119	0.15801
0.3	0.313654	0.319967	0.322934	0.286756	0.280372	0.234297
0.4	0.418531	0.428573	0.432264	0.38157	0.374048	0.317244
0.5	0.522876	0.527888	0.537657	0.47756	0.460333	0.39434
0.6	0.628345	0.637337	0.644066	0.575528	0.558848	0.47359
0.7	0.735566	0.748173	0.752742	0.671235	0.652653	0.551986
0.8	0.840669	0.84967	0.866935	0.767874	0.745637	0.632138
0.9	0.942857	0.95279	0.95965	0.865663	0.842034	0.699513
1.0	1.049555	1.070124	1.086386	0.963795	0.923967	0.784944
1.1	1.149199	1.168573	1.17295	1.057588	1.022129	0.850167
1.2	1.261828	1.274544	1.282971	1.153871	1.113055	0.925446
1.3	1.377075	1.390275	1.400111	1.261696	1.208544	1.02469
1.4	1.48905	1.491054	1.520387	1.358809	1.308554	1.12328
1.5	1.585922	1.600973	1.617019	1.448175	1.393673	1.129653
1.6	1.689838	1.708122	1.706942	1.5509	1.485379	1.217758
1.7	1.799079	1.806959	1.873999	1.64807	1.579655	1.29533
1.8	1.908234	1.901308	1.954876	1.731427	1.68426	1.351689
1.9	2.01281	2.01027	2.061363	1.831409	1.77006	1.441274
2.0	2.12237	2.130897	2.140291	1.937691	1.8553	1.556845
N	2.166033	3.662734	5.463011	1.921625	3.244746	4.668475

The case of small sample size When the sample size is not large, 200 say, we have compared our estimator to the other two also. We have used our resampling procedure. Our simulation results show that the standard deviations of Hill's estimator are getting too large. But our resampling procedure can still help to get very good estimates. The comparison is simulated when $\alpha = 0.3, 0.6, 1.2, 1.6$. 200 repetitions are calculated. The resampling rate here is 15% and the corresponding k –the number of the largest order statistics in Hill's or R&D's estimators is taking to be 20.

Comparison simulation results –with Hill's and R & D's estimators

α	MEAN	MEANh	MEANrd	STD	STDh	STDrd
0.3	0.280791	0.307381	0.280648	0.034833	0.069951	0.097299
0.6	0.568137	0.636665	0.564266	0.065783	0.135139	0.184693
1.2	1.15863	1.287416	1.17409	0.141467	0.285386	0.389861
1.6	1.545009	1.740074	1.548522	0.184832	0.417576	0.551903

Comparison simulation results –with Hill's and R & D's estimators

α	R	Rh	Rrd	Q31	Q31h	Q31rd
0.3	0.175789	0.373111	0.572837	0.046683	0.08476	0.120882
0.6	0.345151	0.826827	1.12011	0.093912	0.180731	0.230578
1.2	0.804151	1.758115	2.141391	0.205987	0.381218	0.534697
1.6	1.127301	2.656967	3.731009	0.213468	0.493332	0.687471

As we have mentioned above, the estimators of Hill and Resnick & De Haan are in fact plotting procedures. It means one can only get an estimate from a given sample by observing the plot. In what follows there are some figures comparing our estimator to that of Hill or Resnick and De Haan. We know from the following pictures that the estimator of Resnick and De Haan is relatively non-sensitive to the choice of the corresponding parameter k. The situation of Hill estimator is not the same. That is also the reason why we take $k = 100$ when the sample size is 1000. In this situation the Hill's estimator performs almost best.

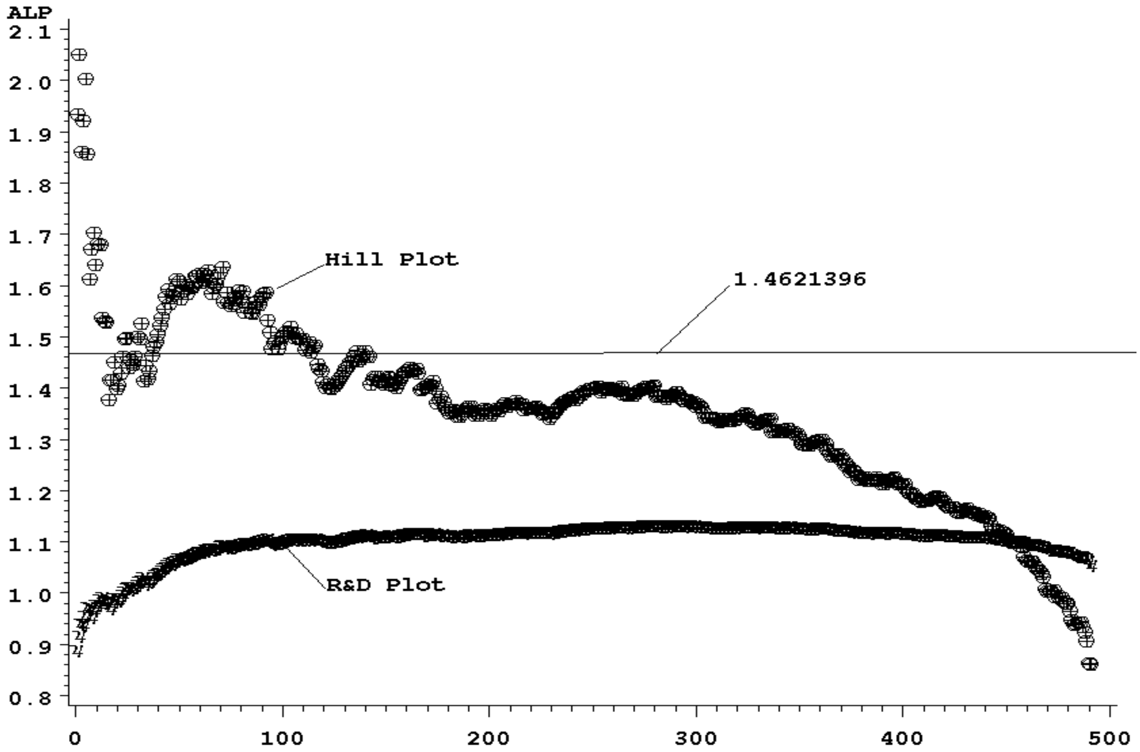

Figure 5.1: DA(α), $n = 1000,\ \alpha = 1.5,\ \widehat{\alpha} = 1.4621396$

In picture (5.1) we have simulated the three estimators when the parent distribution is attracted to a 1.5-stable law. The value of α is supposed to be 1.5. Our estimate is 1.4621396. The estimates from Hill or Resnick and De Haan are smaller.

In picture (5.2), the value $\alpha = 1.5$ of the stable exponent is to be estimated. Our estimator gives value of 1.5235288. From the plots of Hill or Resnick and De Haan we can not get meaningful estimates.

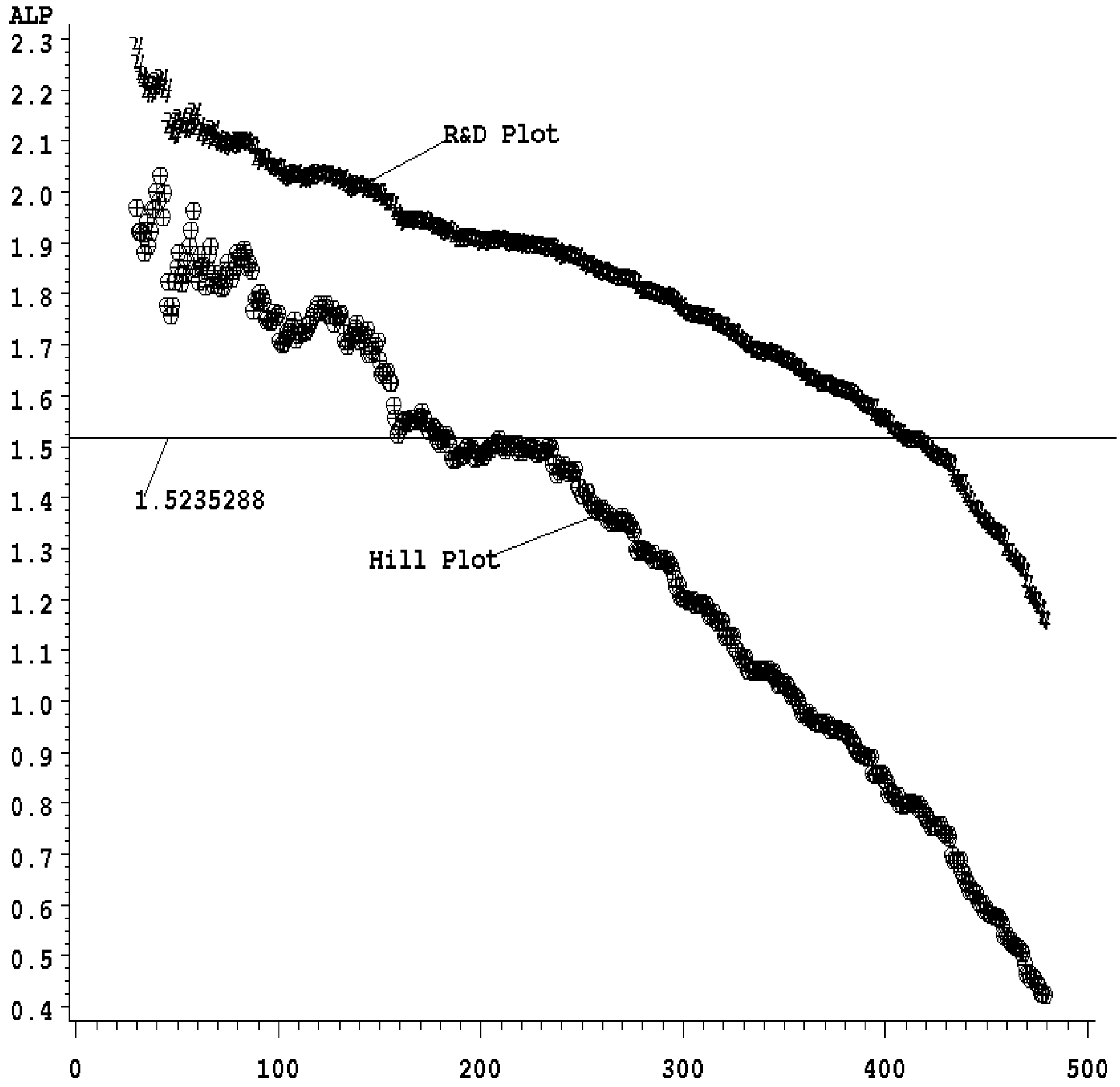

Figure 5.2: Stable(α), $n = 1000$, $\alpha = 1.5$, $\widehat{\alpha} = 1.5235288$

A Notation

From the simulation results above we know that, our estimator preforms always very well. But we want give some further hints how to use it.

When the sample size is large enough, that is we have enough observations, we suggest a procedure of sub-sampling. That is we should partition the sample in to subsamples. Because the rate of convergence of our estimator is only of order $\log n$,

n is the sample size, but at the same time the convergence region is very narrow. It means we can get from every subsample an estimator almost as good as that from the whole sample. When taking average of the sub-estimators we can improve the standard deviations considerably.

When the sample size is not so large, we cannot get much by sub-sampling procedures. In this situation we can use our resampling procedure. By deleting the extremal estimates in the resampling procedure, we can get an relatively robust estimate of the corresponding parameter with small standard deviation. Or we can use our estimator with U-statistic structure. That is the best estimator we can get from the sum preserving property.

5.3.1 The simulation results of $\widehat{\alpha_{md}}$

The minimum-distance estimator $\widehat{\alpha_{md}}$ is simulated 50 times with sample size 200 when $\alpha = 0.3, 0.6, 0.9, 1.2, 1.5, 1.8$. The results show that, even with a sample size of only 200, the estimates are already very good. The estimators of Press and Zolotarev are also simulated. All the three estimators perform good when α is less than 1, and our MD estimator performs the best when α is greater than 1.

simulation–$\widehat{\alpha_{md}}$–50 times

α	MEANmd	MEANp	MEANz	STDmd	STDp	STDp
0.3	0.307	0.300574	0.303741	0.027683	0.079809	0.024684
0.6	0.614	0.58393	0.609143	0.059367	0.08769	0.048202
0.9	0.9285	0.924629	0.9346	0.101017	0.112087	0.082291
1.2	1.2405	1.233229	1.227397	0.132084	0.122718	0.193122
1.5	1.549	1.587643	1.636426	0.198204	0.124703	0.318997
1.8	1.8015	1.845927	1.848307	0.162601	0.101778	0.368223

Acknowledgements

I would like especially to thank Professor Dr. Manfred Denker for his constant encouragement, helpful suggestions and valuable discussions during the preparation and the writing of the dissertation.

I would like to thank Dr. Susanne Koch for her great help in improving my English expression.

Also I would like to thank Dr. Arne Bathke for his help while I was doing the simulations and writing the dissertation.

The discussions with Eduardo Fdez. de Córdoba Martos are always meaningful. I am very grateful for that.

Finally, I would like to express my deep gratefulness to Oliver Schmitt and Manuel Stadlbauer for their constant help in my study also in my daily life.

This research was financially supported by Hanns-Seidel-Stiftung. I am deeply grateful for their kindly help.

Chapter 6

SAS Programme

Programme 1 –when the parent is stable or in DA(α) data al;

```
input ala;
cards;
0
;
run;
data a(keep=ala);
set al;
do i = 1 to 1000;
output;
end;
run;
proc iml;
ala=j(1000,1,0);
do ns = 1 to 1000;
ma=j(20,1,0);
m=j(20,1,0);
do k = 1 to 20;
x1=ranexp(j(50,1,-1));
x2=(ranuni(j(50,1,-1))-1/2)#3.14159;
y=sin(1.9#x2)/(cos(x2))##(1/1.9)#(cos((1-1.9)#x2)/x1 )##(1/1.9-1);
/*x = rannor(j(50.1.-1));
y = abs(x)##(-1/1.9)#sign(x);*/
u = abs(sum(y));
z = abs(max(y));
m[k]= (log(u))/(log(50));
ma[k]=m[k]#(m[k]¿1/2)+(m[k]¡1/2)#log(u+50##(m[k]))/log(50);
al = 20/sum(m);
ala[ns]=20/sum(ma);
end;
```

```
end;
edit a;
replace all varala;
quit;
proc univariate data=a normal;
var ala;
run;
```

Comparison with Hill or Resnick and De Haan estimator

```
data alpa;
input alp alpb;
cards;
0 0
;
run;
data alpha;
set alpa;
do i=30 to 500 by 1;
output;
end;
run;
proc iml;
ma=j(20,1,0);
m=j(20,1,0);
do k = 1 to 20;
x = rannor(j(50,1,-1));
y = abs(x)##(-1/1.9)#sign(x);
/*x1=ranexp(j(50,1,-1));
x2=(ranuni(j(50,1,-1))-1/2)#3.14159;
y=sin(1.9#x2)/(cos(x2))##(1/1.9)#(cos((1-1.9)#x2)/x1 )##(1/1.9-1);*/
u = abs(sum(y));
z = abs(max(y));
m[k]= (log(u))/(log(50));
ma[k]=m[k]#(m[k]¿1/2)+(m[k]¡1/2)#log(u+50##(m[k]))/log(50);
al = 20/sum(m);
ala=20/sum(ma);
w = w//y;
end;
print al ala;
v=w;
w[1001-rank(w),]=v;
a = j(1000,471,0);
b=j(1000,471,0);
do i = 1 to 1000;
```

```
do j = 1 to 450;
a[i,j]=1/(j+20)#(i¡=j+20)+(-1)#(i=j+21);
b[i,j]=1/log(j+20)#((i=1)-(i=j+20));
end;
end;
lv =log(abs(w));
alp = (lv‘*a)##(-1);
alpb= (lv‘*b)##(-1);
edit alpha;
replace all varalp alpb;
quit;
data alph;
set alpha;
j=i;
output;
run;
proc gplot data=alph;
plot alp*j=’+’ alpb*j=’.’/ overlay ;
run;
```

Bibliography

[1] Jon Aaronson and Manfred Denker, Characteristic Functions of Random Variables Attracted to 1-Stable Laws, The Annals of Probability, Vol.26 No.1, pp399-415(1998)

[2] James N. Arvesen, Jackknifing U-statistics, The annals of mathematical statistics, Vol.40, No. 6, 2076-2100(1969)

[3] Manfred Denker, Asymptotic Distribution Theory in Nonparametric Statistics, Friedr. Vieweg& Sohn, 1985

[4] Manfred Denker and Jakubowski, Stable limit distributions for strongly mixing sequences,Statistics & Probability letters 8(1989) 477-483

[5] Bingham, N. H., Goldie, C. M., Teugels, J. L., Regular Variation, Cambridge University Express, 1989

[6] J. M. Chambers, C. L. Mallows and B. W. Stuck, A Method for Simulating Stable Random Variables, JASA, Vol.71, No.354, 340-344(1976)

[7] Chow,Telin & Jozef L. Teugels, The sum and the maximum of i.i.d. random variables, Asymptotic Statistics, Proc. 2nd Prague Symp.1978,81-92(1979)

[8] Compagnone,D. & Denker,M., Nonparametric tests for scale and location, Nonparametric Statistics, Vol. 7, pp.123-154(1996)

[9] Csörgö, S., Deheuvels, P. and Mason, D. Kernel estimates of the tail index of a distribution, Ann. Statist. 13, 1050-1077(1985)

[10] E. F. Fama, The behavior of stock market prices, Journal of Business, 37, 34-105(1965)

[11] E. F. Fama R̃. Roll,Some properties of symmetric stable distributions, JASA, 63, 817-836(1968)

[12] E. F. Fama R̃. Roll, Parameter estimates for symmetric stable distributions, JASA, 66, 331-338(1971)

[13] Feller,W., An introduction to probability theory and its applications, Vol. 2, 2nd edition, Wiley, New York,1971

[14] Andrey Feuerverger & Peter Hall, Estimating a tail exponent by modelling departure from a Pareto distribution, Ann. Statist. 27, no.2, 760-781(1999)

[15] R. L. Hagerman, More evidence on the distribution of security returns, The Journal of Finance, Vol.28, No.4, 1213-1221(1978)

[16] Ibragimov,I.A. & Linnik,Yu.V., Independent and stationary sequences of random variables, Wolters-Noordhoff publishing Groningen,The Netherlands, 1971

[17] Heyde,C.C., On large deviation probabilities in the case of attraction to a non-normal stable law, SANKHY$\overline{A}$, pp253(1968)

[18] Jean-Marc Bardet, Local correlation dimension and Hausdorff dimension of continuous random fields, C. R. Acad .Sci. Paris,t.326,Série I, p.589-594(1998)

[19] L.de Haan, On regular variation and its application to weak convergence of sample extremes, Mathematisch Centrum, Amsterdam 1975

[20] L. de Haan, T. Themido Pereira, Estimating the index of a stable distribution, Statistics & Probability Letters 41, pp39–55(1999)

[21] Laurens de Haan, Sidney Resnick,A simple asymptotic estimate for the index of a stable distribution, J. Roy. Statist. Soc. Ser. B 42, 83-88(1980)

[22] Laurens de Haan, Sidney Resnick, On asymptotic normality of the Hill estimator, Commun.Statis., Stochastic Models 14, No.4, 849-866(1998)

[23] Hill,B. M., A simple general approach to inference about the tail of a distribution, Ann. Statist.3 1163-1174(1975)

[24] Peter Hall, On some simple estimates of an exponent of regular variation, J. R. Statist. Soc. B 44 No. 1, pp. 37-42(1982)

[25] Hill,B. M., Zipf's law and prior distributions for the composition of a population, JASA 65, 1220-1232(1970)

[26] Donald R. Holt and Edwin L. Crow, Tables and graphs of the stable probability density functions, J. Res. Bur. Standards Sect. B 77 143-198(1973)

[27] Hosking, J. R. M., Wallis, J. R. and Wood, E.F. Estimation of the generalised extreme-value distribution by the method of probability-weighted moments, Technometrics 27, 251-261(1985)

[28] Norman I. Johnson & Samuel Kotz, Continuous univariate distributions, Houghton Mifflin Company, Boston, 1970

[29] Matthias Löwe, Exponential inequalities and principles of large deviations for U-statistics and von Misses-statistics, diss. Uni. Bielefeld, 1992

[30] V.S. Koroljuk and Yu.V. Borovskich, Theory of U-Statistics, Kluwer Academic Publishers, 1994

[31] Ioannis A. Kouttrouvelis, Regression-Type Estimation of the Parameters of Stable Laws, JASA, Vol.75, No.372, 918-928(1980)

[32] Benoit Mandelbrot, The Pareto-Lévy law and the distribution of income, Internat. Econ. Rev. 1 79-106(1960)

[33] Benoit Mandelbrot, The variation of certain speculative prices, Journal of Business, 36, 394-419(1963)

[34] Benoit Mandelbrot, New methods in statistical economics, J. Polit. Economy, 71, 421-440(1963)

[35] Mandelbrot,B.,and H. Taylor, On the distribution of stock price differences, Operations Research, 15, 1057-1062(1967)

[36] Frank Marohn, Estimating the index of a stable law via the pot-method, Statistics &P̃robability Letters 41, pp413-423(1999)

[37] Mark M. Meerschaert and Hans-Peter Scheffler, A simple robust estimation method for the thickness of heavy tails, Journal of Statistical Planning and Inference 71, 19-34(1998)

[38] Mittnik,s. & Rachev,S.T., Tail estimation of the stable index α, Appl.Math.Lett. 9, No.3, 53-56(1996)

[39] Dietmar Pfeifer, Einführung in die Extremwertstatistik, B. G. Teubner Stuttgart, 1989

[40] Pickands, J., Statistical inference using extreme order statistics, Ann. Statist. 3, 119-131(1975)

[41] S. James Press, Estimation in univariate and multivariate stable distributions, JASA,Vol.67,No.340,842-846(1972)

[42] Rempala, Grzegorz and Gupta, Arjun, Weak Limits of U-statistics of Infinite Order, Random Operators and Stochastic Equations, 7, No.1, 39-52 (1999)

[43] Sidney I. Resnick, Catalin Starica, Asymptotic behaviour of Hill's estimator for autoregressive data, Commun.Statis., Stochastic Models 13, No.4, 703-721(1997)

[44] Sidney I. Resnick, Heavy tail modeling and teletraffic data, Ann. Statist. Vol.25, No.5, 1805-1869(1997)

[45] Samorodnitsky,G. & Taqqu,M.S., Stable non-Gaussian random processes, Chapman & Hall,1994

[46] S. Rao Jammalamadaka & Svante Janson, Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances, Ann. Prob. Vol. 14, No. 4, 1347-1358(1986)

[47] J. TeichmÏler, A note on the distribution of stock price changes, JASA 66, 282-284(1971)

[48] Teugels, J. L., Limit theorems on order statistics, Ann. Probab. 9, 868-880(1981)

[49] J. Tobin, Liquidity Preference as Behavior Toward Risk, Review of Economic Studies, 26, pp.65-86(1958)

[50] Todorovic, P., Stochastic Models of Floods, Water Resource Research 14, 345-356(1978)

[51] Zipf, G. K., National Unity and Disunity: the nation as a bio-social organism, Principia Press, Bloomington,IN(1941)

[52] V. M. Zolotarev, One-dimensional Stable Distributions, Translations of mathematical monographs,Vol.65, 1986

[53] William C. Parr & William R. Schucany, Minimum-Distance and Robust Estimation, JASA vol.75, no.371,616-624(1980)

Lebenslauf

Familienname :	Fan
Vorname :	Zhaozhi
Geschlecht :	männlich
Geburtsdatum :	18.04.1964
Geburtsort :	Shandong, China
Staatsangehörigkeit :	chinesisch
Familienstand :	verheiratet, ein Kind
09. 1972–07. 1977	Grundschule in Weifang, Provinz Shandong
09. 1977–07. 1981	Mittelschule in Weifang, Provinz Shandong
09. 1981–07. 1985	Studium an der Shandong-Universität Studienfach : Mathematik Abschluss : Bachelor of Science
09. 1985–07. 1988	Aspiranten-Studium an der Shandong-Universität Studienfach : Operations Research & Control Theory Schwerpunkte : Statistik Abschluss : Master of Science
02. 1998–02. 1999	Intensiv-Deutschkurs am Deutsch-Kolleg der Tongji Universität
Seit 03. 1999	Doktorand am Institut für mathematische Stochastik, Georg-August-Universität zu Göttingen
07. 1988–09. 1991	Assistent an der Fakultät für Management Science der Shandong-Universität
10. 1992–01. 1998	Dozent an der Fakultät für Management Science der Shandong-Universität