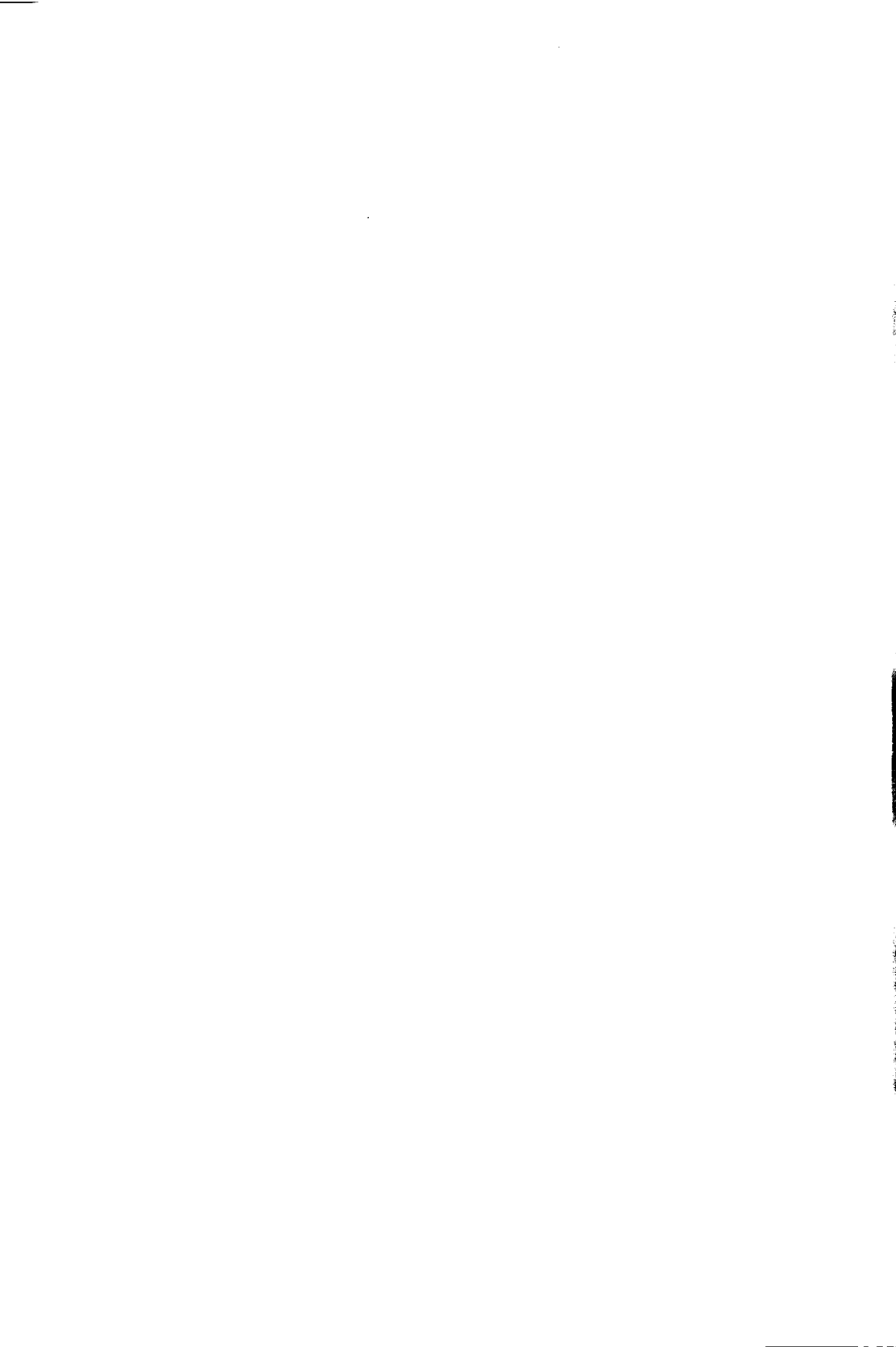


720sb
2007/15 3385

TR3385

Estimating the extreme value index

- tales of tails -



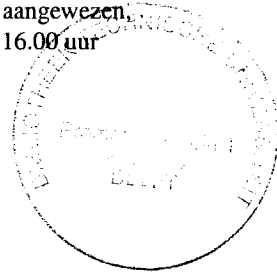
Estimating the extreme value index

– tales of tails –

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. F. Wakker,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College voor Promoties aangewezen,
op dinsdag 5 oktober 1999 te 16.00 uur

door



Peter-Paul DE WOLF

wiskundig ingenieur

geboren te Zwijndrecht

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. P. Groeneboom

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. P. Groeneboom,	Technische Universiteit Delft, promotor
Prof. dr. J. Beirlant,	Katholiek Universiteit Leuven
Prof. dr. F.M. Dekking,	Technische Universiteit Delft
Prof. dr. R.D. Gill,	Universiteit Utrecht
Prof. dr. L. de Haan,	Erasmus Universiteit Rotterdam
Dr. G. Hooghiemstra,	Technische Universiteit Delft
Dr. H.P. Lopuhaä,	Technische Universiteit Delft

Wolf, Peter-Paul de

Estimating the extreme value index, tales of tails / Peter-Paul de Wolf.

Thesis Delft University of Technology. - With ref. - With summary in Dutch.

Subject headings: extreme value estimation / Generalized Pareto / kernel type estimators

Copyright © 1999 by Peter-Paul de Wolf

TR 83855

Stellingen

behorend bij het proefschrift

Estimating the extreme value index

- tails of tails -

door

Peter-Paul de Wolf

Bij de eerste twee stellingen (beide uit: P.P. DE WOLF EN R.H. RENSSSEN (1996), *Asymptotics and the STSI design*, research paper no. 9644, CBS, Voorburg) gaan we uit van een geneste rij eindige populaties die als volgt wordt gedefinieerd: laat $\{y_v\}_{v=1}^{\infty}$ een rij elementen zijn en $\{N_v\}_{v=1}^{\infty}$ een stijgende rij integers. Definieer de v -de populatie U_v als de verzameling van de eerste N_v elementen y_1, \dots, y_{N_v} .

Ook beschouwen we de volgende rij steekproefontwerpen: neem per populatie U_v een gestratificeerde, enkelvoudig aselechte steekproef zonder teruglegging (aan de hand van een vast allocatieschema) ter grootte n_v . Voor $h = 1, \dots, H_v$ worden de strata aangegeven met U_{vh} , hun grootte met N_{vh} en de allocatie met n_{vh} .

Definieer de volgende grootheden: $\bar{Y}_{vh} = (\sum_{i \in U_{vh}} y_i) / N_{vh}$, $S_{vh}^2 = (\sum_{i \in U_{vh}} (y_i - \bar{Y}_{vh})^2) / (N_{vh} - 1)$, $W_{vh} = N_{vh} / N_v$, $\lambda_{vh} = (N_{vh} n_v) / (N_v n_{vh})$ en $f_{vh} = n_{vh} / N_{vh}$.

Stelling 1. Naast de natuurlijke voorwaarde dat de steekproefgrootte n_v naar oneindig moet gaan, zijn in bovengenoemde situatie de voorwaarden

- (a) De populaties U_v moeten zodanig zijn dat $\sum_{h=1}^{H_v} W_{vh} S_{vh}^2$ een eindige limiet heeft als $v \rightarrow \infty$.
- (b) De allocatie moet zodanig zijn dat er een constante $l_u < \infty$ bestaat, zodanig dat als $v \rightarrow \infty$, $\lambda_{vh} \leq l_u$ voor alle h met $S_{vh}^2 \neq 0$.

voldoende voor de consistentie van de Horvitz-Thompson schatter van het populatie gemiddelde, in die zin dat voor iedere $\varepsilon > 0$

$$\lim_{v \rightarrow \infty} \mathbb{P} \left(|\hat{Y}_v - \bar{Y}_v| > \varepsilon \right) = 0$$

waarbij \hat{Y}_v de Horvitz-Thompson schatter is van het populatie gemiddelde \bar{Y}_v van populatie U_v .

Stelling 2. Zij $\eta_v = \sqrt{n_v} (\hat{Y}_v - \mathbb{E} \hat{Y}_v)$. Neem voor v groot genoeg aan dat voor alle h geldt dat $N_{vh} \geq 2$, dat $H_v \leq \mathcal{H} < \infty$ en dat

$$\lim_{v \rightarrow \infty} \frac{\sum_{h=1}^{H_v} \lambda_{vh} W_{vh} \sqrt{(1-f_{vh})/n_{vh}} S_{vh}^2}{\sum_{h=1}^{H_v} \lambda_{vh} W_{vh} (1-f_{vh}) S_{vh}^2} = 0$$

Definieer

$$D_{N_v}^2 = \text{Var}(\eta_v) = \sum_{h=1}^{H_v} \lambda_{vh} W_{vh} (1-f_{vh}) S_{vh}^2$$

Dan is η_v / D_{N_v} asymptotisch normaal verdeeld, dan en slechts dan als,

$$\frac{\sum_{h=1}^{H_v} \lambda_{vh} (1-f_{vh})^2 \sum_{i \in C_{vhr}} (y_i - \bar{Y}_{vh})^2}{\sum_{h=1}^{H_v} \lambda_{vh} (1-f_{vh}) \sum_{i \in U_{vh}} (y_i - \bar{Y}_{vh})^2} + \frac{\sum_{h=1}^{H_v} \lambda_{vh} (1-f_{vh}) f_{vh} \sum_{i \in D_{vhr}} (y_i - \bar{Y}_{vh})^2}{\sum_{h=1}^{H_v} \lambda_{vh} (1-f_{vh}) \sum_{i \in U_{vh}} (y_i - \bar{Y}_{vh})^2} \rightarrow 0$$

als $v \rightarrow \infty$, voor alle $\tau > 0$, waarbij C_{vht} en D_{vht} gegeven zijn door

$$C_{vht} = \mathcal{U}_{vht} \sqrt{\frac{f_{vh}}{1-f_{vh}}} \quad \text{en} \quad D_{vht} = \mathcal{U}_{vht} \sqrt{\frac{1-f_{vh}}{f_{vh}}}$$

met

$$\mathcal{U}_{vht} = \left\{ i \in U_{vh} : |y_i - \bar{Y}_{vh}| > \tau \sqrt{\frac{\sum_{k=1}^{H_v} \lambda_{vk}(1-f_{vk})(N_{vk}-1)S_{vk}^2}{\lambda_{vh}(1-f_{vh})}} \right\}$$

Stelling 3 en Stelling 4 hebben betrekking op de volgende situatie: Zij $\{\xi_i\}_{i=1}^n$ een gegeven rij scores van een categoriale variabele ξ die K categorieën heeft. Definieer de variabele X door middel van de overgangskansen $\mathbb{P}(X = l | \xi = k) = p_{kl}$ voor $k, l = 1, \dots, K$. De 'Post Randomisation Method' (PRAM) bestaat dan uit het uitvoeren van een kansexperiment met genoemde overgangskansen, waardoor de scores $\{X_i\}_{i=1}^n$ worden verkregen. Dat kansexperiment wordt dan volledig bepaald door de $K \times K$ Markov-matrix P met elementen p_{kl} .

Stelling 3. In hoeverre de 'Post Randomisation Method' (PRAM) een bruikbare methode is voor het beveiligen van microdata tegen spontane herkenning, hangt mede af van de statistische kennis van de gebruiker van de microdata.

Stelling 4. Zij P een Markov-matrix met overgangskansen p_{kl} . Pas met deze matrix PRAM toe op de scores $\{\xi_i\}_{i=1}^n$ om de scores $\{X_i\}_{i=1}^n$ te verkrijgen. Definieer de matrix P^* als de Markov-matrix met de elementen

$$p_{ik}^* = \frac{p_{kl} T_{\xi}(k)}{\sum_j p_{jl} T_{\xi}(j)}$$

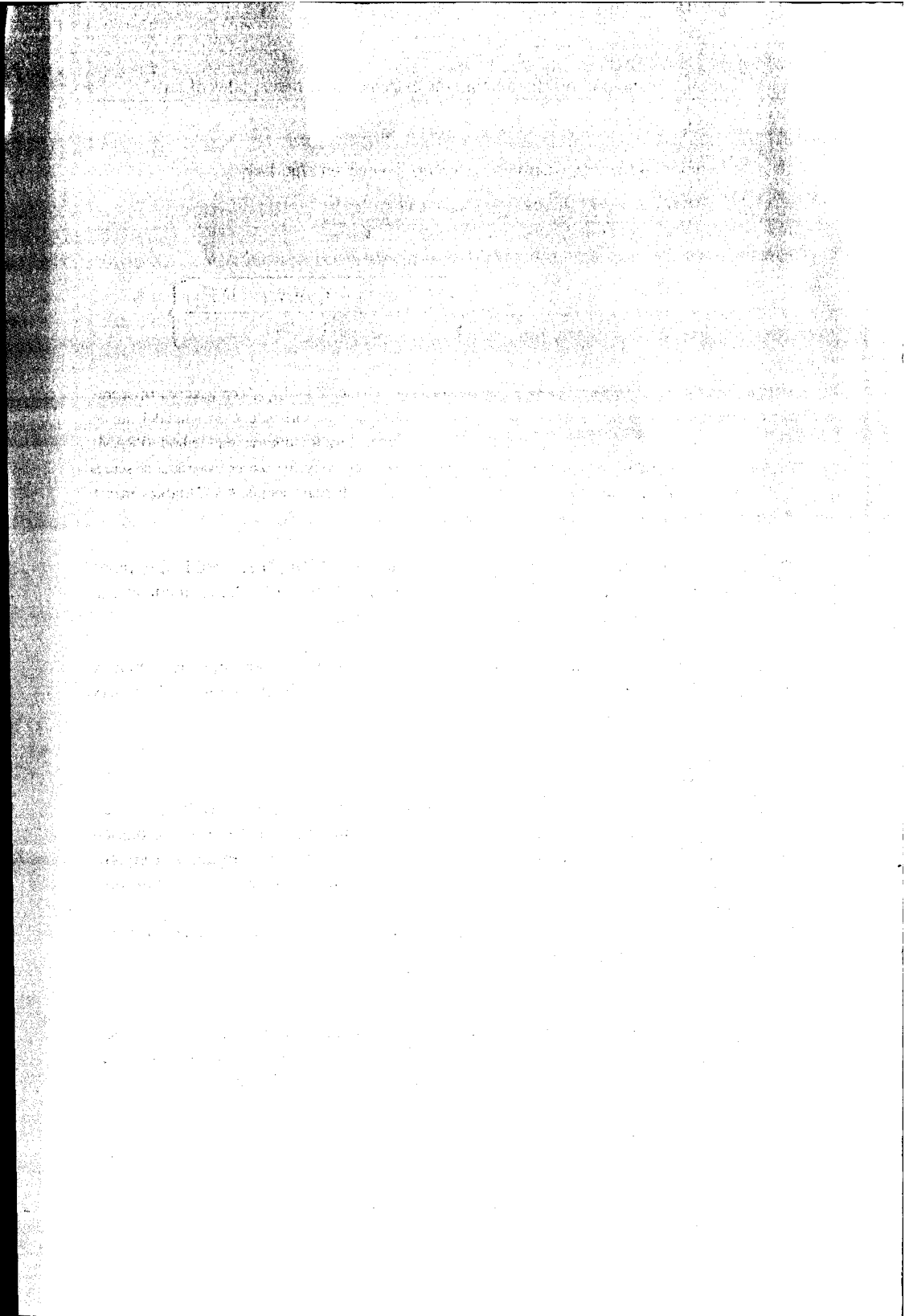
waarbij T_{ξ} de frequentietabel is van de scores van de categoriale variabele ξ , i.e., $T_{\xi}(k) = \sum_{i=1}^n 1_{\{\xi_i=k\}}$ voor $k = 1, \dots, K$. Door met P^* nogmaals PRAM toe te passen op de scores $\{X_i\}_{i=1}^n$ worden scores $\{X_i^*\}_{i=1}^n$ verkregen waarvoor geldt dat de frequentietabel T_{X^*} van de uiteindelijk verkregen scores een zuivere schatter is voor de oorspronkelijke frequentietabel T_{ξ} .

P.P. DE WOLF, J. GOUWEELEEUW, P. KOOIMAN EN L. WILLENBORG (1997), *Reflections on PRAM*, research paper no. 9742, CBS, Voorburg.

Stelling 5. Het gebruik van een afstootcurve ongelijk aan de stapfunctie op $[0, L]$, waarbij L de gemiddelde levensduur van een kapitaalgoed is, is van grotere invloed op de berekening van kapitaalgoederenvoorraad dan menig econometrist dacht.

Stelling 6. Een uitspraak die begint met "Het is statistisch bewezen dat..." is bijna zeker niet waar.

Stelling 7. De toestand van een kwade Schrödinger kat kan worden beschreven met een Cauchy verdeling.



Acknowledgements

Many people have supported me in many ways, before and during the time I was writing this thesis. Since thanking all of them individually for their support would take a lot of time and they must eagerly be looking forward to read all that I have written, I will only mention some of them by name.¹

First of all I would like to thank Piet for the time we spent discussing many mathematical, statistical and musical themes. He introduced me to the field of extreme value theory and together we strolled along some of the winding roads. Even though we had to backtrack several times because some of these roads were dead ends, I think we finally managed to arrive at a promising place.

Another guide during my Ph.D.-trip was Rudolf. At the beginning of my TU Delft era, we had several stimulating discussions on many subjects. Later on, Rudolf was replaced by Rik who helped me to complete my thesis.

At TU Delft it were Geurt, Ronald and René who sometimes stimulated me to continue my research, either by coincidence or by request. Leo helped me whenever I needed system support on the Unix-machines.

When I interchanged the University with Statistics Netherlands as employer, Peter allowed me to revisit the university several times without charging me any time or any money, even though my estimate of the needed time was not that good... However, finally I finished.

If I had not started studying mathematics in the first place, this thesis would never have been written.² My parents have supported me from the beginning of my study, and they still do, so I am very grateful to them and I am glad to be able to present them this thesis.

At some time during the writing of this thesis, Carolien came into my life. I know it must have been difficult for her to deal with such an obsessed man like me, but she did and I am grateful that she supported me all that time as well.

Finally I want to thank Remco who, even though he did not know it, caused the concluding processes to gain enough momentum so I could finish this thesis in time to grant him the time he deserves.

¹Those that do not find their names on this page should know that they have only temporarily escaped my mind.

²At least not by me.

Contents

1	Introduction	1
1.1	Examples of extremes	1
1.2	Questions concerning extreme situations	3
1.3	Some extreme value theory	4
1.3.1	Limiting distributions	5
1.3.2	Alternative formulations of our main assumption	8
1.3.3	The domains of attraction	9
1.3.4	Regular variation	12
1.4	Scope of this thesis	14
2	Estimation of the extreme value index	17
2.1	Minimax lower bounds	17
2.1.1	Minimax risk in the setting of the extreme value index	18
2.1.2	Minimax lower bound: a positive extreme value index	20
2.1.3	Minimax lower bound: a negative extreme value index	27
2.2	Pickands' estimator	28
2.3	A maximum likelihood approach	32
2.4	Hill's estimator	35
2.5	A moment estimator	37
2.6	A kernel type estimator	40
2.7	Sharpness of the lower bounds to the minimax risk	43
3	A kernel type estimator revisited	47
3.1	Introduction	47
3.2	Main results	49
3.3	Basic decomposition	53
3.4	Proofs	56
3.4.1	Proof of Theorem 3.1	56
3.4.2	Proof of Corollary 3.1	59
3.4.3	Proof of Theorem 3.2	59
3.4.4	Proof of Theorem 3.3	62

4	A new type of estimator	65
4.1	Introduction	65
4.2	Main results	66
4.3	Proof of Theorem 4.1	72
4.3.1	The random terms	74
4.3.2	The deterministic part	84
4.4	Proof of Theorem 4.2	85
4.5	Proof of Theorem 4.3	86
4.6	Dependence on the smoothing parameters	89
5	Alternative kernel type estimators	95
5.1	Introduction	95
5.2	Defining the estimator	96
5.3	Main results	98
5.4	Consistency	102
5.4.1	Preliminaries	102
5.4.2	The positive part	110
5.4.3	The negative part	112
5.5	Asymptotic normality	114
5.5.1	Preliminaries	114
5.5.2	The various parts of the estimator	115
5.5.3	The various parts together	120
5.6	Exploring the bias	122
5.6.1	Preliminaries	122
5.6.2	The bias under Condition 5.3	123
5.6.3	The bias under Condition 5.4	126
6	A simulation study	129
6.1	Introduction	129
6.2	Summary of the estimators	130
6.3	Finite sample behaviour	132
6.3.1	The Hall model	133
6.3.2	Generalized Extreme Value distributions	135
6.3.3	Uniform distribution	136
6.3.4	Hall type for negative index	137
6.3.5	Generalized Pareto	138
6.3.6	The Lobith data	139
6.4	Discussion of the results	139
	Bibliography	143
	Summary	147
	Samenvatting (Summary in Dutch)	150

Curriculum Vitae



Chapter 1

Introduction

This thesis focuses on statistical aspects of extreme value theory. In this chapter we will provide some examples of extremes and some background information on extreme value theory. In the last section we will briefly describe the contents of the subsequent chapters of this thesis.

1.1 Examples of extremes

Nowadays it seems common practice to use statistics to describe or even predict the average behaviour of many different phenomena. However, in many situations it is not only that average behaviour that is an interesting subject of research. As a simple illustration, consider a car tire. Under normal, average circumstances, the tire becomes weaker due to the accumulation of many small damages (i.e., due to wear and tear). And, eventually, it might blow and cause an accident. On the other hand, if the car hits an unexpectedly sharp rock in the middle of the road, the tire might blow just that instance. In the latter case, the accident is caused by one extreme damage. This simple example shows, that it seems quite natural that one does not restrict oneself to the analysis of the average situations alone, but that one considers extreme situations as well. Actually, in many different areas, there is a dire need to describe extreme situations in order to try to prevent or at least reduce severe damage, financial or otherwise. Despite the non-average behaviour, statistics is still a useful tool.

At least as far back as the Egyptian era, people have been interested in describing extreme situations. In the early ages of their reign, the Egyptians were already interested in a more or less accurate prediction of the flooding of the river Nile, which is obviously an extreme event: it is caused by an excessively high water level. Their interest was mainly induced by agricultural considerations: inundation of the land immediately surrounding the river could improve the soil. Proper use of this natural fertilization, benefitted their crop.

However, besides this positive agricultural effect, flooding can just as well have a disastrous effect: especially if society is not prepared well enough, extensive human and material loss might be a result of the same phenomenon.

Over the years, the Dutch people have found themselves in such a position that they had to deal with exactly these kind of problems. Since a major part of their country is below mean sea level and the northern and western part is adjacent to the North Sea, it is quite vulnerable to the behaviour of that sea during extreme weather conditions. Indeed, in 1953 a major part of The Netherlands was flooded by the North Sea, due to a severe storm in conjunction with spring tide. This disaster caused a considerable amount of human and material loss and initiated the so called Delta plan that had to safeguard the country against future threats of the North Sea.

Unfortunately, as recently as in 1995, the Dutch people were threatened by yet another source of flooding. In the first part of that year, very high water levels in some major rivers gave rise to the evacuation of thousands of people in nearby villages. Fortunately, this time there were no direct casualties, but it initiated a renewed discussion on the strengthening of the river dikes that was delayed due to environmental arguments.

Flooding is obviously not the only natural phenomenon that can be studied using extreme value theory. Other situations one could think of are e.g., extremely high or low temperatures, extreme atmospheric pressure, excessive rainfall, droughts, etc. In case of extreme temperatures one could think of e.g., the melt-down temperature of nuclear plants, or of temperatures close to 0 Kelvin.

An interesting aspect in the situation of low temperatures and droughts, is the presence of a natural lower bound to the measurements. In these cases the extreme events are the measurements extremely close to this bound. Obviously, these considerations do not only apply to lower bounds but to upper bounds as well. As examples in which an upper bound is present, though its value might be unknown, one could think of e.g., the lifetime of certain species or of sports records such as in high jumping, triple jumping, 100 meter sprint, etc.

In the analysis of strength of materials, it is also desirable to predict the behaviour of constructions like oil-platforms and large buildings under extreme conditions. Often civil engineers use so called 'design conditions' based on the effect of extreme forces and moments induced on those constructions by storms, earthquakes and similar extreme processes.

Corrosion is yet another situation in which extremes are of interest. A surface with a large number of small pits fails due to corrosion if any one of the pits penetrates through the thickness of the surface. The depths of the pits, though initially random, increase in time due to corrosion. Hence the deepest pit, i.e., an extreme event, causes the failure.

In economical and financial analysis, extreme value theory has become a very important issue as well. In non-life insurance, portfolios often contain claims that should be considered extreme rather than average. Actually, the whole area of re-insurance is a very important application of extreme value theory, since it usually has to safeguard an insurance company against excessive claims that may endanger the solvency of that company.

A more theoretical area in which extreme value analysis can be used, is the field of statistical estimators itself. Even though such estimators are ideally accurate, very large or very small realizations are likely to occur. Hence, investigation of the behaviour of

'extreme realizations' of such an estimator is of considerable interest. In a sense it is a measure for the accurateness of the estimator.

In the analysis of (pseudo) random number generators, the extreme numbers that are generated (i.e., the numbers close to the boundaries of the interval in which the numbers are generated) are interesting subjects for extreme value theory.

1.2 Questions concerning extreme situations

In the situations described in the previous section, some interesting questions arise, concerning extreme events. We already mentioned the Delta plan, in which the Dutch government had to answer a very important and quite illustrative question, in order to find the right way to strengthen the sea shore of The Netherlands: a very high quantile had to be estimated, on which they could base the design conditions of the new system that should defend their country against the all-time enemy, the North Sea. More formally, the strengthening had to be such that the probability of flooding was extremely small, say e.g., 0.01% per annum. Loosely speaking that means that the new sea defense should, on average, only be defeated once in ten thousand years.

A typical problem that arises in estimating such a very high quantile is quite apparent in this problem: the quantile to be estimated concerned the event of flooding only once in ten thousand years. However, the available measurements spanned a much shorter time interval. In other words, a value had to be estimated that, with high probability, was far out of the range of the measurements. Note however, that there is a positive probability that an observation larger than that very high quantile was already present in the measurements.

In the situation that a (natural) bound to the behaviour of the investigated phenomenon is present, other interesting statistical questions appear. First of all, the bound could be known beforehand, like e.g., in case of temperatures close to 0 Kelvin. In that case the way in which the distribution tends to that bound is the interesting feature to investigate. In other situations, the bound may be known to be present, even though the exact value is not known. In that case it is not only interesting to estimate the behaviour of the distribution function close to that bound, but estimating that value quite accurately is just as interesting. The third possibility is that one does not know beforehand, whether a bound is present or not. Obviously, detecting that possible bound might be very important.

Actually, in the aforementioned situation of flooding, one would be tempted to say that there is a natural bound to the high water level, even though its value is unknown. Strictly speaking, the fact that there is only a finite amount of water on earth bounds the extreme sea level. A less absurd bound is given by the fact that the (finite) water depth limits the height of waves. However, whether or not it is possible to deduce the existence of such a bound purely based on the measurements, is an interesting question itself. Recently, statistical research did suggest the presence of a bound in the situation of extreme wave heights.

Sometimes, the estimation of very high quantiles on its own, does not provide us

with enough information. In certain situations, the behaviour of the probability distribution of the subject of interest beyond a certain (extremely) high threshold as a whole, is needed in further research. That part of the distribution function is usually called its (extreme) tail. One could argue that being able to estimate an arbitrary (high) quantile should provide us with enough information to estimate the whole tail. Unfortunately, this is not quite right in the same sense that knowledge of the pointwise behaviour of a function does not imply that one knows the uniform behaviour of that function as well. Moreover, though confidence intervals of estimated quantiles do provide some information about the tail, they do not necessarily contain enough information to construct confidence bounds for the whole tail of that distribution.

All the questions raised in this section, tacitly assume that the (extreme) tail of a distribution is well defined. Even though in theory it is very well possible to give a precise definition, in practice that might be much harder. Usually it is not at all clear, above which threshold the extreme tail starts. In this thesis we will try to shed some more light on this fundamental question in extreme value analysis.

1.3 Some extreme value theory

In order to be able to deal with questions like the ones raised in the previous section using statistical and probabilistic methods, we will assume the considered measurements to be realizations of independent random variables, i.e., we will consider the data to be a realization of a sample X_1, \dots, X_n from a common distribution function F . For a discussion on the effect of dependency between the X_i 's see LEADBETTER, LINDGREN AND ROOTZÉN (1983).

Even though both theoretically and conceptually very interesting aspects arise considering higher dimensional distributions, we will only consider the one dimensional case in this thesis.

Obviously, inference about extremely small measurements can be translated into inference about extremely large measurements, using the trivial relation $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$. We shall therefore only be concerned with large extremes and henceforth use the term 'extreme events' to indicate events concerning large extremes.

Intuitively, the distribution of the largest value(s) of a sample should provide us with some information about extreme events. Consequently, we could consider the distribution of the maximum of the sample, i.e., the distribution of $\max(X_1, \dots, X_n)$. However, the limit of that distribution when the sample size n tends to infinity, is degenerate: depending on the underlying distribution function F , the maximum of the sample will tend either to infinity or to a finite constant value. In order to try to obtain a non-degenerate limiting distribution, one could think of applying a transformation to the maximum of the sample. Hence we arrive at the following main assumption used in this thesis: there exist sequences of real numbers $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}$, with $a_n > 0$ and $b_n \in \mathbb{R}$, such that

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \lim_{n \rightarrow \infty} P \left\{ \frac{\max(X_1, \dots, X_n) - b_n}{a_n} \leq x \right\} = G(x) \quad (1.1)$$

for all x , where G is a non-degenerate distribution function¹. Then F is said to be in the domain of attraction of G , denoted by $F \in \mathcal{D}(G)$. This assumption only induces a restriction on the upper tail of the distribution function F . This is quite natural, since we are interested in the region beyond the largest observation and hence we will need some restriction on the way the distribution function could possibly continue beyond that largest observation. Without such an assumption, inference in that region would be impossible: if we allow the tail of the distribution function to change in an arbitrary way, there is no hope we could ever sensibly extrapolate beyond the largest measurement. On the other hand, any assumption on the upper tail of a distribution limits the possible behaviour of that tail. Fortunately, a large class of distribution functions satisfies assumption (1.1), including many of the distribution functions that are commonly used in practice, like e.g., the Normal, Uniform, Exponential, Weibull and Cauchy distribution.

1.3.1 Limiting distributions

In standard central limit theory one obtains an asymptotic normal distribution for the *sum* of many independent and identically distributed random variables, regardless of the underlying common distribution function, provided some mild conditions (e.g., existence of second moment). Moreover, in order to be able to apply the asymptotic theory, one does not need to know the behaviour of the underlying distribution function *in every detail*: e.g., knowledge about the first two moments suffices.

In extreme value theory, a similar situation occurs: the limiting distribution G in (1.1) can only come from a limited class of distribution functions, regardless of the underlying distribution function F . As stated before, it is only the behaviour of the tail of the distribution that determines to which domain of attraction F belongs.

That result in extreme value theory, first discovered by FISHER AND TIPPET (1928) and later proved in complete generality by GNEDENKO (1943), is given by the following theorem:

Theorem 1.1

Suppose distribution function F satisfies assumption (1.1) with limiting distribution G . Then G is the same, up to location and scale, as one of the following distributions:

$$\begin{array}{ll}
 \text{Type I:} & \Lambda(x) = \exp(-e^{-x}) \quad x \in \mathbb{R} \\
 \text{Type II:} & \Phi_{\alpha}(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} \quad \text{with } \alpha > 0 \\
 \text{Type III:} & \Psi_{\alpha}(x) = \begin{cases} \exp(-(-x)^{\alpha}) & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \text{with } \alpha > 0
 \end{array}$$

We will refer to these distributions as the extreme value distributions.

¹We will call this kind of convergence of a sequence of distribution functions either 'weak convergence' (denoted by \xrightarrow{w}) or 'convergence in distribution' (denoted by \xrightarrow{D})

Proof of Theorem 1.1:

See e.g., GNEDENKO (1943), DE HAAN (1970) or RESNICK (1987). ■

Alternatively, Type I is also known as Gumbel or double exponential, Type II as Fréchet or heavy tailed and Type III as (reverse) Weibull type. Note that the right tail of a Weibull distribution is of Type I and the left tail is of Type III, hence the alternative name *reverse* Weibull for Type III. The terminology Type I, II and III is according to Gumbel's characterization in GUMBEL (1958).

The statement that the limiting distribution G in (1.1) must be the same as one of the extreme value distributions *up to location and scale*, introduces some freedom in the choice of the normalizing constants $\{a_n\}$ and $\{b_n\}$. Actually, this is a consequence of the following theorem, known as Khintchine's theorem:

Theorem 1.2 (Khintchine's theorem)

Let $\{F_n\}$, $n \in \mathbb{N}$, be a sequence of distribution functions and H be a non-degenerate distribution function. Let $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}$, with $a_n > 0$ and $b_n \in \mathbb{R}$ be such that

$$F_n(a_n x + b_n) \xrightarrow{w} H(x)$$

Then, for some non-degenerate distribution function H^* and sequences $\{\alpha_n\}$ and $\{\beta_n\}$, $n \in \mathbb{N}$, with $\alpha_n > 0$ and $\beta_n \in \mathbb{R}$,

$$F_n(\alpha_n x + \beta_n) \xrightarrow{w} H^*(x)$$

if and only if

$$\frac{\alpha_n}{a_n} \rightarrow A \quad \text{and} \quad \frac{\beta_n - b_n}{a_n} \rightarrow B$$

for some $A > 0$ and $B \in \mathbb{R}$. Moreover,

$$H^*(x) = H(Ax + B)$$

Proof:

See e.g., LEADBETTER, LINDGREN AND ROOTZÉN (1983) (Theorem 1.2.3) or RESNICK (1987) (Proposition 0.2). ■

An application of this theorem yields the following: using a suitable choice of the sequences $\{a_n\}$ and $\{b_n\}$, the three limiting types may be combined into the single form

$$G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right) \quad \text{for all } x \text{ such that } 1 + \gamma x > 0 \quad (1.2)$$

with $\gamma \in \mathbb{R}$ and the convention that $G_0(x) = \lim_{\gamma \rightarrow 0} G_\gamma(x) = \exp(-e^{-x})$ for $x \in \mathbb{R}$. This G_γ is called the Generalized Extreme Value distribution with extreme value index γ , abbreviated to $GEV(\gamma)$. If $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in \mathbb{R}$, we say that the distribution function F itself has extreme value index γ as well. One of the first papers in which this single form appeared is due to VON MISES (1936).

Moreover, Theorem 1.2 shows that the extreme value index γ as function of the distribution F is well defined, in the sense that it is uniquely determined, independent of the choice of the sequences $\{a_n\}$ and $\{b_n\}$.

Obviously, we have the following relations between Λ , Φ_α , Ψ_α and G_γ :

$$\Lambda(x) = G_0(x)$$

$$\Phi_\alpha(x) = G_{1/\alpha}(\alpha(x-1)) \quad \alpha > 0$$

$$\Psi_\alpha(x) = G_{-1/\alpha}(\alpha(x+1)) \quad \alpha > 0$$

Hence, the three extreme value distributions can be characterized by the sign of γ : Type I corresponds to $\gamma = 0$, Type II to $\gamma > 0$ and Type III to $\gamma < 0$. In Figure 1.3.1 typical examples of these three situations are shown.

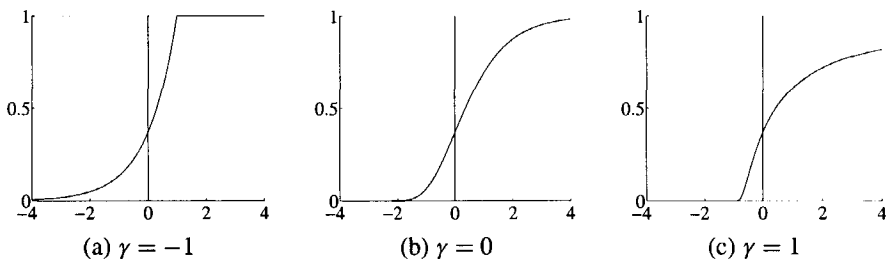


Figure 1.3.1: Three Generalized Extreme Value Distributions

The two cases $\gamma < 0$ and $\gamma > 0$ are related in the following way: Suppose that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma < 0$. Defining the *upper endpoint* of a distribution F by

$$x_F^\circ = \sup\{x : F(x) < 1\} \leq \infty$$

and putting $\tilde{F}(x) = F(x_F^\circ - 1/x)$ then yields that $\tilde{F} \in \mathcal{D}(G_{-\gamma})$. The opposite relation holds as well: if $F \in \mathcal{D}(G_\gamma)$ for some $\gamma > 0$ then the distribution

$$\tilde{F}(x) = \begin{cases} F\left(\frac{1}{x^\circ - y}\right) & y < x^\circ \\ 1 & y \geq x^\circ \end{cases}$$

for some $x^\circ > 0$ is in the domain of attraction of $G_{-\gamma}$ and has upper endpoint x° .

The following properties can easily be derived from the definitions of the three limiting types:

1. If $\gamma < 0$ then $x_{G_\gamma}^\circ < \infty$.
2. If $\gamma \geq 0$ then $x_{G_\gamma}^\circ = \infty$.

3. In case $\gamma > 0$ the integrals

$$\int_1^{\infty} t^{\rho} dG_{\gamma}(t)$$

are finite for all $\rho < 1/\gamma$ and infinite for all $\rho > 1/\gamma$.

In case $\gamma = 0$ these integrals are finite for all ρ .

These properties also hold, to some extent, for distributions F that are in the domain of attraction of one of these extreme value distributions:

1'. If $F \in \mathcal{D}(G_{\gamma})$ for some $\gamma < 0$ then $x_F^{\circ} < \infty$.

2'. If $F \in \mathcal{D}(G_{\gamma})$ for some $\gamma > 0$ then $x_F^{\circ} = \infty$. If $F \in \mathcal{D}(G_0)$ then x_F° can be finite as well as infinite.

3'. If $F \in \mathcal{D}(G_{\gamma})$ for some $\gamma > 0$ the integrals

$$\int_1^{\infty} t^{\rho} dF(t)$$

are finite for all $\rho < 1/\gamma$ and infinite for all $\rho > 1/\gamma$.

If $F \in \mathcal{D}(G_0)$ with $x_F^{\circ} = \infty$ then these integrals are finite for all ρ .

1.3.2 Alternative formulations of our main assumption

As mentioned before, assumption (1.1) induces a restriction on the upper tail of the underlying distribution function F . The following equivalent form of that assumption, cf. BALKEMA AND DE HAAN (1974), illustrates that restriction.

Let x° be the upper endpoint of F . Then $F \in \mathcal{D}(G_{\gamma})$ if and only if, for some positive function $\alpha(\cdot)$

$$\lim_{t \uparrow x^{\circ}} \frac{1 - F(t + x\alpha(t))}{1 - F(t)} = -\log G_{\gamma}(x) = (1 + \gamma x)^{-1/\gamma} \quad (1.3)$$

for all $x > 0$ with $1 + \gamma x > 0$. Intuitively this means that we relate the behaviour of F at the more extreme quantiles $t + x\alpha(t)$ to its behaviour at the moderate quantiles t .

A slightly different formulation was used by PICKANDS (1975): consider the conditional probability of X not exceeding $t + x$ given it has already exceeded t , i.e., consider

$$F_t(x) = \frac{F(t+x) - F(t)}{1 - F(t)} \quad x > 0 \quad (1.4)$$

Then the following holds for any continuous distribution function F : $F \in \mathcal{D}(G_{\gamma})$ for some $\gamma \in \mathbb{R}$ if and only if, for some positive function $\alpha(\cdot)$

$$\lim_{t \uparrow x^{\circ}} \sup_{0 \leq x \leq x^{\circ} - t} |F_t(x) - F_{GPD}(x; \gamma, \alpha(t))| = 0 \quad (1.5)$$

where $F_{GPD}(x; \gamma, \sigma)$ is the distribution function of a Generalized Pareto Distribution, defined for $\sigma > 0$ and $\gamma \in \mathbb{R}$ as

$$F_{GPD}(x; \gamma, \sigma) = 1 - (1 + \gamma x / \sigma)^{-1/\gamma} \quad (1.6)$$

with $x > 0$ and $1 + \gamma x / \sigma > 0$ and the convention that

$$F_{GPD}(x; 0, \sigma) = \lim_{\gamma \rightarrow 0} F_{GPD}(x; \gamma, \sigma) = 1 - e^{-x/\sigma}$$

for $x > 0$. Formally this means that, for a continuous distribution function F , $F \in \mathcal{D}(G_\gamma)$ if and only if the tail of F , normalized to be a distribution in the sense of (1.4), is approximately a Generalized Pareto distribution. A simple calculation shows that, if F is a Generalized Pareto distribution with parameters γ and σ then its tail F_t is also a Generalized Pareto distribution, but with the parameters γ and $(\sigma + \gamma t)$.

Another equivalent formulation of the domain of attraction assumption (1.1) is in terms of the inverse of the distribution function, i.e., in terms of the quantile function. Define the quantile function $Q(\cdot)$ as

$$Q(s) = F^{-1}(s) = \inf\{x : F(x) \geq s\} \quad \text{for } 0 < s < 1$$

Then (1.1) is equivalent to the existence of a positive function a such that

$$\lim_{s \downarrow 0} \frac{Q(1 - sx) - Q(1 - s)}{a(s)} = \frac{x^{-\gamma} - 1}{\gamma} \quad (1.7)$$

for $x > 0$, $\gamma \in \mathbb{R}$ and the convention that $(x^{-\gamma} - 1)/\gamma = -\log x$ for $\gamma = 0$. Quite often one will find this formulation in terms of the function $U(\cdot)$ defined by

$$U(x) = \left(\frac{1}{1 - F} \right)^{-1}(x) \quad 0 < x < \infty$$

Then (1.7) reads

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{\tilde{a}(t)} = \frac{x^\gamma - 1}{\gamma}$$

for $x > 0$, $\gamma \in \mathbb{R}$, with $\tilde{a}(t) = a(1/t)$ and the convention that $(x^\gamma - 1)/\gamma = \log x$ for $\gamma = 0$.

1.3.3 The domains of attraction

Various necessary and sufficient conditions are known, that can be used to determine to which (if any) domain of attraction a specific distribution function belongs. These conditions obviously involve the 'tail-behaviour' of that distribution function, i.e., they are concerned with $1 - F(x)$ as x increases. Rather simple but useful sufficient conditions can be found in VON MISES (1936). These conditions apply whenever the tail of the distribution function has a (second) derivative close to its upper endpoint.

We will state these von Mises conditions in terms of the three limiting distributions Λ , Φ_α and Ψ_α .

Theorem 1.3 (The von Mises conditions)

Let F be a distribution function that is absolutely continuous in a (left) neighbourhood of x_F° , with density f . Then sufficient conditions for F to belong to one of the three possible domains of attraction are:

1. Suppose f has a negative derivative f' for all x in some interval (x_1, x_F°) , $f(x) = 0$ for $x \geq x_F^\circ$ and

$$\lim_{t \uparrow x_F^\circ} \frac{f'(t)(1-F(t))}{(f(t))^2} = -1$$

Then $F \in \mathcal{D}(\Lambda)$.

2. Suppose $f(x) > 0$ for all x in some interval (x_1, ∞) and for some $\alpha > 0$

$$\lim_{t \rightarrow \infty} \frac{tf(t)}{1-F(t)} = \alpha$$

Then $F \in \mathcal{D}(\Phi_\alpha)$.

3. Suppose $f(x) > 0$ for all x in some interval (x_1, x_F°) , $f(x) = 0$ for all $x > x_F^\circ$ and for some $\alpha > 0$

$$\lim_{t \uparrow x_F^\circ} \frac{(x_F^\circ - t)f(t)}{1-F(t)} = \alpha$$

Then $F \in \mathcal{D}(\Psi_\alpha)$.

Proof:

See e.g., DE HAAN (1976). ■

These sufficient conditions are actually quite close to the necessary and sufficient conditions, as is shown in the next theorem:

Theorem 1.4

Necessary and sufficient conditions for a distribution function F to belong to one of the three possible domains of attraction are:

1. $F \in \mathcal{D}(\Lambda)$ if and only if there exists some strictly positive function $g(t)$ such that

$$\lim_{t \uparrow x_F^\circ} \frac{1-F(t+xg(t))}{1-F(t)} = e^{-x} \quad (1.8)$$

for all real x . Moreover, if (1.8) holds for some strictly positive function g , then it also holds for the following choice of g :

$$g(t) = \frac{\int_t^{x_F^\circ} (1-F(u)) du}{1-F(t)}$$

for $t < x_F^\circ$.

2. $F \in \mathcal{D}(\Phi_\alpha)$ for some $\alpha > 0$ if and only if $x_F^\circ = \infty$ and

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}$$

for each $x > 0$.

3. $F \in \mathcal{D}(\Psi_\alpha)$ for some $\alpha > 0$ if and only if $x_F^\circ < \infty$ and

$$\lim_{h \downarrow 0} \frac{1 - F(x_F^\circ - xh)}{1 - F(x_F^\circ - h)} = x^\alpha$$

for each $x > 0$.

Proof:

See e.g., GNEDENKO (1943). ■

Moreover, the von Mises conditions are necessary and sufficient for F to belong to a ‘twice differentiable’ domain of attraction $\mathcal{D}^{(2)}(G)$, i.e., the first and second derivatives of the distribution of the scaled maximum converge to the corresponding first and second derivatives of the limit. More formally, $F \in \mathcal{D}^{(2)}(G)$ if

$$(a_n)^l (F^n)^{(l)}(a_n x + b_n) \rightarrow G^{(l)}(x) \quad \text{as } n \rightarrow \infty \quad l = 0, 1, 2$$

locally uniformly² on the support of the appropriate extreme value distribution G , where $h^{(l)}$ denotes the l -th derivative of the function h with respect to its argument. (See PICKANDS (1986) and SMITH (1988))

Another way to see how well the von Mises conditions describe the needed properties of distributions belonging to one of the three domains of attraction, uses the notion of ‘tail-equivalency’: two distribution functions F_1 and F_2 are said to be tail-equivalent if their upper endpoints coincide (i.e., $x_{F_1}^\circ = x_{F_2}^\circ (= x^\circ)$) and if they satisfy

$$\lim_{t \rightarrow x^\circ} \frac{1 - F_1(t)}{1 - F_2(t)} = C$$

for some $C > 0$.

One can show that, if F_1 is in one of the domains of attraction there exists a tail-equivalent distribution F_2 belonging to the same domain of attraction, that satisfies the appropriate von Mises conditions, as stated in Theorem 1.3.

²Local uniform convergence is convergence on compact subsets.

1.3.4 Regular variation

In the previous subsection we stated necessary and sufficient conditions that characterised the domains of attraction of the three extreme value distributions. These conditions are closely related to the concept of regularly varying functions. Moreover, in the analysis of the behaviour of estimators in the field of extreme value theory, properties of regularly varying functions and so called Π -varying functions are frequently used. We will therefore state some definitions and results concerning these two classes of functions. For a more detailed treatment, we refer to SENETA (1976) and GELUK AND DE HAAN (1987) for theory on regular variation and Π -variation and to DE HAAN (1970) for the connection between regular variation and extreme value theory. The proofs of the theorems we will state, can be found in these references.

Regular variation is a one-sided, local and asymptotic property of an eventually positive, measurable function. Being a local property, it is defined relative to a point:

DEFINITION 1.1

A real valued, measurable function $U : [A, \infty) \rightarrow \mathbb{R}_+$ for some $A > 0$, is regularly varying at infinity with index $\rho \in \mathbb{R}$, notation $U \in \text{RV}_\rho^\infty$, if for all $x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\rho \quad (1.9)$$

◇

Similarly, we may define a function U to be regularly varying at 0, notation $U \in \text{RV}_\rho^0$, if the function $U(1/x)$ is regularly varying at infinity with index $-\rho$. When $\rho = 0$, the function U is said to be *slowly varying*.

Moreover, if one only assumes that the limit in (1.9) exists and is positive for all $x > 0$, it can be shown that that limit is necessarily of the form x^ρ .

It is obvious that a regularly varying function can be considered as a function whose asymptotic behaviour near a point is that of a power function multiplied by a factor that varies 'more slowly' than a power function. Indeed, $U \in \text{RV}_\rho^\infty$ if and only if it can be written in the form $U(x) = x^\rho L(x)$ with L a slowly varying function.

Two basic and important theorems concerning properties of regularly varying functions are the following:

Theorem 1.5 (Uniform Convergence Theorem)

If $U \in \text{RV}_\rho^\infty$, then for every fixed $[a, b]$ with $0 < a < b < \infty$ relation (1.9) holds uniformly for $x \in [a, b]$.

Theorem 1.6 (Representation Theorem)

If U defined on $[A, \infty)$, $A > 0$, is regularly varying at infinity with index ρ , then there exist measurable functions $c(\cdot)$ and $r(\cdot)$ with

$$\lim_{t \rightarrow \infty} c(t) = c_0 \in (0, \infty) \quad \text{and} \quad \lim_{t \rightarrow \infty} r(t) = \rho \quad (1.10)$$

and a constant $B \geq A$, such that for all $x \geq B$

$$U(x) = c(x) \exp \left(\int_B^x \frac{r(t)}{t} dt \right) \quad (1.11)$$

Conversely, if (1.11) holds with $c(\cdot)$ and $r(\cdot)$ satisfying (1.10), then $U \in \text{RV}_\rho^\infty$.

This representation of regularly varying functions is also known as the Karamata representation.

The class of Π -varying functions is obtained considering a generalization of the class of regularly varying functions. First restate the definition of regular variation as follows: A measurable function $U : [A, \infty) \rightarrow \mathbb{R}_+$ is regularly varying at infinity if there exists a positive function $a(\cdot)$ such that for all $x > 0$ the limit $\lim_{t \rightarrow \infty} U(tx)/a(t)$ exists and is positive. (In which case $a(t) = U(t)$ is one possible choice.)

Then an obvious way to proceed is to consider the cases in which a measurable function $U : [A, \infty) \rightarrow \mathbb{R}_+$ is such that there exist real functions $a(\cdot) > 0$ and $b(\cdot)$ such that for all $x > 0$ the limit

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)}$$

exists and the limit function is not constant (to avoid trivialities). This being equivalent to the existence of a non-constant limit

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} \quad (1.12)$$

for all $x > 0$, we get the following theorem:

Theorem 1.7

If $U : [A, \infty) \rightarrow \mathbb{R}$ is measurable, $a(\cdot)$ is positive and (1.12) is not constant, then

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = c \frac{x^\rho - 1}{\rho} \quad x > 0 \quad (1.13)$$

for some $\rho \in \mathbb{R}$ and $c \neq 0$, with the convention that the right hand side of (1.13) reads $c \log x$ if $\rho = 0$. Moreover, we may take $a(\cdot)$ to be a measurable function in RV_ρ^∞ .

Actually, $\rho \neq 0$ corresponds to classes of functions we have met before:

Theorem 1.8

Suppose the function U is such that Theorem 1.7 holds with $\rho \neq 0$ and $c > 0$.³

- a. If $\rho > 0$ then $U \in \text{RV}_\rho^\infty$.
- b. If $\rho < 0$ then $U(\infty) := \lim_{t \rightarrow \infty} U(t)$ exists and $U(\infty) - U(t) \in \text{RV}_\rho^\infty$

Hence, the case of $\rho = 0$ in Theorem 1.7 defines a new class of functions, which is called the class of Π -varying functions. Hence, we get the following definition:

DEFINITION 1.2

A real valued, measurable function $U : [A, \infty) \rightarrow \mathbb{R}$ for some $A > 0$ is Π -varying at infinity with auxiliary function $a(\cdot) > 0$, notation $U \in \Pi$ or $U \in \Pi(a)$, if for all $x > 0$

³It suffices to consider $c > 0$ since replacing U by $-U$ changes the sign of c .

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \log x$$

◇

Again, it is customary to define a function U to be Π -varying at 0, notation $U \in \Pi^0$ if $U(1/x)$ is Π -varying at infinity.

Now we are able to reformulate the necessary and sufficient conditions of Theorem 1.4 in terms of regular variation and Π -variation:

Theorem 1.9 (Reformulation of Theorem 1.4)

Necessary and sufficient conditions for a distribution function F , with its quantile function Q , to belong to one of the three possible domains of attraction are:

1. $F \in \mathcal{D}(\Lambda)$ if and only if the function $s \mapsto Q(1-s)$ is Π -varying at 0.
2. $F \in \mathcal{D}(\Phi_\alpha)$ for some $\alpha > 0$ if and only if its tail function $x \mapsto 1 - F(x)$ is regularly varying at infinity with index $-\alpha$, or equivalently, if and only if the function $s \mapsto Q(1-s)$ is regularly varying at 0 with index $-1/\alpha$.
3. $F \in \mathcal{D}(\Psi_\alpha)$ for some $\alpha > 0$ if and only if $x_F^\circ < \infty$ and the function $x \mapsto 1 - F(x_F^\circ - 1/x)$ is regularly varying at infinity with index $-\alpha$, or equivalently, if and only if its quantile function has the property that $Q(1) < \infty$ and the function $s \mapsto Q(1) - Q(1-s)$ is regularly varying at 0 with index $1/\alpha$.

Note that Theorem 1.8 can be used to change the equivalent formulation (1.7) of our main assumption (1.1) into the just stated conditions on the quantile functions.

1.4 Scope of this thesis

The formulation of the possible limiting distributions of the affinely transformed maximum of a sample using the Generalized Extreme Value distribution, shows that the parameter γ , i.e., the extreme value index, is an important characteristic of the distribution. In the remainder of this thesis we will mainly be concerned with the estimation of that parameter.

The difficulty of estimating that extreme value index, can be quantified using the concept of the minimax risk of an estimator. In Chapter 2, we will first derive lower bounds to the rate with which that minimax risk of estimating the extreme value index tends to zero. In the remaining part of that chapter, we will briefly discuss Pickands' estimator (PICKANDS (1975)), a maximum likelihood estimator (SMITH (1987)), Hill's estimator (HILL (1975)), a moment estimator (DEKKERS, EINMAHL AND DE HAAN (1989)) and a kernel type estimator (CSÖRGŐ, DEHEUVELS AND MASON (1985)). We will state their definitions and some of the (asymptotic) results obtained in the mentioned references. Moreover, in case of the kernel type estimator we will present an alternative interpretation of that estimator. Using that interpretation, some of the results in RESNICK AND STĂRICĂ (1995) can easily be derived in an alternative way.

As stated at the end of section 1.2, in practice it might be very difficult to decide at which point the extreme tail of a distribution starts. Indeed, the asymptotics of the previously mentioned estimators all depend on the value of the threshold above which the extreme tail is defined to start. In 'real life' applications one often calculates and plots the estimator using different thresholds and then uses rather subjective methods to decide what threshold to use. In Chapter 3 we will make a first attempt to develop a data-driven method, i.e., a method that uses (objective) information of the sample at hand, to decide what threshold to use in the case of the kernel type estimator. To that end we will derive a more general asymptotic result than the one in the paper by Csörgő et al. The results of this chapter were derived jointly with Rudolf Grübel and published in GRÜBEL AND DE WOLF (1994).

In Chapter 4 we will introduce a new estimator of the extreme value index and discuss its asymptotic behaviour. For this new estimator, two parameters determine its behaviour. At the end of Chapter 4 a small simulation study will illustrate that dependence and will indicate a possible solution to the problem of choosing these parameters appropriately.

Chapter 5 contains results concerning yet another kernel type estimator of the extreme value index. Actually, a whole class of estimators will be introduced and their consistency and asymptotic normality will be proved. These results were derived jointly with Piet Groeneboom and Rik Lopuszka.

Finally, in Chapter 6 we will present a simulation study of some of the mentioned estimators of the extreme value index and discuss their finite sample properties using the results of these simulations, indicating the advantages of the methods as well as their limitations. Moreover, we will apply the estimators to estimate the extreme value index of the distribution of 211 measurements on water discharges at Lobith, The Netherlands, during the period 1901–1991.

Chapter 2

Estimation of the extreme value index

In this chapter we will derive a lower bound for the minimax risk of estimating the extreme value index γ , over a certain class of distribution functions.

Moreover, we will briefly describe some estimators of that extreme value index. Some of the analytical results concerning these estimators will be stated, in particular their asymptotic properties. At the end of this chapter we will illustrate the sharpness of the minimax lower bounds, by showing that at least one of the in this chapter introduced estimators achieves these bounds in convergence in distribution. A discussion of the finite sample behaviour will be deferred to Chapter 6.

In the subsequent sections of this chapter, we will be concerned with a sample X_1, \dots, X_n , with the $X_i \stackrel{\mathcal{D}}{=} X$ i.i.d. with common distribution function F . We will denote the (ascending) order statistics of a sample by $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

2.1 Minimax lower bounds

Some attempts have been made to measure the difficulty of the estimation of the extreme value index γ . One approach was taken in HALL AND WELSH (1984). They derived the maximum rate of convergence in probability for estimating the extreme value index over a certain subclass of the class of all distribution functions that belong to any of the three domains of attraction of the extreme value distributions. To be more specific, they considered the problem of finding the fastest rate at which a sequence $\{a_n\}_{n=1}^{\infty}$ can tend to 0 and yet satisfy

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{D}} \mathbb{P}(|\hat{\gamma}_n - \gamma| \leq a_n) = 1$$

for a specific class \mathcal{D} of distribution functions with positive extreme value index.

In DREES (1995) the same approach was taken, but for a more general class of distribution functions than the one considered by Hall and Welsh. Moreover, in Drees' paper, a generalization to non-positive γ was given.

Another frequently used approach is given by calculating the minimax risk of the estimation problem. In the setting of extreme value estimation an attempt to that end

was taken in DONOHO AND LIU (1991). They derived (a lower bound to) the minimax rate of the estimation of the extreme value index γ over virtually the same class of distribution functions Hall considered. However, their results are rather difficult to grasp for two reasons: these results are stated in a very implicit way and their derivations depend on results stated in a technical report that is difficult to obtain (DONOHO AND LIU (1987)).

In this section we will derive a lower bound to the minimax risk over a larger class of distribution functions than the one considered in the aforementioned papers. Moreover, we will derive that lower bound in a rather straightforward manner.

First we will briefly introduce the notion of the minimax risk in our setting of the estimation of the extreme value index.

2.1.1 Minimax risk in the setting of the extreme value index

Let \mathcal{F} denote some subclass of the class of all distribution functions that belong to any of the three domains of attraction of the extreme value distributions, whose densities with respect to the Lebesgue measure are assumed to exist. From this point onwards, we will identify a member of the class \mathcal{F} by its density and hence refer to \mathcal{F} as a class of densities.

Let Γ be the functional that maps a density from that class to the corresponding extreme value index, i.e., for all $f \in \mathcal{F}$ define

$$\Gamma f = \gamma \quad \text{if } F \in \mathcal{D}(G_\gamma)$$

where F is the distribution function corresponding to the density f .

Let γ_n ($n \geq 1$) be an estimation procedure, i.e., a sequence of measurable functions that map \mathbb{R}^n into \mathbb{R} in such a way that $\Gamma_n = \gamma_n(X_1, X_2, \dots, X_n)$ is an estimator of Γf for sequences of i.i.d. stochastic variables X_1, X_2, \dots with common density $f \in \mathcal{F}$.

In order to say that an estimation procedure behaves well, the difference $\Gamma_n - \Gamma f$ should at least be close to zero in some sense. To quantify this, define the risk of the estimation procedure γ_n , evaluated at $f \in \mathcal{F}$ by

$$R_l(n, \gamma_n, f; \Gamma) = \mathbb{E}_f l(|\gamma_n(X_1, X_2, \dots, X_n) - \Gamma f|)$$

where l is a loss function, i.e., an increasing non-negative function on $[0, \infty)$ with $l(0) = 0$. Commonly used loss functions are $l(x) = x$ and $l(x) = x^2$.

Obviously, the given definition of risk describes a local property of the estimation procedure: the risk is only evaluated at one specific point of the class of densities \mathcal{F} . To describe the corresponding global property of the procedure, it seems reasonable to take the maximum risk over the class of densities \mathcal{F} , i.e., to consider the worst possible performance of the procedure. More formally:

$$MR_l(n, \gamma_n; \Gamma, \mathcal{F}) = \sup_{f \in \mathcal{F}} R_l(n, \gamma_n, f; \Gamma)$$

To get a measure for the difficulty of the estimation problem, the worst performance is minimized over all estimation procedures:

$$MMR_l(n; \Gamma, \mathcal{F}) = \inf_{\gamma_n} M R_l(n, \gamma_n; \Gamma, \mathcal{F}) = \inf_{\gamma_n} \sup_{f \in \mathcal{F}} R_l(n, \gamma_n, f; \Gamma)$$

This is called the (global) minimax risk for estimating Γf based on a sample of size n , generated by a density $f \in \mathcal{F}$.

If the functional is such that Γf can be estimated consistently, the minimax risk will typically tend to zero as n tends to infinity. The rate of this convergence indicates how difficult it is to estimate that functional: if, for n tending to infinity, $MMR_l(n; \Gamma, \mathcal{F}) \sim \delta_n(l)$ for some sequence of positive constants $\delta_n(l)$ tending to 0 as n tends to infinity, then Γf is said to be δ_n -estimable with respect to the loss function l . The first objective is usually to find a lower bound for the minimax risk of an estimation problem.

Obviously, for any subclass $\mathcal{F}_n \subseteq \mathcal{F}$

$$MMR_l(n; \Gamma, \mathcal{F}) \geq MMR_l(n; \Gamma, \mathcal{F}_n) \quad (2.1)$$

The right hand side of (2.1) is called the *local* minimax risk at f , if \mathcal{F}_n converges to the singleton $\{f\}$.

By Jensen's inequality and the monotonicity of a loss-function l , we have that

$$MMR_l(n; \Gamma, \mathcal{F}) \geq l(MMR_{l_1}(n; \Gamma, \mathcal{F}))$$

where l_1 is the identity, i.e., $l_1(x) = x$. In the remainder of this chapter, we will therefore confine ourselves to the minimax risk with loss function $l_1(x) = x$ and henceforth suppress the subscript l_1 in the notation of the minimax risk.

In JONGBLOED (1995) a lower bound to the minimax risk is derived that in our case translates to: for all sets $\{f_0, f_1\} \subset \mathcal{F}$,

$$MMR(n; \Gamma, \{f_0, f_1\}) \geq \frac{1}{4} |\Gamma f_0 - \Gamma f_1| (1 - H^2(f_0, f_1))^{2n} \quad (2.2)$$

where $H^2(f_0, f_1)$ denotes the squared Hellinger distance between the two densities, i.e., in our situation

$$H^2(f_0, f_1) = \frac{1}{2} \int \left(\sqrt{f_0(x)} - \sqrt{f_1(x)} \right)^2 dx$$

Note that, since both f_0 and f_1 are probability densities,

$$1 - H^2(f_0, f_1) = \int \sqrt{f_0(x)f_1(x)} dx$$

2.1.2 Minimax lower bound: a positive extreme value index

In this subsection we will obtain a lower bound to the minimax risk of estimating a positive extreme value index over a specified class of distribution functions, including, as a special case, the class considered in HALL AND WELSH (1984) and in DONOHO AND LIU (1991).

The result is summarized in the following theorem:

Theorem 2.1

Let $L(\cdot)$ be a slowly varying function at infinity that is positive and eventually monotone. Define $\mathcal{F}_1 = \mathcal{F}_1(\gamma_0, C_0, \varepsilon, \beta, L)$ with $\gamma_0, C_0, \varepsilon, \beta > 0$ to be the class of continuous densities f on (x_0, ∞) for some $x_0 > 0$ that satisfy the following conditions:

1. $f(x) = \frac{C}{\gamma} x^{-\frac{1}{\gamma}-1} (1 + r(x))$
2. $|C - C_0| \leq \varepsilon, \quad \gamma > 0 \quad \text{and} \quad |\gamma - \gamma_0| \leq \varepsilon$
3. $|r(x)| \leq x^{-\beta} L(x)$

Let $g(x) = x^{1+2\beta\gamma_0} (L(x^{\gamma_0}))^{-2}$ and denote its inverse (for x large enough) by $g^{-1}(\cdot)$.

Then the following inequality on the minimax risk of estimating γ holds true:

$$\liminf_{n \rightarrow \infty} \left(\frac{n}{g^{-1}(n)} \right)^{1/2} \text{MMR}(n; \Gamma, \mathcal{F}_1) \geq k$$

for some constant $k > 0$, depending on γ_0, C_0 and β .

Note that a distribution function F corresponding to a density $f \in \mathcal{F}_1$ satisfies

$$1 - F(x) = Cx^{-\frac{1}{\gamma}} (1 + R(x))$$

with

$$R(x) = \frac{x^{\frac{1}{\gamma}}}{\gamma} \int_x^{\infty} y^{-\frac{1}{\gamma}-1} r(y) dy$$

Moreover, using condition 3. on the function $r(\cdot)$, $|R(x)| \leq x^{-\beta} L^*(x)$ with $L^*(\cdot)$ slowly varying at infinity as well.

In order to prove the theorem we will need the following lemma:

Lemma 2.1

For some $L_0 > 0$, let $L: [L_0, \infty) \rightarrow \mathbb{R}^+$ be a monotone function that is slowly varying at infinity. Let $M = M_n$ be a sequence of real numbers tending to infinity and let β be a positive constant.

Define the sequence $\lambda = \lambda_n$ as:

(a) In case $L(\cdot)$ is monotone non-increasing:

$$\lambda = \beta M^{-\beta} L(M)$$

(b) In case $L(\cdot)$ is monotone non-decreasing to a finite limit $L(\infty)$:

$$\lambda = \frac{L(L_0)}{L(\infty)} \beta M^{-\beta} L(M)$$

(c) In case $L(\cdot)$ is monotone non-decreasing to infinity:

$$\lambda = c_0 \beta M^{-\beta} L(M)$$

for arbitrary, fixed $c_0 \in (0, 1)$.

Then the following holds true:

$$y^\beta \left(y^{-\lambda} - 1 \right) \leq M^{-\beta} L(My) \quad \forall y \in \left[\frac{L_0}{M}, 1 \right) \quad (2.3)$$

for M large enough.

Proof of Lemma 2.1:

Using monotonicity of $L(\cdot)$ and defining in case (a) (i.e., $L \downarrow$)

$$\alpha = 1 \quad \text{and} \quad y_M(\alpha) = \frac{L_0}{M}$$

in case (b) (i.e., $L \uparrow L(\infty) < \infty$)

$$\alpha = \frac{L(L_0)}{L(\infty)} \quad \text{and} \quad y_M(\alpha) = \frac{L_0}{M}$$

and in case (c) (i.e., $L \uparrow \infty$)

$$y_M(\alpha) = \frac{L^{-1}(\alpha L(M))}{M} \quad \text{with } \alpha \in (0, 1) \text{ arbitrary}$$

where $L^{-1}(s) = \inf\{y \in [L_0, \infty) : L(y) \geq s\}$, we have that

$$L(My) \geq \alpha L(M) \quad \forall y \in [y_M(\alpha), 1)$$

Obviously, in cases (a) and (b) $y_M(\alpha) \rightarrow 0$ as $M \rightarrow \infty$. In case (c) the notion of rapid variation (see SENETA (1976)) entails that

$$y_M(\alpha) \rightarrow 0 \quad \text{if} \quad 0 < \alpha < 1$$

as M tends to infinity.

Using the definitions of α and $y_M(\alpha)$, (2.3) is satisfied for all $y \in [y_M(\alpha), 1)$ if

$$y^\beta (y^{-\lambda} - 1) \leq \alpha M^{-\beta} L(M) \quad (2.4)$$

holds for all $y \in [y_M(\alpha), 1)$. Note that, in cases (a) and (b), this implies that (2.3) holds for all y to be considered.

The left hand side of inequality (2.4) is maximized for $y_{max} = (1 - \lambda/\beta)^{1/\lambda}$, hence we need

$$\left(1 - \frac{\lambda}{\beta}\right)^{\beta/\lambda} \frac{\lambda}{\beta - \lambda} \leq \alpha M^{-\beta} L(M) \quad (2.5)$$

Note that, since y_{max} tends to $e^{-1/\beta} > 0$ as λ tends to 0 and $y_M(\alpha)$ tends to 0 as M tends to infinity, y_{max} will fall into the interval $[y_M(\alpha), 1)$ for M large enough. Moreover, equation (2.5) is satisfied if we take

$$\lambda = \alpha \beta M^{-\beta} L(M) \quad (2.6)$$

Since $L(\cdot)$ is slowly varying, this λ will indeed tend to zero as M tends to infinity.

In case (c) we still have to consider the inequality on the remaining interval $[L_0/M, y_M(\alpha))$. Using the fact that $L(\cdot)$ now is non-decreasing to infinity, we have to consider

$$M^\beta (y_M(\alpha))^\beta \left[(y_M(\alpha))^{-\lambda} - 1 \right] \leq L(L_0) \quad (2.7)$$

Since $L^{-1}(\cdot)$ is also non-decreasing to infinity and $L^{-1}(\alpha L(M))/M \rightarrow 0$ as $M \rightarrow \infty$,

$$\frac{\log L^{-1}(\alpha L(M))}{\log M} \in (0, 1)$$

for M large enough. Hence, using that both $L(\cdot)$ and $\log(\cdot)$ are slowly varying at infinity,

$$\begin{aligned} & -M^{-\beta} L(M) \log \frac{L^{-1}(\alpha L(M))}{M} \\ & = M^{-\beta} L(M) \log M \left[1 - \frac{\log L^{-1}(\alpha L(M))}{\log M} \right] \rightarrow 0 \end{aligned}$$

and $M^{-\beta} L(M) \rightarrow 0$ as $M \rightarrow \infty$. Hence, substituting (2.6) in the left hand side of inequality (2.7) we obtain

$$\begin{aligned} & (L^{-1}(\alpha L(M)))^\beta \left[\left(\frac{L^{-1}(\alpha L(M))}{M} \right)^{-\alpha \beta M^{-\beta} L(M)} - 1 \right] \\ & = (L^{-1}(\alpha L(M)))^\beta \left[-\alpha \beta M^{-\beta} L(M) \log \frac{L^{-1}(\alpha L(M))}{M} (1 + o(1)) \right] \\ & = \alpha \beta (L^{-1}(\alpha L(M)))^\beta \frac{L(M) \log M}{M^\beta} \left[1 - \frac{\log L^{-1}(\alpha L(M))}{\log M} \right] (1 + o(1)) \quad (2.8) \end{aligned}$$

Moreover, taking logarithms in (2.8), we get

$$-\beta \log M \left[1 - \frac{\log L^{-1}(\alpha L(M))}{\log M} - \frac{\log L(M)}{\beta \log M} - \frac{\log \log M}{\beta \log M} + \frac{\log \left(1 - \frac{\log L^{-1}(\alpha L(M))}{\log M} \right)}{\beta \log M} \right]$$

Since the last four terms inside the square brackets tend to 0 by the arguments following (2.7), this tends to $-\infty$. Hence, (2.4) is satisfied in case (c) as well. ■

Now everything is set up to prove Theorem 2.1.

Proof of Theorem 2.1:

First note that the condition on the support of the densities is without loss of generality, since we are only interested in the right (or upper) tail of these distributions. Secondly note that for the proof we can assume $L(\cdot)$ to be monotone on its entire support.

In view of (2.2) define the densities

$$f_0(x) = \frac{C_0}{\gamma_0} x^{-\frac{1}{\gamma_0}-1} \quad x \in [C_0^{\gamma_0}, \infty) \tag{2.9}$$

and

$$f_1(x) = \begin{cases} \frac{C_1}{\gamma_1} x^{-\frac{1}{\gamma_1}-1} & x \in [M, \infty) \\ \frac{C_2}{\gamma_0} x^{-\frac{1}{\gamma_0}-1} & x \in [C_0^{\gamma_0}, M) \end{cases}$$

with $\gamma_1 > 0$ and $M > C_0^{\gamma_0}$ to be specified lateron. Since f_1 should be a continuous probability density, the following conditions need to be imposed:

$$\frac{C_1}{\gamma_1} M^{-\frac{1}{\gamma_1}} = \frac{C_2}{\gamma_0} M^{-\frac{1}{\gamma_0}} \tag{2.10}$$

and

$$C_1 M^{-\frac{1}{\gamma_1}} + C_2 (C_0^{-1} - M^{-\frac{1}{\gamma_0}}) = 1 \tag{2.11}$$

To use inequality (2.2) to its full strength, the two densities should be such that the distance between γ_0 and γ_1 is large even though the Hellinger distance between the two densities is small. Therefore, we will take $\gamma_1 = \gamma_0 + \eta$ and let $\eta = \eta_n$ tend to 0 with an appropriate rate as n tends to infinity and simultaneously let $M = M_n$ tend to infinity with n .

Using condition (2.10), f_1 can be rewritten into

$$f_1(x) = \frac{C_1}{\gamma_1} x^{-\frac{1}{\gamma_1}-1} (1 + r_1(x))$$

with

$$r_1(x) = \begin{cases} 0 & x \in [M, \infty) \\ \left(\frac{x}{M}\right)^{\frac{1}{\gamma_1} - \frac{1}{\gamma_0}} - 1 & x \in [C_0^{\gamma_0}, M) \end{cases}$$

Moreover, (2.10) and (2.11) can be combined into the equation

$$C_1 = C_0 \frac{\gamma_1}{\gamma_0} M^{\frac{1}{\gamma_1} - \frac{1}{\gamma_0}} \left(1 + C_0 \left(\frac{\gamma_1}{\gamma_0} - 1\right) M^{-\frac{1}{\gamma_0}}\right)^{-1} \quad (2.12)$$

To use inequality (2.2), we will first consider the part involving the Hellinger distance between the two densities f_0 and f_1 . Using the definitions of these functions, we obtain

$$\begin{aligned} 1 - H^2(f_0, f_1) &= \int \sqrt{f_0(x)f_1(x)} dx \\ &= \sqrt{\frac{C_0 C_1}{\gamma_0 \gamma_1}} \left(\int_M^\infty x^{-\frac{1}{2}(\frac{1}{\gamma_0} + \frac{1}{\gamma_1}) - 1} dx + M^{\frac{1}{\gamma_0} - \frac{1}{\gamma_1}} \int_{C_0^{\gamma_0}}^M x^{-\frac{1}{\gamma_0} - 1} dx \right) \\ &= \sqrt{\frac{C_0 C_1}{\gamma_0 \gamma_1}} M^{-\frac{1}{2\gamma_1}} \left(\frac{2\gamma_0 \gamma_1}{\gamma_0 + \gamma_1} M^{-\frac{1}{2\gamma_0}} + \gamma_0 M^{\frac{1}{2\gamma_0}} \left(C_0^{-1} - M^{-\frac{1}{\gamma_0}}\right) \right) \\ &= \sqrt{\frac{C_0 C_1}{\gamma_0 \gamma_1}} M^{-\frac{1}{2\gamma_1}} \left(\frac{\gamma_0(\gamma_1 - \gamma_0)}{\gamma_0 + \gamma_1} M^{-\frac{1}{2\gamma_0}} + \frac{\gamma_0}{C_0} M^{\frac{1}{2\gamma_0}} \right) \end{aligned}$$

Substitution of (2.12) together with $\gamma_1 = \gamma_0 + \eta$ in the last formula then yields

$$1 - H^2(f_0, f_1) = \left(1 + \frac{\eta}{\gamma_0} C_0 M^{-\frac{1}{\gamma_0}}\right)^{-\frac{1}{2}} \left(1 + \frac{\eta}{2\gamma_0 + \eta} C_0 M^{-\frac{1}{\gamma_0}}\right)$$

i.e., inequality (2.2) becomes

$$\begin{aligned} MMR(n; \Gamma, \{f_0, f_1\}) &\geq \frac{1}{4} |\gamma_1 - \gamma_0| (1 - H^2(f_0, f_1))^{2n} \\ &= \frac{1}{4} \eta \left(1 + \frac{\eta}{\gamma_0} C_0 M^{-\frac{1}{\gamma_0}}\right)^{-n} \left(1 + \frac{\eta}{2\gamma_0 + \eta} C_0 M^{-\frac{1}{\gamma_0}}\right)^{2n} \end{aligned} \quad (2.13)$$

Taking logarithms on the right hand side and disregarding the constant 1/4, we obtain

$$\log \eta - n \log \left(1 + \frac{\eta}{\gamma_0} C_0 M^{-1/\gamma_0}\right) + 2n \log \left(1 + \frac{\eta}{2\gamma_0 + \eta} C_0 M^{-1/\gamma_0}\right)$$

Using that we will let η tend to 0 and M tend to infinity, expansion yields

$$\begin{aligned} &\log \eta - n \left(\frac{\eta}{\gamma_0} C_0 M^{-1/\gamma_0} + O\left(\eta^2 M^{-2/\gamma_0}\right) \right) + \\ &\quad + 2n \left(\frac{\eta}{2\gamma_0 + \eta} C_0 M^{-1/\gamma_0} + O\left(\eta^2 M^{-2/\gamma_0}\right) \right) \\ &= \log \eta - C_0 n \frac{\eta^2}{2\gamma_0 + \eta} M^{-1/\gamma_0} + O\left(n \eta^2 M^{-2/\gamma_0}\right) \end{aligned}$$

Hence, the right hand side of (2.13) is, for large M , maximized taking

$$\eta \sim \left(nM^{-\frac{1}{\gamma_0}} \right)^{-1/2} \quad (2.14)$$

The exact conditions on the function $r(\cdot)$ as given in Theorem 2.1, will further restrict the allowable behaviour of η and M and will yield the rate for the lower bound to the minimax risk as mentioned in the theorem.

Obviously, f_0 as defined in (2.9) satisfies this extra condition, since the corresponding $r_0(x) \equiv 0$.

To obtain that f_1 is in the class \mathcal{F}_1 as well (for n large enough) we need to check the following conditions:

1. $\eta \leq \varepsilon$
2. $|r_1(x)| = \left| \left(\frac{x}{M} \right)^{\frac{1}{\eta} - \frac{1}{\gamma_0}} - 1 \right| \leq x^{-\beta} L(x)$ for all $x \in [C_0^{\gamma_0}, M]$
3. $|C_0 - C_1| \leq \varepsilon$

The first condition is easily satisfied by the condition that η should tend to 0 as n tends to infinity. The condition on the function $r_1(\cdot)$ can be rephrased as follows

$$y^\beta \left(y^{-\lambda} - 1 \right) \leq M^{-\beta} L(My) \quad \text{for all } y \in [C_0^{\gamma_0}/M, 1]$$

where $\lambda = \eta / (\gamma_0(\gamma_0 + \eta)) > 0$. Invoke Lemma 2.1 with $L(\cdot)$, λ , β and M as defined in this theorem and $L_0 = C_0^{\gamma_0}$. That is, in order to get that the second condition is satisfied, we have to impose that

$$\eta \sim M^{-\beta} L(M) \quad (2.15)$$

Now consider the condition on $|C_0 - C_1|$. Using that we will let η tend to 0 and M tend to infinity,

$$\begin{aligned} |C_0 - C_1| &= C_0 \left| 1 - \left(1 + \frac{\eta}{\gamma_0} \right) M^{-\frac{\eta}{\gamma_0(\gamma_0 + \eta)}} \left(1 + M^{-\frac{1}{\gamma_0}} C_0 \frac{\eta}{\gamma_0} \right)^{-1} \right| \\ &= C_0 \left| 1 - \left(1 + \frac{\eta}{\gamma_0} \right) M^{-\frac{\eta}{\gamma_0(\gamma_0 + \eta)}} + o \left(\eta M^{-\frac{1}{\gamma_0} - \frac{\eta}{\gamma_0(\gamma_0 + \eta)}} \right) \right| \\ &= C_0 \left| 1 - M^{-\frac{\eta}{\gamma_0(\gamma_0 + \eta)}} + o \left(M^{-\frac{\eta}{\gamma_0(\gamma_0 + \eta)}} \right) \right| \end{aligned}$$

This is small, for M large enough, if

$$1 - \varepsilon/2 < M^{-\frac{\eta}{\gamma_0(\gamma_0 + \eta)}} < 1 + \varepsilon/2$$

i.e., if

$$-\frac{\tilde{\varepsilon}}{2} < -\frac{\eta}{\gamma_0(\gamma_0 + \eta)} \log M < \frac{\tilde{\varepsilon}}{2}$$

However, with (2.15), i.e., $\eta \sim M^{-\beta} L(M)$ we have

$$-\frac{\eta}{\gamma_0(\gamma_0 + \eta)} \log M \sim M^{-\beta} L(M) \log(M)$$

and this tends to zero, since both $L(\cdot)$ and $\log(\cdot)$ are slowly varying.

Finally, combining (2.15) with (2.14) then yields

$$M^{-\frac{1+2\beta\gamma_0}{\gamma_0}} (L(M))^2 \sim \frac{1}{n}$$

Defining $g(x) = x^{1+2\beta\gamma_0} (L(x^{\gamma_0}))^{-2}$ with inverse $g^{-1}(\cdot)$ (for x large enough), we obtain the minimax lower bound as stated in the theorem:

$$\eta \sim \left(\frac{n}{g^{-1}(n)} \right)^{-1/2}$$

■

Substituting $L(x) \equiv A$ for some $A > 0$, we obtain $g^{-1}(s) = (A^2 s)^{1/(1+2\beta\gamma_0)}$ and hence the lower bound to the minimax risk yields the same rate for η as in HALL AND WELSH (1984), that is,

$$\eta \sim n^{-\frac{\beta\gamma_0}{1+2\beta\gamma_0}}$$

Moreover, if we substitute $L(x) = A \log x$ for some $A > 0$, the inverse of $g(\cdot)$ is approximated by

$$g^{-1}(s) \approx \left(\frac{A\gamma_0}{1+2\beta\gamma_0} \right)^2 (s(\log s)^2)^{\frac{1}{1+2\beta\gamma_0}}$$

and we obtain the lower bound

$$\eta \sim n^{-\frac{\beta\gamma_0}{1+2\beta\gamma_0}} (\log n)^{\frac{1}{1+2\beta\gamma_0}}$$

i.e., η tends to 0 slightly slower as in the case Hall and Welsh considered.

2.1.3 Minimax lower bound: a negative extreme value index

In this subsection we will obtain a lower bound to the minimax risk of estimating a negative extreme value index over a specified class of distribution functions. Since distribution functions with a negative extreme value index can be transformed into distribution functions with a positive extreme value index, a lower bound can easily be deduced as a corollary of Theorem 2.1:

Corollary 2.1

Let $L(\cdot)$ be a slowly varying function at infinity that is positive and eventually monotone. Define $\mathcal{F}_2 = \mathcal{F}_2(\gamma_0, C_0, \varepsilon, \beta, L)$ with $\gamma_0 < 0$ and $C_0, \varepsilon, \beta > 0$ to be the class of continuous densities f on (x_0, x_F°) for some $x_0 \in \mathbb{R}$ and upper endpoint $x_F^\circ > 0$, that satisfy the following conditions:

1. $f(x) = -\frac{C}{\gamma}(x_F^\circ - x)^{-\frac{1}{\gamma}-1}(1 + r(x_F^\circ - x))$
2. $|C - C_0| \leq \varepsilon, \quad \gamma < 0 \quad \text{and} \quad |\gamma - \gamma_0| \leq \varepsilon$
3. $|r(s)| \leq s^\beta L(1/s)$

Let $g(x) = x^{1-2\beta\gamma_0}(L(x^{-\gamma_0}))^{-2}$ and denote its inverse (for x large enough) by $g^{-1}(\cdot)$.

Then the following inequality on the minimax risk of estimating γ holds true:

$$\liminf_{n \rightarrow \infty} \left(\frac{n}{g^{-1}(n)} \right)^{1/2} MMR(n; \Gamma, \mathcal{F}_2) \geq k$$

for some constant $k > 0$.

Proof of Corollary 2.1:

First note that, as stated before,

$$MMR(n; \Gamma, \mathcal{F}_2) \geq MMR(n; \Gamma, \{f_1, f_2\})$$

for all sets $\{f_1, f_2\} \subset \mathcal{F}_2$. Fix two densities h_1 and h_2 in $\mathcal{F}_1(-\gamma_0, C_0, \varepsilon, \beta, L)$, with extreme value indices γ_i and constants C_i and where \mathcal{F}_1 is defined in Theorem 2.1. Define the densities f_i ($i = 1, 2$) as

$$f_i(x) = h_i \left(\frac{1}{x^\circ - x} \right) (x^\circ - x)^{-2} \quad \text{for } x^\circ - C_i^{\gamma_i} < x < x^\circ$$

i.e., if $X \sim h_i$ then $x^\circ - 1/X \sim f_i$. Note that these densities f_i are in the class \mathcal{F}_2 with extreme value indices $-\gamma_i$, upper endpoints x° and constants C_i . Moreover,

$$\inf_{\hat{\gamma}} \sup_{f \in \{f_1, f_2\}} \mathbb{E}_f |\hat{\gamma}(X_1, \dots, X_n) - \Gamma f| = \inf_{\hat{\gamma}^*} \sup_{h \in \{h_1, h_2\}} \mathbb{E}_h |\hat{\gamma}^*(Y_1, \dots, Y_n) - \Gamma h|$$

where

$$\hat{\gamma}^*(Y_1, \dots, Y_n) = -\hat{\gamma} \left(x^\circ - \frac{1}{Y_1}, \dots, x^\circ - \frac{1}{Y_n} \right)$$

the X_i i.i.d. f and the Y_i i.i.d. h . Finally,

$$\inf_{\hat{\gamma}^*} \sup_{h \in \{h_1, h_2\}} \mathbb{E}_h |\hat{\gamma}^*(Y_1, \dots, Y_n) - \Gamma h| \geq \inf_{\hat{\gamma}} \sup_{h \in \{h_1, h_2\}} \mathbb{E}_h |\hat{\gamma}(Y_1, \dots, Y_n) - \Gamma h|$$

where the infimum on the right hand side is taken over all possible estimators $\hat{\gamma}$. The corollary then follows from Theorem 2.1. \blacksquare

2.2 Pickands' estimator

The estimator introduced by J. Pickands III in PICKANDS (1975) is based on his formulation of assumption (1.1): A continuous distribution F with upper endpoint x_F° is in the domain of attraction of G_γ for some $\gamma \in \mathbb{R}$ if and only if

$$\lim_{t \uparrow x_F^\circ} \sup_{0 \leq x \leq x_F^\circ - t} |F_t(x) - F_{GPD}(x; \gamma, \alpha(t))| = 0$$

with $F_{GPD}(x; \gamma, \sigma)$ a Generalized Pareto Distribution and F_t the conditional probability of X not exceeding $x + t$ given it has already exceeded t . Formally it means that if the threshold t is large, the conditional distribution of X , given that $X \geq t$ can be approximated by a Generalized Pareto Distribution.

The quantile function of a Generalized Pareto Distribution can easily be calculated:

$$Q_{GPD}(s; \gamma, \sigma) = \sigma \int_0^{-\log(1-s)} e^{\gamma u} du = \begin{cases} \sigma \frac{(1-s)^{-\gamma} - 1}{\gamma} & \gamma \neq 0 \\ -\sigma \log(1-s) & \gamma = 0 \end{cases}$$

for $\gamma \in \mathbb{R}$ and $\sigma > 0$.

Using that formula, we can express the parameters γ and σ in terms of the quantiles $Q_{GPD}(\frac{1}{2}; \gamma, \sigma)$ and $Q_{GPD}(\frac{3}{4}; \gamma, \sigma)$:

$$\gamma = \log \left(\frac{Q_{GPD}(\frac{3}{4}; \gamma, \sigma) - Q_{GPD}(\frac{1}{2}; \gamma, \sigma)}{Q_{GPD}(\frac{1}{2}; \gamma, \sigma)} \right) / \log 2$$

and

$$\sigma = (Q_{GPD}(\frac{1}{2}; \gamma, \sigma)) / \int_0^{\log 2} \exp(\gamma u) du$$

Pickands proposed to estimate the parameters γ and σ by a simple percentile method: let M be an integer much smaller than n and consider the excesses $X_{(n-i+1)} - X_{(n-4M+1)}$, $i = 1, \dots, 4M - 1$, over threshold $X_{(n-4M+1)}$ to be an (ordered) sample from a Generalized Pareto Distribution. Then use the sample quantiles of these excesses to estimate the parameters, i.e.,

$$\hat{\gamma}_{n,M}^P = \log \left(\frac{X_{(n-M+1)} - X_{(n-2M+1)}}{X_{(n-2M+1)} - X_{(n-4M+1)}} \right) / \log 2$$

and

$$\hat{\sigma}_{n,M}^P = \left(X_{(n-2M+1)} - X_{(n-4M+1)} \right) / \int_0^{\log 2} \exp(\hat{\gamma}_{n,M}^P u) du$$

Since we are only interested in the estimation of γ , the definition of Pickands' estimator of σ was stated just for reference.

Obviously, the estimators depend on the choice of the integer M . Pickands argued that M should tend to infinity (in probability) as n tends to infinity, but M/n should tend to 0 (in probability) as n increases. He provided a data adaptive way to choose the integer M and then proved that the Generalized Pareto Distribution with estimated parameters, using that integer M , is a consistent estimator of the tail of the underlying distribution F in the sense that for all $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{0 \leq x < \infty} \left| \frac{1 - F(X_{(n-4M+1)} + x)}{1 - F(X_{(n-4M+1)})} - (1 - \tilde{F}_M(x)) \right| > \varepsilon \right\} = 0$$

with \tilde{F}_M the Generalized Pareto Distribution with estimated parameters, i.e., $\tilde{F}_M(x) = F_{GPD}(x; \hat{\gamma}_{n,M}^P, \hat{\sigma}_{n,M}^P)$. Specifically, he proposed to compute M in the following way: For each $l, l = 1, 2, \dots, \lfloor n/4 \rfloor$ define

$$d_l = \sup_{0 \leq x < \infty} |\hat{F}_l(x) - \tilde{F}_l(x)|$$

where $\hat{F}_l(\cdot)$ is the empirical distribution function of the values $X_{(n-i+1)} - X_{(n-4l+1)}$, $i = 1, \dots, 4l - 1$ and $\tilde{F}_l(\cdot)$ is a Generalized Pareto Distribution with parameters $\hat{\gamma}_{n,l}^P$ and $\hat{\sigma}_{n,l}^P$. Then choose M to be the largest integer solution of

$$d_M = \min_{1 \leq l \leq \lfloor n/4 \rfloor} d_l$$

In DEKKERS AND DE HAAN (1989) a deterministic sequence m is considered. In that case it is not too involved to show weak consistency of Pickands' estimator. For the sake of the argument and to show the connection with an equivalent formulation of our main assumption different from Pickands' formulation, we will give a proof of weak consistency, following DEKKERS AND DE HAAN (1989). To that end we need the following lemma:

Lemma 2.2

Let Z_1, \dots, Z_n be a sample from a standard exponential distribution¹ and $k = k_n$ be a sequence of integers such that $k \leq n$ and $k \rightarrow \infty$ as $n \rightarrow \infty$. Then

$$Z_{(n-k+1)} - Z_{(n-2k+1)} \xrightarrow{P} \log 2$$

as $n \rightarrow \infty$, with \xrightarrow{P} denoting convergence in probability.

¹In the sequel we will denote a random variable X , having an exponential distribution with expectation μ , by $X \sim \mathcal{E}(\mu)$. In particular, $\mathcal{E}(1)$ denotes the standard exponential distribution.

Proof of Lemma 2.2:

It is well known that the spacings of standard exponential variables can be regarded as independent exponential variables with expectation depending on i :

$$\left\{ Z_{(n-i+1)} - Z_{(n-i)} \right\}_{i=1}^n \stackrel{\mathcal{D}}{=} \left\{ \frac{E_i}{i} \right\}_{i=1}^n$$

with $Z_{(0)} = 0$ and E_1, \dots, E_n are i.i.d. $\mathcal{E}(1)$. Hence,

$$Z_{(n-k+1)} - Z_{(n-2k+1)} \stackrel{\mathcal{D}}{=} \sum_{i=k}^{2k-1} \frac{E_i}{i}$$

Note that

$$\sum_{i=k}^{2k-1} \frac{1}{i} = \log 2 + O\left(\frac{1}{k}\right)$$

as $k \rightarrow \infty$. Independency of the E_i then yields

$$\begin{aligned} \mathbb{E} \left(\sum_{i=k}^{2k-1} \frac{E_i}{i} - \log 2 \right)^2 &= \left(\sum_{i=k}^{2k-1} \frac{1}{i} \right)^2 + \sum_{i=k}^{2k-1} \frac{1}{i^2} - 2 \log 2 \sum_{i=k}^{2k-1} \frac{1}{i} + (\log 2)^2 \\ &\leq \frac{C}{k} \end{aligned}$$

for some $C > 0$ and the assertion follows by applying Chebyshev's inequality. ■

Now we are able to prove weak consistency of Pickands' estimator:

Theorem 2.2

Let F be a distribution that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$. Let $m = m_n$ be an integer sequence with $m_n \rightarrow \infty$ and $m_n/n \rightarrow 0$ as $n \rightarrow \infty$. Then $\hat{\gamma}_{n,m}^P \rightarrow \gamma$ in probability as $n \rightarrow \infty$.

Proof of Theorem 2.2:

(cf. the proof in DEKKERS AND DE HAAN (1989)) First note that the order statistics $X_{(i)}$ can be written in terms of the quantile function $Q = F^{-1}$, evaluated at the corresponding order statistics of a uniform sample of size n :

$$\left\{ X_{(i)} \right\}_{i=1}^n \stackrel{\mathcal{D}}{=} \left\{ Q \left(1 - \exp(-Z_{(i)}) \right) \right\}_{i=1}^n$$

with the $Z_{(i)}$ the order statistics of a sample of size n of i.i.d. $\mathcal{E}(1)$ random variables. Note that the condition $m_n/n \rightarrow 0$ implies $\exp(-Z_{(n-2m+1)}) \rightarrow 0$ a.s. as $n \rightarrow \infty$ and hence

$$2\hat{\gamma}_{n,m}^P = \frac{Q \left(1 - e^{-Z_{(n-m+1)}} \right) - Q \left(1 - e^{-Z_{(n-2m+1)}} \right)}{Q \left(1 - e^{-Z_{(n-2m+1)}} \right) - Q \left(1 - e^{-Z_{(n-4m+1)}} \right)}$$

$$\begin{aligned}
&= \frac{Q\left(1 - e^{-Z_{(n-2m+1)}} \cdot e^{-Z_{(n-m+1)} + Z_{(n-2m+1)}}\right) - Q\left(1 - e^{-Z_{(n-2m+1)}}\right)}{Q\left(1 - e^{-Z_{(n-2m+1)}}\right) - Q\left(1 - e^{-Z_{(n-2m+1)}} \cdot e^{-Z_{(n-4m+1)} + Z_{(n-2m+1)}}\right)} \\
&\rightarrow \frac{2^\gamma - 1}{1 - 2^{-\gamma}} = 2^\gamma
\end{aligned}$$

in probability, as $n \rightarrow \infty$, by Lemma 2.2 and a reformulation of (1.7):

$$\lim_{s \downarrow 0} \frac{Q(1 - sx) - Q(1 - s)}{Q(1 - sy) - Q(1 - s)} = \frac{x^{-\gamma} - 1}{y^{-\gamma} - 1}$$

locally uniformly in $x, y > 0$ with $y \neq 1$. Taking logarithms then yields the assertion. \blacksquare

Strong consistency can be established under a slightly stronger restriction on the growth of the sequence m_n : $m_n/n \rightarrow 0$ and $m_n/\log \log n \rightarrow \infty$ as $n \rightarrow \infty$.

Moreover, in DEKKERS AND DE HAAN (1989) asymptotic normality is derived of Pickands' estimator of γ . We will state these results for easy reference. The result is given in two theorems, under different conditions on the distribution function F :

Theorem 2.3

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$. Assume that its quantile function $Q = F^{-1}$ has positive derivative Q' and that there exists a positive function a , such that the function $t \mapsto \pm t^{-\gamma-1} Q'(1 - 1/t)$ (with either choice of sign) is Π -varying at infinity with auxiliary function a . Then, for sequences $m = m_n \rightarrow \infty$ satisfying $m_n = o(n/g^{-1}(n))$, where g^{-1} is the inverse of $g(t) = t^{-2\gamma-1} (Q'(1 - 1/t)/a(t))^2$,

$$\sqrt{m} (\hat{\gamma}_{n,m}^P - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_P^2)$$

as $n \rightarrow \infty$, with $\mathcal{N}(0, \sigma_P^2)$ denoting a normal distribution with mean 0 and variance σ_P^2 given by

$$\sigma_P^2 = \frac{\gamma^2 (2^{2\gamma+1})}{(2(2^\gamma - 1) \log 2)^2}$$

Theorem 2.4

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$. Suppose that the derivative Q' of the quantile function $Q = F^{-1}$ of F exists and that, for $\gamma \neq 0$ there exist constants $\rho > 0$ and $c > 0$ such that the function $t \mapsto \pm (t^{-1-\gamma} Q'(1 - 1/t) - c^\gamma |\gamma|^{\gamma+1})$ (with either choice of sign) is regularly varying at infinity with index $-\rho |\gamma|$. Then, for sequences $m = m_n \rightarrow \infty$ satisfying $m_n = o(n/g^{-1}(n))$, where g^{-1} is the inverse of

$$g(t) = t^{-2\gamma-1} \left(Q'(1 - 1/t) / (t^{-\gamma-1} Q'(1 - 1/t) - c^\gamma |\gamma|^{1+\gamma}) \right)^2,$$

$$\sqrt{m} (\hat{\gamma}_{n,m}^P - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_P^2)$$

as $n \rightarrow \infty$, with $\mathcal{N}(0, \sigma_p^2)$ denoting a normal distribution with mean 0 and variance σ_p^2 given by

$$\sigma_p^2 = \frac{\gamma^2(2^{2\gamma+1})}{(2(2^\gamma - 1) \log 2)^2}$$

Note that the Normal distribution satisfies the conditions of Theorem 2.3 and that the Cauchy distribution satisfies the conditions of Theorem 2.4, but not those of Theorem 2.3. The conditions of the two theorems can be unified conform Theorem 1.7 into the following conditions: $Q'(1 - 1/t) \in \text{RV}_{\gamma+1}^\infty$ and, with either choice of sign,

$$\lim_{t \rightarrow \infty} \frac{(tx)^{-\gamma-1} Q'(1 - 1/(tx)) - t^{-\gamma-1} Q'(1 - 1/t)}{a(t)} = \pm \frac{x^{-\rho|\gamma|} - 1}{-\rho|\gamma|} \quad (2.16)$$

for some $\rho \geq 0$ and some positive function a . Since for $\gamma = 0$ the limit does not depend on ρ , the case $\gamma = 0$ is not present in Theorem 2.4.²

Note that the family of Generalized Pareto Distributions formally does not satisfy condition (2.16), since the function $t \mapsto t^{-\gamma-1} Q'(1 - 1/t)$ is constant. However, a close look at the proof as given in DEKKERS AND DE HAAN (1989) shows that the asymptotic normality result still holds: in that proof, condition (2.16) is used to bound a remainder term that in case of a Generalized Pareto Distribution equals 0. Moreover, the condition on the sequence m_n for asymptotic normality turns out to be that $m_n = o(n)$ only, in addition to $m_n \rightarrow \infty$.

The two theorems concerning the asymptotic normality of Pickands' estimator, show that the asymptotic distribution of the estimator has mean 0 *provided* the sequence m_n is chosen appropriately. The proofs of these theorems, as given in DEKKERS AND DE HAAN (1989), reveal that if one chooses the sequence m_n to be of the same order as $n/g^{-1}(n)$ instead of of smaller order, the estimator will have an asymptotic bias. Hence, the choice of the sequence m_n seems to be of crucial importance. However, the right order of the growth of that sequence is determined by properties of the (tail of the) unknown underlying distribution function F . E.g., in case of a Normal distribution one needs $m_n = o(\log^2 n)$, in case of a Cauchy distribution one needs $m_n = o(n^{4/5})$ and in case of a Generalized Pareto Distribution, as stated before, one needs $m_n = o(n)$ in addition to $m_n \rightarrow \infty$.

2.3 A maximum likelihood approach

Another estimator, based on Pickands' results on the approximation of the tail of a distribution by a Generalized Pareto Distribution, was introduced by R.L. Smith in SMITH (1987). He proposed to use only those values of the sample that exceed a certain threshold: Let X_1, \dots, X_n be a sample with common distribution function F that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$. Fix a (high)

²Note that on the right hand side of (2.16) the absolute value of γ appears, and not just γ itself as stated erroneously in DEKKERS AND DE HAAN (1989).

threshold u and consider the excesses Y_1, \dots, Y_N , where N is the number of exceedences over u , i.e., the number of X_i larger than u , and $Y_i = X_j - u$ where j is the index of the i -th exceedence. Then, conditionally on N , the excesses are i.i.d. with distribution function $F_u(x) = (F(u+x) - F(u))/(1 - F(u))$. Smith's proposal was to approximate that distribution F_u by a Generalized Pareto Distribution whose parameters are estimated by a maximum likelihood method based on the excesses Y_1, \dots, Y_N .

In the original paper by Smith, there is no explicit proof of consistency of the estimators under the single condition of F being in the domain of attraction of an extreme value distribution. Under some extra conditions, Smith did prove the asymptotic normality of the estimators for the three classes of extreme value distributions, provided $\gamma > -1/2$. In case $\gamma \leq -1$, the maximum likelihood estimators do not exist: letting σ go off to infinity will continuously increase the likelihood. In case $-1 < \gamma \leq -1/2$, the problem becomes non-regular and other approaches are proposed, see e.g., SMITH (1985). We will state the results concerning the maximum likelihood estimators in case $\gamma > -1/2$.

Let F be a distribution function in the domain of attraction of an extreme value distribution G_γ for some $\gamma \neq 0$. Denote its upper endpoint by x_F° . Define the function L by

DEFINITION 2.1

$$L(x) = \begin{cases} x^{-1/\gamma}(1 - F(x_F^\circ - 1/x)) & \gamma < 0 \\ x^{1/\gamma}(1 - F(x)) & \gamma > 0 \end{cases}$$

◇

Conform Theorem 1.9, L is a slowly varying function (at infinity). Smith's results can be given using the following two conditions on that function L ('slow variation with remainder'):

SR1 $L(xt)/L(t) = 1 + O(\phi(t))$, as $t \rightarrow \infty$ and for all $x > 0$

SR2 $L(xt)/L(t) = 1 + ch_\rho(x)\phi(t) + o(\phi(t))$, as $t \rightarrow \infty$ and for all $x > 0$

where $\phi(t) > 0$, $\phi(t) \rightarrow 0$ as $t \rightarrow \infty$, $c \in \{-1, 0, 1\}$ and $h_\rho(x) = \int_1^x u^{-\rho-1} du = (1 - x^{-\rho})/\rho$ for some $\rho > 0$.

Smith remarked that in case of SR2, excluding trivial cases, that is excluding $c = 0$, ϕ is necessarily a regularly varying function with index $-\rho$. The results will be given in two theorems, dealing with the cases $\gamma > 0$ and $-1/2 < \gamma < 0$ respectively.

Theorem 2.5

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma > 0$. Assume that the corresponding L -function satisfies SR2. Let Y_1, \dots, Y_N denote the excesses over threshold u_N , where $N \rightarrow \infty$ and $u_N \rightarrow \infty$, such that

$$\sqrt{N}c\phi(u_N) \frac{\gamma}{1 + \gamma\rho} \rightarrow \mu$$

for some $\mu \in \mathbb{R}$. Then, with probability tending to 1, there exists a local maximum $(\hat{\gamma}_N^S, \hat{\sigma}_N^S)$ of the Generalized Pareto Likelihood based on Y_1, \dots, Y_N , such that

$$\sqrt{N} \begin{bmatrix} \frac{\hat{\sigma}_N^S}{\gamma u_N} - 1 \\ \hat{\gamma}_N^S - \gamma \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{bmatrix} \frac{\mu(1+\gamma)(1+2\gamma\rho)}{1+\gamma(1+\rho)} \\ \frac{\mu\gamma(1+\gamma)(1-\rho)}{1+\gamma(1+\rho)} \end{bmatrix}, \mathfrak{Z}_\gamma \right)$$

with

$$\mathfrak{Z}_\gamma = \begin{bmatrix} 2(1+\gamma) & -(1+\gamma) \\ -(1+\gamma) & (1+\gamma)^2 \end{bmatrix} \quad (2.17)$$

If L satisfies only SR1 with ϕ non-increasing, and $\sqrt{N}\phi(u_N) \rightarrow 0$, then the same result holds with $\mu = 0$.

Theorem 2.6

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in (-1/2, 0)$. Denote the (finite) upper endpoint of F by x_F° . Assume that the corresponding L -function satisfies SR2. Let Y_1, \dots, Y_N denote the excesses over threshold u_N , where $N \rightarrow \infty$ and $u_N \rightarrow x_F^\circ$, such that

$$\sqrt{N}c\phi(1/(x_F^\circ - u_N)) \frac{-\gamma}{1-\gamma\rho} \rightarrow \mu$$

for some $\mu \in \mathbb{R}$. Then, with probability tending to 1, there exists a local maximum $(\hat{\gamma}_N^S, \hat{\sigma}_N^S)$ of the Generalized Pareto Likelihood based on Y_1, \dots, Y_N , such that

$$\sqrt{N} \begin{bmatrix} \frac{\hat{\sigma}_N^S}{-\gamma(x_F^\circ - u_N)} - 1 \\ \hat{\gamma}_N^S - \gamma \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{bmatrix} \frac{\mu(1+\gamma)(1-2\gamma\rho)}{1+\gamma(1-\rho)} \\ \frac{\mu\gamma(1+\gamma)(1+\rho)}{1+\gamma(1-\rho)} \end{bmatrix}, \mathfrak{Z}_\gamma \right)$$

with \mathfrak{Z} as in (2.17).

If L satisfies only SR1 with ϕ non-increasing, and $\sqrt{N}\phi(1/(x_F^\circ - u_N)) \rightarrow 0$, then the same result holds with $\mu = 0$.

In case $\gamma = 0$ a different approach is needed to obtain asymptotic normality of the maximum likelihood estimators. In BALKEMA AND DE HAAN (1972) it was proved that in case F is in the domain of attraction of the extreme value distribution Λ (i.e., $\gamma = 0$) there exists a representation

$$1 - F(x) = c(x) \exp \left(- \int_{-\infty}^x \frac{dt}{\phi(t)} \right) \quad x < x_F^\circ \quad (2.18)$$

where $c(x) \rightarrow 1$ as $x \rightarrow x_F^\circ \leq \infty$, ϕ is a positive differentiable function and $\phi'(x) \rightarrow 0$ as $x \rightarrow x_F^\circ$. The additional conditions needed for asymptotic normality are now given using the function ϕ .

Theorem 2.7

Let F be a distribution function that is in the domain of attraction of the extreme value distribution G_0 . Denote its upper endpoint by x_F° . Assume that, as $u \rightarrow x_F^\circ$ and for some $K > 1$, the function ϕ in the representation (2.18) satisfies

$$\frac{\phi'(u + y\phi(u))}{\phi'(u)} \rightarrow 1 \quad \text{uniformly over } 0 \leq y \leq -K \log|\phi'(u)|$$

$$c(u) - 1 \sim s\phi'(u) \quad \text{for finite } s$$

Let Y_1, \dots, Y_N denote the excesses over threshold u_N , where $N \rightarrow \infty$, $u_N \rightarrow x_F^\circ$ such that

$$\sqrt{N}\phi'(u_N) \rightarrow \mu$$

for some $\mu \in \mathbb{R}$. Then, with probability tending to 1, there exists a local maximum $(\hat{\gamma}_N^S, \hat{\sigma}_N^S)$ of the Generalized Pareto Likelihood based on Y_1, \dots, Y_N , such that

$$\sqrt{N} \begin{bmatrix} \frac{\hat{\sigma}_N^S}{\phi'(u_N)} - 1 \\ \hat{\gamma}_N^S \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \right)$$

There is one practical issue concerning the use of the maximum likelihood estimators we would like to mention. In calculating the estimators, it is essential not to try to solve the corresponding score equations, but to use a maximization procedure on the (log-) likelihood itself, since $\gamma = 0$ is always a solution to the score equations. On the other hand, the feasible region for the parameters in the maximization procedure is bounded, hence, near these boundaries, extra attention should be paid on the maximization procedure that is used.

2.4 Hill's estimator

In 1975 another estimator of the extreme value index was introduced by B.M. Hill in his paper 'A simple general approach to inference about the tail of a distribution' (HILL (1975)). This estimator is based on the regular variation property in case the underlying distribution function F is in the domain of attraction of G_γ with $\gamma > 0$, or, as Hill stated it, on the assumption that the tail of the distribution is of Zipf or Pareto form for large x , i.e., $1 - F(x) \sim Cx^{-1/\gamma}$ as $x \rightarrow \infty$ for some $\gamma > 0$ and $C > 0$. Hence, Hill's estimator is only applicable in case the extreme value index is known to be positive, i.e., only in case the underlying distribution function exhibits a 'heavy' tail.

One way to introduce this estimator is to use a maximum likelihood approach: Assume that the distribution function F is of the form $1 - Cx^{-1/\gamma}$ for all $x \geq u$, where u is some threshold. I.e., assume that

$$\frac{1 - F(x)}{1 - F(u)} = \left(\frac{x}{u}\right)^{-1/\gamma} \quad x \geq u \quad (2.19)$$

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

Then it seems plausible to make inference on γ using only those order statistics that exceed the threshold u . Say that there are k of those order statistics: $X_{(n)}, \dots, X_{(n-k+1)}$. The log-likelihood function, using only the k upper order statistics is then given by

$$L(\gamma; X_{(n-k+1)}, \dots, X_{(n)}) = -k \log(\gamma u) + k \log(1 - F(u)) - \frac{\gamma + 1}{\gamma} \sum_{i=1}^k (\log X_{(n-i+1)} - \log u)$$

Solving the corresponding score equation and substituting $X_{(n-k)}$ for the threshold u , then yields the Hill estimator of γ :

$$\hat{\gamma}_{n,k}^H = \frac{1}{k} \sum_{i=1}^k (\log X_{(n-i+1)} - \log X_{(n-k)})$$

Note that this estimator can also be written in terms of the (normalized) spacings of the logarithms of the observations:

$$\hat{\gamma}_{n,k}^H = \frac{1}{k} \sum_{i=1}^k i (\log X_{(n-i+1)} - \log X_{(n-i)}) \quad (2.20)$$

In the original paper HILL (1975), Hill did not derive asymptotic normality results for his estimator, nor did he prove consistency of his estimator. He discussed the effect on the maximum likelihood approach of the conditioning on the order statistics $X_{(n)}, \dots, X_{(n-k)}$ and on the event that $X_{(n-k)} > u$, both from a Bayesian and a frequentists point of view.

MASON (1982a) proved weak consistency of Hill's estimator for any sequence $k = k_n$, satisfying $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$. Strong consistency was proved in DEHEUVELS, HAEUSLER AND MASON (1988) under the condition that $k/\log \log n \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$. Asymptotic normality can be established under certain extra conditions, see e.g., HAEUSLER AND TEUGELS (1985). In this section however, we will state the asymptotic normality result that is obtained as special case of the proof in DEKKERS, EINMAHL AND DE HAAN (1989) of asymptotic normality of their moment estimator (see also Section 2.5 of this thesis).

Theorem 2.8

Let F be a distribution function that is in the domain of attraction of some extreme value distribution G_γ with $\gamma > 0$. Let Q denote its quantile function. Assume that there exists a positive function b , such that the function $t \mapsto \pm t^{-\gamma} Q(1 - 1/t)$ (with either choice of sign) is Π -varying at infinity with auxiliary function b . Then, for sequences $k = k_n \rightarrow \infty$ satisfying $k_n = o(n/g^{-1}(n))$, where g^{-1} is the inverse of $g(t) = t^{1-2\gamma} (Q(1 - 1/t)/b(t))^2$,

$$\sqrt{k} (\hat{\gamma}_{n,k}^H - \gamma) \xrightarrow{D} \mathcal{N}(0, \sigma_H^2)$$

as $n \rightarrow \infty$, with $\mathcal{N}(0, \sigma_H^2)$ denoting a normal distribution with mean 0 and variance σ_H^2 given by $\sigma_H^2 = \gamma^2$.

Some generalizations of Hill's estimator were proposed in e.g., DEKKERS, EINMAHL AND DE HAAN (1989), CSÖRGŐ, DEHEUVELS AND MASON (1985) and GRÜBEL AND DE WOLF (1994). The asymptotic properties of these estimators yield, as a special case, the asymptotic properties of the Hill estimator. The first two generalizations will be discussed in the sequel of this chapter and the latter one will be discussed in Chapter 3 of this thesis.

There are a few important remarks we would like to make on the Hill estimator. Even though the likelihood was based on the assumption that the distribution function F satisfies property (2.19) *exactly* for some threshold u , it can be shown that Hill's estimate is still consistent under the assumption that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma > 0$, provided that the number of order statistics k that is used, tends to infinity at the right rate.

As in the case of Pickands' estimator, the choice of the rate with which the sequence $k = k_n$ tends to infinity, is an important issue. Indeed, not only the question of asymptotic normality depends on that rate, but the estimator can be heavily biased, depending on the rate of k .

A disadvantage of Hill's estimator is the fact that even though the extreme value index is by definition translation invariant, i.e., the distributions of the variables X and $X + u$ have the same extreme value index, the Hill estimator is not: calculation of the Hill estimator based on X_1, \dots, X_n yields a different value as calculated based on $X_1 + u, \dots, X_n + u$. However, it seems quite natural to use the logarithms of the observations when studying the extreme behaviour of the underlying distribution function F and, obviously, Hill's estimator is invariant under translation of the logarithms of the observations.

The short discussion in this section of Hill's estimator, deliberately mixes up two different approaches that are frequently used in the estimation of the extreme value index:

1. Fix some high threshold u and use the order statistics that exceed that threshold. The number of order statistics used in the estimator then is a random number, N say, with a Binomial distribution with parameters n and $1 - F(u)$, i.e., $N \sim \mathcal{B}(n, 1 - F(u))$.
2. Fix the number of order statistics that are used in the estimator, k say. Now the threshold can be viewed to be a random number U , whose distribution depends on k and the underlying distribution function F . One choice of defining U could be to take U to be equal to $X_{(n-k)}$.

Often, one does not pay much attention to the effect of which procedure is used. In fact, it often plays a minor role in the (asymptotic) properties of the estimator in question.

2.5 A moment estimator

One generalization of Hill's estimator was introduced in DEKKERS, EINMAHL AND DE HAAN (1989). Their main goal was to find an estimator, in some sense similar to Hill's

estimator, that was consistent for all $\gamma \in \mathbb{R}$. To that end, they introduced the following two quantities:

$$M_{n,k}^{(r)} = \frac{1}{k} \sum_{i=1}^k \left(\log X_{(n-i+1)} - \log X_{(n-k)} \right)^r \quad r = 1, 2$$

Note that $M_{n,k}^{(1)}$ equals Hill's estimator. Then they proposed the following estimator of the extreme value index:

$$\hat{\gamma}_{n,k}^M = M_{n,k}^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_{n,k}^{(1)})^2}{M_{n,k}^{(2)}} \right)^{-1} \quad (2.21)$$

We will give an heuristic derivation of this moment estimator, using the following lemma (cf. DEKKERS, EINMAHL AND DE HAAN (1989), Lemma 2.5):

Lemma 2.3

Assume that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in \mathbb{R}$ and that $x_F^\circ > 0$. Denote the corresponding quantile function by Q . Then, for some positive function a ,

$$\lim_{s \downarrow 0} \frac{\log Q(1-sx) - \log Q(1-s)}{a(s)/Q(1-s)} = \begin{cases} -\log x & \gamma \geq 0 \\ \frac{x^{-\gamma} - 1}{\gamma} & \gamma < 0 \end{cases}$$

for $x > 0$.

Moreover, in case $\gamma > 0$ we may take $a(s)/Q(1-s) = \gamma$ and in case $\gamma < 0$ we may take $a(s)/Q(1-s) = -\gamma(\log Q(1) - \log Q(1-s))$.

Note that $M_{n,k}^{(r)}$ can be viewed as a sample analogue of the expectation of $(\log(X/t))^r$ conditional on the event that $X > t$, with threshold t taken to be the $(n-k)$ -th order statistic of the sample. Indeed,

$$\begin{aligned} \mathbb{E} \left((\log(X/t))^r \mid X > t \right) &= \int_t^{x_F^\circ} (\log(x/t))^r d \left(\frac{F(x) - F(t)}{1 - F(t)} \right) \\ &= - \int_t^{x_F^\circ} (\log(x/t))^r d \left(\frac{1 - F(x)}{1 - F(t)} \right) \end{aligned}$$

Substitution of $y = 1 - F(x)$ and $s = 1 - F(t)$ then yields

$$\begin{aligned} \int_0^s (\log Q(1-y) - \log Q(1-s))^r d(y/s) \\ = \int_0^1 (\log Q(1-sx) - \log Q(1-s))^r dx \end{aligned}$$

Denoting the last integral by $M_s^{(r)}$, $M_{n,k}^{(r)}$ could be considered to be an estimate of $M_s^{(r)}$ with $s = k/n$.

Now use Lemma 2.3 to deduce that, in case $\gamma > 0$,

$$M_s^{(r)} \approx \gamma^r \int_0^1 (-\log x)^r dx = r\gamma^r \quad r = 1, 2$$

and in case $\gamma < 0$,

$$\begin{aligned} M_s^{(r)} &\approx (-\gamma)^r (\log Q(1) - \log Q(1-s))^r \int_0^1 \left(\frac{x^{-\gamma} - 1}{\gamma} \right)^r dx \\ &= (\log Q(1) - \log Q(1-s))^r \frac{r(-\gamma)^r}{(1-r\gamma)(1-(r-1)\gamma)} \quad r = 1, 2 \end{aligned}$$

Then substitution of these expressions into the equivalent of (2.21), i.e., into

$$\gamma_s := M_s^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_s^{(1)})^2}{M_s^{(2)}} \right)^{-1}$$

yields, for $\gamma > 0$

$$\gamma_s \approx \gamma + 1 - \frac{1}{2} \left(1 - \frac{\gamma^2}{2\gamma^2} \right)^{-1} = \gamma$$

and in case $\gamma < 0$

$$\begin{aligned} \gamma_s &\approx -\gamma (\log Q(1) - \log Q(1-s)) \frac{1}{1-\gamma} + 1 - \frac{1}{2} \left(1 - \frac{1-2\gamma}{2(1-\gamma)} \right)^{-1} \\ &= \gamma + \frac{\gamma}{\gamma-1} (\log Q(1) - \log Q(1-s)) \\ &\approx \gamma \end{aligned}$$

for small s . Actually, the preceding arguments can be made more rigorous, yielding consistency of the moment estimator.

Asymptotic normality of the moment estimator was derived in the same paper DEKKERS, EINMAHL AND DE HAAN (1989). We will state the results in the following theorem:

Theorem 2.9

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$. Denote its quantile function by Q . Assume that there exist positive functions b_1 , b_2 , b_3 and b_4 such that for all $x > 0$ (with either choice of sign),

(a) in case $\gamma > 0$, the function $t \mapsto \pm t^{-\gamma} Q(1-1/t)$ is Π -varying at infinity with auxiliary function b_1

(b) in case $\gamma = 0$,

$$\lim_{t \rightarrow \infty} \frac{\log Q(1-1/(tx)) - \log Q(1-1/t) + b_2(t) \log x}{b_3(t)} = \pm \frac{(\log x)^2}{2}$$

(c) in case $\gamma < 0$, the function $t \mapsto \mp t^{-\gamma}(Q(1) - Q(1 - 1/t))$ is Π -varying at infinity with auxiliary function b_4

Then, for sequences $k = k_n \rightarrow \infty$ satisfying $k_n = o(n/g^{-1}(n))$, where g^{-1} is the inverse of

$$g(t) = \begin{cases} t^{1-2\gamma} \left(\frac{Q(1-1/t)}{b_1(t)} \right)^2 & \gamma > 0 \\ t \left(\frac{b_2(t)}{b_3(t)} \right)^2 & \gamma = 0 \\ t^{1-2\gamma} \left(\frac{\log Q(1) - \log Q(1-1/t)}{b_4(t)} \right)^2 & \gamma < 0 \end{cases}$$

and additionally in case $\gamma = 0$, $k_n = o(n/g_1^{-1}(n))$ where $g_1^{-1}(\cdot)$ is the inverse of $g_1(t) = t(Q(1 - 1/t)/a(1/t))^2$ with the function a as defined in Lemma 2.3,

$$\sqrt{k}(\hat{\gamma}_{n,k}^M - \gamma) \xrightarrow{D} \mathcal{N}(0, \sigma_M^2)$$

as $n \rightarrow \infty$, with $\mathcal{N}(0, \sigma_M^2)$ denoting a normal distribution with mean 0 and variance σ_M^2 given by

$$\sigma_M^2 = \begin{cases} 1 + \gamma^2 & \gamma \geq 0 \\ (1 - \gamma)^2(1 - 2\gamma) \left(4 - 8 \frac{1 - 2\gamma}{1 - 3\gamma} + \frac{(5 - 11\gamma)(1 - 2\gamma)}{(1 - 3\gamma)(1 - 4\gamma)} \right) & \gamma < 0 \end{cases}$$

Again we would like to stress the fact that the sequence k_n has to be chosen appropriately, in order to obtain a centred limiting distribution. If one would take $k_n \sim cn/g^{-1}(n)$ for some constant c , then the limiting distribution would have mean $\pm\sqrt{c}$, where the sign corresponds with the sign in conditions (a), (b) and (c) in Theorem 2.9. A more detailed result on the asymptotic bias of the moment estimator under slightly different conditions, can be found in the final chapter ‘A Simulation Study in Extreme-Value Estimation’ of the Ph.D.-thesis by A.L.M. Dekkers (DEKKERS (1991)).

2.6 A kernel type estimator

Another attempt to generalize Hill’s estimator was made in the paper ‘Kernel estimates of the tail index of a distribution’ (CSÖRGŐ, DEHEUVELS AND MASON (1985)). In that paper the authors still only considered the case $\gamma > 0$. They proposed to use a smoother version of the Hill estimator in the sense that the abrupt cutoff at the threshold in Hill’s estimator should be replaced by a smoother one.

To define their estimator we need a kernel function K that satisfies the following conditions:

KC1 $K(u) \geq 0$ for $0 < u < \infty$

KC2 $K(\cdot)$ is non-increasing and right continuous on $(0, \infty)$

KC3 $\int_0^\infty K(u) du = 1$

KC4 $\int_0^\infty u^{-1/2} K(u) du < \infty$

Then the estimator is defined as

$$\hat{\gamma}_{n,h}^K = \left(\int_0^{1/h} \log_+ Q_n(1 - hu) d(uK(u)) \right) / \left(\int_0^{1/h} K(u) du \right)$$

where $h > 0$ is called the bandwidth, Q_n is the empirical quantile function and $\log_+ x = \log(x \vee 1)$. Routine manipulations show that $\hat{\gamma}_{n,h}^K$ can be written in the equivalent form

$$\left(\sum_{i=1}^n \frac{i}{nh} K\left(\frac{i}{nh}\right) \left(\log_+ X_{(n-i+1)} - \log_+ X_{(n-i)} \right) \right) / \left(\int_0^{1/h} K(u) du \right) \quad (2.22)$$

where we define $X_{(0)} = 1$. Note that using the uniform kernel, i.e., $K(u) = 1$ if $u \in (0, 1]$ and $K(u) = 0$ elsewhere, and putting $h = k/n$, the estimator coincides with Hill's estimator $\hat{\gamma}_{n,k}^H$ (see (2.20)).

To be able to state the asymptotic normality of the kernel type estimator, we will need some additional conditions on the kernel K :

KC5 There exists an $M_1 < \infty$ such that $K(u) = 0$ for $u > M_1$

KC6 There exists an $M_2 < \infty$ such that the kernel K has a derivative $k(u)$ for $u > M_2$ and such that $\lim_{u \rightarrow \infty} u^{3/2} k(u) = 0$

as well as some restrictions on the underlying distribution function F :

FC1 1. The function $Q(1 - \cdot)$ is regularly varying at 0 with index $-\gamma$, i.e., the quantile function satisfies the representation

$$Q(1 - s) = s^{-\gamma} c(s) \exp \left(\int_s^1 \frac{b(u)}{u} du \right) \quad 0 < s < 1 \quad (2.23)$$

where c is a function with $c(s) \rightarrow c \in (0, \infty)$ as $s \rightarrow 0$ and b a function with $b(s) \rightarrow 0$ as $s \rightarrow 0$.

2. Without loss of generality, $Q(0) = 1$.

FC2 1. In the representation (2.23), one has that either KC5 is satisfied and $c(s) = c$ (constant) for $0 < s < \varepsilon$ for some $\varepsilon > 0$ or $c(s) = c$ (constant) for $0 < s \leq 1$.

2. One has either KC6 is satisfied or the function b in (2.23) may be chosen to be bounded on $(0, 1)$.

For a discussion of these conditions on F we refer to CSÖRGŐ, DEHEUVELS AND MASON (1985). The asymptotic normality result for $\hat{\gamma}_{n,h}^K$ is given by

Theorem 2.10

Let $KC1, \dots, KC4$ and $FC1$ be satisfied. Then, as $h = h_n \rightarrow 0$ and $nh \rightarrow \infty$,

$$\hat{\gamma}_{n,h}^K \xrightarrow{P} \gamma$$

Moreover, if in addition $FC2$ is satisfied, then, for $h \rightarrow 0$ and $nh \rightarrow \infty$,

$$\sqrt{nh} (\hat{\gamma}_{n,h}^K - \gamma - \beta_C(h)) \xrightarrow{D} \mathcal{N}(0, \gamma^2 \sigma_K^2)$$

with β_C and σ_K^2 given by

$$\beta_C(h) = \left(\int_0^{1/h} b(hu)K(u) du \right) / \left(\int_0^{1/h} K(u) du \right)$$

$$\sigma_K^2 = \int_0^\infty K^2(u) du$$

A similar estimator was discussed in DE WOLF (1991). In that masters thesis, a boundary kernel was used that was allowed to be negative on parts of its support. Moreover, a first attempt was made to provide a data-driven method to find the optimal bandwidth that minimizes the asymptotic mean squared error. In GRÜBEL AND DE WOLF (1994) (see also Chapter 3 of this thesis) some more extensions were presented.

Another way to interpret the kernel type estimator, is to view it as a weighted average of Hill estimators, averaged over the number of order statistics used in each Hill estimator. To see this, we will assume $FC1$, i.e., $\log_+ X = \log X$ a.s., and we will disregard the normalizing constant $\int_0^{1/h} K(u) du$ in (2.22). (Alternatively, one could absorb that normalizing constant in the weights of the weighted average.)

Consider the following weighted average of Hill estimators:

$$H_n = \sum_{j=1}^n w_j \hat{\gamma}_{n,j}^H$$

where the w_j are the weights and $\hat{\gamma}_{n,j}^H$ as in (2.20). Then observe that this can be rewritten into

$$H_n = \sum_{i=1}^n \left(\sum_{j=i}^n \frac{w_j}{j} \right) i \left(\log X_{(n-i+1)} - \log X_{(n-i)} \right)$$

Compared to the kernel type estimator, we get the relations:

$$w_i = \frac{i}{nh} \left(K \left(\frac{i}{nh} \right) - K \left(\frac{i+1}{nh} \right) \right) \quad \text{and} \quad K \left(\frac{i}{nh} \right) = nh \sum_{j=i}^n \frac{w_j}{j} \quad (2.24)$$

Note that these relations are indeed satisfied in case of Hill's estimator itself using k order statistics, i.e., in case $w_j = 0$ for $j \neq k$, $w_k = 1$, $K(u) = 1$ if $u \in (0, 1]$, $K(u) = 0$ elsewhere and identifying h with k/n .

In RESNICK AND STÄRICĂ (1995) an averaged Hill estimator was proposed as well:

$$\hat{\gamma}_{n,k}^R = \frac{1}{k(t-s)} \sum_{p=\lceil ks \rceil}^{\lceil kt \rceil} \hat{\gamma}_{n,p}^H$$

for some $0 < s < t$. In view of the previous remarks this is in some sense equivalent to a kernel type estimator. Indeed, in their case (2.24) yields the following kernel:

$$K\left(\frac{i}{k}\right) = \begin{cases} k \sum_{j=\lceil ks \rceil}^{\lceil kt \rceil} \frac{1}{j} \frac{1}{t-s} & \lceil ks \rceil > i \\ k \sum_{j=i}^{\lceil kt \rceil} \frac{1}{j} \frac{1}{t-s} & \lceil ks \rceil < i < \lceil kt \rceil \\ 0 & i > \lceil kt \rceil \end{cases}$$

and continuously extended elsewhere. Moreover, for the asymptotic properties of this estimator we may approximate this kernel by

$$K(x) = \begin{cases} \frac{\log(t/s)}{t-s} & 0 \leq x \leq s \\ \frac{\log(t/x)}{t-s} & s \leq x \leq t \\ 0 & \text{elsewhere} \end{cases} \quad (2.25)$$

Note that the definition of σ_k^2 in Theorem 2.10, using the (approximating) kernel in case of RESNICK AND STÄRICĂ (1995), i.e., the kernel as given in (2.25), yields the same asymptotic variance as Resnick and Stărică derived in their own paper. I.e., the asymptotic variance is then given by

$$\sigma_{RS}^2 = \gamma^2 \frac{2}{t-s} \left(1 - \frac{s \log(t/s)}{t-s} \right)$$

2.7 Sharpness of the lower bounds to the minimax risk

Whence lower bound theorems concerning the minimax risk in estimating the extreme value index are derived, as in Section 2.1, the question remains whether these bounds are attainable or not. Usually, after deriving such theorems on lower bounds, an estimator is defined that attains the lower bound but then in the sense of convergence in distribution. This type of convergence however, is not strong enough to ensure that the minimax risk of that estimator converges to zero with the same rate as the lower bound to the minimax risk. On the other hand, it neither ensures that there exists no estimator of the extreme value index that converges in distribution with a faster rate than the lower bound to the minimax risk.

In this section, we will show that, considering a few specific classes of distribution functions, the rate of convergence in distribution of the Pickands estimator equals the

rate of the lower bound as stated in Theorem 2.1 for positive index and Corollary 2.1 for negative index. The Pickands estimator was taken because of computational convenience.

The first class of distribution functions we will consider, is the class \mathcal{P}_1 , consisting of all distribution functions that satisfy

$$1 - F(x) = Cx^{-\frac{1}{\gamma}}(1 + Ax^{-\beta}) \quad x \geq x_0$$

for $\gamma, \beta > 0$, $C, A \in \mathbb{R}$ and some suitably chosen $x_0 > 0$. Note that the corresponding quantile function satisfies

$$Q(1-s) \sim C^\gamma s^{-\gamma} \left(1 + \gamma AC^{-\beta\gamma} s^{\beta\gamma}\right) \quad \text{as } s \downarrow 0$$

For distribution functions in the class \mathcal{P}_1 , Theorem 2.4 applies, since

$$t^{-\gamma-1} Q'(1-1/t) - \gamma C^\gamma \sim (\beta-1)\gamma^2 AC^{-\gamma(\beta-1)} t^{-\beta\gamma} \in \text{RV}_{-\beta\gamma}^\infty$$

The rate of convergence in distribution, as given in Theorem 2.4, is related to $\delta_n = \sqrt{n/g^{-1}(n)}$ where $g^{-1}(\cdot)$ is the inverse of

$$g(t) = t^{-2\gamma-1} \left(\frac{Q'(1-1/t)}{t^{-\gamma-1} Q'(1-1/t) - \gamma C^\gamma} \right)^2$$

For the distribution functions in class \mathcal{P}_1 this function satisfies

$$g(t) \sim \left(\frac{C^{\beta\gamma}}{\gamma(\beta-1)A} \right)^2 t^{1+2\beta\gamma} \quad \text{as } t \rightarrow \infty$$

hence, $\delta_n \sim n^{\beta\gamma/(1+2\beta\gamma)}$ as $n \rightarrow \infty$ for these distribution functions. In Theorem 2.4 the rate is assumed to be $o(\delta_n)$ to obtain a centred limiting normal distribution, whereas the rate of the lower bound to the minimax risk is $O(\delta_n)$. However, as stated at the end of Section 2.2, that rate will still yield asymptotic normality of the estimator even though then an asymptotic bias will be present.

Another class of distribution functions we will consider is the class \mathcal{P}_2 consisting of all distribution functions satisfying

$$1 - F(x) = Cx^{-\frac{1}{\gamma}}(1 + Ax^{-\beta} \log x) \quad x \geq x_0$$

for $\gamma, \beta > 0$, $A, C \in \mathbb{R}$ and some suitably chosen $x_0 > 0$. Note that the corresponding quantile function now satisfies

$$Q(1-s) \sim C^\gamma s^{-\gamma} \left(1 - \gamma^2 AC^{-\beta\gamma} s^{\beta\gamma} \log s\right) \quad \text{as } s \downarrow 0$$

The distribution functions of this class again satisfy the condition of Theorem 2.4:

$$t^{-\gamma-1} Q'(1-1/t) - \gamma C^\gamma \sim \gamma^2 AC^{-\gamma(\beta-1)} t^{-\beta\gamma} (1 - \gamma(\beta-1) \log t) \in \text{RV}_{-\beta\gamma}^\infty$$

As before, the rate of convergence in distribution is hence related to $\delta_n = \sqrt{n/g^{-1}(n)}$ where $g^{-1}(\cdot)$ in this case is the inverse of

$$g(t) \sim t^{1+2\beta\gamma} (\log t)^{-2} \left(\gamma^2 (\beta - 1) AC^{-\beta\gamma} \right)^{-2} \quad \text{as } t \rightarrow \infty$$

Allowing for an asymptotic bias then yields the rate

$$\delta_n = O \left(\left(\frac{n^{\beta\gamma}}{\log n} \right)^{\frac{1}{1+2\beta\gamma}} \right) \quad \text{as } n \rightarrow \infty$$

which is again the same rate with which the lower bound to the minimax risk tends to zero for this class of distribution functions.

The third class of distribution functions concerns the negative index equivalent of \mathcal{P}_1 , i.e., class \mathcal{P}_3 consists of functions satisfying

$$1 - F(x) = C(x_F^\circ - x)^{-\frac{1}{\gamma}} \left(1 + A(x_F^\circ - x)^\beta \right) \quad x_o \leq x \leq x_F^\circ$$

for some $\gamma < 0$, $\beta, x_F^\circ > 0$, $A, C \in \mathbb{R}$ and some suitably chosen $x_o \in \mathbb{R}$. The corresponding quantile function then satisfies

$$Q(1-s) \sim x_F^\circ - C^\gamma s^{-\gamma} \left(1 + \gamma AC^{-\beta\gamma} s^{-\beta\gamma} \right) \quad \text{as } s \downarrow 0$$

Again Theorem 2.4 is applicable to the functions in this class:

$$t^{-\gamma-1} Q'(1-1/t) - |\gamma| C^\gamma \sim \gamma^2 (\beta + 1) AC^{\gamma(\beta+1)} t^{\beta\gamma} \quad \text{as } t \rightarrow \infty$$

Moreover, similar calculations as in the case of the class \mathcal{P}_1 yield that, allowing for asymptotic bias in Pickands' estimator, $\delta_n \sim n^{-\beta\gamma/(1-2\beta\gamma)}$ for n tending to infinity, as given in Corollary 2.1.

Indeed, defining the negative index equivalent of \mathcal{P}_2 by the class \mathcal{P}_4 consisting of distribution functions satisfying

$$1 - F(x) = C(x_F^\circ - x)^{-\frac{1}{\gamma}} \left(1 + A(x_F^\circ - x)^{-\beta} \log(x_F^\circ - x) \right) \quad x_o \leq x \leq x_F^\circ$$

for some $\gamma < 0$, $\beta, x_F^\circ > 0$, $A, C \in \mathbb{R}$ and some suitably chosen $x_o \in \mathbb{R}$, analogue computations again yield the rate of the lower bound to the minimax risk as the rate of convergence of Pickands' estimator when allowing for asymptotic bias:

$$\delta_n = O \left(\left(\frac{n^{-\beta\gamma}}{\log n} \right)^{\frac{1}{1-2\beta\gamma}} \right) \quad \text{as } n \rightarrow \infty$$

Chapter 3

A kernel type estimator revisited

In this chapter we will be concerned with the kernel type estimator introduced by CSÖRGŐ, DEHEUVELS AND MASON (1985). This estimator requires the choice of a bandwidth parameter which, roughly, controls the fraction of upper order statistics that is used for the estimate. We will provide an alternative and more general proof of the asymptotic normality of the estimator. As a corollary we will be able to describe the asymptotic behaviour of a combined estimator, i.e., the estimator of the extreme value index that arises when a data-dependent bandwidth is used. Moreover, we will introduce a bootstrap inspired bandwidth selection method in an attempt to provide a fully data driven estimation procedure.

The results given in this chapter were published as a TU Delft report: 'Estimation of the tail index of a distribution'. (GRÜBEL AND DE WOLF (1994))

3.1 Introduction

We will be concerned with a sample X_1, \dots, X_n of i.i.d random variables with common distribution function F that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma > 0$. Denote the ascending order statistics of such a sample by $X_{(1)} \leq \dots \leq X_{(n)}$. Let $h > 0$ and let $K : \mathbb{R}^+ \rightarrow \mathbb{R}$ be a fixed function with properties to be specified. We define the extreme value index estimator $\hat{\gamma}_{n,h}$ with *bandwidth* h and *kernel function* K by

$$\hat{\gamma}_{n,h} = \sum_{i=1}^n \frac{i}{n} K_h \left(\frac{i}{n} \right) \left(\log_+ X_{(n-i+1)} - \log_+ X_{(n-i)} \right) \quad (3.1)$$

where $K_h(x) = K(x/h)/h$, $\log_+ x = \log(1 \vee x)$ and $X_{(0)} = 1$.

This estimator arises quite naturally by rewriting the von Mises condition (see Theorem 1.3) in terms of the quantile function $Q(s) = F^{-1}(s)$:

$$\lim_{s \downarrow 0} \left[-s \frac{d}{ds} \log Q(1-s) \right] = \gamma$$

Then substitution of a smoothed version of the empirical quantile function of the log-data, $Y_i = \log X_i$, which has jumps of height $\log X_{(i)} - \log X_{(i-1)}$ at i/n , leads to the estimator given in (3.1).

Throughout the rest of this chapter we will assume that the quantile function of the log-data, which equals the logarithm of the quantile function Q , can be written in the following form:

$$\log Q(1-s) = -\gamma \log s + \int_s^1 \frac{b(u)}{u} du \quad 0 < s \leq 1 \quad (3.2)$$

with b being such that the integrals exist and, moreover, $b(s) \rightarrow 0$ as $s \downarrow 0$. This assumption is slightly stronger than the Karamata representation of regularly varying functions as given in Theorem 1.6. It amounts to two additional requirements: the support condition $F(1) = 0$, $F(x) > 0$ if $x > 1$ (without loss of generality), and a smoothness condition, which holds if, for example, F has a positive density that itself is regularly varying at infinity. We refer to CSÖRGŐ, DEHEUVELS AND MASON (1985) for a discussion of these additional assumptions.

Note that the problem has been rewritten to a situation that can be viewed in the following two ways:

- a semi-parametric situation, with γ the 1-dimensional parameter of interest and b the infinite dimensional nuisance parameter
- a situation of estimating a curve at one of its boundaries, i.e., estimation of $\lim_{s \downarrow 0} \phi(s)$ with $\phi(s) = -s \frac{d}{ds} \log Q(1-s) = \gamma + b(s)$

As specified before, the kernel estimator (3.1) includes a bandwidth parameter h . It turns out that this parameter, roughly, determines the amount of order statistics that is used in the calculation of the estimator. It may not be surprising, especially in the curve estimation point of view mentioned before, that the choice of this bandwidth also determines the bias and variance of $\hat{\gamma}_{n,h}$: h too small will result in a large variance and h too large will result in a large bias. Hence, like in curve estimation, one would like to choose h in an optimal way, i.e., such that, asymptotically, the squared bias and the variance are of the same order. This particular choice of h however, depends on unknown properties of (the tail of) the underlying distribution function. Indeed, this may even depend on the γ one is trying to estimate. One way to avoid this problem, is to use a data dependent bandwidth: let the data decide which bandwidth is optimal, i.e., construct an estimate of the optimal bandwidth based on the sample in question.

In the sequel of this chapter, we will not only derive asymptotic properties for our estimator using deterministic bandwidths, but we will consider the combined estimator (i.e., the estimator with use of a data dependent bandwidth) as well. Furthermore, we will introduce a bootstrap inspired bandwidth selection procedure in an attempt to close the gap between theory and practice. In this chapter we will discuss the theoretical properties of this procedure and in Chapter 6 the finite sample size behaviour will be discussed using a simulation study.

3.2 Main results

Let X_1, \dots, X_n be a sample from a distribution function F such that equation (3.2) is satisfied. Let $K: \mathbb{R}^+ \rightarrow \mathbb{R}$ be such that

$$\begin{cases} K(x) = 0 & x \notin [0, 1] \\ |K(x) - K(y)| \leq C|x - y| & x, y \in (0, 1) \\ \int K(x) dx = 1 \end{cases} \quad (3.3)$$

for some $C > 0$, and define the kernel type estimator of γ as

$$\hat{\gamma}(n, h) = \sum_{i=1}^n \frac{i}{n} K_h \left(\frac{i}{n} \right) \left(\log X_{(n-i+1)} - \log X_{(n-i)} \right) \quad (3.4)$$

with $K_h(x) = K(x/h)/h$ and $h > 0$.

Note, using the support condition on K in (3.3) and the definition of $\hat{\gamma}(n, h)$, that nh is, indeed, approximately the number of order statistics used in the calculation of the estimate.

The asymptotic normality result of Csörgő *et al.* as given in section 2.6 of this thesis, shows that slow convergence of the bandwidth h to zero yields a fast rate for the estimator (i.e., small variance), but, unless the bias term $\beta_C(h)$ is of smaller order than \sqrt{nh} , the estimator will be asymptotically biased.

This brings us back to one of the questions raised in the introduction of this chapter: how to choose h optimally? Let us define the optimality in the following way: let, with some abuse of terminology, the *asymptotic mean squared error* of $\hat{\gamma}(n, h)$ be given by

$$\text{AMSE}(n, h) = \beta(h)^2 + \frac{1}{nh} \gamma^2 \sigma_K^2 \quad (3.5)$$

with $\beta(h)$ and σ_K^2 defined as

$$\beta(h) = \int K(u)b(hu) du \quad (3.6)$$

and

$$\sigma_K^2 = \int K^2(u) du$$

and let h_n^o be a value that minimizes the function $h \mapsto \text{AMSE}(n, h)$. So h_n^o leads to a situation in which the squared bias and the variance are asymptotically of the same order.

Assume that we have a consistent estimate \hat{h}_n for the optimal bandwidth, in the sense that $\hat{h}_n/h_n^o \rightarrow 1$ in probability, as $n \rightarrow \infty$. Then the next question that arises is, how will the combined estimator $\hat{\gamma}(n, \hat{h}_n)$ behave?

In order to answer that question, we will need a 'simultaneous' asymptotic normality result for a range of h values. That result is given in our first theorem: a functional central limit theorem.

Theorem 3.1

Assume that F and K satisfy (3.2) and (3.3) respectively. Let $h_n \rightarrow 0$ be such that $nh_n \rightarrow \infty$ and let T be the interval $[t_l, t_r]$, with $0 < t_l \leq t_r < \infty$. Define the stochastic processes Z_n with index set T by

$$Z_n(t) = \sqrt{nh_n}(\hat{\gamma}(n, th_n) - \gamma - \beta(th_n))$$

with $\beta(\cdot)$ and $\hat{\gamma}(\cdot, \cdot)$ as in (3.6) and (3.4) respectively.

Then $Z_n \xrightarrow{D} Z$ with Z a Gaussian process with continuous sample paths, mean function 0 and covariance structure

$$\text{cov}(Z_s, Z_t) = \gamma^2 \int K(su)K(tu)du.$$

Obviously, substitution of $t = 1$ yields asymptotic normality, similar to the results in CSÖRGŐ, DEHEUVELS AND MASON (1985). Our Theorem 3.1 does not only cover the more general situation $t \in [t_l, t_r]$ as well, but the proof also differs from the proof in the paper by Csörgő *et al.*, who based their proof on the Hungarian Construction. Our proof is quite short and uses standard central limit theory and a well known property of the spacings of samples from an exponential distribution.

Theorem 3.1, together with the Skorohod-Dudley representation theorem and an easy pathwise analysis, will lead us to the following corollary concerning the combined estimator. A similar idea has been used in CSÖRGŐ (1984).

Corollary 3.1

Let F , K and h_n be as in Theorem 3.1. If a sequence \hat{h}_n of random variables satisfies $\hat{h}_n/h_n \rightarrow 1$ in probability, then

$$\sqrt{nh_n}(\hat{\gamma}(n, \hat{h}_n) - \gamma - \beta(\hat{h}_n)) \xrightarrow{D} \mathcal{N}(0, \gamma^2 \sigma_K^2).$$

Qualitative assumptions on b will imply closeness of $\beta(\hat{h}_n)$ to $\beta(h_n)$.

A similar result was derived by HALL AND WELSH (1985) for a subclass of the class of distribution functions we consider in this paper, and in case of Hill's estimator of γ , see HILL (1975). Hill's estimator can be considered as a special case of our estimator, using $K(x) \equiv 1$ on $[0, 1]$.

Note that Corollary 3.1 can be interpreted in the following way: if \hat{h}_n is an estimator of h_n^0 with the property that \hat{h}_n/h_n^0 tends to 1 in probability, then the combined estimator $\hat{\gamma}(n, \hat{h}_n)$ has the same asymptotic behaviour (up to the order terms considered here) as the estimator that we would obtain if we knew and used the optimal bandwidth.

However, if we want to use the corollary in practice, we will need to find an estimator for h_n^0 which is consistent in the above sense. Under certain assumptions on b , the order of the bias term $\beta(h)$ will be known, and a limit result can be obtained for h_n^0 . Although there might be several minimizers of $\text{AMSE}(n, h)$, this is a valid goal, since the ratio of two such minimizers will tend to 1 as $n \rightarrow \infty$. Unfortunately, as can be deduced from (3.5), this will involve the unknown γ and other unknown quantities, related to the b -function (for an example, see discussion following Theorem 3.2). One

could try to estimate these quantities and to plug these estimates into the asymptotic formula. Instead, we propose a different approach: a bootstrap-inspired bandwidth selection procedure. In its final form it does not involve taking resamples.

The curve estimation point of view mentioned in the introduction, inspired us to consider the following procedure: first estimate the whole curve $\phi(\cdot)$, use that to construct an estimator of the quantile function of the log-data. Then resample from the distribution function $\hat{F}(\cdot)$ derived from that estimate.

To be more specific, following the kernel type approach of non-parametric curve estimation, regarding $\{(i/n, i(\log X_{(n-i+1)} - \log X_{(n-i)}))\}_{i=1}^n$ as the data points, introduce the following estimator for the function ϕ :

$$\hat{\phi}_{n,g}(s) = \sum_{i=1}^n \frac{i}{n} K_g \left(\frac{i}{n} - s \right) (\log X_{(n-i+1)} - \log X_{(n-i)})$$

for some bandwidth $g > 0$. Note that $\hat{\phi}_{n,g}(0) = \hat{\gamma}(n, g)$.

Then an estimator for the logarithm of the quantile function can be obtained, using the relation

$$\phi(s) = -s \frac{d}{ds} (\log Q(1-s))$$

which leads us to the estimate

$$\log \hat{Q}_{n,g}(1-s) = (\log \widehat{Q}(1-s))_{n,g} = \int_s^1 \frac{\hat{\phi}_{n,g}(u)}{u} du$$

We can use this estimate of the quantile function of the log-data to generate pseudo-random samples $\log X_1^*, \dots, \log X_n^*$. For each of these resamples we can compute the estimates $\hat{\gamma}_g^*(n, h)$. After a large number of repetitions we could use the mean of the squared differences $\hat{\gamma}_g^*(n, h) - \hat{\gamma}(n, g)$ as an estimator of $\text{AMSE}(n, h)$. By construction, the estimate of the associated distribution function has, conditionally on X_1, \dots, X_n , a regularly varying tail with index $-1/\hat{\gamma}(n, g)$. Hence Theorem 3.1 motivates the following estimator for the asymptotic mean squared error (which dispenses with resamples):

$$\widehat{\text{AMSE}}_{n,g}(n, h) = \hat{\beta}_{n,g}(h)^2 + \frac{1}{nh} \hat{\gamma}(n, g)^2 \sigma_K^2 \quad (3.7)$$

with

$$\hat{\beta}_{n,g}(h) = \int K(s) \hat{b}_{n,g}(hs) ds$$

where

$$\begin{aligned} \hat{b}_{n,g}(s) &= \hat{\phi}_{n,g}(s) - \hat{\gamma}(n, g) \\ &= \sum_{i=1}^n \frac{i}{n} \left(K_g \left(\frac{i}{n} - s \right) - K_g \left(\frac{i}{n} \right) \right) (\log X_{(n-i+1)} - \log X_{(n-i)}) \end{aligned}$$

The following theorem gives conditions for the above procedure to work in the sense that it produces an estimator \hat{h}_n that is sufficiently close to h_n° .

Theorem 3.2

Let F , K and h_n be as in Theorem 3.1. Assume additionally that b has a non-zero right derivative at 0, and let g_n be such that $g_n \downarrow 0$ and $h_n/g_n \rightarrow 0$. Let \hat{h}_n^o be a minimizer of $h \mapsto \text{AMSE}(n, h)$ and let \hat{h}_n be a value that minimizes $h \mapsto \widehat{\text{AMSE}}_{n, g_n}(n, h)$, with $\widehat{\text{AMSE}}_{n, g}(n, h)$ as in (3.7).

Then $\hat{h}_n/h_n^o \rightarrow 1$ in probability.

The additional condition imposed on the b -function of having a non-zero right derivative at 0, can be translated to the following additional condition on the distribution function F itself: The second derivative of F exists on some interval (\tilde{t}, ∞) and

$$\lim_{t \rightarrow \infty} \frac{1}{F'(t)} \frac{d}{dt} \left(\frac{tF'(t)}{1-F(t)} \right) = \frac{C}{\gamma^2}$$

with $C \neq 0$. Then we can write $b'(s) = C(1 + o(1))$ as $s \downarrow 0$.

Although this condition on b of having a non-zero right derivative at 0 arises quite naturally from the curve estimation point of view, it is a rather strict condition in the situation of extreme value index estimation. As mentioned before, the optimal bandwidth may depend on characteristics of the function b . For example, consider a shifted version of the family of Generalized Pareto Distributions (for computational convenience, we shifted the support of the distributions to $[1, \infty)$):

$$F_{SGPD}(x; \gamma, \sigma) = 1 - \left(1 + \frac{\gamma}{\sigma}(x-1) \right)^{-1/\gamma} \quad x \geq 1$$

with $\gamma \in \mathbb{R}$ and $\sigma > 0$.

The associated b -function is then given by

$$b(s) = \gamma \frac{(\sigma - \gamma)s^\gamma}{\sigma - (\sigma - \gamma)s^\gamma} \quad 0 < s \leq 1$$

Hence Theorem 3.2 now only applies to the case $\gamma = 1$, $\sigma \neq \gamma$. (Remember that we restricted ourselves to positive γ .)

So even in this parametric setting our theorem only applies to a small subclass. Actually our theorem reduces the class of possible distributions to the class in which the b -function is (approximately) linear near 0. This fixes the rate with which the optimal bandwidth should tend to 0 to $n^{-1/3}$, though in general this rate depends on the underlying distribution.

A class of distribution functions with regularly varying tail that is quite frequently used to study the behaviour of estimators of the tail index γ , is known as Hall's model:

$$1 - F(x) = C_1 x^{-1/\gamma} \left(1 + C_2 x^{-\beta} (1 + o(1)) \right) \quad \text{as } x \rightarrow \infty$$

with $C_1 > 0$, $C_2 \neq 0$, $\gamma, \beta > 0$. Our assumption of linearity of b near 0 then resembles Hall's model with $\beta = 1/\gamma$. As discussed in HALL AND WELSH (1985), this is the case e.g., when F is a power of a smooth function with $\lim_{x \rightarrow \infty} F''(x) \neq 0$, when F itself is an

extreme value distribution $\exp(-x^{-1/\gamma})$ with index $\gamma > 1$ or when F is a (max-)stable distribution with index $\gamma > 1$.

Whether or not this bootstrap-inspired bandwidth selection procedure is valid in, or can be extended to, a more general situation, still remains to be investigated.

The above mentioned estimator for the bias, $\hat{\beta}_{n,g}(h)$, can also be used to construct bias-corrected confidence intervals. Let p_α denote the α -quantile of the standard normal distribution. We only give the result for upper confidence bounds, lower bounds and two-sided intervals can be constructed similarly.

Theorem 3.3

Under the assumptions of Theorem 3.2,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \gamma \leq -\hat{\beta}_{n,g_n}(\hat{h}_n) + \hat{\gamma}(n, \hat{h}_n) \left(1 + \frac{\sigma_K p_\alpha}{\sqrt{n \hat{h}_n}} \right) \right\} = 1 - \alpha$$

3.3 Basic decomposition

In the proofs of Theorem 3.1 and Theorem 3.2 we make use of a decomposition of our estimator $\hat{\gamma}(n, h)$ into a weighted sum of i.i.d. random variables and a remainder term. In this section we will introduce that decomposition and we will derive a bound for the remainder term.

Enlarging, if necessary, the basic probability space on which the X_i 's are defined, we can write $\log X_{(i)} = \log Q(U_{(i)})$ with $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}$ the order statistics from a sample from the uniform distribution on the interval $(0, 1)$ and $\log Q(s)$ satisfying equation (3.2). Hence, we can rewrite our estimator as $\hat{\gamma}(n, h) = \hat{\gamma}^{(1)}(n, h) + \hat{\gamma}^{(2)}(n, h)$ with

$$\begin{aligned} \hat{\gamma}^{(1)}(n, h) &= \gamma \sum_{i=1}^n \frac{i}{n} K_h \left(\frac{i}{n} \right) \left(-\log(1 - U_{(n-i+1)}) + \log(1 - U_{(n-i)}) \right) \\ \hat{\gamma}^{(2)}(n, h) &= \sum_{i=1}^n \frac{i}{n} K_h \left(\frac{i}{n} \right) \int_{1-U_{(n-i+1)}}^{1-U_{(n-i)}} \frac{b(s)}{s} ds \end{aligned}$$

with the convention that $U_{(0)} = 0$. Note that the variables $W_{n,1}, \dots, W_{n,n}$ given by

$$W_{n,i} = i \left(-\log(1 - U_{(n-i+1)}) + \log(1 - U_{(n-i)}) \right) \quad (3.8)$$

are the scaled spacings of a sample from an exponential distribution with mean 1. It is well known, see e.g., PYKE (1965) p.400, that these normalized spacings are again independent and exponentially distributed with mean 1. Hence, the first part in the decomposition of our estimator is just a weighted sum of i.i.d. standard exponential variables:

$$\hat{\gamma}^{(1)}(n, h) = \gamma \sum_{i=1}^n \frac{1}{n} K_h \left(\frac{i}{n} \right) W_{n,i}$$

Therefore, standard central limit theory in classical or functional form, applies directly to this part.

To handle the second part of the decomposition, let $V_i = 1 - U_i$ and let the empirical distribution function of V_1, \dots, V_n be denoted by H_n , i.e.,

$$H_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[V_i, \infty)}(t)$$

Note that the V_i again form a sample from a uniform $(0, 1)$ distribution. As H_n takes the value i/n on the interval $[V_{(i)}, V_{(i+1)}) = [1 - U_{(n-i+1)}, 1 - U_{(n-i)})$ we have

$$\hat{\gamma}^{(2)}(n, h) = \int_{V_{(1)}}^1 H_n(s) K_h(H_n(s)) \frac{b(s)}{s} ds \quad (3.9)$$

Next we will derive a bound for $\hat{\gamma}^{(2)}(n, th_n)$ uniformly in $t \in T$, with T and h_n as in Theorem 3.1.

It is well known that the uniform empirical process

$$B_n(t) = \sqrt{n}(H_n(t) - t) \quad 0 \leq t \leq 1$$

converges in distribution to a Brownian bridge. Although this fact is not explicitly used in the following analysis, it did motivate us to consider the following further decomposition of equation (3.9).

Use

$$H_n(s)K_h(H_n(s)) = sK_h(s) + (H_n(s) - s)K_h(H_n(s)) + s(K_h(H_n(s)) - K_h(s))$$

to obtain the decomposition

$$\hat{\gamma}^{(2)}(n, h) = \beta(h) - I_{n,1}(h) + I_{n,2}(h) + I_{n,3}(h)$$

with $\beta(h)$ as in (3.6) and

$$\begin{aligned} I_{n,1}(h) &= \int_0^{V_{(1)}} K_h(s) b(s) ds \\ I_{n,2}(h) &= \int_{V_{(1)}}^1 \frac{1}{\sqrt{n}} B_n(s) K_h\left(s + \frac{1}{\sqrt{n}} B_n(s)\right) \frac{b(s)}{s} ds \\ I_{n,3}(h) &= \int_{V_{(1)}}^1 \left(K_h\left(s + \frac{1}{\sqrt{n}} B_n(s)\right) - K_h(s) \right) b(s) ds \end{aligned}$$

It follows from the assumptions (3.3) on K , that $\|K\|_\infty = \sup_x |K(x)| < \infty$. From $V_{(1)} = O_P(1/n)$ we obtain for the $I_{n,1}$ -term

$$\begin{aligned} |I_{n,1}(th_n)| &\leq \sup_{0 < s \leq V_{(1)}} |b(s)| \int_0^{V_{(1)}} |K_{th_n}(s)| ds \\ &\leq o_P(1) \frac{1}{t_1 h_n} \|K\|_\infty O_P(1/n) \end{aligned}$$

hence,

$$\sup_{t \in T} |I_{n,1}(th_n)| = o_P\left(\frac{1}{nh_n}\right)$$

To handle the other $I_{n,i}$ -terms, we will need the support condition on K , i.e., that K_h vanishes outside $[0, h]$. Then

$$\begin{aligned} |I_{n,2}(th_n)| &\leq \int_0^{2t, h_n} \frac{1}{\sqrt{n}} |B_n(s)| \left| K_{th_n}\left(s + \frac{1}{\sqrt{n}} B_n(s)\right) \right| |b(s)| \frac{1}{s} ds \\ &\quad + \int_{2t, h_n}^{2t, h_n \vee H_n^{-1}(th_n)} \frac{1}{\sqrt{n}} |B_n(s)| \left| K_{th_n}\left(s + \frac{1}{\sqrt{n}} B_n(s)\right) \right| |b(s)| \frac{1}{s} ds \end{aligned}$$

Note that the second term in this bound is needed to include the case $s + B_n(s)/\sqrt{n} = H_n(s) > s$, in which case we have to integrate upto $H_n^{-1}(th_n)$. However, the probability that this second integral differs from zero is bounded from above by $\mathbb{P}(H_n^{-1}(t, h_n) > 2t, h_n)$, which tends to 0, so that this term is of order $o_P(1/\sqrt{nh_n})$. For the first integral use the crude bound

$$\frac{1}{\sqrt{n}} \frac{1}{t, h_n} \|K\|_\infty \sup_{0 < s \leq 2t, h_n} |b(s)| \int_0^{2t, h_n} \frac{1}{s} |B_n(s)| ds$$

and the fact that

$$\begin{aligned} \mathbb{E} \int_0^{2t, h_n} \frac{1}{s} |B_n(s)| ds &\leq \int_0^{2t, h_n} \frac{1}{s} (\mathbb{E}(B_n(s)^2))^{1/2} ds \\ &= \int_0^{2t, h_n} \frac{1}{s} \sqrt{s(1-s)} ds \\ &\leq \int_0^{2t, h_n} \frac{1}{\sqrt{s}} ds = \frac{1}{2} \sqrt{2t, h_n} \end{aligned}$$

to deduce, using Markov's inequality, that

$$\sup_{t \in T} |I_{n,2}(th_n)| = o_P\left(\frac{1}{\sqrt{nh_n}}\right)$$

Next consider $I_{n,3}$:

$$\begin{aligned} |I_{n,3}(th_n)| &\leq \int_0^{t, h_n \wedge H_n^{-1}(th_n)} \left| K_{th_n}\left(s + \frac{1}{\sqrt{n}} B_n(s)\right) - K_{th_n}(s) \right| |b(s)| ds \\ &\quad + \int_{t, h_n \wedge H_n^{-1}(th_n)}^{t, h_n \vee H_n^{-1}(th_n)} \left| K_{th_n}\left(s + \frac{1}{\sqrt{n}} B_n(s)\right) - K_{th_n}(s) \right| |b(s)| ds \end{aligned}$$

Use the Lipschitz continuity of K , i.e., $|K_h(x) - K_h(y)| \leq C|x - y|/h^2$ for $x, y \in (0, h)$ to obtain the following bound for the first integral:

$$\frac{C}{t, h_n^2} \sup_{0 < s \leq t, h_n} |b(s)| \int_0^{t, h_n} \frac{1}{\sqrt{n}} |B_n(s)| ds$$

Since

$$\begin{aligned} \mathbb{E} \int_0^{t_r h_n} \frac{1}{\sqrt{n}} |B_n(s)| ds &\leq \int_0^{t_r h_n} \frac{1}{\sqrt{n}} (\mathbb{E}(B_n(s))^2)^{1/2} ds \\ &= \int_0^{t_r h_n} \frac{1}{\sqrt{n}} \sqrt{s(1-s)} ds \\ &\leq \int_0^{t_r h_n} \frac{1}{\sqrt{n}} \sqrt{s} ds = \frac{2}{3\sqrt{n}} (t_r h_n)^{3/2} \end{aligned}$$

the first integral is indeed of order $o_P(1/\sqrt{nh_n})$.

Note that in the second integral one of the K -functions is identically zero. Hence we may derive the following bound:

$$\|K\|_\infty \sup_{s \in \Delta_n(t)} |b(s)| \frac{1}{th_n} |H_n^{-1}(th_n) - th_n|$$

with $\Delta_n(t) = \{s : (th_n \wedge H_n^{-1}(th_n)) < s < (th_n \vee H_n^{-1}(th_n))\}$. Then use a slightly modified version of Theorem 6.5 of EINMAHL (1994) to get weak convergence of the tail empirical quantile process $v_n(s) = \sqrt{nh_n}(H_n^{-1}(sh_n) - sh_n)/h_n$ for $s \in [0, 1]$ to a Brownian Motion. A simple rescaling then yields

$$\frac{1}{th_n} |H_n^{-1}(th_n) - th_n| = O_P(1/\sqrt{nh_n})$$

uniformly in $t \in T$.

Finally, use that

$$\begin{aligned} \sup_{s \in \Delta_n(t)} |b(s)| &\leq \max\left\{ \sup_{0 < s < t_r h_n} |b(s)|, \sup_{0 < s < H_n^{-1}(t_r h_n)} |b(s)| \right\} \\ &= \max\{o(1), o_P(1)\} = o_P(1) \end{aligned}$$

to obtain $o_P(1/\sqrt{nh_n})$ -behaviour of $I_{n,3}$, uniformly over T .

Hence, we may conclude that

$$\sup_{t \in T} \sqrt{nh_n} \left| \hat{\gamma}^{(2)}(n, th_n) - \beta(th_n) \right| = o_P(1)$$

which is the bound on the remainder term we will need in the next section.

3.4 Proofs

3.4.1 Proof of Theorem 3.1

In Theorem 3.1 we stated the convergence of the processes Z_n with index set T and

$$Z_n(t) = \sqrt{nh_n} (\hat{\gamma}(n, th_n) - \gamma - \beta(th_n))$$

to a Gaussian limit process Z . However, since the previous section resulted in a decomposition of our estimator in two parts, with the second part satisfying

$$\sup_{t \in T} \sqrt{nh_n} \left(\hat{\gamma}^{(2)}(n, th_n) - \beta(th_n) \right) = o_P(1)$$

we arrived in the more familiar situation of i.i.d. random variables. Hence, to prove Theorem 3.1 we now only have to deal with $\hat{\gamma}^{(1)}(n, th_n)$ and its convergence.

Define the stochastic processes $A_n : T \rightarrow \mathbb{R}$, $n \in \mathbb{N}$ by

$$\begin{aligned} A_n(t) &= \sqrt{nh_n} \left(\hat{\gamma}^{(1)}(n, th_n) - \mathbb{E} \hat{\gamma}^{(1)}(n, th_n) \right) \\ &= \frac{\gamma}{t\sqrt{nh_n}} \sum_{i=1}^n K \left(\frac{i}{nh_n t} \right) (W_{n,i} - 1) \end{aligned}$$

then it suffices for the proof of Theorem 3.1, to show that the A_n converge in distribution to the Gaussian process Z mentioned in the theorem and that $\mathbb{E} \hat{\gamma}^{(1)}(n, th_n) = \gamma + o(1/\sqrt{nh_n})$ uniformly in $t \in T$.

For the convergence in distribution, note that the A_n 's can be considered as the row sums of an array of random functions, independent within rows. To establish the desired convergence in distribution, we could use the functional central limit theorem as formulated in POLLARD (1990). However, as the random functions are continuous functions on a compact interval of the real line, we do not have to use Pollard's theorem. In our, rather simple case we can use the techniques of BILLINGSLEY (1968).

First we will show the convergence of the finite dimensional distributions. Therefore define

$$Y_n = \sum_{j=1}^m a_j A_n(t_j)$$

with $t_1, \dots, t_m \in T$ and $a_1, \dots, a_m \in \mathbb{R}$. Then, interchanging the order of summations, we may rewrite Y_n into $Y_n = \sum_{i=1}^n \eta_{n,i}$ with

$$\eta_{n,i} = \frac{\gamma}{\sqrt{nh_n}} \left(\sum_{j=1}^m \frac{a_j}{t_j} K \left(\frac{i}{nh_n t_j} \right) \right) (W_{n,i} - 1)$$

By construction, the $\eta_{n,i}$'s are independent, have zero means and variances given by

$$\text{var } \eta_{n,i} = \frac{\gamma^2}{nh_n} \left(\sum_{j=1}^m \frac{a_j}{t_j} K \left(\frac{i}{nh_n t_j} \right) \right)^2$$

hence $\mathbb{E} Y_n = 0$ and

$$\text{var } Y_n = \gamma^2 \sum_{j=1}^m \sum_{k=1}^m a_j a_k \sum_{i=1}^n \frac{1}{nh_n t_j t_k} K \left(\frac{i}{nh_n t_j} \right) K \left(\frac{i}{nh_n t_k} \right)$$

As K is a Lipschitz continuous function, we have

$$\begin{aligned} \int K(t_j s) K(t_k s) ds &= \frac{1}{t_j t_k} \int K\left(\frac{s}{t_j}\right) K\left(\frac{s}{t_k}\right) ds \\ &= \sum_{i=1}^n \frac{1}{nh_n} \frac{1}{t_j t_k} K\left(\frac{i}{nh_n t_j}\right) K\left(\frac{i}{nh_n t_k}\right) + O\left(\frac{1}{nh_n}\right) \end{aligned}$$

so

$$\text{var } Y_n \rightarrow \gamma^2 \sum_{j=1}^m \sum_{k=1}^m a_j a_k \int K(t_j s) K(t_k s) ds$$

which is the variance of $\sum_{j=1}^m a_j Z(t_j)$. Moreover, $\eta_{n,i} = 0$ for $i > t_r nh_n$ and, for $i \leq t_r nh_n$

$$\begin{aligned} \mathbb{E} \eta_{n,i}^4 &= \frac{\gamma^4}{n^2 h_n^2} \left(\sum_{j=1}^m \frac{a_j}{t_j} K\left(\frac{i}{nh_n t_j}\right) \right)^4 \mathbb{E} (W_{n,i} - 1)^4 \\ &= O\left(\frac{1}{n^2 h_n^2}\right) \end{aligned}$$

uniformly in i , so that $\sum_{i=1}^n \mathbb{E} \eta_{n,i}^4 = O(1/(nh_n))$. Hence Liapunov's condition is satisfied and the central limit theorem applies. So, for all $(a_1, \dots, a_m) \in \mathbb{R}^m$ we have

$$\sum_{j=1}^m a_j A_n(t_j) \xrightarrow{\mathcal{D}} \sum_{j=1}^m a_j Z(t_j)$$

and, using the Cramér-Wold device, we may conclude that the finite dimensional distributions converge.

Rests us to prove tightness to obtain the convergence in distribution of the processes A_n . Following BILLINGSLEY (1968), eq. (12.51), it suffices to show that, for some $C < \infty$

$$\mathbb{E} (A_n(t) - A_n(s))^2 \leq C(t-s)^2$$

for all $s, t \in T$. Since the variables $W_{n,1}, \dots, W_{n,n}$, as defined in (3.8) are independent, it follows that

$$\mathbb{E} (A_n(t) - A_n(s))^2 = \frac{\gamma^2}{nh_n} \sum_{i=1}^n \left(\frac{1}{t} K\left(\frac{i}{nh_n t}\right) - \frac{1}{s} K\left(\frac{i}{nh_n s}\right) \right)^2$$

Note that, for all $s, t \in T$ the summands vanish for $i > nh_n t_r$. On the other hand

$$\begin{aligned} &\left| \frac{1}{t} K\left(\frac{i}{nh_n t}\right) - \frac{1}{s} K\left(\frac{i}{nh_n s}\right) \right| \\ &\leq \frac{1}{t} \left| K\left(\frac{i}{nh_n t}\right) - K\left(\frac{i}{nh_n s}\right) \right| + \left| K\left(\frac{i}{nh_n s}\right) \right| \left| \frac{1}{t} - \frac{1}{s} \right| \\ &\leq C'|t-s| \end{aligned}$$

for all $s, t \in T$, with some constant C' not depending on n or i . These arguments, put together, imply the tightness criterion, and $A_n \xrightarrow{\mathcal{D}} Z$ follows.

As stated at the beginning of this section, to complete the proof of Theorem 3.1 we still have to show that, uniformly in $t \in T$

$$\mathbb{E} \left(\hat{\gamma}^{(1)}(n, th_n) \right) = \frac{\gamma}{inh_n} \sum_{i=1}^n K \left(\frac{i}{nh_{nt}} \right) = \gamma + o(1/\sqrt{nh_n})$$

However, again using the Lipschitz continuity of K , we get

$$\begin{aligned} \frac{\gamma}{inh_n} \sum_{i=1}^n K \left(\frac{i}{inh_n} \right) &= \gamma \sum_{i=1}^{\lfloor inh_n \rfloor} \int_{\frac{i-1}{n}}^{\frac{i}{n}} (K_{th_n}(u) + (K_{th_n}(i/n) - K_{th_n}(u))) \, du \\ &= \gamma \int_0^{\lfloor inh_n \rfloor/n} K_{th_n}(u) \, du + O \left(\frac{\lfloor inh_n \rfloor}{t^2 n^2 h_n^2} \right) \\ &= \gamma + O(1/(nh_n)) \end{aligned}$$

uniformly in $t \in T$.

3.4.2 Proof of Corollary 3.1

Let F, K, h_n, Z_n, Z and $T = [t_l, t_r]$ be as in Theorem 3.1. Assume that the random sequence \hat{h}_n is such that $\hat{h}_n/h_n \rightarrow 1$ in probability and define $t_n = \hat{h}_n/h_n$. Theorem 3.1 then gives that the random vector $(Z_n, t_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow C(t_l, t_r) \times \mathbb{R}$ converges in distribution to the vector $(Z, 1)$. Then use the Skorohod-Dudley representation theorem, to deduce that there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ carrying the random variables $(\tilde{Z}_n, \tilde{t}_n)$ and \tilde{Z} , with the property that $(\tilde{Z}_n, \tilde{t}_n) \stackrel{\mathcal{D}}{=} (Z_n, t_n)$, $\tilde{Z} \stackrel{\mathcal{D}}{=} Z$ and $(\tilde{Z}_n, \tilde{t}_n) \rightarrow (\tilde{Z}, 1)$ a.s. Since \tilde{Z} has continuous sample paths, we obtain that $\tilde{Z}_n(\tilde{t}_n) \xrightarrow{\mathcal{D}} \tilde{Z}(1)$. Finally return to the original probability space to obtain that

$$Z_n(t_n) \xrightarrow{\mathcal{D}} Z(1) = \mathcal{N}(0, \gamma^2 \sigma_K^2)$$

3.4.3 Proof of Theorem 3.2

Under the assumptions of Theorem 3.2, the bias part of the actual $AMSE(n, h)$ is of the form

$$\beta(h) = \int_0^1 K(s)b(hs) \, ds = hb'(0)C_1 + o(h) \tag{3.10}$$

with $C_1 = \int sK(s) \, ds$. We will first show that the estimate $\hat{\beta}_{n,g}$ is close to this equation:

Lemma 3.1

Under the assumptions of Theorem 3.2 we have

$$\hat{\beta}_{n,g}(h) = hb'(0)C_1 + o(h) + o_P(1/\sqrt{nh})$$

with $C_1 = \int sK(s) \, ds$.

Proof of Lemma 3.1:

Following similar arguments and using the same notation as in the basic decomposition of $\hat{\gamma}(n, h)$, we can decompose $\hat{b}_{n,g}$ into

$$\begin{aligned} \hat{b}_{n,g}(x) &= \gamma \sum_{i=1}^n \frac{1}{n} \left[K_g \left(\frac{i}{n} - x \right) - K_g \left(\frac{i}{n} \right) \right] W_{n,i} \\ &\quad + \sum_{j=0}^3 (I_{n,j}(x; g) - I_{n,j}(0; g)) \end{aligned}$$

where

$$I_{n,0}(x; g) = \int_0^1 K_g(s-x)b(s) ds$$

$$I_{n,1}(x; g) = - \int_0^{V(1)} K_g(s-x)b(s) ds$$

$$I_{n,2}(x; g) = \int_{V(1)}^1 \frac{1}{\sqrt{n}} B_n(s) K_g \left(s + \frac{1}{\sqrt{n}} B_n(s) - x \right) \frac{b(s)}{s} ds$$

$$I_{n,3}(x; g) = \int_{V(1)}^1 \left(K_g \left(s + \frac{1}{\sqrt{n}} B_n(s) - x \right) - K_g(s-x) \right) b(s) ds$$

Note that it is sufficient to consider $\hat{b}_{n,g}(x)$ uniformly in $x \in [0, h]$.

The weighted sum of i.i.d. random variables can be dealt with using the assumptions on the kernel K :

$$\begin{aligned} \text{var} \left\{ \frac{1}{n} \sum_{i=1}^n \left[K_g \left(\frac{i}{n} - x \right) - K_g \left(\frac{i}{n} \right) \right] W_{n,i} \right\} \\ &= \frac{1}{n^2} \sum_{i=1}^n \left[K_g \left(\frac{i}{n} - x \right) - K_g \left(\frac{i}{n} \right) \right]^2 \\ &= O \left(\frac{ng}{n^2 g^2} \right) = o \left(\frac{1}{nh} \right) \end{aligned}$$

and

$$\begin{aligned} \text{IE} \left\{ \frac{1}{n} \sum_{i=1}^n \left[K_g \left(\frac{i}{n} - x \right) - K_g \left(\frac{i}{n} \right) \right] W_{n,i} \right\} \\ &= \left[\int_0^1 K_g(u-x) du + O \left(\frac{1}{ng} \right) \right] - \left[\int_0^1 K_g(u) du + O \left(\frac{1}{ng} \right) \right] \\ &= o \left(\frac{1}{nh} \right) \end{aligned}$$

both bounds being uniform in $x \in [0, h]$.

Now we will show that the difference $I_{n,0}(hs; g) - I_{n,0}(0; g)$ leads to the desired equation for the bias:

$$\begin{aligned} I_{n,0}(hs; g) - I_{n,0}(0; g) &= \int_0^1 (K_g(x - hs) - K_g(x)) b(x) dx \\ &= \int_0^1 K(z) (b(gz + hs) - b(gz)) dz \end{aligned}$$

since $h \downarrow 0$ and $g \downarrow 0$, hence $(1 - hs)/g > 1$ for n large enough and all $s \in [0, 1]$. Using the assumptions on $b(\cdot)$ then yields

$$\begin{aligned} \int_0^1 K(z) (b'(gz)hs + o(h)) dz &= hs \int_0^1 K(z) (b'(0) + o(1)) dz \\ &= hsb'(0) + o(h) \end{aligned}$$

with $o(h)$ uniform in $s \in [0, 1]$. Hence

$$\int_0^1 K(s) (I_{n,0}(hs; g) - I_{n,0}(0; g)) ds = hsb'(0)C_1 + o(h)$$

as required.

The remaining terms $I_{n,j}(hs; g)$ with $j = 1, \dots, 3$ will give, using similar arguments as we used in finding a bound for the remainder term in the basic decomposition of $\hat{y}(h, n)$, that $I_{n,j}(hs; g) = o_P(1/\sqrt{nh})$ for $j = 2, 3$ and that $I_{n,1}(hs; g) = o_P(1/(nh))$, again, uniform in $s \in [0, 1]$. Thus

$$\hat{\beta}_{n,g}(h) = hsb'(0)C_1 + o(h) + o_P(1/\sqrt{nh})$$

■

From equation (3.10) it follows that a minimizing value h_n^o of $h \mapsto \text{AMSE}(n, h)$ is of the order $n^{-1/3}$ and the corresponding $\text{AMSE}(n, h_n^o)$ of the order $n^{-2/3}$. Thus, as a consequence of Lemma 3.1 we have that

$$\sup_{h \in \mathcal{H}_n} n^{2/3} \left| \widehat{\text{AMSE}}_{n,g_n}(n, h) - \text{AMSE}(n, h) \right| \rightarrow 0 \quad \text{in probability}$$

with $\mathcal{H}_n = \{h_n | an^{-1/3} < h_n < bn^{-1/3}\}$ for arbitrary $0 < a \leq b < \infty$.

To complete the proof of Theorem 3.2, we will follow the lines of the proof of Theorem 3 in RICE (1984).

Define

$$\begin{aligned} l(z) &= \lim_{n \rightarrow \infty} n^{2/3} \text{AMSE}(n, zn^{-1/3}) \\ &= (zb'(0)C_1)^2 + \frac{\gamma^2}{z} C_2 \end{aligned}$$

with the constants C_1 and C_2 only depending on the kernel K . Then

$$\sup_{a \leq z \leq b} \left| n^{2/3} \left(\widehat{\text{AMSE}}(n, zn^{-1/3}) - l(z) \right) \right| \rightarrow 0$$

for arbitrary $0 < a \leq b < \infty$. Using that $l(z)$ is strictly convex and infinitely often differentiable on $(0, \infty)$, we obtain that $z_o = \operatorname{argmin}_z l(z)$ is unique. Indeed, $n^{1/3}h_n^o \rightarrow z_o$ as $n \rightarrow \infty$, provided that $a < z_o < b$.

Define $z_n = n^{1/3}h_n^o$ and $\hat{z}_n = n^{1/3}\hat{h}_n$. Fix $\delta > 0$ and define the function $D(\cdot)$ by $D(\delta) = \inf_{|z-z_o|>\delta} (l(z) - l(z_o))$ and $\hat{l}(z) = n^{2/3}\widehat{\text{AMSE}}_{n,g_n}(n, zn^{-1/3})$ with minimizer \hat{z}_n . Then

$$\begin{aligned} \mathbb{P}(|\hat{z}_n - z_o| > \delta) &\leq \mathbb{P}(l(\hat{z}_n) - l(z_o) > D(\delta)) \\ &\leq \mathbb{P}(l(\hat{z}_n) - \hat{l}(\hat{z}_n) + \hat{l}(z_o) - l(z_o) > D(\delta)) \\ &\leq \mathbb{P}(l(\hat{z}_n) - \hat{l}(\hat{z}_n) > D(\delta)/2) + \mathbb{P}(\hat{l}(z_o) - l(z_o) > D(\delta)/2) \end{aligned}$$

However, these terms converge to 0 by the remarks following the proof of Lemma 3.1. Hence

$$n^{1/3}(\hat{h}_n - h_n^o) \rightarrow 0 \quad \text{in probability}$$

or

$$\frac{\hat{h}_n}{h_n^o} \rightarrow 1 \quad \text{in probability}$$

3.4.4 Proof of Theorem 3.3

By Theorem 3.2 we have that $\hat{h}_n/h_n^o \rightarrow 1$ in probability and hence, by Corollary 3.1

$$\sqrt{nh_n^o}(\hat{\gamma}(n, \hat{h}_n) - \gamma - \beta(\hat{h}_n)) \xrightarrow{D} \mathcal{N}(0, \gamma^2 \sigma_K^2)$$

Moreover $\hat{\gamma}(n, \hat{h}_n)/\gamma \rightarrow 1$ in probability.

Since $h_n^o \in \mathcal{H}_n$ for some $0 < a \leq b < \infty$, see remarks following the proof of Lemma 3.1, and $\hat{h}_n/h_n^o \rightarrow 1$ in probability, it follows that, $\forall \varepsilon > 0$

$$\mathbb{P}\left(an^{-1/3}(1 - \varepsilon) < \hat{h}_n < bn^{-1/3}(1 + \varepsilon) \right) \rightarrow 1$$

hence, $\hat{h}_n \in \tilde{\mathcal{H}}_n$ with probability tending to 1, where $\tilde{\mathcal{H}}_n = \{h_n | \frac{1}{2}an^{-1/3} < h_n < \frac{3}{2}bn^{-1/3}\}$. Using Lemma 3.1 we then derive that

$$\sup_{h \in \tilde{\mathcal{H}}_n} n^{1/3} \left| \beta(h) - \hat{\beta}_{n,g_n}(h) \right| \rightarrow 0 \quad \text{in probability}$$

and hence, $\forall \varepsilon > 0$

$$\mathbb{P}\left(n^{1/3} \left| \beta(\hat{h}_n) - \hat{\beta}_{n,g_n}(\hat{h}_n) \right| > \varepsilon\right) \rightarrow 0$$

So

$$\frac{\sqrt{n\hat{h}_n}}{\hat{\gamma}(n, \hat{h}_n)} \left(\hat{\gamma}(n, \hat{h}_n) - \gamma - \hat{\beta}_{n,g_n}(\hat{h}_n) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_K^2)$$

which one can use to construct asymptotic (bias-corrected) confidence intervals.

Chapter 4

A new type of estimator

In this chapter a new type of estimator of the extreme value index will be introduced and its asymptotic properties will be discussed. In the calculation of this new estimator two parameters have to be chosen appropriately. When the proof of the theorems concerning the introduced estimator is given, we will illustrate the behaviour of the estimator as a function of these two parameters by means of a small simulation study.

This new estimator is a generalization of the estimator of the previous chapter in the sense that it is a consistent estimator for both positive and negative extreme value indices, whereas the estimator of the previous chapter could only be used in case of a positive extreme value index. In the final section of the present chapter, the new estimator will be related to other existing estimators.

4.1 Introduction

Throughout this chapter, we will be concerned with a sample X_1, \dots, X_n of i.i.d. random variables with common distribution function F that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \neq 0$. Denote the ascending order statistics of such a sample by $X_{(1)} \leq \dots \leq X_{(n)}$. Consider the function $\phi(\cdot)$ defined by

$$\phi(s) = -s \frac{d}{ds} \log Q(1-s) \quad (4.1)$$

with $Q(s) = F^{-1}(s)$ the quantile function of the variables X_i , assuming existence and differentiability of $\log Q(1-\cdot)$. In case of a Generalized Pareto Distribution $F_{GPD}(\cdot; \gamma, \sigma)$, as defined in (1.6), the corresponding ϕ -function is given by

$$\phi_{GPD}(s; \gamma) = \frac{\gamma}{1-s^\gamma} \quad \text{for } \gamma \neq 0, \quad s \in (0, 1)$$

Note that this function does not depend on the parameter σ present in the distribution function $F_{GPD}(\cdot; \gamma, \sigma)$.

Results in BALKEMA AND DE HAAN (1974) and PICKANDS (1975) indicate that the tail of a distribution function $F \in \mathcal{D}(G_\gamma)$ can be approximated by a Generalized Pareto

Distribution with the same γ but with a σ depending on the threshold above which the tail is defined to start. (See also (1.3) and (1.5) in this thesis.) Since $\phi_{GPD}(\cdot; \gamma)$ does not depend on the parameter σ , it is appealing to try to approximate, in a neighbourhood of 0, the ϕ -function corresponding to F by one corresponding to a Generalized Pareto Distribution with the same γ but with arbitrary $\sigma > 0$. The main idea behind the new estimator of the extreme value index is based on that observation.

Indeed, we will use a kernel type method to estimate the ϕ -function corresponding to F on the interval $[s_o, 1]$ non-parametrically, with s_o a value close to 0 and call that estimator $\hat{\phi}_h(s)$. Then we will continuously extend that estimator to the interval $(0, s_o]$ using $\phi_{GPD}(\cdot; \gamma)$. Note that the continuity of the extension yields an estimator of γ : continuity at s_o implies

$$\hat{\phi}_h(s_o) = \phi_{GPD}(s_o; \gamma) = \frac{\gamma}{1 - s_o^\gamma}$$

and hence yields an estimator for γ , since for each fixed s_o , $\phi_{GPD}(s_o; \gamma)$ is a monotone function of γ and is therefore invertible.

Intuitively, the closeness of s_o to zero should depend on the sample size n , for s_o is related to the threshold above which the approximation of the tail of the underlying distribution function by one of a Generalized Pareto Distribution is considered to start to be effective. In the next section we will see that the rate at which s_o tends to 0 depends on the non-parametric method that is used to estimate $\phi(\cdot)$.

Obviously, if one is only interested in the estimation of the extreme value index, it suffices to consider the $\hat{\phi}_h(\cdot)$ only at the point s_o . Estimating $\phi(\cdot)$ on a non-empty subinterval of $(0, 1)$ could be used to estimate (extreme) quantiles. Moreover, an estimator of the quantile function can be constructed in a similar way. Such an estimator could be used to draw bootstrap-samples in order to try to find a mean squared error optimal smoothing-parameter s_o . In the present chapter however, only the estimation of the extreme value index itself will be considered.

4.2 Main results

Let X_1, \dots, X_n be a sample from a distribution function F in the domain of an extreme value distribution G_γ and let ϕ be defined as in (4.1). Let $K: \mathbb{R} \rightarrow \mathbb{R}$ be such that

$$[\text{KC1}] \quad K(x) = 0 \quad x \notin (-1, 1)$$

$$[\text{KC2}] \quad 0 < K(x) < \infty \quad x \in (-1, 1)$$

$$[\text{KC3}] \quad \text{The derivative } dK(x)/dx \text{ exists and is uniformly bounded on } (-1, 1)$$

$$[\text{KC4}] \quad \int K(x) dx = 1$$

Define the boundary kernel $\tilde{K}_t(\cdot)$ by

$$\tilde{K}_t(x) = (\alpha_t + x\beta_t)K(x)1_{(-1,t)}(x)$$

where α_t and β_t are determined by the equations

$$\int_{-1}^t (\alpha_t + u\beta_t)K(u) du = 1 \quad \text{and} \quad \int_{-1}^t u(\alpha_t + u\beta_t)K(u) du = 0 \quad (4.2)$$

Note that condition [KC2] implies that both α_t and β_t are bounded uniform in $t \in [0, 1]$, hence that the boundary kernel is bounded, uniform in $t \in [0, 1]$ and $x \in (-1, 1)$.

We will make frequent use of a function that is closely related to this boundary kernel. Define for $t \in (0, 1)$ the function $\Psi_t(\cdot)$ as

$$\Psi_t(u) = u\tilde{K}_t(t-u) \quad (4.3)$$

Note that equations (4.2) imply that

$$\int_0^{t+1} \frac{\Psi_t(u)}{u} du = 1 \quad \text{and} \quad \int_0^{t+1} (t-u) \frac{\Psi_t(u)}{u} du = 0 \quad (4.4)$$

moreover, these equalities can be combined to obtain

$$\int_0^{t+1} \Psi_t(u) du = t \quad (4.5)$$

As stated in the previous section, the closeness to zero of the point s_o at which the transition of a non-parametric estimator to a parametric one is made, should intuitively depend on the sample size n . Hence it seems reasonable to estimate $\phi(\delta_n)$ for some sequence $\delta_n \rightarrow 0$ as $n \rightarrow \infty$. However, in order to be able to use a boundary kernel, as implicitly defined in (4.3), the δ_n should be of the same order as the bandwidth h that is used, i.e., the function $\phi(\cdot)$ is estimated at the point ht , for some $t \in (0, 1)$. Therefore, define the estimator of $\phi(ht)$ by

$$\hat{\phi}_h(ht) = \sum_{i=1}^n \Psi_t\left(\frac{i}{nh}\right) \left(\log X_{(n-i+1)} - \log X_{(n-i)}\right) \quad (4.6)$$

Note that by definition of $\Psi_t(\cdot)$, approximately the $nh(t+1)$ largest order statistics will be used in the calculation of the estimate. However, the start of the parametric assumption on the underlying distribution function is at the threshold y_t where $1 - F(y_t) = ht$, hence more than just the nh values of the sample above the threshold y_t will be used. This differs from the situation of other estimators like e.g., the moment estimator and the maximum likelihood estimator (for both estimators, see Chapter 2). For these estimators only the values of the sample above a certain threshold y_t are used in the calculation of the estimator. Consequently, this new estimator depends on the chosen threshold in a much smoother way as is the case with e.g., the moment estimator and the maximum likelihood estimator.

As stated in the introduction of this chapter, we will define an estimator of γ by equating the estimated ϕ -function corresponding to the underlying distribution function

to the ϕ -function corresponding to a Generalized Pareto Distribution. To that end, first define the function $\Phi_y(\cdot)$ for a fixed $y \in (0, 1)$ by:

$$\Phi_y(x) = \begin{cases} \frac{x}{1-y^x} & \text{if } x \neq 0 \\ -\frac{1}{\log y} & \text{if } x = 0 \end{cases} \quad (4.7)$$

Note that, for each fixed $y \in (0, 1)$ this function is monotone and continuous in x and hence has an inverse $\Phi_y^{-1}(\cdot)$, although this inverse can not be expressed explicitly. Therefore, we can define the new estimator of the extreme value index:

$$\hat{\gamma}_{n,h}^G = \Phi_{ht}^{-1}(\hat{\phi}_h(ht)) \quad (4.8)$$

for some fixed $t \in (0, 1)$. I.e., this estimator is the γ that solves the equation

$$\hat{\phi}_h(ht) = \frac{\gamma}{1-(ht)^\gamma}$$

Note that, in case of a positive extreme value index and choosing $t = 0$, this estimator coincides with the kernel type estimator discussed in Chapter 3 of this thesis. Indeed, this is easily seen by equating $\Psi_0(u)$ with $uK^*(u)$ where $K^*(\cdot)$ is the kernel used in Chapter 3, and using equations (3.1) and (4.6).

Since first we will be estimating the ϕ -function rather than the underlying distribution function F itself, the conditions needed to derive asymptotic normality will be expressed in terms of this $\phi(\cdot)$ -function¹:

[PC1] If $\gamma > 0$,

$$\phi(s) = \gamma(1 + b(s))$$

with $b(s) \rightarrow 0$ as $s \rightarrow 0$, uniform on some interval $(0, \varepsilon)$.

[PC2] If $\gamma < 0$,

$$\phi(s) = -\gamma s^{-\gamma} L(s)$$

with $L(\cdot)$ slowly varying at zero and $L(s) \rightarrow L_0 > 0$ as $s \rightarrow 0$, uniform on some interval $(0, \varepsilon)$.

The conditions arise quite naturally if one considers the expansion of the ϕ -function of a Generalized Pareto Distribution in the neighbourhood of zero:

$$\phi_{GPD}(s; \gamma) = \begin{cases} \gamma(1 + O(s^{-\gamma})) & \gamma > 0 \\ -\gamma s^{-\gamma}(1 + O(s^{-\gamma})) & \gamma < 0 \end{cases} \quad \text{as } s \downarrow 0$$

¹Note that the case $\gamma = 0$ is not covered by these conditions. This type of limiting extreme value distribution is not yet considered, since the corresponding ϕ_{GPD} -function $-1/\log s$ is not as easily generalized.

More generally, we want to consider functions of the following type:

$$\phi(s) = \begin{cases} \gamma L_1(s) & \gamma > 0 \\ -\gamma s^{-\gamma} L_2(s) & \gamma < 0 \end{cases} \quad \text{as } s \downarrow 0$$

with $L_i(\cdot)$ ($i = 1, 2$) slowly varying functions at zero. The specific choices of the slowly varying functions in [PC1] and [PC2], stem from the observation that even though the extreme value index does not change under a translation of the underlying distribution function, the corresponding ϕ -function does:

$$F_a(x) = F(x - a) \quad a > 0 \quad \implies \quad \phi_a(s) = \phi(s) \frac{Q(1-s)}{Q(1-s) + a}$$

Hence, since the quantile function $Q(1 - \cdot)$ tends to the (possibly infinite) upper endpoint x_F^o of the underlying distribution function, this yields

$$\begin{aligned} \phi_a(s) &\sim \phi(s) & \gamma > 0 \\ \phi_a(s) &\sim \phi(s) \frac{x_F^o}{x_F^o + a} & \gamma < 0 \end{aligned} \quad \text{as } s \downarrow 0$$

and this is represented in [PC1] and [PC2], by taking $L_1(s) = 1 + b(s)$ with $b(s) \rightarrow 0$ as $s \rightarrow 0$ and $L_2(s) = L(s)$ respectively.

The asymptotic behaviour of the estimator of $\phi(ht)$, as process in t , is stated in the following theorem:

Theorem 4.1

Assume that the ϕ -function corresponding to the underlying distribution function F satisfies either condition [PC1] or [PC2]. Moreover, assume that the kernel K satisfies conditions [KC1]–[KC4]. Let $h = h_n = H(n)$ where, in case $\gamma > 0$, $H \in \text{RV}_{-\rho}^\infty$ for some $\rho \in (0, 1)$ and in case $\gamma < 0$, $\sqrt{nH(n)} \phi(H(n)) \rightarrow \infty$ and $H(n) \rightarrow 0$ as $n \rightarrow \infty$. Define the stochastic processes Z_n with index set $(0, 1)$ by

$$Z_n(t) = \frac{\sqrt{nh}}{\phi(h)} (\hat{\phi}_h(ht) - D_{n,h}(t))$$

where

$$D_{n,h}(t) = \int_{1/nh}^{t+1} \log Q(1 - hu) d\Psi_t(u)$$

Then $Z_n \xrightarrow{\mathcal{D}} Z$ with

$$Z(t) = \int_0^{t+1} x^{-\bar{\gamma}-1} W(x) d\Psi_t(x)$$

where $\bar{\gamma} = \gamma \wedge 0$ and $W(\cdot)$ is standard Brownian Motion. Moreover,

$$D_{n,h}(t) - \phi_{GPD}(ht; \gamma) = \int_0^{t+1} (\phi(hu) - \phi_{GPD}(ht; \gamma)) \frac{\Psi_t(u)}{u} du + o\left(\frac{\phi(h)}{\sqrt{nh}}\right)$$

REMARK 4.1

Note that Z is a Gaussian process with continuous sample paths, mean function 0 and covariance structure

$$\text{cov}(Z(t), Z(v)) = \int_0^{t+1} \int_0^{v+1} \frac{x \wedge y}{xy} (xy)^{-\bar{\gamma}} d\Psi_v(y) d\Psi_t(x)$$

REMARK 4.2

An alternative representation of the limiting process $Z(\cdot)$ is obtained using partial integration:

$$\begin{aligned} Z(t) &= [x^{-\bar{\gamma}-1} \Psi_t(x) W(x)]_0^{t+1} - \int_0^{t+1} \Psi_t(x) d(x^{-\bar{\gamma}-1} W(x)) \\ &= [x^{-\bar{\gamma}} \tilde{K}_t(x) W(x)]_0^{t+1} - \int_0^{t+1} x \tilde{K}_t(x) d(x^{-\bar{\gamma}-1} W(x)) \\ &= - \int_0^{t+1} x \tilde{K}_t(t-x) d(x^{-\bar{\gamma}-1} W(x)) \\ &= - \int_0^{t+1} \Psi_t(x) d(x^{-\bar{\gamma}-1} W(x)) \end{aligned}$$

where the last equality follows from the fact that $\bar{\gamma} \leq 0$, $W(0) = 0$ a.s., $\tilde{K}_t(0)$ is bounded, $\tilde{K}_t(t+1) = 0$ and $W(t+1) = O_P(1)$.

REMARK 4.3

In case of positive index, the asymptotic covariance function can be rewritten as

$$\text{cov}(Z(t), Z(v)) = \int_0^{(t \wedge v)+1} \frac{\Psi_t(x)}{x} d\Psi_v(x) + \int_0^{(t \wedge v)+1} \frac{\Psi_v(x)}{x} d\Psi_t(x)$$

Moreover, applying partial integration, this is equivalent to

$$\text{cov}(Z(t), Z(v)) = \int_0^{(t \wedge v)+1} \frac{\Psi_t(x) \Psi_v(x)}{x^2} dx$$

REMARK 4.4

In case $\gamma = -1$, the limiting covariance structure can be written as

$$\text{cov}(Z(t), Z(v)) = \int_0^{(t \wedge v)+1} \Psi_t(x) \Psi_v(x) dx$$

REMARK 4.5

In case of positive index, the conditions on the sequence h_n imply that $nh_n \rightarrow \infty$ as $n \rightarrow \infty$, since $H \in \text{RV}_{-\rho}^\infty$ for some $\rho \in (0, 1)$.

In case of negative index, the conditions on the sequence h_n also imply that $nh_n \rightarrow \infty$ as $n \rightarrow \infty$, since $\phi(h) \rightarrow 0$ as $h \rightarrow 0$.

Theorem 4.1 will be the starting point for proving the asymptotic normality of the general kernel type estimator of the extreme value index. The presented results of the asymptotic properties can be applied for any fixed $t \in [0, 1]$ in (4.8) in case $\gamma > 0$, whereas in case $\gamma < 0$ the chosen $t \in [0, 1]$ should satisfy a certain condition.

Theorem 4.2 (positive index)

Let ϕ , F , K and h be as in Theorem 4.1 for some $\gamma > 0$. Additionally, let t_0 be any fixed value in the interval $[0, 1]$ and let $\hat{\gamma}_{n,h}^G$ be given by (4.8) with $t = t_0$. Then

$$\frac{\sqrt{nh}}{\phi(h)} \left(\hat{\gamma}_{n,h}^G - B_{n,h}(t_0) \right) \xrightarrow{\mathcal{D}} Z_{t_0}^+$$

where $B_{n,h}(t_0) = \Phi_{ht_0}^{-1}(D_{n,h}(t_0))$ with $D_{n,h}(t_0)$ as in Theorem 4.1 and $Z_{t_0}^+$ a normally distributed random variable with mean 0 and variance

$$\int_0^{t_0+1} \int_0^{t_0+1} \frac{x \wedge y}{xy} d\Psi_{t_0}(y) d\Psi_{t_0}(x)$$

REMARK 4.6

Note that, using remark 4.3, the variance of the limiting normally distributed variable $Z_{t_0}^+$ can be rewritten as

$$2 \int_0^{t_0+1} \frac{\Psi_{t_0}(x)}{x} d\Psi_{t_0}(x) = \int_0^{t_0+1} \left(\frac{\Psi_{t_0}(x)}{x} \right)^2 dx$$

Theorem 4.3 (negative index)

Let ϕ , F , K and h be as in Theorem 4.1 for some $\gamma < 0$. Additionally, define

$$\kappa_\gamma(t) = \int_0^{t+1} u^{-\gamma-1} \Psi_t(u) du$$

Let t_0 be a value in the interval $[0, 1]$ that satisfies $L_{ot_0}^\gamma \kappa_\gamma(t_0) > 0$ and let $\hat{\gamma}_{n,h}^G$ be given by (4.8) with $t = t_0$. Then

$$\sqrt{nh} \log(ht_0) \kappa_\gamma(t_0) \left(\hat{\gamma}_{n,h}^G - B_{n,h}(t_0) \right) \xrightarrow{\mathcal{D}} Z_{t_0}^-$$

where $B_{n,h}(t_0) = \Phi_{ht_0}^{-1}(D_{n,h}(t_0))$ with $D_{n,h}(t_0)$ as in Theorem 4.1 and $Z_{t_0}^-$ a normally distributed random variable with mean 0 and variance

$$\int_0^{t_0+1} \int_0^{t_0+1} \frac{x \wedge y}{xy} (xy)^{-\gamma} d\Psi_{t_0}(y) d\Psi_{t_0}(x)$$

REMARK 4.7

Note that in case $\gamma = -1$, the constant $\kappa_\gamma(t)$ equals t for all $t \in (0, 1)$ by (4.5). Hence, the condition on the chosen t_0 is satisfied for all $t_0 \in (0, 1)$ by [PC2].

REMARK 4.8

As the proof of the theorem will show, $\log(ht_0) (B_{n,h}(t_0) - \gamma) \rightarrow \log(L_{ot_0}^\gamma \kappa_\gamma(t_0))$ as $n \rightarrow \infty$.

4.3 Proof of Theorem 4.1

We will first present an outline of the proof, before going too much into detail. The estimator of ϕ can be represented in the following way:

$$\hat{\phi}_h(ht) = \int_0^{t+1} \log Q_n(1-hu) d\Psi_t(u)$$

with $Q_n(\cdot)$ the empirical quantile function and $\Psi_t(\cdot)$ as defined by equation (4.4). This follows by rearranging the terms in the definition of $\hat{\phi}_h(\cdot)$ and noting that $Q_n(1-u)$ equals $X_{(n-k)}$ for $k/n \leq u < (k+1)/n$.

Since the behaviour of the empirical quantile function near the boundaries of the interval $[0, 1]$ differs from the behaviour on the interior of that interval, we will decompose the estimator into two parts:

$$\hat{\phi}_h(ht) = \int_{1/nh}^{t+1} \log Q_n(1-hu) d\Psi_t(u) + \int_0^{1/nh} \log Q_n(1-hu) d\Psi_t(u)$$

Assuming that the empirical quantile function will approximate the underlying quantile function, the first term is likely to be close to its deterministic equivalent. Therefore we decompose the estimator further into

$$\begin{aligned} \hat{\phi}_h(ht) &= \int_{1/nh}^{t+1} \log Q(1-hu) d\Psi_t(u) + \\ &+ \int_{1/nh}^{t+1} (\log Q_n(1-hu) - \log Q(1-hu)) d\Psi_t(u) \\ &+ \int_0^{1/nh} \log Q_n(1-hu) d\Psi_t(u) \\ &= D_{n,h}(t) + Y_{n,h}^{(1)}(t) + R_{n,h}^{(1)}(t) \end{aligned} \quad (4.9)$$

i.e., into a deterministic part $D_{n,h}(t)$ (as given in the theorem) and two random parts $Y_{n,h}^{(1)}(t)$ and $R_{n,h}^{(1)}(t)$. The term $R_{n,h}^{(1)}(t)$ turns out to be negligible with respect to the other terms, while the other random term will, properly scaled, lead to the limiting process mentioned in the theorem.

To obtain that limiting process, we use that $Q_n(1-x) \stackrel{D}{=} Q(\Gamma_n(1-x))$ where $\Gamma_n(\cdot)$ is the empirical quantile function of a $\mathcal{U}(0, 1)$ sample of size n , to rewrite $Y_{n,h}^{(1)}(t)$ into

$$\int_{1/nh}^{t+1} (\log Q(\Gamma_n(1-hu)) - \log Q(1-hu)) d\Psi_t(u)$$

By the symmetry of a uniform sample we have that $\Gamma_n(1-x) \stackrel{D}{=} 1 - \Gamma_n(x)$, hence $Y_{n,h}^{(1)}(t)$ is equal in distribution to

$$\int_{1/nh}^{t+1} (\log Q(1 - \Gamma_n(hu)) - \log Q(1-hu)) d\Psi_t(u)$$

Using a first order Taylor expansion on the function $u \mapsto \log Q(1-u)$ and noting that $\frac{d}{dx}(\log Q(1-x)) = -\phi(x)/x$, yields the further decomposition

$$\begin{aligned} -Y_{n,h}^{(1)}(t) &= \int_{1/nh}^{t+1} \phi(hu) \frac{\Delta_n(hu)}{hu} d\Psi_t(u) + R_{n,h}^{(2)}(t) \\ &= \int_{1/n}^{h(t+1)} \phi(u) \frac{\Delta_n(u)}{u} d\Psi_t\left(\frac{u}{h}\right) + R_{n,h}^{(2)}(t) \\ &= Y_{n,h}^{(2)}(t) + R_{n,h}^{(2)}(t) \end{aligned}$$

where

$$\Delta_n(u) = \Gamma_n(u) - u \tag{4.10}$$

and $R_{n,h}^{(2)}(t)$ is a remainder term.

It is well known that the scaled empirical quantile process of a uniform $(0, 1)$ sample, $\sqrt{n}\Delta_n(u)$, converges in distribution to a Brownian Bridge. Moreover, a sequence $B^{(n)}$ of Brownian Bridges can be constructed, a so called Hungarian Embedding, such that the weighted difference $(\sqrt{n}\Delta_n(u) - B^{(n)}(u))/u^{1/2-\nu}$ converges uniformly to 0 in probability, for any $\nu \in [0, 1/2)$. We can make use of that result by rewriting the scaled $Y_{n,h}^{(2)}(t)$ into

$$\begin{aligned} \frac{\sqrt{nh}}{\phi(h)} Y_{n,h}^{(2)}(t) &= \sqrt{h} \int_{1/n}^{h(t+1)} \frac{\phi(u) B^{(n)}(u)}{\phi(h) u} d\Psi_t\left(\frac{u}{h}\right) \\ &\quad + \sqrt{h} \int_{1/n}^{h(t+1)} \frac{\phi(u) \sqrt{n}\Delta_n(u) - B^{(n)}(u)}{\phi(h) u} d\Psi_t\left(\frac{u}{h}\right) \\ &= Y_{n,h}^{(3)}(t) + R_{n,h}^{(3)}(t) \end{aligned} \tag{4.11}$$

The mentioned result of the Hungarian Embedding indeed yields that $R_{n,h}^{(3)}(t)$ is of negligible order.

The term $Y_{n,h}^{(3)}(t)$ is completely determined by the behaviour of the sequence of Brownian Bridges. Since a Brownian Bridge $B(\cdot)$ is in distribution equal to $W(u) - uW(1)$ where $W(\cdot)$ is standard Brownian Motion, we can use $W(hu) \stackrel{D}{=} \sqrt{h}W(u)$ (Brownian scaling), to show that $Y_{n,h}^{(3)}(t)$ indeed converges to the limiting process mentioned in the theorem.

The actual proof of Theorem 4.1 is organized as follows: in section 4.3.1 the random terms will be considered. That section starts with stating some miscellaneous results that we will be needing when dealing with the random terms. In subsection 4.3.1.1, the first remainder term $R_{n,h}^{(1)}(t)$ will be shown to be of negligible order. Subsections 4.3.1.2

and 4.3.1.3 will deal with the other remainder terms $R_{n,h}^{(2)}(t)$ and $R_{n,h}^{(3)}(t)$ respectively. In subsection 4.3.1.4 the limiting process itself will be derived for the remaining random term $Y_{n,h}^{(3)}(t)$. Finally, in section 4.3.2 the deterministic term $D_{n,h}(t)$ will be considered.

4.3.1 The random terms

The following results are taken from WELLNER (1978).

Lemma 4.1

Let $\Gamma_n(u)$ be the empirical quantile function of a uniform $(0, 1)$ sample of size n and let $\Delta_n(\cdot)$ be given by (4.10). Then the following holds:

$$\sup_{\frac{1}{n} \leq u \leq 1} \left| \frac{\Gamma_n(u)}{u} \right| = O_P(1), \quad \sup_{\frac{1}{n} \leq u \leq 1} \left| \frac{u}{\Gamma_n(u)} \right| = O_P(1) \quad (4.12)$$

and

$$\sup_{b_n \leq u \leq 1} \left| \frac{\Delta_n(u)}{u} \right| = o_P(1) \quad (4.13)$$

where b_n is any sequence of positive numbers satisfying $nb_n \rightarrow \infty$ as $n \rightarrow \infty$.

Proof of Lemma 4.1:

From Lemma 2 in WELLNER (1978) we have for all $\lambda \geq 1$ that

$$\mathbb{P} \left(\sup_{\frac{1}{n} \leq u \leq 1} \left| \frac{\Gamma_n(u)}{u} \right| \geq \lambda \right) \leq e\lambda e^{-\lambda}$$

and

$$\mathbb{P} \left(\sup_{\frac{1}{n} \leq u \leq 1} \left| \frac{u}{\Gamma_n(u)} \right| \geq \lambda \right) \leq e\lambda^{-1}$$

Hence (4.12) follows.

The second assertion, (4.13), is Theorem 0 in WELLNER (1978). ■

Another result concerning the empirical quantile function $\Gamma_n(\cdot)$ is presented in the following lemma:

Lemma 4.2

Let $\Delta_n(\cdot)$ be given by (4.10). Then the following holds with probability 1:

$$u + t\Delta_n(u) \geq 0 \quad \text{for all } t \in [0, 1] \text{ and } u \in [0, 1] \quad (4.14)$$

and

$$\sup_{0 < u < 1} |\Delta_n(u)| \rightarrow 0 \quad (4.15)$$

Proof of Lemma 4.2:

The first statement, (4.14), follows simply by the definition of $\Delta_n(u)$.

For the second statement note that this is the quantile equivalent of the Glivenko-Cantelli result for the empirical distribution function. Indeed, since Γ_n is the inverse of the empirical distribution function F_n , we have that $|\Gamma_n(F_n(s)) - s| \leq 1/n$ and hence

$$\begin{aligned} \sup_{0 < u < 1} |\Delta_n(u)| &= \sup_{0 < s < 1} |\Gamma_n(F_n(s)) - F_n(s)| \\ &\leq \sup_{0 < s < 1} |F_n(s) - s| + \sup_{0 < s < 1} |\Gamma_n(F_n(s)) - s| \\ &= \sup_{0 < s < 1} |F_n(s) - s| + O(1/n) \end{aligned}$$

where the first term of the last line tends to 0 almost surely by Glivenko-Cantelli as $n \rightarrow \infty$. ■

The next result will be stated without proof and concerns the Hungarian Embedding as mentioned in the outline of the proof of Theorem 4.1 and is reformulated as:

In CSÖRGŐ, CSÖRGŐ, HORVÁTH AND MASON (1986) a probability space is constructed on which there exists a sequence U_1, U_2, \dots of i.i.d. $\mathcal{U}(0, 1)$ variables, and a sequence of Brownian Bridges $\{B^{(n)}(s)\}$, which has, among others, the following property (cf. Theorem 2.1 in CSÖRGŐ, CSÖRGŐ, HORVÁTH AND MASON (1986)):

Lemma 4.3

For any $0 \leq \nu < 1/2$

$$\sup_{1/n \leq u \leq 1 - 1/n} \frac{|\sqrt{n}\Delta_n(u) - B^{(n)}(u)|}{u^{\frac{1}{2} - \nu}} = O_P(n^{-\nu})$$

as $n \rightarrow \infty$, where $\Delta_n(u) = \Gamma_n(u) - u$ with $\Gamma_n(\cdot)$ the quantile function of U_1, \dots, U_n .

In view of this result, we will assume that any uniform $(0, 1)$ sample and sequence of Brownian Bridges we will be using, are defined on the above mentioned probability space.

Finally, the next lemma will be of use in dealing with the first remainder term $R_{n,h}^{(1)}(t)$.

Lemma 4.4

Let $\nu(\cdot)$ and $\mu(\cdot)$ be regularly varying at infinity with index $\lambda_\nu < 0$ and $\lambda_\mu > 0$ respectively. Moreover, let $\mu(\cdot)$ be non-decreasing. Then the following holds:

$$\lim_{x \rightarrow \infty} x\nu(a^{\mu(x)}) = 0 \quad \forall a > 1$$

Proof of Lemma 4.4:

Recall the following properties of regularly varying functions (see e.g., Chapter 1 of BINGHAM, GOLDIE AND TEUGELS (1987)):

(i) Suppose $f \in \text{RV}_\rho^\infty$. Then

$$f(x) \rightarrow \begin{cases} 0 & \rho < 0 \\ \infty & \rho > 0 \end{cases}$$

as $x \rightarrow \infty$.

(ii) Suppose f is non-decreasing and $f \in \text{RV}_\rho^\infty$ with $0 < \rho < \infty$. Then $f^{-1} \in \text{RV}_{1/\rho}^\infty$, where f^{-1} denotes the inverse of f .

(iii) Suppose $f_1 \in \text{RV}_{\rho_1}^\infty$, $f_2 \in \text{RV}_{\rho_2}^\infty$ with $f_2(t) \rightarrow \infty$ as $t \rightarrow \infty$. Then $f_1 \circ f_2 \in \text{RV}_{\rho_1 \rho_2}^\infty$.

Obviously,

$$\begin{aligned} \lim_{x \rightarrow \infty} x v(a^{\mu(x)}) &= \lim_{y \rightarrow \infty} \mu^{-1}(y) v(a^y) \\ &= \lim_{x \rightarrow \infty} \mu^{-1} \left(\frac{\log x}{\log a} \right) v(x) \end{aligned}$$

where in the first equality we substituted $y = \mu(x)$ and in the last equality we substituted $x = a^y$ for $a > 1$. Noting that $\log(\cdot) \in \text{RV}_0^\infty$ we get

$$\begin{aligned} \mu(\cdot) \in \text{RV}_{\lambda_\mu}^\infty &\implies \mu^{-1}(\cdot) \in \text{RV}_{1/\lambda_\mu}^\infty \\ &\implies \mu^{-1} \left(\frac{\log(\cdot)}{\log(a)} \right) \in \text{RV}_0^\infty \\ &\implies \mu^{-1} \left(\frac{\log(\cdot)}{\log(a)} \right) v(\cdot) \in \text{RV}_{\lambda_v}^\infty \\ &\implies \mu^{-1} \left(\frac{\log(x)}{\log(a)} \right) v(x) \rightarrow 0 \qquad \text{as } x \rightarrow \infty \end{aligned} \tag{4.16}$$

where the first implication follows from property (ii), the second from property (iii) and the last from property (i). Equation (4.16) then yields the assertion. ■

4.3.1.1 The random term $R_{n,h}^{(1)}(t)$

In this subsection we will prove that the remainder term $R_{n,h}^{(1)}(t)$ is of negligible order, i.e., we will prove the following proposition.

Proposition 4.1

Let $R_{n,h}^{(1)}(t)$ be defined by (4.9). Then, as $n \rightarrow \infty$

$$\frac{\sqrt{nh}}{\phi(h)} R_{n,h}^{(1)}(t) = o_P(1)$$

uniform in $t \in (0, 1)$.

Proof of Proposition 4.1:

Note that, since $Q_n(1-u) = X_{(n)}$ for $0 \leq u < 1/n$,

$$R_{n,h}^{(1)}(t) = \log X_{(n)} \int_0^{1/nh} d\Psi_t(u) = \log X_{(n)} \Psi_t\left(\frac{1}{nh}\right)$$

using that $\Psi_t(0) = 0$ by assumption. Moreover, using the boundedness assumption on the kernel $K(\cdot)$ that defined the function $\Psi_t(\cdot)$ (i.e., condition [KC2]), we obtain

$$R_{n,h}^{(1)}(t) = O_P\left(\frac{\log X_{(n)}}{nh}\right)$$

uniform in $t \in (0, 1)$.

In case $\gamma > 0$

Fix $\varepsilon > 0$ and note that

$$\begin{aligned} 0 \leq \mathbb{P}\left(\log X_{(n)} \geq \varepsilon \sqrt{nh_n}\right) &= 1 - F^n\left(\exp(\varepsilon \sqrt{nh_n})\right) \\ &\leq n\left(1 - F\left(\exp(\varepsilon \sqrt{nh_n})\right)\right) \end{aligned}$$

where the latter inequality follows from the fact that $1 - x^n \leq n(1 - x)$ for all $x \in [0, 1]$ and $n \geq 1$.

Moreover, since $F \in \mathcal{D}(G_\gamma)$ for $\gamma > 0$ is equivalent to $1 - F(\cdot) \in \text{RV}_{-1/\gamma}^\infty$ we can invoke Lemma 4.4: take $a = \exp(\varepsilon)$, $\nu(x) = 1 - F(x)$, $\lambda_\nu = -1/\gamma$, and $\mu(x) = \sqrt{xH(x)}$ with $H(\cdot)$ as defined in Theorem 4.1, i.e., $\lambda_\mu = (1 + \rho)/2$. In the result of Lemma 4.4 replace x by n and define $h = h_n = H(n)$ to obtain that $\log X_{(n)} = o_P(\sqrt{nh})$ as n tends to infinity, i.e., $R_{n,h}^{(1)}(t) = o_P(1/\sqrt{nh})$, uniform in $t \in (0, 1)$, as n tends to infinity. Moreover, since $\phi(h) \rightarrow \gamma$ as $h \rightarrow 0$ we have $\sqrt{nh}R_{n,h}^{(1)}(t)/\phi(h) = o_P(1)$ uniform in $t \in (0, 1)$, as n tends to infinity.

In case $\gamma < 0$

The underlying distribution function now has a finite upper endpoint, i.e., since x_F° is assumed to be positive, $\log X_{(n)} = O_P(1)$ as n tends to infinity. That is, $R_{n,h}^{(1)}(t) = O_P(1/(nh))$, uniform in $t \in (0, 1)$ as n tends to infinity. Moreover, by the conditions on h as given in the theorem, $\sqrt{nh}R_{n,h}^{(1)}(t)/\phi(h) = o_P(1)$ uniform in $t \in (0, 1)$, as n tends to infinity. ■

4.3.1.2 The random term $R_{n,h}^{(2)}(t)$

The term $R_{n,h}^{(2)}(t)$ was the remainder term resulting from an application of a first order Taylor expansion. Using the Lagrange representation of the remainder term, i.e.,

$$f(x) - f(x_0) = (x - x_0)f'(x_0) + (x - x_0) \int_0^1 (f'(x_0 + \zeta(x - x_0)) - f'(x_0)) d\zeta$$

we get, with $\Delta_n(u) = \Gamma_n(u) - u$,

$$\begin{aligned} -(\log Q(1 - \Gamma_n(u)) - \log Q(1 - u)) &= \\ &= \frac{\phi(u)}{u} \Delta_n(u) + \Delta_n(u) \int_0^1 \left(\frac{\phi(u + \zeta \Delta_n(u))}{u + \zeta \Delta_n(u)} - \frac{\phi(u)}{u} \right) d\zeta \\ &= \frac{\phi(u)}{u} \Delta_n(u) + \frac{\phi(u)}{u} \Delta_n(u) \int_0^1 \left(\frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} \frac{u}{u + \zeta \Delta_n(u)} - 1 \right) d\zeta \end{aligned}$$

Integrating the last term over u yields $R_{n,h}^{(2)}(t)$. We hence obtain

$$\frac{\sqrt{nh}}{\phi(h)} R_{n,h}^{(2)}(t) = \sqrt{nh} \int_{1/n}^{h(t+1)} \frac{\phi(u)}{\phi(h)} \frac{\Delta_n(u)}{u} R_n(u) d\Psi_t\left(\frac{u}{h}\right) \quad (4.17)$$

where

$$R_n(u) = \int_0^1 \left(\frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} \left(1 + \zeta \frac{\Delta_n(u)}{u} \right)^{-1} - 1 \right) d\zeta \quad (4.18)$$

Note that by (4.14) both $\phi(u + \zeta \Delta_n(u))$ and $(1 + \zeta \Delta_n(u)/u)^{-1}$ are well defined for all $\zeta \in [0, 1]$ and $u \in [1/n, 2h]$.

In the remainder of this subsection we will show that $R_{n,h}^{(2)}(t)$ is of negligible order, i.e., we will prove the following proposition.

Proposition 4.2

Let $R_{n,h}^{(2)}(t)$ be defined by (4.17). Then, as $n \rightarrow \infty$,

$$\frac{\sqrt{nh}}{\phi(h)} R_{n,h}^{(2)}(t) = o_P(1)$$

uniform in $t \in (0, 1)$.

In the proof of this proposition, the next two lemmas will be needed.

Lemma 4.5

Let $\phi(\cdot)$ be defined by (4.1), satisfying either condition [PC1] or condition [PC2] and let $\Delta_n(\cdot)$ be defined by (4.10). Then the following holds:

$$\frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} = O_P(1)$$

uniform in $\zeta \in [0, 1]$ and $u \in [1/n, 2h]$.

Proof of Lemma 4.5:

In case $\gamma > 0$, we have by (4.15) and condition [PC1] that

$$\frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} = \frac{1 + b(u + \zeta \Delta_n(u))}{1 + b(u)} = O_P(1)$$

uniform in $\zeta \in [0, 1]$ and $u \in [1/n, 2h]$, since $b(s) \rightarrow 0$ as $s \rightarrow 0$ uniform on some interval $(0, \varepsilon)$.

In case $\gamma < 0$, by condition [PC2] and results (4.13) and (4.15) we have

$$\frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} = \left(1 + \zeta \frac{\Delta_n(u)}{u}\right)^{-\gamma} \frac{L(u + \zeta \Delta_n(u))}{L(u)} = O_P(1)$$

uniform in $\zeta \in [0, 1]$ and $u \in [1/n, 2h]$, since $L(s) = L_o(1 + o(1))$ as $s \rightarrow 0$ uniform on some interval $(0, \varepsilon)$. ■

Lemma 4.6

Let $R_n(u)$ be defined by (4.18) and let b_n be a sequence of positive numbers satisfying $nb_n \rightarrow \infty$ as $n \rightarrow \infty$. Then the following holds:

$$\sup_{\frac{1}{n} \leq u \leq 2h} |R_n(u)| = O_P(1) \tag{4.19}$$

and

$$\sup_{b_n \leq u \leq 2h} |R_n(u)| = o_P(1) \tag{4.20}$$

Proof of Lemma 4.6:

Note that, by Lemma 4.5,

$$\begin{aligned} |R_n(u)| &\leq O_P(1) \int_0^1 \left(1 + \zeta \frac{\Delta_n(u)}{u}\right)^{-1} d\zeta + 1 \\ &= O_P(1) \frac{u}{\Delta_n(u)} \log \left(1 + \frac{\Delta_n(u)}{u}\right) + 1 \end{aligned}$$

with the $O_P(1)$ -term uniform in $u \in [1/n, 2h]$. Use the definition of $\Delta_n(\cdot)$ and the fact that $\log x$ has a monotone derivative to obtain that,

$$\begin{aligned} |R_n(u)| &\leq O_P(1) u \frac{\log \Gamma_n(u) - \log u}{\Gamma_n(u) - u} + 1 \\ &\leq O_P(1) \left(1 \vee \frac{u}{\Gamma_n(u)}\right) + 1 \end{aligned}$$

Applying (4.12) then yields (4.19).

For the second assertion, i.e., (4.20), rewrite $R_n(u)$ as

$$R_n(u) = \int_0^1 \left(\frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} - 1 \right) \left(1 + \zeta \frac{\Delta_n(u)}{u} \right)^{-1} d\zeta + \\ + \int_0^1 \left(\left(1 + \zeta \frac{\Delta_n(u)}{u} \right)^{-1} - 1 \right) d\zeta$$

By similar arguments as with the derivation of the first assertion,

$$\sup_{b_n \leq u \leq 2h} \int_0^1 \left(1 + \zeta \frac{\Delta_n(u)}{u} \right)^{-1} d\zeta = O_P(1)$$

Moreover, by the conditions [PC1] and [PC2] on ϕ and results (4.13) and (4.15),

$$\sup_{b_n \leq u \leq 2h} \left| \frac{\phi(u + \zeta \Delta_n(u))}{\phi(u)} - 1 \right| = o_P(1)$$

and finally, again by (4.13) and using that $\log(1+x) = x + O(x^2)$ as $x \rightarrow 0$,

$$\sup_{b_n \leq u \leq 2h} \left| \int_0^1 \left(1 + \zeta \frac{\Delta_n(u)}{u} \right)^{-1} d\zeta - 1 \right| = \\ = \sup_{b_n \leq u \leq 2h} \left| \frac{u}{\Delta_n(u)} \log \left(1 + \frac{\Delta_n(u)}{u} \right) - 1 \right| = o_P(1)$$

and the second assertion follows. ■

Now everything is set up for the proof of proposition 4.2.

Proof of Proposition 4.2:

In view of Lemma 4.6 rewrite equation (4.17) into

$$\frac{\sqrt{nh}}{\phi(h)} R_{n,h}^{(2)}(t) = \sqrt{nh} \int_{1/n}^{b_n} \frac{\phi(u) \Delta_n(u)}{\phi(h) u} R_n(u) d\Psi_t \left(\frac{u}{h} \right) + \\ + \sqrt{nh} \int_{b_n}^{h^{t+1}} \frac{\phi(u) \Delta_n(u)}{\phi(h) u} R_n(u) d\Psi_t \left(\frac{u}{h} \right) \\ = R_{n,h}^{(2.1)}(t) + R_{n,h}^{(2.2)}(t)$$

where b_n is a sequence of positive numbers satisfying $nb_n \rightarrow \infty$. Denoting the derivative of $\Psi_t(u)$ with respect to u by $\Psi'_t(u)$, the first term then satisfies

$$\left| R_{n,h}^{(2.1)}(t) \right| \leq \sqrt{nh} \sup_{1/n \leq u \leq 2h} |R_n(u)| \sup_{1/n \leq u \leq 2h} \left| \frac{\Delta_n(u)}{u} \right| \int_{1/nh}^{b_n/h} \frac{\phi(hu)}{\phi(h)} \left| \Psi'_t(u) \right| du \\ = \sqrt{nh} O_P(1) O_P(1) O(1) \frac{b_n - 1/n}{h}$$

where the last equality follows from (4.19), (4.12) and the boundedness of $|\Psi'_t(u)|$ and $\phi(hu)/\phi(h)$ on the interval $(1/nh, b_n/h)$, uniform in $t \in [0, 1]$. If, additionally to the assumption that $nb_n \rightarrow \infty$ as $n \rightarrow \infty$, we assume that $\sqrt{nh}(b_n - 1/n)/h \rightarrow 0$ as $n \rightarrow \infty$, e.g., by taking $b_n = h(nh)^{-1/2-\lambda}$ for some $0 < \lambda < 1/2$, we obtain that $|R_{n,h}^{(2.1)}(t)| = o_P(1)$.

For $R_{n,h}^{(2.2)}(t)$, we will make use of Lemma 4.3 and equation (4.20). Rewrite $R_{n,h}^{(2.2)}(t)$ into

$$\begin{aligned} R_{n,h}^{(2.2)}(t) &= \sqrt{h} \int_{b_n}^{h(t+1)} \frac{\phi(u)}{\phi(h)} \frac{\sqrt{n}\Delta_n(u) - B^{(n)}(u)}{u} R_n(u) d\Psi_t\left(\frac{u}{h}\right) + \\ &\quad + \sqrt{h} \int_{b_n}^{h(t+1)} \frac{\phi(u)}{\phi(h)} \frac{B^{(n)}(u)}{u} R_n(u) d\Psi_t\left(\frac{u}{h}\right) \\ &= R_{n,h}^{(2.3)}(t) + R_{n,h}^{(2.4)}(t) \end{aligned}$$

Note that by Lemma 4.3, for any $0 < \nu < 1/2$,

$$\sup_{b_n \leq u \leq 2h} \frac{|\sqrt{n}\Delta_n(u) - B^{(n)}(u)|}{u^{1/2-\nu}} = O_P(n^{-\nu})$$

and that, on the interval $(b_n/h, t+1)$, $|\Psi'_t(u)|$ and $\phi(hu)/\phi(h)$ are bounded and $|R_n(hu)| = O_P(1)$. So, for $R_{n,h}^{(2.3)}(t)$ we obtain that

$$\begin{aligned} |R_{n,h}^{(2.3)}(t)| &\leq O_P(n^{-\nu}) O_P(1) O(1) \sqrt{h} \int_{b_n/h}^{t+1} (hu)^{-\frac{1}{2}-\nu} du \\ &= O_P(n^{-\nu}) O_P(1) O\left(h^{\frac{1}{2}-\frac{1}{2}-\nu}\right) = O_P((nh)^{-\nu}) = o_P(1) \end{aligned}$$

Now consider $R_{n,h}^{(2.4)}(t)$:

$$\begin{aligned} |R_{n,h}^{(2.4)}(t)| &\leq \sup_{b_n \leq u \leq 2h} |R_n(u)| \sup_{b_n \leq u \leq 2h} \frac{\phi(u)}{\phi(h)} \sup_{b_n \leq u \leq 2h} |\Psi'_t(u)| \int_0^{2h} \frac{|B^{(n)}(u)|}{u} du \\ &= o_P(1) O(1) \int_0^{2h} \frac{|B^{(n)}(u)|}{u} du \end{aligned} \quad (4.21)$$

where in the last equality we used (4.20). Moreover, by the Markov inequality:

$$\mathbb{P}\left(\int_0^{2h} \frac{|B^{(n)}(u)|}{u} du \geq \varepsilon\right) \leq \frac{\mathbb{E} \int_0^{2h} \frac{|B^{(n)}(u)|}{u} du}{\varepsilon} \leq \frac{\int_0^{2h} \frac{\sqrt{\mathbb{E}(B^{(n)}(u)^2)}}{u} du}{\varepsilon}$$

$$\begin{aligned}
 &= \frac{\int_0^{2h} \frac{\sqrt{u(1-u)}}{u} du}{\varepsilon} \leq \frac{\int_0^{2h} u^{-1/2} du}{\varepsilon} \\
 &= \frac{O(h^{1/2})}{\varepsilon} \rightarrow 0
 \end{aligned}$$

hence, the last factor in (4.21) is $O_P(1)$ and consequently $|R_{n,h}^{(2.4)}(t)| = o_P(1)$ uniform in $t \in [0, 1]$. ■

4.3.1.3 The random term $R_{n,h}^{(3)}(t)$

Again we will show that this remainder term is of negligible order. Since $R_{n,h}^{(3)}(t)$ is already scaled properly by definition, we will prove the following proposition.

Proposition 4.3

Let $R_{n,h}^{(3)}(t)$ be defined by (4.11). Then, as $n \rightarrow \infty$,

$$R_{n,h}^{(3)}(t) = o_P(1)$$

uniform in $t \in (0, 1)$.

Proof of Proposition 4.3:

For this term we can use similar arguments as for the term $R_{n,h}^{(2.3)}(t)$ that appeared in the proof of Proposition 4.2: for any $0 < \nu < 1/2$ we have by Lemma 4.3 that

$$\begin{aligned}
 |R_{n,h}^{(3)}(t)| &\leq \sup_{\frac{1}{n} \leq u \leq 2h} \frac{|\sqrt{n}\Delta_n(u) - B^{(n)}(u)|}{u^{1/2-\nu}} \sqrt{h} \int_{1/nh}^{t+1} \frac{\phi(hu)}{\phi(h)} (hu)^{-\frac{1}{2}-\nu} |\Psi'_t(u)| du \\
 &= O_P(n^{-\nu}) \sqrt{h} \int_{1/nh}^{t+1} \frac{\phi(hu)}{\phi(h)} (hu)^{-\frac{1}{2}-\nu} |\Psi'_t(u)| du \tag{4.22}
 \end{aligned}$$

Again using that both $\Psi'_t(u)$ and $\phi(hu)/\phi(h)$ are bounded on $(1/nh, t+1)$ uniform in $t \in [0, 1]$, we get that (4.22) equals

$$O_P(n^{-\nu}) O\left(h^{\frac{1}{2}-\frac{1}{2}-\nu}\right) = O_P((nh)^{-\nu}) = o_P(1)$$

■

4.3.1.4 The random term $Y_{n,h}^{(3)}(t)$

Finally, the random term $Y_{n,h}^{(3)}(t)$ should lead to the asserted limiting process. Indeed, since $B^{(n)}(u) \stackrel{D}{=} (W(u) - uW(1))$ with $W(\cdot)$ standard Brownian Motion, the behaviour of $Y_{n,h}^{(3)}(t)$ is completely determined by the behaviour of Brownian Motion.

Using that $W(hu) \stackrel{\mathcal{D}}{=} \sqrt{h} W(u)$, rewrite $Y_{n,h}^{(3)}(t)$ into

$$\begin{aligned}
 Y_{n,h}^{(3)}(t) &\stackrel{\mathcal{D}}{=} \sqrt{h} \int_{1/n}^{h(t+1)} \frac{\phi(u)}{\phi(h)} \left(\frac{W(u)}{u} - W(1) \right) d\Psi_t \left(\frac{u}{h} \right) \\
 &\stackrel{\mathcal{D}}{=} \int_{1/nh}^{t+1} \frac{\phi(hu)}{\phi(h)} \frac{W(u)}{u} d\Psi_t(u) - \sqrt{h} W(1) \int_{1/nh}^{t+1} \frac{\phi(hu)}{\phi(h)} d\Psi_t(u)
 \end{aligned}$$

It is easily seen that the last term in the previous equation is $O_P(\sqrt{h}) = o_P(1)$, by the boundedness of the integral and the fact that $W(1) \sim \mathcal{N}(0, 1)$. Since $\phi(hu)/\phi(h) \sim u^{-\bar{\gamma}}$ with $\bar{\gamma} = \gamma \wedge 0$ by the assumptions [PC1] and [PC2] on $\phi(\cdot)$, the other term is first written as

$$\begin{aligned}
 \int_0^{t+1} u^{-\bar{\gamma}} \frac{W(u)}{u} d\Psi_t(u) + \int_0^{t+1} \left(\frac{\phi(hu)}{\phi(h)} - u^{-\bar{\gamma}} \right) \frac{W(u)}{u} d\Psi_t(u) + \\
 - \int_0^{1/nh} \frac{\phi(hu)}{\phi(h)} \frac{W(u)}{u} d\Psi_t(u)
 \end{aligned} \tag{4.23}$$

Note that by the boundedness of $\phi(hu)/\phi(h)$ and the fact that $\mathbb{E}(W(u)^2) = u$, the Markov inequality yields, for any $\varepsilon > 0$,

$$\begin{aligned}
 \mathbb{P} \left(\left| \int_0^{1/nh} \frac{\phi(hu)}{\phi(h)} \frac{W(u)}{u} d\Psi_t(u) \right| \geq \varepsilon \right) &\leq \frac{\mathbb{E} \int_0^{1/nh} \frac{\phi(hu)}{\phi(h)} \frac{|W(u)|}{u} |\Psi'_t(u)| du}{\varepsilon} \\
 &\leq \frac{\int_0^{1/nh} \frac{\phi(hu)}{\phi(h)} \frac{\sqrt{\mathbb{E}(W(u)^2)}}{u} |\Psi'_t(u)| du}{\varepsilon} \\
 &= \frac{O(1) \int_0^{1/nh} u^{-1/2} du}{\varepsilon} \\
 &= \frac{O((nh)^{-1/2})}{\varepsilon} = o(1)
 \end{aligned}$$

Hence the last term in (4.23) is $o_P(1)$. The second term is dealt with similarly, using that the assumptions [PC1] and [PC2] on $\phi(\cdot)$ imply that

$$\sup_{0 < u < 2} \left| \frac{\phi(hu)}{\phi(h)} - u^{-\bar{\gamma}} \right| = o(1)$$

The first term of (4.23), denoted by $Z(t)$, is a Gaussian process with expectation zero and covariance structure satisfying

$$\text{cov}(Z(t), Z(v)) = \mathbb{E} \int_0^{t+1} \int_0^{v+1} x^{-\bar{\gamma}} y^{-\bar{\gamma}} \frac{W(x)W(y)}{xy} d\Psi_v(x) d\Psi_t(y)$$

$$= \int_0^{t+1} \int_0^{v+1} \frac{x \wedge y}{xy} (xy)^{-\tilde{\gamma}} d\Psi_v(x) d\Psi_t(y)$$

4.3.2 The deterministic part

In this section we will consider the deterministic part $D_{n,h}(t)$ as given in the decomposition of $\hat{\phi}_h(ht)$:

$$\begin{aligned} D_{n,h}(t) &= \int_{1/nh}^{t+1} \log Q(1-hu) d\Psi_t(u) \\ &= [\Psi_t(u) \log Q(1-hu)]_{u=1/nh}^{t+1} + \int_{1/nh}^{t+1} \phi(hu) \frac{\Psi_t(u)}{u} du \\ &= -\Psi_t\left(\frac{1}{nh}\right) \log Q\left(1-\frac{1}{n}\right) + \int_{1/nh}^{t+1} \phi(hu) \frac{\Psi_t(u)}{u} du \\ &= D_{n,h}^{(1)}(t) + D_{n,h}^{(2)}(t) \end{aligned}$$

where the third equality follows from the condition that $\Psi_t(t+1) = 0$.

The first term, $D_{n,h}^{(1)}(t)$, is dealt with using the fact that $\Psi_t(1/nh) = O(1/nh)$ uniform in $t \in [0, 1]$ and

in case $\gamma > 0$

that the quantile function $Q(1-\cdot)$ is regularly varying at zero, hence $\log Q(1-\cdot)$ is slowly varying at zero and that $\phi(h) \rightarrow \gamma > 0$ as $h \downarrow 0$. I.e., since $h = H(n)$ with $H \in \text{RV}_{-\rho}^\infty$ for some $\rho \in (0, 1)$,

$$\frac{\sqrt{nh}}{\phi(h)} D_{n,h}^{(1)}(t) = O\left(\frac{\log Q(1-1/n)}{\gamma \sqrt{nH(n)}}\right) = O(D(n))$$

with $D \in \text{RV}_{(\rho-1)/2}^\infty$. Since $\rho - 1 < 0$, $D(n) \rightarrow 0$ as $n \rightarrow \infty$, see e.g., property (i) in Lemma 4.4 and the assertion follows.

In case $\gamma < 0$

the distribution function has a finite positive upper endpoint, i.e., $|\log Q(1)| < \infty$ and hence

$$\frac{\sqrt{nh}}{\phi(h)} D_{n,h}^{(1)}(t) = O\left(\frac{1}{\sqrt{nh}\phi(h)}\right) = o(1)$$

by the assumptions on h .

The second term, $D_{n,h}^{(2)}(t)$, should be close to the ϕ -function of the corresponding Generalized Pareto Function. Indeed, using

$$\left| \int_0^{1/nh} \frac{\Psi_t(u)}{u} du \right| \leq \frac{1}{nh} \sup_{0 \leq u \leq 1/nh} \left| \frac{\Psi_t(u)}{u} \right| = O\left(\frac{1}{nh}\right)$$

we obtain

$$\begin{aligned}
 D_{n,h}^{(2)}(t) &= \phi_{GPD}(ht; \gamma) + \int_0^{t+1} (\phi(hu) - \phi_{GPD}(ht; \gamma)) \frac{\Psi_t(u)}{u} du + \\
 &\quad - \int_0^{1/nh} \phi(hu) \frac{\Psi_t(u)}{u} du \\
 &= \phi_{GPD}(ht; \gamma) + \int_0^{t+1} (\phi(hu) - \phi_{GPD}(ht; \gamma)) \frac{\Psi_t(u)}{u} du + \\
 &\quad + \sup_{0 \leq u \leq 1/n} |\phi(u)| O\left(\frac{1}{nh}\right) \\
 &= \phi_{GPD}(ht; \gamma) + D_{n,h}^{(3)}(t) + D_{n,h}^{(4)}(t)
 \end{aligned}$$

Note that, in case $\gamma > 0$, $\phi(u) \rightarrow \gamma$ as $u \downarrow 0$ and in case $\gamma < 0$, $\phi(u) \rightarrow 0$ as $u \downarrow 0$. Hence, by similar arguments as for $D_{n,h}^{(1)}(\cdot)$, $\sqrt{nh}D_{n,h}^{(4)}(t)/\phi(h) = o(1)$ as $n \rightarrow \infty$. Thus,

$$D_{n,h}(t) - \phi_{GPD}(ht; \gamma) = \int_0^{t+1} (\phi(hu) - \phi_{GPD}(ht; \gamma)) \frac{\Psi_t(u)}{u} du + o\left(\frac{\phi(h)}{\sqrt{nh}}\right)$$

4.4 Proof of Theorem 4.2

We have that

$$\begin{aligned}
 \frac{\sqrt{nh}}{\phi(h)} (\hat{\gamma}_{n,h}^G - B_{n,h}(t)) &= \frac{\sqrt{nh}}{\phi(h)} (\Phi_{h^{-1}}^{-1}(\hat{\phi}_h(ht)) - \Phi_{h^{-1}}^{-1}(D_{n,h}(t))) \\
 &= Z_n(t)Y_n(t)
 \end{aligned}$$

where

$$\begin{aligned}
 Z_n(t) &= \frac{\sqrt{nh}}{\phi(h)} (\hat{\phi}_h(ht) - D_{n,h}(t)) \\
 Y_n(t) &= \int_0^1 (\Phi_{h^{-1}}^{-1})' (D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t))) du
 \end{aligned}$$

with $(\Phi_{h^{-1}}^{-1})'(u)$ the derivative of $\Phi_{h^{-1}}^{-1}(u)$ with respect to u .

From Theorem 4.1 we know that $Z_n(t) \xrightarrow{\mathcal{D}} Z(t)$ as $n \rightarrow \infty$ with $Z(\cdot)$ as defined in that theorem. For $Y_n(t)$ note that

$$(\Phi_h^{-1})'(u) = \frac{1}{\Phi_h'(\Phi_h^{-1}(u))}$$

with $\Phi'_h(y)$ the derivative of $\Phi_h(y)$ with respect to y , i.e.,

$$\Phi'_h(y) = \frac{1 - h^y + y h^y \log h}{(1 - h^y)^2}$$

Moreover, for any $0 < a \leq b < \infty$,

$$\sup_{a \leq y \leq b} |\Phi'_h(y) - 1| \rightarrow 0 \quad \text{as } h \downarrow 0$$

by the continuity of $\Phi'_h(\cdot)$ on $[a, b]$ and the fact that $\Phi'_h(y) \rightarrow 1$ for all $y \in [a, b]$ as $h \rightarrow 0$. Therefore, since $\Phi_h^{-1}(u) > 0 \Leftrightarrow u > -1/\log h$,

$$\sup_{a \leq u \leq b} |(\Phi_h^{-1})'(u) - 1| \rightarrow 0 \quad \text{as } h \downarrow 0 \quad (4.24)$$

for any $0 < a \leq b < \infty$.

For fixed $t \in (0, 1)$ define the event $A_n(t)$ as

$$A_n(t) = \left\{ \omega : \frac{1}{2}\gamma \leq D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t)) \leq \frac{3}{2}\gamma, \forall u \in [0, 1] \right\}$$

Theorem 4.1 yields that $D_{n,h}(t) \rightarrow \gamma > 0$ and that $\hat{\phi}_h(ht) - D_{n,h}(t) = o_P(1)$ as $n \rightarrow \infty$. Hence, $\mathbb{P}(A_n(t)) \rightarrow 1$ as $n \rightarrow \infty$. Moreover, on $A_n(t)$, $Y_n(t) \rightarrow 1$ by (4.24), thus $Y_n(t) \xrightarrow{P} 1$ and the assertion follows.

4.5 Proof of Theorem 4.3

Before giving the proof of Theorem 4.3, we will state and prove the following lemma:

Lemma 4.7

For $x < 0$ and $h \in (0, 1)$, let the function $\Phi_h(x)$ be defined by (4.7) with $y = h$. For fixed h , denote its inverse by $\Phi_h^{-1}(\cdot)$. Let $F_h(x)$ be a function that satisfies

$$F_h(x) = \Phi_h(x) + ch^{-x} + o(h^{-x}) \quad \text{as } h \downarrow 0 \quad (4.25)$$

for any $x < 0$ and some constant $c > x$. Then the following holds:

$$(\log h)(\Phi_h^{-1}(F_h(x)) - x) \rightarrow -\log(1 - c/x) \quad \text{as } h \downarrow 0 \quad (4.26)$$

Proof of Lemma 4.7:

First note that, for $x < 0$, $\Phi_h(x)$ can be rewritten into

$$\Phi_h(x) = \frac{-xh^{-x}}{1 - h^{-x}}$$

Equation (4.25) is then rewritten as,

$$\frac{-\bar{x}h^{-\bar{x}}}{1 - h^{-\bar{x}}} = \frac{-xh^{-x}}{1 - h^{-x}} + ch^{-x} + o(h^{-x}) \quad (4.27)$$

where $\bar{x} = \bar{x}_h = \Phi_h^{-1}(F_h(x))$. Multiplying (4.27) by $h^{\bar{x}}$ gives

$$\frac{-\bar{x}h^{x-\bar{x}}}{1-h^{-\bar{x}}} = \frac{-x+c+o(1)}{1-h^{-x}}$$

using that $ch^{-x} = o(1)$ as $h \rightarrow 0$. Taking logarithms on both sides then yields

$$\log(-\bar{x}) + (x-\bar{x})\log h - \log(1-h^{-\bar{x}}) - \log(-x+c) + \log(1-h^{-x}) = o(1)$$

provided $\bar{x} < 0$ and $c-x > 0$. Note that $\bar{x} < 0$ follows immediately from (4.27), for if $\bar{x} > 0$ then the left hand side of that equation tends to \bar{x} whereas the right hand side tends to 0, i.e., a contradiction and if $\bar{x} \rightarrow 0$ then the left hand side is of order $1/\log h$ whereas the right hand side is of order h^{-x} , i.e., again a contradiction.

Rearranging terms and noting that $\log(1-h^{-x}) = o(1)$ we get

$$(x-\bar{x})\log h - \log(1-c/x) + \log(\bar{x}/x) - \log(1-h^{-\bar{x}}) = o(1) \quad (4.28)$$

Define $\varepsilon_h = (x-\bar{x})\log h - \log(1-c/x)$. Then

$$h^{-\bar{x}} = h^{-x} \left(1 - \frac{c}{x}\right) e^{\varepsilon_h}$$

and

$$\bar{x} = x \left(1 + \frac{\log(1-c/x) + \varepsilon_h}{-x\log h}\right)$$

Substituting this in (4.28) we obtain

$$\varepsilon_h + \log \left[1 + \frac{\log(1-c/x) + \varepsilon_h}{-x\log h}\right] - \log [1 - h^{-x}(1-c/x)e^{\varepsilon_h}] = o(1) \quad (4.29)$$

However, since in (4.29) the second and the third term are of smaller order than the first term, the left hand side can only tend to zero if ε_h tends to zero. This in turn implies (4.26). ■

Now we can proceed with the proof of Theorem 4.3. Write

$$\begin{aligned} \sqrt{nh}\log(ht) \left(\hat{\gamma}_{n,h}^G - B_{n,h}(t)\right) &= \sqrt{nh}\log(ht) \left(\Phi_{ht}^{-1}(\hat{\phi}_h(ht)) - \Phi_{ht}^{-1}(D_{n,h}(t))\right) \\ &= Z_n(t)Y_n(t) \end{aligned}$$

with

$$Z_n(t) = \frac{\sqrt{nh}}{\phi(h)} \left(\hat{\phi}_h(ht) - D_{n,h}(t)\right)$$

$$Y_n(t) = \phi(h)\log(ht) \int_0^1 (\Phi_{ht}^{-1})' (D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t))) du$$

where $(\Phi_{ht}^{-1})'(u)$ is the derivative of $\Phi_{ht}^{-1}(u)$ with respect to u .

Theorem 4.1 yields that $Z_n(t) \xrightarrow{D} Z(t)$ with $Z(\cdot)$ as defined in that theorem. For $Y_n(t)$ note that

$$(\Phi_h^{-1})'(u) = \frac{1}{\Phi_h'(\Phi_h^{-1}(u))}$$

with $\Phi_h'(x)$ the derivative of $\Phi_h(x)$ with respect to x , i.e., for $x < 0$, as $h \rightarrow 0$

$$\Phi_h'(x) = \frac{xh^{-x} \log h - h^{-x} + h^{-2x}}{(1-h^{-x})^2} = O(xh^{-x} \log h)$$

Hence, for any $x < 0$ and $0 \leq \delta < -x$, as $h \rightarrow 0$

$$\frac{xh^{-x} \log h}{\Phi_h'(x+\delta)} = \frac{x}{x+\delta} h^\delta (1+o(1)) \quad (4.30)$$

From Theorem 4.1 we know that

$$D_{n,h}(t) = \Phi_{ht}(\gamma) + \int_0^{t+1} (\phi(hu) - \Phi_{ht}(\gamma)) \frac{\Psi_t(u)}{u} du + o\left(\frac{\phi(h)}{\sqrt{nh}}\right)$$

Substituting $\phi(s) = -\gamma s^{-\gamma} L_o(1+o(1))$ uniform on some interval $(0, \varepsilon)$, we obtain for fixed $t \in (0, 1)$

$$\begin{aligned} D_{n,h}(t) - \Phi_{ht}(\gamma) &= \\ &= \left[-\gamma \int_0^{t+1} (hu)^{-\gamma} L_o(1+o(1)) \frac{\Psi_t(u)}{u} du - \frac{-\gamma(ht)^{-\gamma}}{1-(ht)^{-\gamma}} \right] + o\left(\frac{\phi(h)}{\sqrt{nh}}\right) \\ &= -\gamma \left[L_o t^\gamma \int_0^{t+1} u^{-\gamma-1} \Psi_t(u) du - 1 \right] (ht)^{-\gamma} + o((ht)^{-\gamma}) \\ &= -\gamma [L_o t^\gamma \kappa_\gamma(t) - 1] (ht)^{-\gamma} + o((ht)^{-\gamma}) \\ &= c(ht)^{-\gamma} + o((ht)^{-\gamma}) \end{aligned}$$

with $\kappa_\gamma(t) = \int_0^{t+1} u^{-\gamma-1} \Psi_t(u) du$ and $c = \gamma [1 - L_o t^\gamma \kappa_\gamma(t)]$.

Assuming that we can fix $t \in (0, 1)$ such that $c > \gamma$, Lemma 4.7 then yields that

$$\log(ht) (\Phi_{ht}^{-1}(D_{n,h}(t)) - \gamma) \rightarrow -\log\left(1 - \frac{c}{\gamma}\right)$$

as $n \rightarrow \infty$. Moreover, since from Theorem 4.1 we know that, for fixed t ,

$$\hat{\phi}_h(ht) - D_{n,h}(t) = O_P\left(\frac{\phi(h)}{\sqrt{nh}}\right) = o_P((ht)^{-\gamma})$$

so

$$D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t)) = \Phi_{ht}(\gamma) + c(ht)^{-\gamma} + o_P((ht)^{-\gamma})$$

uniform in $u \in [0, 1]$.

Hence, substituting $x = \gamma$ and $\delta = D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t)) - \gamma$ in (4.30) and noting that, when $\delta \rightarrow 0$, $x/(x + \delta) \rightarrow 1$,

$$\begin{aligned} & \frac{\gamma(ht)^{-\gamma} \log(ht)}{\Phi'_{ht}(D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t)))} \\ &= (ht)^{D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t)) - \gamma} (1 + o_P(1)) \end{aligned}$$

and, uniform in $u \in [0, 1]$, this converges by Lemma 4.7 in probability to

$$\exp\left(-\log\left(1 - \frac{c}{\gamma}\right)\right) = \left(1 - \frac{c}{\gamma}\right)^{-1} = (L_o t^\gamma \kappa_\gamma(t))^{-1}$$

Hence, again using that $\phi(s) = -\gamma s^{-\gamma} L_o(1 + o(1))$, we get for $Y_n(t)$ the expression

$$\begin{aligned} & -\gamma h^{-\gamma} \log(ht) L_o(1 + o(1)) \int_0^1 (\Phi_{ht}^{-1})'(D_{n,h}(t) + u(\hat{\phi}_h(ht) - D_{n,h}(t))) du \\ & \xrightarrow{P} -L_o t^\gamma \left(1 - \frac{c}{\gamma}\right)^{-1} = -\frac{1}{\kappa_\gamma(t)} \end{aligned}$$

Thus,

$$\sqrt{nh} \log(ht) \kappa_\gamma(t) \left(\hat{\gamma}_{n,h}^G - B_{n,h}(t)\right) \xrightarrow{\mathcal{D}} Z_t^-$$

as asserted.

4.6 Dependence on the smoothing parameters

All the estimators mentioned in this thesis use some kind of a smoothing parameter that determines the number of order statistics that is to be used in the calculation of the estimator. It has been common practice to choose that value by plotting the estimator as a function of the number of used order statistics and taking any value in a region for which the estimator was more or less on a constant level. Recently, some attempts to find a data-driven procedure have been made. In HALL (1990) a procedure was developed for the Hill estimator, using some version of the bootstrap procedure. Hall proposed to reduce the size of the bootstrap samples as compared to the original sample. The amount by which it was to be reduced however, depended on (in practice unknown) parameters of the underlying distribution function.

Another attempt to find a data-adaptive procedure can be found in Chapter 3 of this thesis, for the kernel type estimator of Csörgő et al. (see also GRÜBEL AND DE WOLF

(1994)) This procedure does not use any unknown parameters, but is limited in its range of distribution functions it can handle.

The two papers DANIELSSON, DE HAAN, PENG AND DE VRIES (1997) and DE HAAN, PENG AND PEREIRA (1997) introduce yet another bootstrap-based method to determine the smoothing parameters in case of the Hill estimator, Pickands' estimator and the moment estimator. This method does not use any unknown parameters either. Even though results of a small simulation for large sample sizes (20000) are given in DANIELSSON, DE HAAN, PENG AND DE VRIES (1997), additional simulation studies are needed to show how this procedure performs for moderate sample sizes.

Before being able to develop any method to determine the required smoothing parameters in the case of the general kernel type estimator introduced in this chapter, equation (4.8), the dependence of this estimator on the two smoothing parameters should be investigated. Therefore we will perform a small simulation study that will illustrate that dependence for moderate sample sizes.

The definition of the estimator $\hat{\gamma}_{n,h}^G$ includes a kernel that is used to estimate the underlying ϕ -function. In this section we will use the triweight kernel defined by

$$K(x) = \begin{cases} \frac{35}{32}(1-x^2)^3 & \text{for } x \in [-1, 1] \\ 0 & \text{elsewhere} \end{cases}$$

Note that this kernel does indeed satisfy the conditions [KC1]–[KC4].

In our simulation we will use two instances of the generalized extreme value distribution $\text{GEV}(\cdot; \gamma)$. This distribution is defined for all $\gamma \in \mathbb{R}$ by:

$$\text{GEV}(x; \gamma) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right) \quad \text{for } 1 + \gamma x > 0$$

with the convention that $\text{GEV}(x; 0) = \exp(-\exp(-x))$ (see also (1.2) in Chapter 1 of this thesis). We will use this distribution with extreme value indices $+0.2$ and -0.1 . For both instances of this distribution, samples of size 1000 and 10000 will be generated. To show the dependence on the two smoothing parameters h and t simultaneously, the estimator is first plotted as a function of t and h in figures 4.6.1 and 4.6.2.

To facilitate comparison of the behaviour of this new estimator and the behaviour of the moment estimator, both estimators are plotted as a function of the fraction of order statistics that is used in the calculation of the estimator, i.e., for the moment estimator as a function of k/n and for the general kernel type estimator as a function of $h(t+1)$. Moreover, to show the dependence of the latter estimator on the parameter t , this estimator is plotted for several values of t .

As can be seen from figures 4.6.1 and 4.6.2, there seems to be a region on which the estimator is more or less constant. Choosing the values for h and t in that region will yield a quite accurate estimate. Moreover, comparing the general kernel type estimator with the moment estimator, see figures 4.6.3 and 4.6.4, the major advantage of this new estimator over the moment estimator is its smooth behaviour: whereas a small change in the choice of k for the moment estimator can result in a major change of value of the estimate, a small change in the combination of h and t changes the value of the estimate

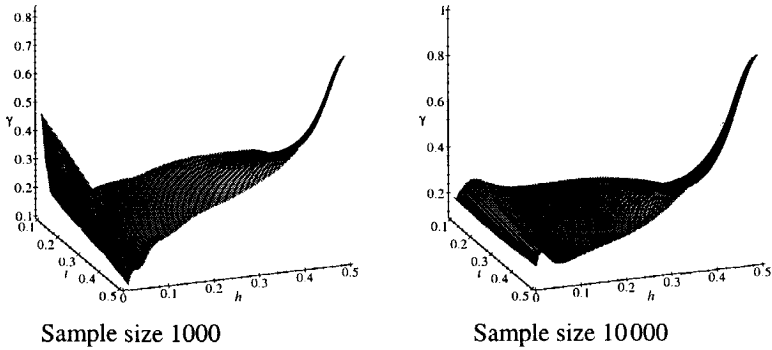


Figure 4.6.1: General kernel type estimator as function of h and t for $GEV(\cdot; 0.2)$

only slightly. This feature of the new estimator ensures that, if a method to approximate or estimate the optimal smoothing parameters produces values in the neighbourhood of the true optimal pair of parameters, the resulting estimate will virtually be the same. In case of the moment estimator, choosing a value of k close to the true optimal value could still result in a very different estimate.

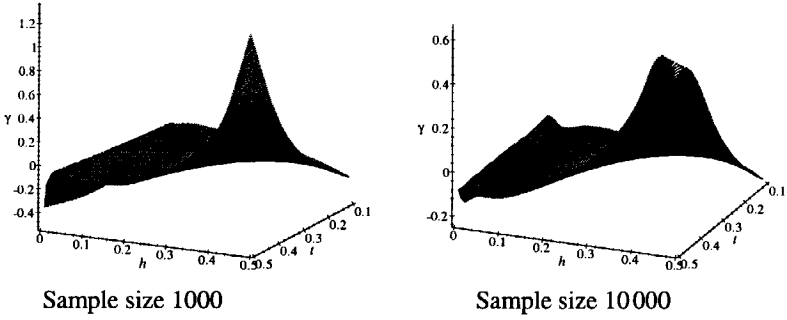


Figure 4.6.2: General kernel type estimator as function of h and t for $GEV(\cdot; -0.1)$

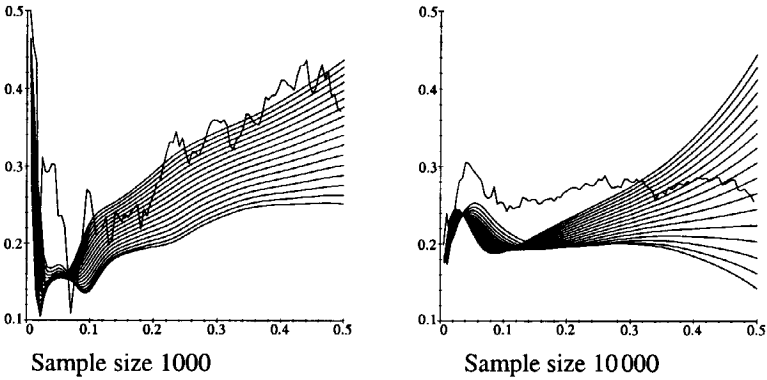


Figure 4.6.3: General kernel type estimator and moment estimator as a function of the fraction of used order statistics, for $GEV(\cdot; 0.2)$

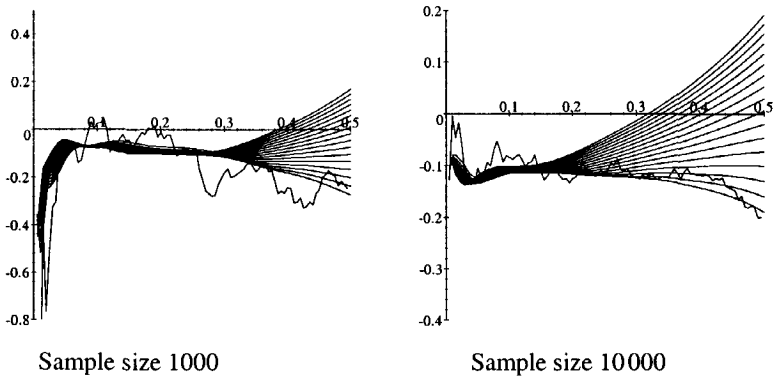


Figure 4.6.4: General kernel type estimator and moment estimator as a function of the fraction of used order statistics, for $GEV(\cdot; -0.1)$

Chapter 5

Alternative kernel type estimators

In this chapter an alternative attempt to extend the applicability of the kernel type estimator discussed in Chapter 3 to the more general situation of estimating any real-valued extreme value index is taken. A new estimator will be introduced and its asymptotic properties will be discussed. In Chapter 6 the finite sample properties will be considered.

5.1 Introduction

Throughout this chapter we will be concerned with a sample X_1, \dots, X_n of i.i.d. random variables with common distribution function F that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$. Moreover, we will denote the ascending order statistics of such a sample by $X_{(1)} \leq \dots \leq X_{(n)}$.

In Chapter 4 the function $\phi(\cdot)$ was introduced:

$$\phi(s) = -s \frac{d}{ds} \log Q(1-s) \quad (5.1)$$

with $Q(s) = F^{-1}(s)$ the quantile function of the variables X_i , assuming existence and differentiability of $\log Q(1-\cdot)$.

The link between the ϕ -function and the extreme value estimator that was used to derive the general kernel type estimator as introduced in that chapter, was based on the (limiting) parametric assumption that the tail of a distribution that is in the domain of attraction of an extreme value distribution can be approximated by the tail of a Generalized Pareto Distribution. This was represented by defining the new estimator as the value γ that solved the equation

$$\hat{\phi}_h(ht) = \frac{\gamma}{1 - (ht)^\gamma} \quad (5.2)$$

where $\hat{\phi}_h(ht)$ was a non-parametric estimate of the ϕ -function corresponding to the underlying distribution function and the right hand side of (5.2) the ϕ -function of a Generalized Pareto Distribution. Since for negative γ the ϕ -function of a Generalized

Pareto Distribution tends to 0 along with its argument, the function $\phi(\cdot)$ was estimated 'near' 0 instead of 'at' 0 and this introduced the extra smoothing-parameter t .

The alternative approach taken in the present chapter, will make use of the same function $\phi(\cdot)$. However, we will no longer use the explicit formula of that function in case of a Generalized Pareto Distribution.

In the following heuristic derivation of the new estimator, we will assume differentiability of any function whenever it is needed. In the final results, the precise conditions will be formulated.

5.2 Defining the estimator

One form of the well-known von Mises conditions, which are sufficient though not necessary for a distribution F to belong to the domain of attraction of an extreme value distribution, is given by

$$\lim_{t \uparrow x_F^0} \left(\frac{d}{dt} \frac{1 - F(t)}{F'(t)} \right) = \gamma$$

where x_F^0 is the upper endpoint¹ of $F \in \mathcal{D}(G_\gamma)$. Expanding the argument and replacing t by $Q(1-s)$, where $Q(\cdot)$ denotes the quantile function corresponding to F , the von Mises condition can be rewritten into

$$\lim_{s \downarrow 0} \left(-1 - \frac{sF''(Q(1-s))}{(F'(Q(1-s)))^2} \right) = \gamma$$

Moreover, noting that

$$\frac{\phi(s)}{s} = -\frac{d}{ds} \log Q(1-s)$$

the limit relation can be translated into

$$\gamma = \lim_{s \downarrow 0} \left(-1 + \phi(s) + \frac{s \frac{d^2}{ds^2} \log Q(1-s)}{\frac{d}{ds} \log Q(1-s)} \right) \quad (5.3)$$

The first non-constant term of the argument of this limit, $\phi(s)$, can be estimated using the kernel estimator already introduced in Chapter 3. The numerator and the denominator of the last term in (5.3) can be estimated separately, using kernel type estimators as well. In defining these estimators, we note that both numerator and denominator can be multiplied by any power of s , without changing the limit. Simulations showed that doing so will lead to more stable estimators. Moreover, as known from the literature on kernel density estimation, in estimating the derivative of a function, using the derivative

¹For a definition of an upper endpoint of a distribution, see Chapter 1 of this thesis.

of the kernel instead of the derivative of the empirical version of the estimable function, will often result in a more stable estimator as well. As a result, we will define the following estimator:

$$\hat{\gamma}_{n,h} = \hat{\gamma}_{n,h}^{(pos)} - 1 + \frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}} \quad (5.4)$$

where, with kernel $K: [0, 1] \rightarrow \mathbb{R}^+$ a fixed function with properties to be specified later, $K_h(u) = K(u/h)/h$, $Q_n(\cdot)$ the empirical quantile function² and $\alpha > 0$,

$$\begin{aligned} \hat{\gamma}_{n,h}^{(pos)} &= - \int_0^h u K_h(u) d \log Q_n(1-u) \\ &= \sum_{i=1}^{n-1} \frac{i}{n} K_h\left(\frac{i}{n}\right) (\log X_{(n-i+1)} - \log X_{(n-i)}) \\ \hat{q}_{n,h}^{(1)} &= - \int_0^h u^\alpha K_h(u) d \log Q_n(1-u) \\ &= \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^\alpha K_h\left(\frac{i}{n}\right) (\log X_{(n-i+1)} - \log X_{(n-i)}) \\ \hat{q}_{n,h}^{(2)} &= - \int_0^h \frac{d}{du} [u^{\alpha+1} K_h(u)] d \log Q_n(1-u) \\ &= \sum_{i=1}^{n-1} \frac{d}{du} [u^{\alpha+1} K_h(u)]_{u=i/n} (\log X_{(n-i+1)} - \log X_{(n-i)}) \end{aligned}$$

Note that, by rearranging terms and using that $Q_n(1-u)$ equals $X_{(n-k)}$ for $k/n \leq u < (k+1)/n$, we can also write

$$\hat{\gamma}_{n,h}^{(pos)} = \int_0^1 \log Q_n(1-hu) d(uK(u)) \quad (5.5)$$

$$\hat{q}_{n,h}^{(1)} = h^{\alpha-1} \int_0^1 \log Q_n(1-hu) d(u^\alpha K(u)) \quad (5.6)$$

$$\hat{q}_{n,h}^{(2)} = h^{\alpha-1} \int_0^1 \log Q_n(1-hu) d\left(\frac{d}{du} [u^{1+\alpha} K(u)]\right) \quad (5.7)$$

In the definition of our estimator, the continuous parameter h is used. This bandwidth determines the number of order statistics that is used in the calculation of the estimator. The continuous nature of the bandwidth ensures that the estimator is a smooth function of the fraction of used order statistics, as opposed to the more discrete nature of e.g., the moment estimator as defined in DEKKERS, EINMAHL AND DE HAAN (1989). However, although the behaviour of our estimator is thus rather smooth, the proofs of its asymptotic properties become a bit more complicated.

²The empirical quantile function is the generalized inverse of the empirical distribution function, i.e., $Q_n(u) = \inf \{x: F_n(x) \geq u\}$. Note that hence $Q_n(u) = X_{(k)}$ for $(k-1)/n < u \leq k/n$ with $k = 1, \dots, n$.

5.3 Main results

Let X_1, \dots, X_n denote a sample from a distribution function F that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$, denoted by $F \in \mathcal{D}(G_\gamma)$. This assumption is equivalent to the existence of $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}$, with $a_n > 0$ and $b_n \in \mathbb{R}$, such that

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\left(- (1 + \gamma x)^{-1/\gamma}\right)$$

for all x with $1 + \gamma x > 0$. We will make the convention that $G_0(x) = \exp(-e^{-x})$ for $x \in \mathbb{R}$. For more information on domains of attraction, we refer to Chapter 1 of this thesis.

We will derive asymptotic properties of the estimator of the extreme value index as defined in (5.4). In the definition of that estimator, the kernel K is used. In the theorems that will follow, that kernel will need to satisfy some or all of the following conditions:

Condition 5.1 (Kernel conditions)

- [CK1] $K(x) = 0$ whenever $x \notin [0, 1)$ and $K(x) \geq 0$ whenever $x \in [0, 1)$
- [CK2] $K(1) = K'(1) = 0$
- [CK3] $\int_0^1 K(x) dx = 1$
- [CK4] $K(\cdot), K'(\cdot)$ and $K''(\cdot)$ are bounded
- [CK5] $\int_0^1 u^{\alpha-1} K(u) du \neq 0$

The first asymptotic result concerns the (weak) consistency of the estimator under the domain of attraction condition. Note that the differentiability of the quantile function is not needed.

Theorem 5.1 (Consistency)

Assume that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in \mathbb{R}$. For arbitrary $\alpha > 0$, let K be a kernel satisfying conditions [CK1]–[CK5] and let $\hat{\gamma}_{n,h}$ be defined by (5.4).

If $h = h_n$ is such that $h \downarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$, then $\hat{\gamma}_{n,h} \rightarrow \gamma$ in probability as n tends to infinity.

Using additional conditions on the underlying distribution function, it is possible to derive asymptotic normality of our estimator, as stated in the next theorem. The additional conditions are similar to the conditions we used in Chapter 4 and we hence refer to that chapter for a discussion of these conditions.

Condition 5.2 (Additional conditions on F)

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$ and assume that $\phi(\cdot)$ as defined in (5.1) exists and is well defined. Moreover,

- [CP1] In case $\gamma \geq 0$, assume that $\phi(s) \rightarrow \gamma$ as $s \downarrow 0$.

[CP2] In case $\gamma < 0$, assume that, for some constant $c > 0$, $s^\gamma \phi(s) \rightarrow -c\gamma$ as $s \downarrow 0$.

[CP3] In case $\gamma = 0$, in addition to [CP1], assume that $\phi(hs)/\phi(s) \rightarrow 1$ as $s \downarrow 0$.

The condition that $\phi(\cdot)$ exists, appears to be rather strict. It is most likely that alternative conditions can be found that avoid the differentiability of the quantile function. However, assuming differentiability, the proof of asymptotic normality will be more intuitive and less complicated.

The asymptotic normality will be stated using the deterministic equivalent of the estimator, i.e.,

$$\gamma_h = \gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 \quad (5.8)$$

with (cf. equations (5.5)–(5.7))

$$\gamma_h^{(pos)} = \int_0^1 \log Q(1-hu) d(uK(u)) \quad (5.9)$$

$$q_h^{(i)} = h^{\alpha-1} \int_0^1 \log Q(1-hu) dK^{(i)}(u) \quad (5.10)$$

where $K^{(1)}(u) = u^\alpha K(u)$ and $K^{(2)}(u) = d(uK^{(1)}(u))/du$ for a kernel $K(\cdot)$.

Theorem 5.2 (Asymptotic normality)

Let X_1, \dots, X_n be a sample from F with F satisfying Condition 5.2. Moreover, for arbitrary $\alpha > 1/2$, let K be a kernel satisfying conditions [CK1]–[CK5] and let $\hat{\gamma}_{n,h}$ be defined as in (5.4). Then, for any $h = h_n$ with $h \downarrow 0$ and $(nh)^{-(\alpha \wedge 1)} \log n = O((nh)^{-1/2})$ as $n \rightarrow \infty$,

$$\sqrt{nh}(\hat{\gamma}_{n,h} - \gamma_h) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{\tilde{K}}^2)$$

where γ_h is defined in (5.8) and $\sigma_{\tilde{K}}^2 = \int_0^1 (a_0 \tilde{K}(u) + a_1 \tilde{K}^{(2)}(u) - a_2 \tilde{K}^{(1)}(u))^2 du$ with

$$\tilde{K}(u) = \int_u^1 x^{-1} d(xK(x)), \quad u \in (0, 1]$$

$$\tilde{K}^{(i)}(u) = \int_u^1 x^{-1-(\gamma \wedge 0)} dK^{(i)}(x), \quad u \in (0, 1]$$

and

$$a_0 = \gamma \vee 0$$

$$a_1 = 1 / \int_0^1 x^{-1-(\gamma \wedge 0)} K^{(1)}(x) dx$$

$$a_2 = (1 + (\gamma \wedge 0)) a_1$$

Note that a_i and $\tilde{K}^{(i)}(\cdot)$ for $i = 1, 2$ depend on both α and γ . Moreover, the functions $\tilde{K}^{(i)}(\cdot)$ may have a singularity at zero if $\alpha \leq 1$. However, they are square integrable for $\alpha > 1/2$ and all values of γ , i.e., σ_K^2 is well defined.

The formulation of this theorem on asymptotic normality implies that our estimator might have an asymptotic bias, defined as $\sqrt{nh}(\gamma_h - \gamma)$. In DEKKERS AND DE HAAN (1991) conditions are stated that cover all possible second order behaviour of quantile functions corresponding to distribution functions that are in the domain of attraction of an extreme value distribution.³ Under these additional conditions we can derive asymptotic expressions for the bias. The conditions mentioned in DEKKERS AND DE HAAN (1991) on the quantile function can be formulated in the following way:

Condition 5.3 (Second order regular variation)

In case $\gamma > 0$ there exist $\rho > 0$ and $c > 0$ such that

$$\pm (\log Q(1-s) + \gamma \log s - \log c) \in \text{RV}_{\gamma\rho}^0$$

In case $\gamma < 0$ there exist $\rho > 0$ and $c > 0$ such that (assume without loss of generality that $Q(1) > 0$)

$$\pm \left(s^\gamma (\log Q(1) - \log Q(1-s)) - \frac{c}{Q(1)} \right) \in \text{RV}_{-\gamma\rho}^0$$

Note that there is no second order regular variation condition in case of $\gamma = 0$. Moreover, the \pm signs are added because of the definition of regular variation. •

Condition 5.4 (Second order Π -variation)

In case $\gamma > 0$ there exists a positive function $b_1(\cdot)$ such that

$$\pm (\log Q(1-s) + \gamma \log s) \in \Pi^0 \left(\frac{b_1(s)}{s^\gamma Q(1-s)} \right)$$

In case $\gamma = 0$, there exist positive functions $b_2(\cdot)$ and $b_3(\cdot)$ with $b_2(s) \rightarrow 0$ as $s \downarrow 0$, such that

$$\lim_{s \downarrow 0} \frac{\log Q(1-sy) - \log Q(1-s) + b_2(s) \log y}{b_3(s)} = -\frac{(\log y)^2}{2}$$

In case $\gamma < 0$ there exists a positive function $b_4(s)$ such that (assuming without loss of generality that $Q(1) > 0$)

$$\pm s^\gamma (\log Q(1) - \log Q(1-s)) \in \Pi^0(b_4(s))$$

where $\Pi^0(f)$ stands for the class of Π -varying functions at zero with auxiliary function f . See Chapter 1 of this thesis for a definition of such a class.

Note that the \pm signs in case of $\gamma \neq 0$ are added to ensure the positiveness of the functions $b_1(\cdot)$ and $b_4(\cdot)$. •

Denoting $\int_0^1 x^\lambda K(x) dx$ by κ_λ , the results concerning the asymptotic bias can be formulated in the following way:

³It is most likely that these conditions can be used to replace the conditions of differentiability of the quantile function in Theorem 5.2.

Theorem 5.3 (Bias under Condition 5.3)

Let γ_n be given by (5.8). Assume that K satisfies conditions [CK1]–[CK5] for some $\alpha > 0$, that Q satisfies Condition 5.3 and that $h = h_n$ is such that $h_n \downarrow 0$ as n tends to infinity.

Then, as n tends to infinity, in case $\gamma > 0$,

$$\gamma_n - \gamma = \mu_1 \bar{a}_1(h) + o(\bar{a}_1(h))$$

where

$$\bar{a}_1(h) = \log Q(1-h) + \gamma \log h - \log c$$

$$\mu_1 = \gamma \rho \left(\rho \frac{\kappa_{\alpha+\gamma\rho-1}}{\kappa_{\alpha-1}} - \kappa_{\gamma\rho} \right)$$

and in case $\gamma < 0$

$$\gamma_n - \gamma = \mu_2 \bar{a}_2(h) + \mu_3 h^{-\gamma} + O(h^{-\gamma} \bar{a}_2(h)) + o(\bar{a}_2(h))$$

where

$$\bar{a}_2(h) = h^\gamma (\log Q(1) - \log Q(1-h)) - c/Q(1)$$

$$\mu_2 = \frac{\rho(\rho+1)Q(1)}{c} \frac{\kappa_{\alpha-\gamma(1+\rho)-1}}{\kappa_{\alpha-\gamma-1}}$$

$$\mu_3 = -\frac{\gamma c}{Q(1)\kappa_{-\gamma}}$$

Combining Theorem 5.3 and Theorem 5.2 then yields the following corollary:

Corollary 5.1

Assume the conditions of Theorem 5.2, additionally assume that Condition 5.3 is satisfied and that $h = h_n$ is such that, as n tends to infinity, $h \downarrow 0$ and

$$nh(\bar{a}_1(h))^2 = nh(\log Q(1-h) + \gamma \log h - \log c)^2 \rightarrow 0$$

in case $\gamma > 0$ or

$$nh(\bar{a}_2(h))^2 = nh^{1+2\gamma}(\log Q(1) - \log Q(1-h)) - c/Q(1))^2 \rightarrow 0$$

and $nh^{1-2\gamma} \rightarrow 0$ in case $\gamma < 0$. Then, as n tends to infinity,

$$\sqrt{nh}(\hat{\gamma}_{n,h} - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_K^2)$$

with σ_K^2 as defined in Theorem 5.2.

Under Condition 5.4, the situation is a little different and we will only state the conditions under which the asymptotic bias vanishes:

Theorem 5.4 (Bias under Condition 5.4)

Let γ_h be given by (5.8). Assume that K satisfies conditions [CK1]–[CK5] for some $\alpha > 0$ and that Q satisfies Condition 5.4.

If $h = h_n$ is such that, when $n \rightarrow \infty$, $h \downarrow 0$, $nh \rightarrow \infty$ and in case $\gamma > 0$

$$nh^{1-2\gamma} \left(\frac{b_1(h)}{Q(1-h)} \right)^2 \rightarrow 0$$

in case $\gamma = 0$

$$nh(b_2(h))^2 \rightarrow 0 \quad \text{and} \quad nh \left(\frac{b_3(h)}{b_2(h)} \right)^2 \rightarrow 0$$

and in case $\gamma < 0$

$$nh^{1-2\gamma} \left(\frac{b_4(h)}{\log Q(1) - \log Q(1-h)} \right)^2 \rightarrow 0$$

and

$$nh(\log Q(1) - \log Q(1-h))^2 \rightarrow 0$$

then $\sqrt{nh}(\gamma_h - \gamma) \rightarrow 0$ as n tends to infinity.

Note that these conditions on h resemble the corresponding ones on k in case of the moment estimator as defined in DEKKERS, EINMAHL AND DE HAAN (1989), see also Theorem 2.9 of this thesis.

5.4 Consistency

In this section we will prove the weak consistency of the newly introduced estimator under the single condition that the underlying distribution function is in the domain of attraction of an extreme value distribution. The proof is naturally divided in two: a positive part, proving that $\hat{\gamma}_{n,h}^{(pos)} \rightarrow (\gamma \vee 0)$ in probability, and a negative part, proving that $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \rightarrow 1 + (\gamma \wedge 0)$ in probability.

5.4.1 Preliminaries

In both parts of the proof of consistency we will make use of several lemmas that will be stated and proved in this subsection.

The first two lemmas are needed as an alternative to the differentiability of the quantile function.

Lemma 5.1

Suppose $F \in \mathcal{D}(G_\gamma)$ with $x_F^0 > 0$. Denote the corresponding quantile function by $Q(s) = F^{-1}(s)$. Then, for some positive function $a(\cdot)$,

$$\lim_{s \downarrow 0} \frac{\log Q(1-sy) - \log Q(1-s)}{a(s)/Q(1-s)} = \begin{cases} -\log y & \gamma \geq 0 \\ \frac{y^{-\gamma} - 1}{\gamma} & \gamma < 0 \end{cases}$$

for all $y > 0$. Moreover, for each $\varepsilon > 0$ there exists s_0 such that, for $0 < s \leq s_0$ and $0 < y \leq 1$,

$$(1-\varepsilon) \frac{1-y^\varepsilon}{\varepsilon} - \varepsilon < \frac{\log Q(1-sy) - \log Q(1-s)}{a(s)/Q(1-s)} < (1+\varepsilon) \frac{y^{-\varepsilon} - 1}{\varepsilon} + \varepsilon$$

provided $\gamma \geq 0$, and

$$1 - (1+\varepsilon)y^{-\gamma-\varepsilon} < \frac{\log Q(1-sy) - \log Q(1-s)}{\log Q(1) - \log Q(1-s)} < 1 - (1-\varepsilon)y^{-\gamma+\varepsilon}$$

provided $\gamma < 0$.

Proof of Lemma 5.1:

Rewrite Lemma 2.5 from DEKKERS, EINMAHL AND DE HAAN (1989), using that $Q(1-s) = U(1/s)$, where $U(\cdot)$ is the inverse of $1/(1-F)$. Essentially, the inequalities of the lemma are properties of regularly varying functions for $\gamma < 0$ and of Π -varying functions for $\gamma \geq 0$. ■

REMARK 5.1

In case $\gamma > 0$, we know that $Q(1-\cdot)$ is regularly varying at 0 with index $-\gamma$ (see e.g., Theorem 1.9 of this thesis), i.e.,

$$\lim_{s \downarrow 0} \frac{Q(1-sy)}{Q(1-s)} = y^{-\gamma}$$

Taking logarithms on both sides, we obtain

$$\lim_{s \downarrow 0} (\log Q(1-sy) - \log Q(1-s)) = -\gamma \log y$$

Hence, in case $\gamma > 0$, we can take $a(s)/Q(1-s) = \gamma$ in Lemma 5.1.

REMARK 5.2

In case $\gamma < 0$, we know that $Q(1) < \infty$ and that $Q(1) - Q(1-\cdot)$ is regularly varying at 0 with index $-\gamma$ (see e.g., Theorem 1.9 of this thesis), i.e.,

$$\lim_{s \downarrow 0} \frac{Q(1) - Q(1-sy)}{Q(1) - Q(1-s)} = y^{-\gamma}$$

For each $y > 0$, note that since $Q(1 - sy)/Q(1)$ tends to 1 as s tends to 0,

$$\begin{aligned} \lim_{s \downarrow 0} \frac{\log Q(1 - sy) - \log Q(1)}{\log Q(1 - s) - \log Q(1)} &= \lim_{s \downarrow 0} \frac{\left(\frac{Q(1 - sy)}{Q(1)} - 1\right)(1 + o(1))}{\left(\frac{Q(1 - s)}{Q(1)} - 1\right)(1 + o(1))} \\ &= \lim_{s \downarrow 0} \frac{Q(1 - sy) - Q(1)}{Q(1 - s) - Q(1)} \\ &= y^{-\gamma} \end{aligned}$$

Hence,

$$\begin{aligned} \lim_{s \downarrow 0} \frac{\log Q(1 - sy) - \log Q(1 - s)}{-\gamma(\log Q(1) - \log Q(1 - s))} &= \lim_{s \downarrow 0} \frac{\log Q(1) - \log Q(1 - sy)}{\gamma(\log Q(1) - \log Q(1 - s))} - \frac{1}{\gamma} \\ &= \frac{y^{-\gamma} - 1}{\gamma} \end{aligned}$$

i.e., we can take $a(s)/Q(1 - s) = -\gamma(\log Q(1) - \log Q(1 - s))$ in Lemma 5.1.

Lemma 5.2

Suppose $F \in \mathcal{D}(G_0)$ with $x_F^0 > 0$. Denote the corresponding quantile function by $Q(s) = F^{-1}(s)$. If for some positive function $a(\cdot)$,

$$\lim_{s \downarrow 0} \frac{\log Q(1 - sy) - \log Q(1 - s)}{a(s)/Q(1 - s)} = -\log y \quad (5.11)$$

for all $y > 0$, then $a(s) = o(Q(1 - s))$ as $s \downarrow 0$.

Proof of Lemma 5.2:

Several situations may occur which we will discuss separately. In case $\gamma = 0$, the upper endpoint of the underlying distribution function can be finite as well as infinite. In case of a finite upper endpoint, i.e., $Q(1) < \infty$, the assertion follows as a property of Π -varying functions (see e.g., GELUK AND DE HAAN (1987)). That property yields that equation (5.11) implies $a(s)/Q(1 - s) = o(\log Q(1 - s))$ as $s \downarrow 0$. This in turn implies that $a(s) = o(Q(1 - s))$ as well, since $Q(1) < \infty$.

Now consider the situation that the upper endpoint is infinite, i.e., $Q(1 - s) \rightarrow \infty$ as $s \downarrow 0$. If in that case $a(s) \rightarrow 0$ as $s \downarrow 0$ the assertion follows trivially. If $a(s) \not\rightarrow 0$, note that using $\log x = (x - 1)(1 + o(1))$ as $x \rightarrow 1$, equation (5.11) yields

$$\begin{aligned} -\log y &= \lim_{s \downarrow 0} \frac{\log Q(1 - sy) - \log Q(1 - s)}{a(s)/Q(1 - s)} \\ &= \lim_{s \downarrow 0} \frac{\log(Q(1 - sy)/Q(1 - s))}{a(s)/Q(1 - s)} \\ &= \lim_{s \downarrow 0} \frac{(Q(1 - sy)/Q(1 - s) - 1)(1 + o(1))}{a(s)/Q(1 - s)} \\ &= \lim_{s \downarrow 0} \frac{(Q(1 - sy) - Q(1 - s))(1 + o(1))}{a(s)} \end{aligned}$$

hence, $Q(1 - \cdot)$ is also Π -varying at 0 with auxiliary function $a(\cdot)$. The assertion then follows using the aforementioned property of Π -varying functions applied to $Q(1 - \cdot)$ instead of $\log Q(1 - \cdot)$. ■

If we want to make use of the bounds of the inequalities of Lemma 5.1 as arguments of an integral and take limits with respect to ε , we will need a majorant for these bounds. The following lemma provides that bound.

Lemma 5.3

For any $\delta > 0$ there exists $\varepsilon_0 > 0$ such that,

$$(1 + \varepsilon) \frac{x^{-\varepsilon} - 1}{\varepsilon} \leq \begin{cases} -x^{-\delta} \log x & 0 < x < \frac{1}{2} \\ -2^\delta \log x & \frac{1}{2} \leq x < 1 \end{cases}$$

for all $\varepsilon \in (0, \varepsilon_0)$.

Proof of Lemma 5.3:

First consider $x \in (0, 1/2)$. Define

$$g_\varepsilon(x) = (1 + \varepsilon) \frac{x^{-\varepsilon} - 1}{\varepsilon} + x^{-\delta} \log x$$

then we have to show that $g_\varepsilon(x) \leq 0$. Note that

$$g_\varepsilon(1/2) = (1 + \varepsilon) \frac{2^\varepsilon - 1}{\varepsilon} - 2^\delta \log 2 \rightarrow (1 - 2^\delta) \log 2$$

as $\varepsilon \downarrow 0$, hence $g_\varepsilon(1/2) < 0$ for ε small enough. Moreover,

$$\begin{aligned} \frac{d}{dx} g_\varepsilon(x) &= -(1 + \varepsilon)x^{-\varepsilon-1} + x^{-\delta-1} - \delta x^{-\delta-1} \log x \\ &= x^{-\varepsilon-1} \left[-(1 + \varepsilon) + x^{\varepsilon-\delta} (1 - \delta \log x) \right] \end{aligned}$$

Since for all $0 < \varepsilon < \delta$ and $x \in (0, 1/2)$ the function $x \mapsto x^{\varepsilon-\delta} (1 - \delta \log x)$ is decreasing, we get that

$$-(1 + \varepsilon) + x^{\varepsilon-\delta} (1 - \delta \log x) \geq -(1 + \varepsilon) + 2^{\delta-\varepsilon} (1 + \delta \log 2)$$

and as $\varepsilon \rightarrow 0$ this tends to $-1 + 2^\delta (1 + \delta \log 2) > 0$. Hence, the derivative of $g_\varepsilon(x)$ is positive for ε small enough, so $g_\varepsilon(x)$ is increasing on $(0, 1/2)$ and thus negative.

Next consider $x \in [1/2, 1)$. Define

$$\tilde{g}_\varepsilon(x) = (1 + \varepsilon) \frac{x^{-\varepsilon} - 1}{\varepsilon} + 2^\delta \log x$$

Then $\tilde{g}_\varepsilon(1) = 0$ and

$$\frac{d}{dx} \tilde{g}_\varepsilon(x) = -(1 + \varepsilon)x^{-\varepsilon-1} + \frac{2^\delta}{x} = x^{-\varepsilon-1} \left[-(1 + \varepsilon) + x^\varepsilon 2^\delta \right]$$

Since the function $x \mapsto x^\epsilon$ is increasing for all $\epsilon > 0$,

$$-(1 + \epsilon) + x^\epsilon 2^\delta \geq -(1 + \epsilon) + 2^{\delta - \epsilon}$$

and this tends to $2^\delta - 1 > 0$ as $\epsilon \rightarrow 0$. Hence, $\tilde{g}_\epsilon(x) \leq 0$ for $x \in [1/2, 1)$ and ϵ small enough. ■

The following lemma shows that a properly scaled vector of order statistics of a uniform $(0, 1)$ sample of size n is another vector of uniform order statistics of a sample with different sample size. This property is needed in the proof of Lemma 5.5.

Lemma 5.4

Let U_1, \dots, U_n be i.i.d. $\mathcal{U}(0, 1)$ variables. Then, for any integer $1 \leq k \leq n - 1$,

$$\left(\frac{U_{(1)}}{U_{(k+1)}}, \dots, \frac{U_{(k)}}{U_{(k+1)}} \right) \stackrel{D}{=} (V_{(1)}, \dots, V_{(k)})$$

where $V_{(1)}, \dots, V_{(k)}$ are the order statistics of k i.i.d. $\mathcal{U}(0, 1)$ variables.

Proof of Lemma 5.4:

Note that the density of $U = (U_{(1)}, \dots, U_{(k+1)})$ is given by

$$f_U(u_1, \dots, u_{k+1}) = \frac{n!}{(n - k - 1)!} (1 - u_{k+1})^{n - k - 1}$$

for $0 \leq u_1 \leq \dots \leq u_{k+1} \leq 1$.

Define $Z_i = U_{(i)} / U_{(k+1)}$ for $i = 1, \dots, k$ and $Z_{k+1} = U_{(k+1)}$. The inverse transformation is thus given by $U_{(i)} = Z_i Z_{k+1}$ for $i = 1, \dots, k$ and $U_{(k+1)} = Z_{k+1}$. The density of $Z = (Z_1, \dots, Z_{k+1})$ is then given by

$$f_Z(z_1, \dots, z_{k+1}) = f_U(z_1 z_{k+1}, \dots, z_k z_{k+1}, z_{k+1}) |J(z)|$$

for $0 \leq z_1 \leq \dots \leq z_k \leq 1$ and $0 \leq z_{k+1} \leq 1$ and where $J(z)$ is the Jacobian of the transformation, i.e., $|J(z)| = z_{k+1}^k$. Hence,

$$f_Z(z_1, \dots, z_{k+1}) = \frac{n!}{(n - k - 1)!} (1 - z_{k+1})^{n - k - 1} z_{k+1}^k$$

for $0 \leq z_1 \leq \dots \leq z_k \leq 1$ and $0 \leq z_{k+1} \leq 1$. Finally, integrating over z_{k+1} , we get that the density of (Z_1, \dots, Z_k) equals $k!$ for $0 \leq z_1 \leq \dots \leq z_k \leq 1$ and hence is the (joint) density of the order statistics of a $\mathcal{U}(0, 1)$ sample of size k . ■

In the definition of our estimator, the term $\log Q_n(1 - hu)$ appears frequently. Since this term is in distribution equal to $\log Q(1 - \Gamma_n(hu))$ where $\Gamma_n(\cdot)$ is the empirical quantile function of a uniform $(0, 1)$ sample of size n , we would like to use Lemma 5.1 with y replaced by a properly scaled version of $\Gamma_n(hu)$. The next lemma shows that the inequalities of Lemma 5.1 with y replaced by $\Gamma_n(hu) / U_{(\lfloor nh \rfloor + 1)}$ behave like the inequalities with y replaced by u , uniformly in u . The continuous nature of the bandwidth h now yields that we will have to act very carefully. However, when we have shown that $\Gamma_n(hu) / U_{(\lfloor nh \rfloor + 1)}$ is close to $\Gamma_{\lfloor nh \rfloor}(u)$ (the empirical quantile function of $\lfloor nh \rfloor$ i.i.d. $\mathcal{U}(0, 1)$ variables), the remaining part of the proof of Lemma 5.5 is rather straightforward.

Lemma 5.5

Let $\Gamma_n(\cdot)$ denote the empirical quantile function of U_1, \dots, U_n with U_i i.i.d. $\mathcal{U}(0, 1)$, h be a sequence of positive numbers with $h = h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as n tends to infinity and $L(\cdot)$ be an integrable, bounded and positive function on $(0, 1)$. Define $k = \lfloor nh \rfloor$ and $\bar{\lambda} = (\lambda \wedge 0)$ for $\lambda > -1$. Then, for each $\beta > (-1 - \bar{\lambda})$,

$$\int_0^1 \left[\left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^\beta - u^\beta \right] u^\lambda L(u) du \xrightarrow{P} 0 \quad (5.12)$$

as $n \rightarrow \infty$.

Proof of Lemma 5.5:

The case $\beta = 0$ is trivial, hence consider $\beta \neq 0$. First write equation (5.12) as

$$\begin{aligned} \int_0^1 \left[\left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^\beta - \left(\frac{\Gamma_n(ku/n)}{U_{(k+1)}} \right)^\beta \right] u^\lambda L(u) du + \\ + \int_0^1 \left[\left(\frac{\Gamma_n(ku/n)}{U_{(k+1)}} \right)^\beta - u^\beta \right] u^\lambda L(u) du \end{aligned} \quad (5.13)$$

For the first term, note that for $j = 1, \dots, k$, by definition

$$(\Gamma_n(hu))^\beta - (\Gamma_n(ku/n))^\beta = \begin{cases} 0 & \frac{j-1}{k} < u \leq \frac{j}{nh} \\ U_{(j+1)}^\beta - U_{(j)}^\beta & \frac{j}{nh} < u \leq \frac{j}{k} \end{cases}$$

hence, the first term equals

$$U_{(k+1)}^{-\beta} \sum_{j=1}^k \int_{j/nh}^{j/k} (U_{(j+1)}^\beta - U_{(j)}^\beta) u^\lambda L(u) du$$

I.e., using that $nh - k < 1$, that

$$|x^{\lambda+1} - y^{\lambda+1}| \leq (\lambda + 1)(x - y)(x^\lambda \vee y^\lambda) \leq (\lambda + 1)(x - y)y^{\bar{\lambda}}$$

for all $0 \leq y \leq x \leq 1$ and denoting the supremum of the function $L(\cdot)$ on $(0, 1)$ by $\|L\|$, we get

$$\begin{aligned} \left| U_{(k+1)}^{-\beta} \sum_{j=1}^k \int_{j/nh}^{j/k} (U_{(j+1)}^\beta - U_{(j)}^\beta) u^\lambda L(u) du \right| \leq \\ \leq \frac{\|L\|}{\lambda + 1} U_{(k+1)}^{-\beta} \sum_{j=1}^k |U_{(j+1)}^\beta - U_{(j)}^\beta| \left| \left(\frac{j}{k} \right)^{\lambda+1} - \left(\frac{j}{nh} \right)^{\lambda+1} \right| \end{aligned}$$

$$\begin{aligned} &\leq \|L\| U_{(k+1)}^{-\beta} \sum_{j=1}^k \left| U_{(j+1)}^\beta - U_{(j)}^\beta \right| \left(\frac{j}{k} - \frac{j}{nh} \right) \left(\frac{1}{nh} \right)^\lambda \\ &< \|L\| U_{(k+1)}^{-\beta} \sum_{j=1}^k \left| U_{(j+1)}^\beta - U_{(j)}^\beta \right| \frac{j}{knh} \left(\frac{1}{nh} \right)^\lambda \\ &= \|L\| U_{(k+1)}^{-\beta} \left| \sum_{j=1}^k \left(U_{(j+1)}^\beta - U_{(j)}^\beta \right) \frac{j}{knh} \right| \left(\frac{1}{nh} \right)^\lambda \end{aligned}$$

The last equality follows from the fact that either all terms of the summation are positive (in case $\beta > 0$) or all terms are negative (in case $\beta < 0$). Continuing with the last expression, we obtain

$$\begin{aligned} &\|L\| U_{(k+1)}^{-\beta} \left| \sum_{j=1}^k \left(U_{(j+1)}^\beta - U_{(j)}^\beta \right) \frac{j}{knh} \right| \left(\frac{1}{nh} \right)^\lambda = \\ &= \frac{\|L\|}{k} U_{(k+1)}^{-\beta} \left| \sum_{j=2}^{k+1} (j-1) U_{(j)}^\beta - \sum_{j=1}^k j U_{(j)}^\beta \right| (nh)^{-(1+\bar{\lambda})} \\ &= \frac{\|L\|}{k} U_{(k+1)}^{-\beta} \left| (k+1) U_{(k+1)}^\beta - U_{(1)}^\beta - \sum_{j=1}^k U_{(j+1)}^\beta \right| (nh)^{-(1+\bar{\lambda})} \\ &= \frac{\|L\|}{k} \left| (k+1) - \left(\frac{U_{(1)}}{U_{(k+1)}} \right)^\beta - \sum_{j=1}^k \left(\frac{U_{(j+1)}}{U_{(k+1)}} \right)^\beta \right| (nh)^{-(1+\bar{\lambda})} \\ &= \frac{\|L\|}{k} \left| \left(1 - \left(\frac{U_{(1)}}{U_{(k+1)}} \right)^\beta \right) + \left(k - \sum_{j=1}^k \left(\frac{U_{(j+1)}}{U_{(k+1)}} \right)^\beta \right) \right| (nh)^{-(1+\bar{\lambda})} \end{aligned}$$

Note that, for $j = 1, \dots, k+1$, in case $\beta > 0$, $U_{(j)}^\beta \geq U_{(1)}^\beta$ and in case $\beta < 0$, $U_{(j)}^\beta \leq U_{(1)}^\beta$. I.e., the last expression is bounded by

$$\frac{\|L\|}{k} \left| 1 - \left(\frac{U_{(1)}}{U_{(k+1)}} \right)^\beta \right| \frac{1+k}{(nh)^{1+\bar{\lambda}}} \leq 2 \|L\| \left| 1 - \left(\frac{U_{(1)}}{U_{(k+1)}} \right)^\beta \right| (nh)^{-(1+\bar{\lambda})} \quad (5.14)$$

In case $\beta > 0$, we know that with probability 1, $1 - (U_{(1)}/U_{(k+1)})^\beta$ is bounded between 0 and 1, hence (5.14) tends to 0 as $n \rightarrow \infty$. In case $\beta < 0$, observe that by

Lemma 5.4, $U_{(1)}/U_{(k+1)} \stackrel{D}{=} V_{(1)}$, using the notation of that lemma. I.e., for any $\delta > 0$,

$$\begin{aligned} \mathbb{P} \left(\left[\left(\frac{U_{(1)}}{U_{(k+1)}} \right)^\beta - 1 \right] > \delta (nh)^{1+\bar{\lambda}} \right) &= \mathbb{P} \left(V_{(1)} < (\delta (nh)^{1+\bar{\lambda}} + 1)^{1/\beta} \right) \\ &= 1 - \left(1 - (\delta (nh)^{1+\bar{\lambda}} + 1)^{1/\beta} \right)^k \end{aligned} \tag{5.15}$$

However, since $\bar{\lambda} > -1$ and $\beta < 0$, we have that

$$k \log \left(1 - (\delta (nh)^{1+\bar{\lambda}} + 1)^{1/\beta} \right) = -k \left(\delta (nh)^{1+\bar{\lambda}} + 1 \right)^{1/\beta} (1 + o(1))$$

as n tends to infinity. I.e., using that by definition $k \sim nh$, (5.15) tends to 0 as n tends to infinity, whenever $1 + (1 + \bar{\lambda})/\beta < 0$ and hence (5.14) tends to 0 in probability as $n \rightarrow \infty$ whenever $-1 - \bar{\lambda} < \beta < 0$.

Finally consider the second term (5.13). Note that Lemma 5.4 yields that all finite dimensional distributions of the process $u \mapsto \Gamma_n(hu)/U_{(k+1)}$ equal in distribution the finite dimensional distributions of the process $u \mapsto \Gamma_k(u)$, where $\Gamma_k(u)$ is the empirical quantile function of a $\mathcal{U}(0, 1)$ sample V_1, \dots, V_k . I.e., the two processes themselves are equal in distribution. Hence, instead of (5.13), consider

$$\int_0^1 \left((\Gamma_k(u))^\beta - u^\beta \right) u^\lambda L(u) du \tag{5.16}$$

Obviously,

$$\begin{aligned} \left| \int_0^1 \left((\Gamma_k(u))^\beta - u^\beta \right) u^\lambda L(u) du \right| &\leq \\ &\sup_{0 < u < 1} \left[u^{\nu_1} (1-u)^{\nu_2} \left| (\Gamma_k(u))^\beta - u^\beta \right| \right] \|L\| \int_0^1 u^{\lambda - \nu_1} (1-u)^{-\nu_2} du \end{aligned}$$

For $\beta > 0$ the right hand side has the same distribution as

$$\sup_{0 < u < 1} \left[u^{\nu_1} (1-u)^{\nu_2} \left| F_{\beta, k}^{-1}(u) - F_\beta^{-1}(u) \right| \right] \|L\| \int_0^1 u^{\lambda - \nu_1} (1-u)^{-\nu_2} du \tag{5.17}$$

where $F_\beta^{-1}(\cdot)$ is the quantile function corresponding to the distribution function $F_\beta(x) = x^{1/\beta}$ for $0 < x < 1$ and $F_{\beta, k}^{-1}(\cdot)$ denotes the empirical quantile function of a sample X_1, \dots, X_k drawn from F_β .

Note that, since $0 < |X_1| < 1$ almost surely,

$$\mathbb{E}|X_1 \wedge 0|^{1/\nu_1} = 0$$

whenever $\nu_1 > 0$ and $\beta > 0$ and

$$\mathbb{E}(X_1 \vee 0)^{1/\nu_2} < \infty$$

whenever $\nu_2 > 0$ and $\beta > 0$. Theorem 3 in MASON (1982b) then yields that the supremum in (5.17) tends to 0 almost surely as $k \rightarrow \infty$. Taking $\nu_1 < (1 + \lambda)$ and $\nu_2 < 1$ gives that the integral in (5.17) is finite. I.e., in case $\beta > 0$ taking $0 < \nu_1 < (1 + \lambda)$ and $0 < \nu_2 < 1$ yields that (5.16) tends to 0 almost surely as k tends to infinity.

In case $\beta < 0$ note that

$$\sup_{0 < u < 1} \left[u^{\nu_1} (1 - u)^{\nu_2} \left| (\Gamma_k(u))^\beta - u^\beta \right| \right] \stackrel{\mathcal{D}}{=} \sup_{0 < u < 1} \left[(1 - u)^{\nu_1} u^{\nu_2} \left| G_{\beta, k}^{-1}(u) - G_\beta^{-1}(u) \right| \right]$$

where $G_\beta^{-1}(\cdot)$ is the quantile function corresponding to the distribution function $G_\beta(x) = 1 - x^{1/\beta}$ for $x \geq 1$ and $G_{\beta, k}^{-1}(\cdot)$ denotes the empirical quantile function of a sample X_1, \dots, X_k drawn from G_β . Again use Theorem 3 in MASON (1982b) together with

$$\mathbb{E}|X_1 \wedge 0|^{1/\nu_2} = 0$$

whenever $\nu_2 > 0$ and $\beta < 0$ and

$$\mathbb{E}(X_1 \vee 0)^{1/\nu_1} = -\frac{1}{\beta} \int_1^\infty x^{\frac{1}{\nu_1} + \frac{1}{\beta} - 1} dx < \infty$$

whenever $\nu_1 > -\beta$. Hence, (5.16) tends to 0 almost surely as k tends to infinity, taking $-\beta < \nu_1 < (1 + \lambda)$ and $0 < \nu_2 < 1$. ■

5.4.2 The positive part

In this subsection we will prove that, under the conditions of Theorem 5.1, we have that $\hat{\gamma}_{n, h}^{(pos)} \xrightarrow{P} (\gamma \vee 0)$ as n tends to infinity.

First observe that $Q_n(t) \stackrel{\mathcal{D}}{=} Q(\Gamma_n(t))$ and $\Gamma_n(1 - t) \stackrel{\mathcal{D}}{=} 1 - \Gamma_n(t)$ where $\Gamma_n(t)$ is the empirical quantile function of a $\mathcal{U}(0, 1)$ sample of size n (denoted by U_1, \dots, U_n) and that conditions [CK2] and [CK4] imply that $\int_0^1 d(uK(u)) = 0$. I.e., note that we can write

$$\hat{\gamma}_{n, h}^{(pos)} \stackrel{\mathcal{D}}{=} \int_0^1 \left(\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)}) \right) d(u(K(u)))$$

Moreover, by definition, $U_{(k+1)} \geq \Gamma_n(hu)$ with probability 1 for all $u \in (0, 1)$, and $U_{(k+1)} \rightarrow 0$ with probability 1 as $h \rightarrow 0$.

Consider the case $\gamma > 0$. By the just stated remarks, we can apply Lemma 5.1 with $y = \Gamma_n(hu)/U_{(k+1)}$, $s = U_{(k+1)}$ and $a(s)/Q(1 - s) = \gamma$ (see Remark 5.1), to get that, with

probability 1, for each $\varepsilon > 0$ there exists an n_0 such that for all $n \geq n_0$, with probability 1,

$$(1 - \varepsilon) \frac{1 - \left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^\varepsilon}{\varepsilon} - \varepsilon <$$

$$\frac{\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})}{\gamma} <$$

$$(1 + \varepsilon) \frac{\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^{-\varepsilon} - 1}{\varepsilon} + \varepsilon$$

for all $u \in (0, 1)$.

Defining $k(u) = d(uK(u))/du$ we get by the boundedness of both K and K' that $k(u) = k^+(u) - k^-(u)$ where $k^\pm(u)$ are positive and bounded functions. Hence, for $\gamma > 0$,

$$\hat{\gamma}_{n,h}^{(pos)} < \gamma \int_0^1 \left[(1 + \varepsilon) \frac{\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^{-\varepsilon} - 1}{\varepsilon} + \varepsilon \right] k^+(u) du$$

$$- \gamma \int_0^1 \left[(1 - \varepsilon) \frac{1 - \left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^\varepsilon}{\varepsilon} - \varepsilon \right] k^-(u) du$$

However, applying Lemma 5.5 twice (once with $\beta = -\varepsilon$, $\lambda = 0$ and $L(u) = k^+(u)$ and once with $\beta = \varepsilon$, $\lambda = 0$ and $L(u) = k^-(u)$) yields that, for any $0 < \varepsilon < 1$, this upperbound tends to

$$\gamma \int_0^1 \left[(1 + \varepsilon) \frac{u^{-\varepsilon} - 1}{\varepsilon} + \varepsilon \right] k^+(u) du - \gamma \int_0^1 \left[(1 - \varepsilon) \frac{1 - u^\varepsilon}{\varepsilon} - \varepsilon \right] k^-(u) du$$

in probability, as $n \rightarrow \infty$. Letting ε tend to 0, using dominated convergence and the majorant given in Lemma 5.3 for the first integral, this tends to

$$\gamma \int_0^1 (-\log u) k^+(u) du - \gamma \int_0^1 (-\log u) k^-(u) du = \gamma \int_0^1 (-\log u) d(uK(u)) = \gamma$$

Similar arguments lead to a lower bound for $\hat{\gamma}_{n,h}^{(pos)}$ that tends to γ in probability as well. I.e., for $\gamma > 0$, we have that $\hat{\gamma}_{n,h}^{(pos)}$ tends to γ in probability as $n \rightarrow \infty$.

In case $\gamma = 0$, the scaling by γ in the inequalities used for positive γ , should be replaced by $a(U_{(k+1)})/Q(1 - U_{(k+1)})$. Since $U_{(k+1)}$ tends to 0 almost surely, Lemma 5.2 yields that this factor tends to 0 with probability 1 as n tends to infinity. This, together with

similar arguments as in the case of positive γ , gives that $\hat{\gamma}_{n,h}^{(pos)}$ tends to 0 in probability as n tends to infinity.

Finally consider the case $\gamma < 0$. Lemma 5.1 now yields the inequalities

$$1 - (1 + \varepsilon) \left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^{-\gamma - \varepsilon} < \frac{\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})}{\log Q(1) - \log Q(1 - U_{(k+1)})} < 1 - (1 - \varepsilon) \left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^{-\gamma + \varepsilon}$$

Thus

$$\frac{\hat{\gamma}_{n,h}^{(pos)}}{\log Q(1) - \log Q(1 - U_{(k+1)})} < \int_0^1 \left[1 - (1 - \varepsilon) \left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^{-\gamma + \varepsilon} \right] k^+(u) du - \int_0^1 \left[1 - (1 + \varepsilon) \left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^{-\gamma - \varepsilon} \right] k^-(u) du$$

Again by two applications of Lemma 5.5 (once with $\beta = -\gamma + \varepsilon$, $\lambda = 0$ and $L(u) = k^+(u)$ and once with $\beta = -\gamma - \varepsilon$, $\lambda = 0$ and $L(u) = k^-(u)$) we get that, for any $0 < \varepsilon < 1 - \gamma$, the integrals tend to

$$\int_0^1 [1 - (1 - \varepsilon)u^{-\gamma + \varepsilon}] k^+(u) du \quad \text{and} \quad \int_0^1 [1 - (1 + \varepsilon)u^{-\gamma - \varepsilon}] k^-(u) du$$

respectively. Since these integrals are bounded for $0 < \varepsilon < 1 - \gamma$ and $\log Q(1) - \log Q(1 - U_{(k+1)})$ tends to 0 with probability 1, we get (with a similar lower bound) that $\hat{\gamma}_{n,h}^{(pos)}$ tends to 0 in probability as n tends to infinity.

I.e., for $\gamma \in \mathbb{R}$ we have that $\hat{\gamma}_{n,h}^{(pos)} \xrightarrow{P} (\gamma \vee 0)$ as $n \rightarrow \infty$.

5.4.3 The negative part

In this subsection we will show that the quotient $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \xrightarrow{P} 1 + (\gamma \wedge 0)$ as n tends to infinity.

Since we will consider $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)}$, we can scale both numerator and denominator by the same factor, without changing the quotient. Moreover, by condition [CK2], we have that, for any $\alpha > 0$,

$$\int_0^1 d(u^\alpha K(u)) = [u^\alpha K(u)]_0^1 = 0$$

and

$$\int_0^1 d \left[\frac{d}{du} u^{\alpha+1} K(u) \right] = [(\alpha + 1)u^\alpha K(u) + u^{\alpha+1} K'(u)]_0^1 = 0$$

i.e., we can shift both terms in the same way as in we did with the positive part as well without changing the quotient.

First consider $\gamma \geq 0$. By the previous remarks, we can rewrite the denominator as

$$\frac{h^{1-\alpha} \hat{q}_{n,h}^{(1)}}{a(U_{(k+1)})/Q(1-U_{(k+1)})} \stackrel{\mathcal{D}}{=} \int_0^1 \frac{\log Q(1-\Gamma_n(hu)) - \log Q(1-U_{(k+1)})}{a(U_{(k+1)})/Q(1-U_{(k+1)})} d(u^\alpha K(u))$$

By similar⁴ arguments as used in the positive part, this will, as n tends to infinity, in probability tend to

$$\int_0^1 (-\log u) d(u^\alpha K(u)) = \int_0^1 u^{\alpha-1} K(u) du \quad (5.18)$$

On the other hand, the numerator can be rewritten as

$$\frac{h^{1-\alpha} \hat{q}_{n,h}^{(2)}}{a(U_{(k+1)})/Q(1-U_{(k+1)})} \stackrel{\mathcal{D}}{=} \int_0^1 \frac{\log Q(1-\Gamma_n(hu)) - \log Q(1-U_{(k+1)})}{a(U_{(k+1)})/Q(1-U_{(k+1)})} d \left(\frac{d}{du} u^{\alpha+1} K(u) \right)$$

and this tends in probability to

$$\begin{aligned} \int_0^1 (-\log u) d \left(\frac{d}{du} u^{\alpha+1} K(u) \right) &= \int_0^1 \frac{d}{du} (u^{\alpha+1} K(u)) u^{-1} du \\ &= \int_0^1 ((\alpha + 1)u^{\alpha-1} K(u) + u^\alpha K'(u)) du \\ &= \int_0^1 u^{\alpha-1} K(u) du + \int_0^1 d(u^\alpha K(u)) \\ &= \int_0^1 u^{\alpha-1} K(u) du \end{aligned}$$

I.e., combining this with (5.18), we obtain that $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \xrightarrow{P} 1$ as n tends to infinity, whenever $\gamma \geq 0$.

⁴Note that in Lemma 5.5 we now have $\lambda = \alpha - 1$

Next consider $\gamma < 0$. Similar arguments lead to

$$\begin{aligned} \frac{h^{1-\alpha} \hat{q}_{n,h}^{(1)}}{\log Q(1) - \log Q(1 - U_{(k+1)})} &\xrightarrow{P} \int_0^1 (1 - u^{-\gamma}) d(u^\alpha K(u)) \\ &= -\gamma \int_0^1 u^{\alpha-\gamma-1} K(u) du \end{aligned}$$

and

$$\begin{aligned} \frac{h^{1-\alpha} \hat{q}_{n,h}^{(2)}}{\log Q(1) - \log Q(1 - U_{(k+1)})} &\xrightarrow{P} \int_0^1 (1 - u^{-\gamma}) d\left(\frac{d}{du} u^{\alpha+1} K(u)\right) \\ &= -\gamma(1 + \gamma) \int_0^1 u^{\alpha-\gamma-1} K(u) du \end{aligned}$$

as $n \rightarrow \infty$, hence $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \xrightarrow{P} 1 + \gamma$ as n tends to infinity.

I.e., for $\gamma \in \mathbb{R}$ we have that $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \xrightarrow{P} 1 + (\gamma \wedge 0)$ as n tends to infinity.

5.5 Asymptotic normality

In this section we will prove the asymptotic normality of the new estimator, as stated in Theorem 5.2.

5.5.1 Preliminaries

The following lemmas will be needed in our proof of the asymptotic normality.

Lemma 5.6

Let F be a distribution function that is in the domain of attraction of an extreme value distribution G_γ . Moreover, without loss of generality, assume that its upper endpoint x_F° is positive and denote the associated quantile function by Q . Then

$$\frac{\log Q(1-s)}{-\log s} \rightarrow 0 \quad \text{as } s \downarrow 0.$$

Proof of Lemma 5.6:

In case $\gamma < 0$, we have that $x_F^\circ < \infty$ and the assertion follows trivially.

In case $\gamma > 0$, the domain of attraction condition is equivalent to the condition that $Q(1 - \cdot) \in \text{RV}_{-\gamma}^0$, see e.g., Theorem 1.9 of this thesis. Hence, $\log Q(1 - \cdot)$ is slowly varying at zero and the assertion follows by property 1 of Proposition 1.7 in GELUK AND DE HAAN (1987).

In case $\gamma = 0$ and $x_F^\circ < \infty$ the assertion again follows trivially. Moreover, $F \in \mathcal{D}(G_0)$

implies that $Q(1 - \cdot) \in \Pi^0$ (see e.g., Theorem 1.9 of this thesis). Corollary 1.18 in GELUK AND DE HAAN (1987) yields that in case $x_F^0 = Q(1) = \infty$ the function $Q(1 - \cdot)$ is then slowly varying at 0. Hence the assertion follows by the same argument as in case of positive γ . ■

The next results are taken from WELLNER (1978) and are stated without proof.

Lemma 5.7

Let $\Gamma_n(\cdot)$ denote the empirical quantile function of a uniform $(0, 1)$ sample of size n . Then, as n tends to infinity,

$$\sup_{\frac{1}{n} \leq u \leq 1} \left| \frac{\Gamma_n(u)}{u} \right| = O_P(1) \quad \text{and} \quad \sup_{\frac{1}{n} \leq u \leq 1} \left| \frac{u}{\Gamma_n(u)} \right| = O_P(1)$$

and

$$\sup_{b_n \leq u \leq 1} \left| \frac{\Gamma_n(u) - u}{u} \right| = o_P(1) \tag{5.19}$$

where b_n is any sequence of positive numbers satisfying $nb_n \rightarrow \infty$ as $n \rightarrow \infty$.

The following result is taken from CSÖRGŐ, CSÖRGŐ, HORVÁTH AND MASON (1986) and is again stated without proof. In CSÖRGŐ, CSÖRGŐ, HORVÁTH AND MASON (1986) a probability space is constructed on which there exists a sequence U_1, U_2, \dots of i.i.d. uniform $(0, 1)$ variables and a sequence of Brownian Bridges $\{B^{(n)}(s)\}$, which has, among others, the following property (cf. Theorem 2.1 in CSÖRGŐ, CSÖRGŐ, HORVÁTH AND MASON (1986)):

Lemma 5.8

For any $0 \leq \nu < 1/2$

$$\sup_{1/n \leq u \leq 1 - 1/n} \frac{|\sqrt{n}(\Gamma_n(u) - u) - B^{(n)}(u)|}{u^{\frac{1}{2} - \nu}} = O_P(n^{-\nu})$$

as $n \rightarrow \infty$, where $\Gamma_n(\cdot)$ is the quantile function of U_1, \dots, U_n .

In view of this result, we will assume that any uniform $(0, 1)$ sample and sequence of Brownian Bridges we will be using, are defined on the above mentioned probability space.

5.5.2 The various parts of the estimator

The estimator consists of three stochastic ingredients: $\hat{\gamma}_{n,h}^{(pos)}$, $\hat{q}_{n,h}^{(1)}$ and $\hat{q}_{n,h}^{(2)}$. However, $\hat{\gamma}_{n,h}^{(pos)}$ can be considered to be a special case $\hat{q}_{n,h}^{(1)}$ with $\alpha = 1$. In this subsection we will prove the following proposition:

Proposition 5.1

Under the assumptions of Theorem 5.2, as n tends to infinity,

$$\sqrt{nh}h^{1-\alpha} \left(\hat{q}_{n,h}^{(i)} - q_h^{(i)} \right) \stackrel{D}{=} \int_0^1 \frac{W(u)}{u} \phi(hu) dK_h^{(i)}(u) + o_P(1) \tag{5.20}$$

for $i = 1, 2$, where $W(\cdot)$ is standard Brownian Motion,

$$K_h^{(1)}(u) = u^\alpha K_h(u)$$

$$K_h^{(2)}(u) = \frac{d}{du} (u^{\alpha+1} K_h(u))$$

$\hat{q}_{n,h}^{(i)}$ are defined in Theorem 5.2 and $q_h^{(i)}$ are given in (5.10).

Proof of Proposition 5.1:

We will only present the proof for $\hat{q}_{n,h}^{(1)}$ (and hence, as a special case, for $\hat{\gamma}_{n,h}^{(pos)}$), since the proof for $\hat{q}_{n,h}^{(2)}$ is similar.

In view of the lemmas of subsection 5.5.1, the left hand side of (5.20) is decomposed into four parts:

$$h^{1-\alpha} \left(\hat{q}_{n,h}^{(1)} - q_h^{(1)} \right) = \tag{5.21}$$

$$\begin{aligned} & h^{1-\alpha} \int_0^{1/n} \log Q_n(1-u) dK_h^{(1)}(u) - h^{1-\alpha} \int_0^{1/n} \log Q(1-u) dK_h^{(1)}(u) \\ & + h^{1-\alpha} \int_{1/n}^{b_n} \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) + h^{1-\alpha} \int_{b_n}^h \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) \end{aligned}$$

where b_n is a sequence of positive real numbers that at the moment only satisfies the property that $1/n < b_n < h$.

For the first part of (5.21) note that $Q_n(1-u)$ is constant for $0 \leq u < 1/n$, i.e.,

$$h^{1-\alpha} \int_0^{1/n} \log Q_n(1-u) dK_h^{(1)}(u) \stackrel{D}{=} \log Q(1-U_{(1)}) \int_0^{1/nh} dK^{(1)}(u)$$

with $U_{(1)}$ the first order statistic from a uniform $(0, 1)$ sample of size n . By Lemma 5.6 and the fact that $U_{(1)} \rightarrow 0$ almost surely as n tends to infinity, we hence have that

$$h^{1-\alpha} \int_0^{1/n} \log Q_n(1-u) dK_h^{(1)}(u) \stackrel{D}{=} o_P \left(\frac{-\log U_{(1)}}{(nh)^\alpha} \right)$$

Using that for any $M > 0$

$$\begin{aligned} \mathbb{P} \left(-(nh)^{\frac{1}{2}-\alpha} \log U_{(1)} \geq M \right) &= 1 - \left(1 - \exp \left(-(nh)^{\alpha-\frac{1}{2}} M \right) \right)^n \\ &\leq n \exp \left(-(nh)^{\alpha-\frac{1}{2}} M \right) \end{aligned}$$

the conditions on h of Theorem 5.2 then yield that

$$h^{1-\alpha} \int_0^{1/n} \log Q_n(1-u) dK_h^{(1)}(u) = o_P\left((nh)^{-1/2}\right)$$

For the second part of (5.21) observe that

$$\begin{aligned} h^{1-\alpha} \int_0^{1/n} \log Q(1-u) dK_h^{(1)}(u) = \\ h^{1-\alpha} \left[\log Q \left(1 - \frac{1}{n} \right) K_h^{(1)} \left(\frac{1}{n} \right) \right] + \int_0^{1/nh} \phi(hu) K^{(1)}(u) du \end{aligned}$$

The conditions on F of Condition 5.2 yield that $\phi(s) \rightarrow (\gamma \vee 0)$ as s tends to 0. I.e., together with another application of Lemma 5.6 then leads to

$$h^{1-\alpha} \int_0^{1/n} \log Q(1-u) dK_h^{(1)}(u) = o\left(\frac{\log n}{(nh)^\alpha}\right) + O((nh)^{-1}) = o\left((nh)^{-1/2}\right)$$

where the last equality follows from the conditions on h .

The third part of (5.21) is dealt with using Lemma 5.7. First observe that

$$\begin{aligned} \int_{1/n}^{b_n} \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) \stackrel{\mathcal{D}}{=} \\ \int_{1/n}^{b_n} [\log Q(1-\Gamma_n(u)) - \log Q(1-u)] dK_h^{(1)}(u) \end{aligned}$$

where $\Gamma_n(\cdot)$ is the empirical quantile function of a uniform $(0, 1)$ sample of size n . Using the mean value theorem, we then get that

$$\int_{1/n}^{b_n} \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) \stackrel{\mathcal{D}}{=} \int_{1/n}^{b_n} \frac{\phi(u + \xi_{n,u})}{u + \xi_{n,u}} (\Gamma_n(u) - u) dK_h^{(1)}(u)$$

with $\xi_{n,u}$ satisfying $|\xi_{n,u}| \leq |\Gamma_n(u) - u|$.

Since $\sup_{0 < u < 1} |\Gamma_n(u) - u| \rightarrow 0$ almost surely as n tends to infinity (see e.g., Lemma 4.2 of this thesis) we get using the conditions on F and Lemma 5.7 that for all $\gamma \in \mathbb{R}$ and n large enough

$$\sup_{1/n \leq u \leq b_n} \left| \phi(u + \xi_{n,u}) \frac{u}{u + \xi_{n,u}} \right| = O_P(1)$$

and we thus obtain that

$$\begin{aligned} \left| \int_{1/n}^{b_n} \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) \right| &= \left| \int_{1/n}^{b_n} \frac{\phi(u + \xi_{n,u})}{u + \xi_{n,u}} (\Gamma_n(u) - u) dK_h^{(1)}(u) \right| \\ &\leq O_P(1) \int_{1/n}^{b_n} \left| \frac{\Gamma_n(u) - u}{u} \right| \left| \frac{dK_h^{(1)}(u)}{du} \right| du \\ &\leq O_P(1) \int_{1/n}^{b_n} \left| \frac{dK_h^{(1)}(u)}{du} \right| du \\ &= h^{\alpha-1} O_P(1) \int_{1/nh}^{b_n/h} \left| \frac{dK^{(1)}(u)}{du} \right| du \end{aligned}$$

where the last inequality follows from another application of Lemma 5.7. Taking $b_n = h(nh)^{-(1/2+\lambda)/\alpha}$ for some $0 < \lambda < \alpha - 1/2$, we get that

$$h^{1-\alpha} \int_{1/n}^{b_n} \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) = O_P \left(\frac{b_n^\alpha}{h} \right) = o_P \left((nh)^{-1/2} \right)$$

as n tends to infinity.

Finally consider the fourth part of the decomposition (5.21). Following the same arguments as for the third part, we arrive at

$$\int_{b_n}^h \log \left(\frac{Q_n(1-u)}{Q(1-u)} \right) dK_h^{(1)}(u) \stackrel{D}{=} \int_{b_n}^h \frac{\phi(u + \xi_{n,u})}{u + \xi_{n,u}} (\Gamma_n(u) - u) dK_h^{(1)}(u)$$

for some $|\xi_{n,u}| \leq |\Gamma_n(u) - u|$. Now use (5.19), together with the conditions on F , to obtain that, for any sequence b_n satisfying $nb_n \rightarrow \infty$ as $n \rightarrow \infty$ and any $\gamma \in \mathbb{R}$

$$\sup_{b_n \leq u \leq h} \left| \frac{\phi(u + \xi_{n,u})}{\phi(u)} \frac{u}{u + \xi_{n,u}} \right| = 1 + o_P(1)$$

This implies that

$$\int_{b_n}^h \frac{\phi(u + \xi_{n,u})}{u + \xi_{n,u}} (\Gamma_n(u) - u) dK_h^{(1)}(u) = (1 + o_P(1)) \int_{b_n}^h \phi(u) \frac{\Gamma_n(u) - u}{u} dK_h^{(1)}(u)$$

Applying Lemma 5.8, we get that the right hand side then equals

$$(1 + o_P(1)) n^{-1/2} \int_{b_n}^h \phi(u) \frac{B_n(u)}{u} dK_h^{(1)}(u) + R_{n,h}$$

where the Brownian Bridges $B_n(\cdot)$ are defined as in Lemma 5.8 and, for arbitrary $0 \leq \nu < 1/2$,

$$\begin{aligned} |R_{n,h}| &\leq O_P\left(n^{-1/2-\nu}\right) \int_{b_n}^h u^{-1/2-\nu} |\phi(u)| \left| \frac{dK_h^{(1)}(u)}{du} \right| du \\ &\leq h^{\alpha-1} O_P\left((nh)^{-1/2-\nu}\right) \int_{b_n/h}^1 u^{-1/2-\nu} |\phi(hu)| \left| \frac{dK^{(1)}(u)}{du} \right| du \\ &= h^{\alpha-1} O_P\left((nh)^{-1/2-\nu}\right) \end{aligned}$$

Using that $B_n(u) \stackrel{\mathcal{D}}{=} W_n(u) + \zeta_n u$ where $W_n(\cdot)$ is standard Brownian Motion and ζ_n is a standard normal variable, independent of $W_n(\cdot)$, we obtain as $h \downarrow 0$ and $nh \rightarrow \infty$,

$$\begin{aligned} n^{-1/2} \int_{b_n}^h \phi(u) \frac{B_n(u)}{u} dK_h^{(1)}(u) &\stackrel{\mathcal{D}}{=} n^{-1/2} \int_{b_n}^h \phi(u) \frac{W_n(u)}{u} dK_h^{(1)}(u) + n^{-1/2} \zeta_n \int_{b_n}^h \phi(u) dK_h^{(1)}(u) \\ &= n^{-1/2} \int_{b_n}^h \phi(u) \frac{W_n(u)}{u} dK_h^{(1)}(u) + h^{\alpha-1} n^{-1/2} \zeta_n \int_{b_n/h}^1 \phi(hu) dK^{(1)}(u) \\ &\stackrel{\mathcal{D}}{=} h^{\alpha-1} (nh)^{-1/2} \int_{b_n/h}^1 \phi(hu) \frac{W_n(u)}{u} dK^{(1)}(u) + h^{\alpha-1} O_P\left(n^{-1/2}\right) \end{aligned}$$

where in the last equality we used the well known scaling property of Brownian Motion that $W(hu) \stackrel{\mathcal{D}}{=} \sqrt{h}W(u)$. Finally, noting that $\mathbb{E}|W(u)| \leq \sqrt{\mathbb{E}W(u)^2} = \sqrt{u}$, hence

$$\int_0^{b_n/h} \phi(hu) \frac{W(u)}{u} dK^{(1)}(u) = O_P\left(\left(\frac{b_n}{h}\right)^{\alpha-1/2}\right)$$

we obtain, taking $b_n = h(nh)^{-(1/2+\lambda)/\alpha}$ for some $0 < \lambda < \alpha - 1/2$, that

$$\begin{aligned} h^{1-\alpha} \int_{b_n}^h \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) &\stackrel{\mathcal{D}}{=} (nh)^{-1/2} (1 + o_P(1)) \int_0^1 \phi(hu) \frac{W(u)}{u} dK^{(1)}(u) + o_P\left((nh)^{-1/2}\right) \end{aligned}$$

Putting the results concerning the four parts of decomposition (5.21) together, we obtain the assertion of Proposition 5.1. ■

5.5.3 The various parts together

In this subsection we will show that the results for the various parts of the estimator as given in Proposition 5.1 lead to the assertion of Theorem 5.2.

First note that the result of Proposition 5.1 can also be written in the following way:

$$h^{1-\alpha} \hat{q}_{n,h}^{(i)} = h^{1-\alpha} q_h^{(i)} + (nh)^{-1/2} A_n^{(i)} + o_P\left((nh)^{1/2}\right)$$

where

$$A_n^{(i)} = \int_0^1 \phi(hu) \frac{W(u)}{u} dK^{(i)}(u)$$

which implies that

$$\frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}} = \frac{q_h^{(2)}}{q_h^{(1)}} + \frac{(nh)^{-1/2} A_n^{(2)}}{h^{1-\alpha} q_h^{(1)}} - \frac{(nh)^{-1/2} A_n^{(1)} h^{1-\alpha} q_h^{(2)}}{\left(h^{1-\alpha} q_h^{(1)}\right)^2} + o_P\left((nh)^{-1/2}\right)$$

since $h^{1-\alpha} q_h^{(i)} = \int_0^1 (\phi(hu)/u) K^{(i)}(u) du = O(1)$.

Moreover, since for each $\varepsilon > 0$ and n large enough $|\phi(hu) - (\gamma \vee 0)| < \varepsilon$ uniformly for $0 < u < 1$ by the conditions on the underlying distribution function F , we have that

$$\int_0^1 \phi(hu) \frac{W(u)}{u} d(uK(u)) = (\gamma \vee 0) \int_0^1 \frac{W(u)}{u} d(uK(u)) (1 + o_P(1))$$

Hence

$$\begin{aligned} \sqrt{nh} (\hat{\gamma}_{n,h} - \gamma_h) &= \sqrt{nh} \left(\hat{\gamma}_{n,h}^{(pos)} - \gamma_h^{(pos)} \right) + \sqrt{nh} \left(\frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}} - \frac{q_h^{(2)}}{q_h^{(1)}} \right) \\ &= (\gamma \vee 0) \int_0^1 \frac{W(u)}{u} d(uK(u)) + \frac{A_n^{(2)}}{h^{1-\alpha} q_h^{(1)}} - \frac{A_n^{(1)} h^{1-\alpha} q_h^{(2)}}{\left(h^{1-\alpha} q_h^{(1)}\right)^2} + o_P(1) \end{aligned}$$

To deal with the $A_n^{(i)}$, $i = 1, 2$, note that in case $\gamma \neq 0$, with $\bar{\gamma} = \gamma \wedge 0$,

$$\sup_{0 < u < 1} \left| \frac{\phi(hu)}{\phi(h)} - u^{-\bar{\gamma}} \right| \rightarrow 0$$

as h tends to 0, by the conditions on F . I.e., by Markov's inequality, the conditions on K and the fact that $\alpha > 1/2$ we have that

$$\left| \int_0^1 \left(\frac{\phi(hu)}{\phi(h)} - u^{-\bar{\gamma}} \right) \frac{W(u)}{u} dK^{(i)}(u) \right| = o_P(1) \quad (5.22)$$

In case $\gamma = 0$, the ϕ -function is slowly varying. Hence the following inequality from the proof of the proposition in the appendix of DE HAAN AND PEREIRA (1999) can be

used to extend the result of (5.22) to all γ : for each $\varepsilon, \varepsilon_1 > 0$ there exists an h_0 such that for all $h \leq h_0$ and $hu \leq h_0$,

$$\left| \frac{\phi(hu)}{\phi(h)} - 1 \right| < \varepsilon e^{\varepsilon_1 |\log u|}$$

Now use (5.22) to obtain

$$\begin{aligned} \frac{A_n^{(2)}}{h^{1-\alpha} q_h^{(1)}} &\xrightarrow{P} \frac{\int_0^1 u^{-1-\bar{\gamma}} W(u) dK^{(2)}(u)}{\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u) du} \\ &= \frac{-\int_0^1 W(u) d\bar{K}^{(2)}(u)}{\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u) du} \end{aligned}$$

and

$$\begin{aligned} \frac{A_n^{(1)} h^{1-\alpha} q_h^{(2)}}{(h^{1-\alpha} q_h^{(1)})^2} &\xrightarrow{P} \frac{\int_0^1 u^{-1-\bar{\gamma}} W(u) dK^{(1)}(u) \int_0^1 u^{-1-\bar{\gamma}} K^{(2)}(u) du}{\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u) du \int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u) du} \\ &= \frac{\int_0^1 W(u) d\bar{K}^{(1)}(u) \int_0^1 u^{-1-\bar{\gamma}} K^{(2)}(u) du}{\left(\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u) du \right)^2} \end{aligned}$$

where for $i = 1, 2$ and $0 < u < 1$,

$$\bar{K}^{(i)}(u) = \int_u^1 x^{-1-\bar{\gamma}} dK^{(i)}(x)$$

Hence, by partial integration,

$$\sqrt{nh} (\hat{\gamma}_{n,h} - \gamma_h) \xrightarrow{D} \int_0^1 \left[a_0 \bar{K}(u) + a_1 \bar{K}^{(2)}(u) - a_2 \bar{K}^{(1)}(u) \right] dW(u)$$

where

$$\bar{K}(u) = \int_u^1 x^{-1} d(xK(x))$$

$$a_0 = \gamma \vee 0$$

$$a_1 = 1 / \int_0^1 x^{-1-\bar{\gamma}} K^{(1)}(x) dx$$

$$a_2 = a_1^2 \int_0^1 x^{-1-\bar{\gamma}} K^{(2)}(x) dx = (1 + (\gamma \wedge 0)) a_1$$

The assertion of Theorem 5.2 then follows.

5.6 Exploring the bias

In this section we will explore the asymptotic bias of our new estimator, with the asymptotic bias defined as

$$\sqrt{nh}(\gamma_h - \gamma) = \sqrt{nh} \left(\gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 - \gamma \right)$$

where $\gamma_h^{(pos)}$, $q_h^{(1)}$ and $q_h^{(2)}$ are defined in (5.9) and (5.10). Note that $\gamma_h^{(pos)}$ is actually a special case of $q_h^{(1)}$ with $\alpha = 1$.

5.6.1 Preliminaries

The following lemmas will be needed when exploring the bias of our estimator.

Lemma 5.9

Assume that Condition 5.3 holds with the + sign. For any $\varepsilon > 0$ there exists $s_o > 0$ such that, for all $0 < s < s_o$ and $0 < y < 1$:

In case $\gamma > 0$, defining $\tilde{a}_1(s) = \log Q(1-s) + \gamma \log s - \log c$,

$$(1 - \varepsilon)y^{\gamma\rho + \varepsilon} < \frac{\tilde{a}_1(sy)}{\tilde{a}_1(s)} < (1 + \varepsilon)y^{\gamma\rho - \varepsilon}$$

In case $\gamma < 0$, assuming without loss of generality that $Q(1) > 0$ and defining $\tilde{a}_2(s) = s^\gamma(\log Q(1) - \log Q(1-s)) - c/Q(1)$,

$$(1 - \varepsilon)y^{-\gamma\rho + \varepsilon} < \frac{\tilde{a}_2(sy)}{\tilde{a}_2(s)} < (1 + \varepsilon)y^{-\gamma\rho - \varepsilon}$$

Proof of Lemma 5.9:

The inequalities are the well known inequalities of regularly varying functions, see e.g., GELUK AND DE HAAN (1987). ■

Similar inequalities can be derived in case of second order Π -variation, as stated in the next lemma that is a reformulation of Lemma 3.5 in DEKKERS, EINMAHL AND DE HAAN (1989) in terms of the quantile function.

Lemma 5.10

Assume that Condition 5.4 holds with the + sign for $\gamma \geq 0$ and the - sign for $\gamma < 0$.

For any $\varepsilon > 0$ there exists $s_o > 0$ such that, for all $0 < s < s_o$ and $0 < y < 1$:

In case $\gamma > 0$, defining $a_1(s) = \log Q(1-s) + \gamma \log s$,

$$(1 - \varepsilon) \frac{1 - y^\varepsilon}{\varepsilon} - \varepsilon < \frac{a_1(sy) - a_1(y)}{s^{-\gamma} b_1(s)/Q(1-s)} < (1 + \varepsilon) \frac{y^{-\varepsilon} - 1}{\varepsilon} + \varepsilon$$

In case $\gamma = 0$, defining $a_2(s) = \log Q(1-s)$,

$$\frac{(1-\varepsilon)^2 y^\varepsilon (\log y)^2}{2} + 2\varepsilon \log y - \varepsilon <$$

$$\frac{a_2(sy) - a_2(s) + b_2(s) \log y}{b_3(s)} <$$

$$\frac{(1+\varepsilon)^2 y^{-\varepsilon} (\log y)^2}{2} - 2\varepsilon \log y + \varepsilon$$

In case $\gamma < 0$, assuming without loss of generality that $Q(1) > 0$ and defining $a_3(s) = s^\gamma (\log Q(1) - \log Q(1-s))$,

$$(1-\varepsilon) \frac{1-y^\varepsilon}{\varepsilon} - \varepsilon < \frac{a_3(s) - a_3(sy)}{b_4(s)/Q(1)} < (1+\varepsilon) \frac{y^{-\varepsilon} - 1}{\varepsilon} + \varepsilon$$

Proof of Lemma 5.10:

In case $\gamma \neq 0$ the inequalities are just the well known inequalities for Π -varying functions, see e.g., GELUK AND DE HAAN (1987) page 27. In case $\gamma = 0$ the inequalities follow using OMEY AND WILLEKENS (1987) to obtain an asymptotic expression for $b_2(\cdot)$ and applying the inequalities for Π -varying functions to that expression, see proof of Lemma 3.5 in DEKKERS, EINMAHL AND DE HAAN (1989) ■

When applying dominated convergence using the inequalities stated in Lemma 5.9 and Lemma 5.10, a majorant is needed. In case $\gamma \neq 0$, Lemma 5.3 can be used. In case $\gamma = 0$, the next lemma provides a majorant.

Lemma 5.11

For any $\delta > 0$ there exists ε_0 such that

$$\frac{(1+\varepsilon)^2}{2} x^{-\varepsilon} (\log x)^2 - 2\varepsilon \log x < x^{-\delta} (\log x)^2 - 2\delta \log x$$

for all $0 < x < 1$ and all $0 < \varepsilon < \varepsilon_0$.

Proof of Lemma 5.11:

For all $0 < \varepsilon < \sqrt{2} - 1$ we have that

$$\frac{(1+\varepsilon)^2}{2} x^{-\varepsilon} (\log x)^2 - 2\varepsilon \log x < x^{-\varepsilon} (\log x)^2 - 2\varepsilon \log x$$

However, since $\varepsilon \mapsto x^{-\varepsilon} (\log x)^2 - 2\varepsilon \log x$ is increasing for all $0 < x < 1$, we have the assertion whenever $0 < \varepsilon < \varepsilon_0$ with $\varepsilon_0 = (\sqrt{2} - 1) \wedge \delta$. ■

5.6.2 The bias under Condition 5.3

We will only consider the conditions with the + sign. Results under the other conditions follow by symmetry.

The case $\gamma > 0$

To handle both $\gamma_h^{(pos)}$ and $q_h^{(1)}$, consider

$$\int_0^1 \log Q(1-hu) d(u^\alpha K(u)) = \int_0^1 \tilde{a}_1(hu) d(u^\alpha K(u)) - \int_0^1 (\gamma \log(hu) - \log c) d(u^\alpha K(u)) \quad (5.23)$$

where $\tilde{a}_1(s) = \log Q(1-s) + \gamma \log s - \log c$, as defined in Lemma 5.9. Note that $|\tilde{a}_1| \in \text{RV}_{\gamma\rho}^0$ for some $\rho > 0$ by Condition 5.3 and hence $\tilde{a}_1(s) \rightarrow 0$ as $s \rightarrow 0$.

Since for any $\alpha > 0$ we have that $\int_0^1 d(u^\alpha K(u)) = 0$, (5.23) equals

$$\tilde{a}_1(h) \int_0^1 \frac{\tilde{a}_1(hu)}{\tilde{a}_1(h)} d(u^\alpha K(u)) - \gamma \int_0^1 \log u d(u^\alpha K(u))$$

Using the inequalities of Lemma 5.9 and dominated convergence, we have that

$$\int_0^1 \frac{\tilde{a}_1(hu)}{\tilde{a}_1(h)} d(u^\alpha K(u)) \rightarrow \int_0^1 u^{\gamma\rho} d(u^\alpha K(u))$$

as h tends to 0. I.e., as $h \downarrow 0$,

$$\begin{aligned} \int_0^1 \log Q(1-hu) d(u^\alpha K(u)) &= \gamma \kappa_{\alpha-1} + \tilde{a}_1(h) \left(\int_0^1 u^{\gamma\rho} d(u^\alpha K(u)) + o(1) \right) \\ &= \gamma \kappa_{\alpha-1} + \tilde{a}_1(h) (-\gamma\rho \kappa_{\alpha+\gamma\rho-1} + o(1)) \end{aligned}$$

Similarly, using that $\int_0^1 d\left(\frac{d}{du}u^{\alpha+1}K(u)\right) = 0$ for any $\alpha > 0$, we get for $q_h^{(2)}$ that

$$\begin{aligned} h^{1-\alpha} q_h^{(2)} &= -\gamma \int_0^1 \log u d\left(\frac{d}{du}u^{\alpha+1}K(u)\right) + \\ &\quad + \tilde{a}_1(h) \left(\int_0^1 u^{\gamma\rho} d\left(\frac{d}{du}u^{\alpha+1}K(u)\right) + o(1) \right) \\ &= \gamma \kappa_{\alpha-1} + \tilde{a}_1(h) (\gamma\rho(\gamma\rho-1)\kappa_{\alpha+\gamma\rho-1} + o(1)) \end{aligned}$$

as h tends to 0. Combining these two expressions then yields

$$\gamma_h = \gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 = \gamma + \tilde{a}_1(h) \left(\gamma\rho \left(\rho \frac{\kappa_{\alpha+\gamma\rho-1}}{\kappa_{\alpha-1}} - \kappa_{\gamma\rho} \right) + o(1) \right)$$

as $n \rightarrow \infty$.

The case $\gamma < 0$

Again consider

$$\int_0^1 \log Q(1-hu) d(u^\alpha K(u)) = - \int_0^1 (hu)^{-\gamma} \left(\tilde{a}_2(hu) + \frac{c}{Q(1)} \right) d(u^\alpha K(u)) \quad (5.24)$$

where $\tilde{a}_2(s) = s^\gamma (\log Q(1) - \log Q(1-s)) - c/Q(1)$, as defined in Lemma 5.9. By Condition 5.3 $|\tilde{a}_2| \in \text{RV}_{-\gamma\rho}^0$ for some $\rho > 0$ and hence $\tilde{a}_2(s) \rightarrow 0$ as $s \rightarrow 0$.

Obviously, (5.24) equals

$$-h^{-\gamma} \tilde{a}_2(h) \int_0^1 u^{-\gamma} \frac{\tilde{a}_2(hu)}{\tilde{a}_2(h)} d(u^\alpha K(u)) - \frac{h^{-\gamma} c}{Q(1)} \int_0^1 u^{-\gamma} d(u^\alpha K(u))$$

The inequalities of Lemma 5.9, together with dominated convergence, yield that

$$\int_0^1 u^{-\gamma} \frac{\tilde{a}_2(hu)}{\tilde{a}_2(h)} d(u^\alpha K(u)) \rightarrow \int_0^1 u^{-\gamma(1+\rho)} d(u^\alpha K(u))$$

as h tends to 0. I.e., as $h \downarrow 0$,

$$\int_0^1 \log Q(1-hu) d(u^\alpha K(u)) = - \frac{\gamma ch^{-\gamma}}{Q(1)} \kappa_{\alpha-\gamma-1} - \gamma(1+\rho) h^{-\gamma} \tilde{a}_2(h) \left(\kappa_{\alpha-\gamma(1+\rho)-1} + o(1) \right)$$

Similar arguments yield for $q_h^{(2)}$ that

$$h^{1-\alpha} q_h^{(2)} = - \frac{\gamma(\gamma+1)ch^{-\gamma}}{Q(1)} \kappa_{\alpha-\gamma-1} + - \gamma(1+\rho)(1+\gamma(1+\rho)) h^{-\gamma} \tilde{a}_2(h) \left(\kappa_{\alpha-\gamma(1+\rho)-1} + o(1) \right)$$

Combining these two expressions, we obtain

$$\begin{aligned} \gamma_h &= \gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 \\ &= \gamma + \frac{\rho(\rho+1)Q(1)}{c} \frac{\kappa_{\alpha-\gamma(1+\rho)-1}}{\kappa_{\alpha-\gamma-1}} \tilde{a}_2(h) - \frac{\gamma c}{Q(1) \kappa_{-\gamma}} h^{-\gamma} + \\ &\quad + O(h^{-\gamma} \tilde{a}_2(h)) + o(\tilde{a}_2(h)) \end{aligned}$$

as $h \downarrow 0$.

5.6.3 The bias under Condition 5.4

We will only consider the conditions with a + sign in case $\gamma \geq 0$ and with a - sign in case $\gamma < 0$. Results under the other conditions follow by symmetry.

The case $\gamma > 0$

For both $\gamma_h^{(pos)}$ and $q_h^{(1)}$ consider $\int_0^1 \log Q(1-hu) d(u^\alpha K(u))$. Since $\int_0^1 d(u^\alpha K(u)) = 0$ for all $\alpha > 0$, we can rewrite this as

$$\int_0^1 (a_1(hu) - a_1(h)) d(u^\alpha K(u)) + \int_0^1 (-\gamma \log u) d(u^\alpha K(u)) = \\ \frac{h^{-\gamma} b_1(h)}{Q(1-h)} \int_0^1 \frac{a_1(hu) - a_1(h)}{h^{-\gamma} b_1(h)/Q(1-h)} d(u^\alpha K(u)) + \gamma \int_0^1 u^{\alpha-1} K(u) du$$

where $a_1(s) = \log Q(1-s) + \gamma \log s$, as defined in Lemma 5.10.

Using the inequalities of Lemma 5.10 and dominated convergence with majorant given in Lemma 5.3, we obtain by similar arguments as in the proof of the consistency, that

$$\int_0^1 \frac{a_1(hu) - a_1(h)}{h^{-\gamma} b_1(h)/Q(1-h)} d(u^\alpha K(u)) \rightarrow \int_0^1 (-\log u) d(u^\alpha K(u))$$

as h tends to 0. I.e., as $h \downarrow 0$,

$$\int_0^1 \log Q(1-hu) d(u^\alpha K(u)) = \gamma \int_0^1 u^{\alpha-1} K(u) du + O\left(\frac{h^{-\gamma} b_1(h)}{Q(1-h)}\right)$$

Similarly, using that $\int_0^1 d\left(\frac{d}{du} u^{\alpha+1} K(u)\right) = 0$ for any $\alpha > 0$, we get for $q_h^{(2)}$ that

$$h^{1-\alpha} q_h^{(2)} = \gamma \int_0^1 u^{\alpha-1} K(u) du + O\left(\frac{h^{-\gamma} b_1(h)}{Q(1-h)}\right)$$

as h tends to 0. Combining things then yields that

$$\gamma_h = \gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 = \gamma + O\left(\frac{h^{-\gamma} b_1(h)}{Q(1-h)}\right)$$

In order to have an asymptotically vanishing bias, i.e., that $\sqrt{nh}(\gamma_h - \gamma) \rightarrow 0$, we thus have to impose the condition on $h = h_n$ that

$$nh^{1-2\gamma} \left(\frac{b_1(h)}{Q(1-h)}\right)^2 \rightarrow 0$$

as n tends to infinity. Note that this condition resembles the condition on the parameter k in case of the moment estimator as defined in DEKKERS, EINMAHL AND DE HAAN (1989). (See also Theorem 2.9 of this thesis.)

The case $\gamma < 0$

Again, for both $\gamma_h^{(pos)}$ and $q_h^{(1)}$ consider $\int_0^1 \log Q(1 - hu) d(u^\alpha K(u))$. Since for any $\alpha > 0$ we have $\int_0^1 d(u^\alpha K(u)) = 0$, we can rewrite this as

$$\int_0^1 (hu)^{-\gamma} (a_3(h) - a_3(hu)) d(u^\alpha K(u)) - h^{-\gamma} a_3(h) \int_0^1 u^{-\gamma} d(u^\alpha K(u)) =$$

$$\frac{h^{-\gamma} b_4(h)}{Q(1)} \int_0^1 u^{-\gamma} \frac{a_3(h) - a_3(hu)}{b_4(h)/Q(1)} d(u^\alpha K(u)) + \gamma h^{-\gamma} a_3(h) \int_0^1 u^{\alpha-\gamma-1} K(u) du$$

where $a_3(s) = s^\gamma (\log Q(1) - \log Q(1 - s))$, as defined in Lemma 5.10. Using the inequalities of that same lemma, dominated convergence with the majorant given in Lemma 5.3 and similar arguments as in the proof of consistency, we obtain that

$$\int_0^1 u^{-\gamma} \frac{a_3(h) - a_3(hu)}{b_4(h)/Q(1)} d(u^\alpha K(u)) \rightarrow \int_0^1 u^{-\gamma} (-\log u) d(u^\alpha K(u))$$

as h tends to 0. I.e., as $h \downarrow 0$,

$$\int_0^1 \log Q(1 - hu) d(u^\alpha K(u)) =$$

$$\gamma h^{-\gamma} a_3(h) \int_0^1 u^{\alpha-\gamma-1} K(u) du + O\left(\frac{h^{-\gamma} b_4(h)}{Q(1)}\right)$$

Similarly, for $q_h^{(2)}$ we obtain

$$h^{1-\alpha} q_h^{(2)} = -h^{-\gamma} a_3(h) \int_0^1 u^{-\gamma} d\left(\frac{d}{du} [u^{\alpha+1} K(u)]\right) + O\left(\frac{h^{-\gamma} b_4(h)}{Q(1)}\right)$$

$$= \gamma(\gamma + 1) h^{-\gamma} a_3(h) \int_0^1 u^{\alpha-\gamma-1} K(u) du + O\left(\frac{h^{-\gamma} b_4(h)}{Q(1)}\right)$$

as h tends to 0. Combining things then yields that

$$\gamma_h = \gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 = \gamma + O(h^{-\gamma} a_3(h)) + O(h^{-\gamma} b_4(h)) + O\left(\frac{b_4(h)}{a_3(h)}\right)$$

In order to have an asymptotically vanishing bias we thus have to impose the conditions on $h = h_n$ that

$$nh \left(\frac{b_4(h)}{a_3(h)}\right)^2 = nh^{1-2\gamma} \left(\frac{b_4(h)}{\log Q(1) - \log Q(1 - h)}\right)^2 \rightarrow 0 \tag{5.25}$$

and

$$nh^{1-2\gamma} (a_3(h))^2 = nh (\log Q(1) - \log Q(1 - h))^2 \rightarrow 0$$

as n tends to infinity. Note that these two conditions imply that $\sqrt{nh} h^{-\gamma} b_4(h) \rightarrow 0$ as $n \rightarrow \infty$ as well. Moreover, condition (5.25) resembles the condition on the parameter k in case of the moment estimator as defined in DEKKERS, EINMAHL AND DE HAAN (1989). (See also Theorem 2.9 of this thesis.)

The case $\gamma = 0$

Similar arguments, using Lemma 5.10 and Lemma 5.11, now yield the asymptotic relations

$$\int_0^1 \log Q(1-hu) d(u^\alpha K(u)) = b_2(h) \int_0^1 (-\log u) d(u^\alpha K(u)) + O(b_3(h))$$

and

$$h^{1-\alpha} q_h^{(2)} = b_2(h) \int_0^1 (-\log u) d\left(\frac{d}{du} [u^{\alpha+1} K(u)]\right) + O(b_3(h))$$

as h tends to 0. Combining these expressions we obtain

$$\gamma_h = \gamma_h^{(pos)} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1 = b_2(h) + O(b_3(h)) + O\left(\frac{b_3(h)}{b_2(h)}\right)$$

In order to have an asymptotically vanishing bias we thus have to impose the conditions on $h = h_n$ that

$$nh(b_2(h))^2 \rightarrow 0 \quad \text{and} \quad nh\left(\frac{b_3(h)}{b_2(h)}\right)^2 \rightarrow 0$$

as n tends to infinity. Note that these two conditions imply that $\sqrt{nh}b_3(h) \rightarrow 0$ as n tends to infinity as well.

Chapter 6

A simulation study

For each of the estimators of the extreme value index as presented in this thesis, the number of upper order statistics that is used, is in some sense reflected in the behaviour of the estimator. Using too many observations will result in a bias, since then the smaller observations, typically not falling in the upper tail of the underlying distribution function, will be included in the calculation. Using too few order statistics on the other hand, will obviously result in a substantial variance.

In this chapter, we will present the results of a simulation study, concerning most of the estimators mentioned in this thesis, in order to describe their finite sample behaviour as well as the above mentioned dependence on the number of used upper order statistics. We will apply the estimators on samples generated from several different distributions with sample sizes 100 and 1000. Moreover, we will present the behaviour of the estimators on a data set consisting of 211 measurements on water discharges at Lobith, The Netherlands, in the period 1901–1991. We will refer to this data set by ‘the Lobith data’.

6.1 Introduction

We will be concerned with a sample X_1, \dots, X_n of i.i.d. random variables with common distribution function F that is in the domain of attraction of an extreme value distribution G_γ , for some $\gamma \in \mathbb{R}$. The ascending order statistics will be denoted by $X_{(1)} \leq \dots \leq X_{(n)}$. Moreover, we will assume that the upper endpoint of the distribution x_F^o is not less than 1. This can always be achieved by shifting the support of the distribution function, a transformation that does not change the extreme value index.

The estimators of the extreme value index γ as mentioned in the previous chapters, are not all applicable to the situation of estimating any real γ . E.g., the Hill estimator will only give a reasonable estimate in case of a positive extreme value index, whereas Pickands’ estimator can be applied for any real γ .

In case one knows the sign of the extreme value index, it might be more appropriate to choose an estimator best suited for that particular range of extreme value indices. An alternative approach to using an estimator applicable for all real γ , might thus be to first

test for the sign of the index and then use an estimator best suited for that situation. We will not deal with that approach in this thesis. For more information on tests for the domain of attraction of extreme value distributions, we refer to HASOFER AND WANG (1992) and FRAGA ALVES AND GOMES (1996).

In the next section, the estimators of the extreme value index we have introduced in this thesis, are presented again for easy reference. For more details on these estimators we refer to the appropriate chapters. In the subsequent sections we will discuss several distribution functions and show the average behaviour of the estimators. For each of these distribution functions, we generated samples of size 100 and 1000. A number of 1000 samples was generated for each of these sample sizes, in order to examine the mean behaviour of the estimators as well as the Mean Squared Error of the estimators.

The Random Number Generator we used was not a linear congruential generator but a 'subtract-with-borrow' generator (see MARSAGLIA AND ZAMAN (1991)) which has an extremely long period. All the quantile functions of the distribution functions we will be using are known explicitly, hence the samples can be generated by applying the quantile functions on a sample of standard uniform variables. To increase computational speed, the following sequential method to generate an ordered sample was used. Let U_1, \dots, U_n be a standard uniform sample of size n and define $T_n = U_1^{1/n}$ and $T_{n-i} = T_{n-i+1} U_{i+1}^{1/(n-i)}$ for $i = 1, \dots, n$. Then $T_1 \leq \dots \leq T_n$ and they are distributed as the order statistics of a sample of size n from a $\mathcal{U}(0, 1)$ distribution.

Moreover, some estimators will be applied to a data set containing 211 measurements on water discharges at Lobith, The Netherlands, in the period 1901–1991. Since this is a single data set it will reveal the major advantage of the kernel type estimators as compared to the other estimators.

Finally, in the last section, we will discuss all the results in some more detail.

6.2 Summary of the estimators

Percentile approach

Let $M = M_n$ be a sequence of integers with $M \leq \lfloor n/4 \rfloor$ and $M_n/n \rightarrow 0$ as $n \rightarrow \infty$. In case of a Generalized Pareto Distribution, observe that

$$\gamma = \log \frac{Q_P(3/4; \gamma, \sigma) - Q_P(1/2; \gamma, \sigma)}{Q_P(1/2; \gamma, \sigma)} / \log 2 \quad (6.1)$$

where $Q_P(\cdot; \gamma, \sigma)$ is the quantile function of a Generalized Pareto Distribution. Define Pickands' estimator as the sample analogue of (6.1) assuming that the excesses $X_{(n-i+1)} - X_{(n-4M+1)}$ stem from a Generalized Pareto Distribution, i.e.,

$$\hat{\gamma}_{n,M}^P = \log \left(\frac{X_{(n-M+1)} - X_{(n-2M+1)}}{X_{(n-2M+1)} - X_{(n-4M+1)}} \right) / \log 2$$

This estimator is applicable for all $\gamma \in \mathbb{R}$.

Maximum Likelihood approach

Define the excesses over threshold $u = u_n$ as $Y_i = X_j - u$ where j is the index of the i -th exceedence and $u_n \rightarrow x_p^*$. Assuming that these excesses come from a Generalized Pareto Distribution with parameters γ and $\sigma(u)$, the estimator is defined by maximizing the likelihood of Y_1, \dots, Y_N based on that parametric assumption and denoted by $\hat{\gamma}_N^3$, where N denotes the number of excesses over threshold u .

Without modification, asymptotic normality of this estimator can be derived for all $\gamma > -1/2$. Moreover, the procedure can be extended to include the range $(-1, -1/2]$ of extreme value indices.

In the simulations, the estimator was not calculated by equating the score functions with zero, but by maximizing the likelihood directly, using a simple maximization method called 'amoeba', as given in Numerical Recipes (PRESS ET AL. (1992)). As a starting value for this iterative procedure, we used the moment estimator, to be defined shortly. Using this approach, we did not encounter any problems concerning the convergence of the numerical procedure in calculating the estimator.

Moment approach

Let k be a sequence of integers tending to infinity, with $k < n$ and $k/n \rightarrow 0$. Define the quantities $M_{n,k}^{(r)}$ for $r = 1, 2$ by

$$M_{n,k}^{(r)} = \frac{1}{k} \sum_{i=1}^k \left(\log X_{(n-i+1)} - \log X_{(n-k)} \right)^r$$

The Hill estimator is then defined by

$$\hat{\gamma}_{n,k}^H = M_{n,k}^{(1)}$$

and the moment estimator by

$$\hat{\gamma}_{n,k}^M = M_{n,k}^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_{n,k}^{(1)})^2}{M_{n,k}^{(2)}} \right)^{-1}$$

The Hill estimator is applicable only when $\gamma > 0$, whereas the moment estimator is applicable for all $\gamma \in \mathbb{R}$.

Kernel type approach

Let $K(\cdot)$ be a kernel with support $[0, 1]$, i.e., $K(x) = 0$ whenever $x \notin [0, 1]$, and let $h = h_n$ be a sequence with $h \rightarrow 0$ and $nh \rightarrow \infty$. The first kernel type estimator is then defined by

$$\hat{\gamma}_{n,h}^K = \sum_{i=1}^{n-1} \frac{i}{n} K_h \left(\frac{i}{n} \right) \left(\log X_{(n-i+1)} - \log X_{(n-i)} \right)$$

where $K_h(x) = K(x/h)/h$. This estimator is applicable only for $\gamma > 0$.

For a more general kernel type estimator, i.e., a kernel type estimator applicable to any real valued extreme value index, additionally define the quantities

$$\hat{q}_{n,h}^{(1)} = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^\alpha K_h\left(\frac{i}{n}\right) (\log X_{(n-i+1)} - \log X_{(n-i)})$$

and

$$\hat{q}_{n,h}^{(2)} = \sum_{i=1}^{n-1} \left[\frac{d}{du} u^{\alpha+1} K_h(u) \right]_{u=i/n} (\log X_{(n-i+1)} - \log X_{(n-i)})$$

for any $\alpha > 0$. Define the general kernel type estimator by

$$\hat{\gamma}_{n,h}^G = \hat{\gamma}_{n,h}^K + \frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}} - 1$$

In the simulations we will use the kernel

$$K(x) = \begin{cases} \frac{315}{128}(1-x^2)^4 & x \in [0, 1] \\ 0 & \text{elsewhere} \end{cases} \tag{6.2}$$

and we will take $\alpha = 0.6$ in the definitions of $\hat{q}_{n,h}^{(1)}$ and $\hat{q}_{n,h}^{(2)}$, in order to ensure asymptotic normality of the general kernel type estimator, see Theorem 5.2. Moreover, note that this kernel satisfies the conditions of Theorem 3.1 as well as those of Theorem 5.2.

6.3 Finite sample behaviour

In this section the finite sample behaviour of the estimators is shown for several distribution functions. In order to be able to compare the estimators, each will be plotted as a function of the fraction of order statistics that is used in their calculation. I.e., if h denotes that fraction, the following estimators will be plotted: $\hat{\gamma}^P(h) = \hat{\gamma}_{n, \lfloor nh/4 \rfloor}^P$, $\hat{\gamma}^H(h) = \hat{\gamma}_{n, \lfloor nh \rfloor}^H$, $\hat{\gamma}^M(h) = \hat{\gamma}_{n, \lfloor nh \rfloor}^M$, $\hat{\gamma}^K(h) = \hat{\gamma}_{n,h}^K$, $\hat{\gamma}^G(h) = \hat{\gamma}_{n,h}^G$ and $\hat{\gamma}^S(h) = \hat{\gamma}_N^S$ with threshold $u_N = X_{(n-\lfloor nh \rfloor)}$. Moreover, in the plots each line corresponding to a particular estimator is labelled by the character in the superscript of the estimator. E.g., a line representing the maximum likelihood estimator $\hat{\gamma}^S(h)$ is indicated by the character S .

In each of the following subsections the mean behaviour of the various estimators will be shown for a specific type of distribution function by plotting the average of each estimator over 1000 samples. Following these subsections on artificially generated samples, we will apply the estimators to a single real-life dataset called the Lobith data. In Section 6.4 we will discuss the behaviour of the various estimators.

6.3.1 The Hall model

A whole class of distribution functions with positive extreme value index γ that is frequently used when studying extreme value distributions is given in e.g., HALL AND WELSH (1984) and is sometimes referred to by 'the Hall model'. The distribution functions of that class are defined to satisfy

$$F(x) \sim 1 - Cx^{-1/\gamma}(1 + Dx^{-\rho}) \quad \text{as } x \rightarrow \infty \quad (6.3)$$

with $C, \gamma, \rho > 0$ and $D \neq 0$. The quantile function hence satisfies

$$Q(s) \sim C^\gamma(1-s)^{-\gamma} (1 + \gamma DC^{-\rho\gamma}(1-s)^{\rho\gamma}) \quad \text{as } s \rightarrow 1$$

and the associated ϕ -function

$$\phi(s) \sim \frac{\gamma(C^{\rho\gamma} + \gamma(1-\rho)Ds^{\rho\gamma})}{C^{\rho\gamma} + \gamma Ds^{\rho\gamma}} \sim \gamma(1 + O(s^{\rho\gamma})) \quad \text{as } s \rightarrow 0$$

6.3.1.1 Special case: Cauchy

A well known distribution function that is contained in this class, is the Cauchy distribution with the extreme value index $\gamma = 1$:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x = 1 - \frac{1}{\pi} \left(\frac{1}{x} - \frac{1}{3x^3} + o(x^{-3}) \right) \quad \text{as } x \rightarrow \infty$$

Its quantile function is given by

$$\begin{aligned} Q(s) &= \cot(\pi(1-s)) \\ &= \frac{1}{\pi(1-s)} \left(1 - \frac{1}{3}(\pi(1-s))^2 + o((1-s)^2) \right) \quad \text{as } s \rightarrow 1 \end{aligned}$$

and the associated ϕ -function by

$$\phi(s) = \frac{\pi s}{\sin(\pi s) \cos(\pi s)} = 1 + \frac{2}{3}(\pi s)^2 + o(s^2) \quad \text{as } s \rightarrow 0$$

Since the extreme value index of the Cauchy distribution is positive, we can use all the mentioned estimators.

6.3.1.2 Another case of the Hall model

Taking equality in (6.3) and substituting $\gamma = 1/3$, $\rho = 3$, $C = 2$ and $D = -1/2$ yields

$$F(x) = 1 - 2x^{-3} \left(1 - \frac{1}{2}x^{-3} \right) \quad x \geq 1$$

$$Q(s) = (1 - \sqrt{s})^{-1/3}$$

$$\phi(s) = \frac{1}{6} \frac{s}{\sqrt{1-s}(1 - \sqrt{1-s})} = \frac{1}{3} + \frac{1}{12}s + o(s) \quad \text{as } s \rightarrow 0$$

I.e., for these particular choices of the parameters, the quantile function can be calculated explicitly and hence used to generate samples from this distribution. Moreover, since $\gamma = 1/3$, all estimators still apply.

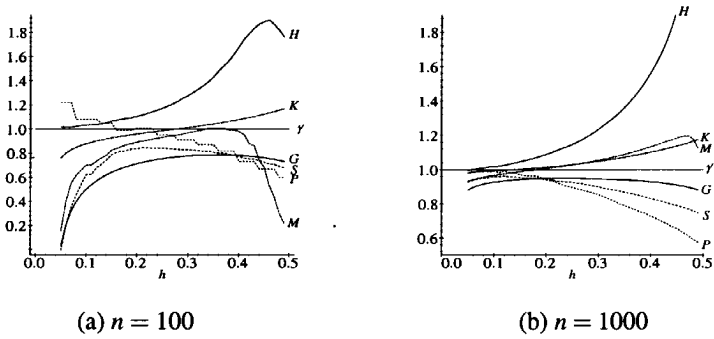


Figure 6.3.1: Cauchy, $\gamma = 1$

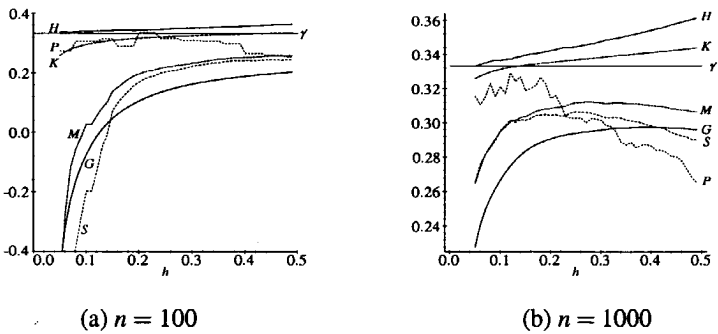


Figure 6.3.2: Hall model, $\gamma = 1/3$

6.3.2 Generalized Extreme Value distributions

The Generalized Extreme Value distribution was introduced in Chapter 1 of this thesis as a way of describing the three possible limiting types of distribution functions, appearing in the extreme value setting, in one single formula. The distribution function is given by

$$F(x) = \exp\left(- (1 + \gamma x)^{-1/\gamma}\right) \quad \text{for all } x \text{ such that } 1 + \gamma x > 0$$

where γ is obviously the extreme value index of the distribution. Its quantile function is given by

$$Q(s) = - \frac{(1 - (-\log s))^{-\gamma}}{\gamma}$$

and the associated ϕ -function is hence given by

$$\phi(s) = - \frac{s}{(1-s)\log(1-s)} \frac{\gamma}{(1 - (-\log(1-s))^\gamma)}$$

Moreover, as $s \rightarrow 0$,

$$\phi(s) = \begin{cases} -\gamma s^{-\gamma} (1 + s^{-\gamma} + o(s^{-\gamma})) & -1 < \gamma < 0 \\ -\frac{1}{\log s} (1 + o(1)) & \gamma = 0 \\ \gamma (1 + s^\gamma + o(s^\gamma)) & 0 < \gamma < 1 \end{cases}$$

In practice, often the extreme value index is close to 0. Therefore we will consider three instances of the Generalized Extreme Value distribution: $GEV(-0.1)$, $GEV(0)$ and $GEV(0.1)$. Obviously, in case of $GEV(-0.1)$, the Hill estimator $\hat{\gamma}^H(h)$ and the kernel estimator $\hat{\gamma}^K(h)$ are no longer applicable. In the other two instances all estimators can be applied.

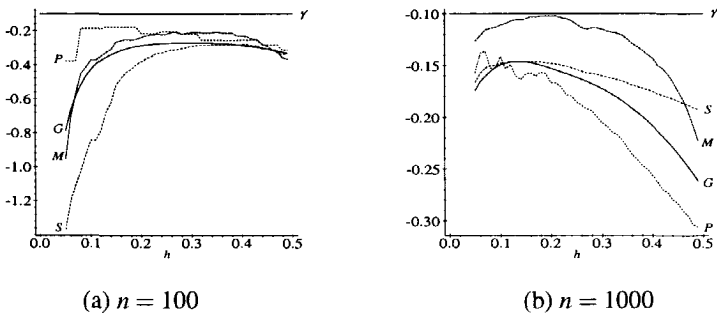


Figure 6.3.3: GEV, $\gamma = -0.1$

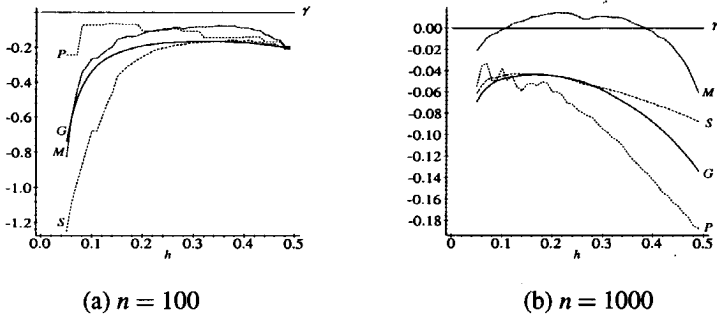


Figure 6.3.4: GEV, $\gamma = 0$

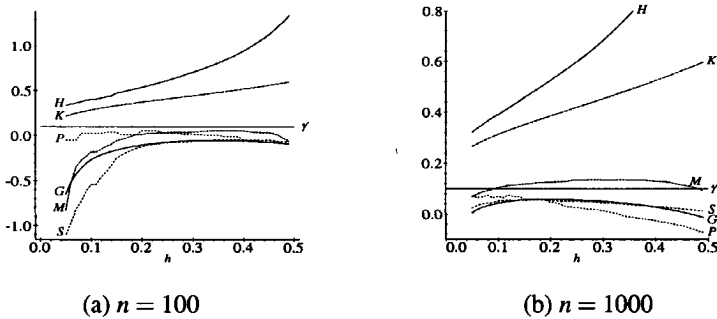


Figure 6.3.5: GEV, $\gamma = 0.1$

6.3.3 Uniform distribution

A typical example of a distribution with a finite upper endpoint is the uniform distribution $\mathcal{U}(a, b)$, defined by

$$F(x) = \frac{x - a}{b - a} \quad a \leq x \leq b$$

with extreme value index $\gamma = -1$. The quantile function is given by

$$Q(s) = a + (b - a)s$$

and the associated ϕ -function by

$$\phi(s) = \frac{(b - a)s}{b - (b - a)s} = \left(1 - \frac{a}{b}\right)s + \left(1 - \frac{a}{b}\right)^2 s^2 + o(s^2) \quad \text{as } s \rightarrow 0$$

In the simulations we will use the uniform distribution with $a = 2$ and $b = 5$, i.e., $\mathcal{U}(2, 5)$. Again, since $\gamma = -1$, $\hat{\gamma}^H(h)$ and $\hat{\gamma}^K(h)$ are not applicable. Moreover, asymptotic normality of the maximum likelihood estimator $\hat{\gamma}^S(h)$ is not covered by Theorem 2.6. However, we did calculate this estimator, just to show its finite sample behaviour in this case as well.

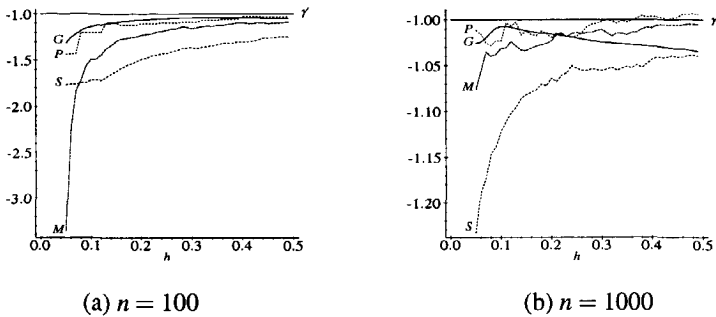


Figure 6.3.6: Uniform on (2,5), $\gamma = -1$

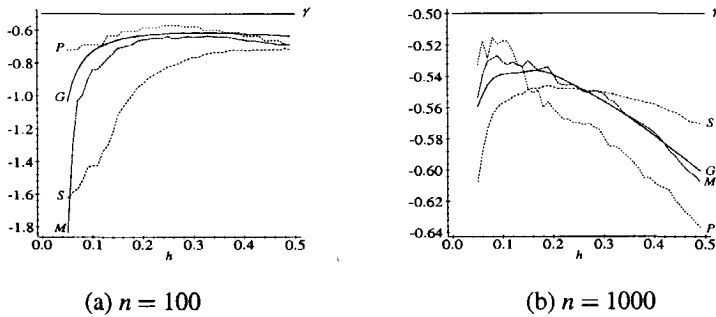


Figure 6.3.7: Hall type, $\gamma = -0.5$

6.3.4 Hall type for negative index

The Hall model, as used before for positive extreme value index, can be transformed to a similar model for negative extreme value indices. Again we will choose the parameters in such a way that the quantile function is explicitly given and can easily be used to generate samples. To be more specific, we will use

$$F(x) = 1 - 2(1-x)^2 + (1-x)^4 \quad 0 < x < 1$$

In this case, the extreme value index γ equals $-1/2$ and the quantile function is given by

$$Q(s) = 1 - \sqrt{1 - \sqrt{s}}$$

and the associated ϕ -function by

$$\phi(s) = \frac{1}{4} \frac{s}{\sqrt{1-s}(\sqrt{1-\sqrt{1-s}})(1-\sqrt{1-\sqrt{1-s}})} = \frac{\sqrt{2}}{4} s^{1/2} + \frac{1}{4} s + o(s)$$

as $s \rightarrow 0$.

6.3.5 Generalized Pareto

The Generalized Pareto Distribution is the limiting form of the properly scaled tale of a distribution that is in the domain of attraction of an extreme value distribution. See e.g., Chapter 1 of this thesis for a discussion of the approximation of the tail of a distribution. The distribution function for $\gamma \in \mathbb{R}$ and $\sigma > 0$ is given by

$$F(x) = 1 - \left(1 + \frac{\gamma}{\sigma}x\right)^{-1/\gamma}$$

with $x > 0$ and $1 + \gamma x/\sigma > 0$. In case of $\gamma = 0$ the distribution is defined as the limit over γ tending to 0, i.e., in that case $F(x) = 1 - \exp(-x/\sigma)$ for all $x > 0$. Note that the parameter γ is the extreme value index of the distribution.

The corresponding quantile function is given by

$$Q(s) = \frac{\sigma}{\gamma} \left((1-s)^{-\gamma} - 1 \right)$$

for $\gamma \neq 0$ and by $Q(s) = -(\log(1-s))/\sigma$ for $\gamma = 0$. The associated ϕ -function is given by

$$\phi(s) = \frac{\gamma}{1-s^\gamma}$$

for $\gamma \neq 0$ and by $\phi(s) = -1/\log s$ in case $\gamma = 0$. Note that, as s tends to zero,

$$\phi(s) = \begin{cases} \gamma(1 + O(s^{-\gamma})) & \gamma > 0 \\ -\gamma s^{-\gamma}(1 + O(s^{-\gamma})) & \gamma < 0 \end{cases}$$

In the simulations we took $\gamma = 0.1$ and $\sigma = 2$.

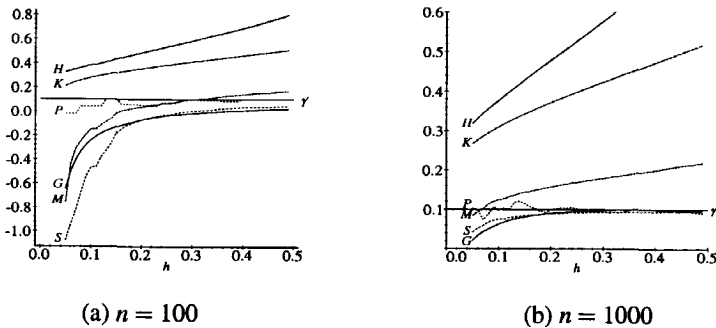


Figure 6.3.8: GPD, $\gamma = 0.1$ and $\sigma = 2$

6.3.6 The Lobith data

The Lobith data were obtained during the period 1901–1991 at a municipality in the Netherlands, called Lobith. The data represent peaks in the water discharges at that particular place along the river Rhine. During the mentioned period of time, the maximum water discharge at Lobith was measured on a daily basis. These maxima were plotted against time, and only those maxima above a certain threshold and at least a fortnight apart were recorded in the Lobith data. Whenever several values appeared above the threshold but within a fortnight of each other, the maximum of those values was recorded. This resulted in 211 measurements.

Considering the above mentioned construction of the data set, it seems reasonable to assume that the measurements represent an i.i.d. sample from a distribution that is in the domain of attraction of an extreme value distribution. The same assumption was made in GROENEBOOM (1993) where this data set was analysed as well. The results from that paper were used by the Dutch government in the ‘evaluation of the underlying assumptions for river dike enforcements’.

The moment estimator $\hat{\gamma}_{n,k}^M$ as well as the general kernel type estimator $\hat{\gamma}_{n,h}^G$ suggest that the extreme value index of the distribution underlying the Lobith data is non-positive. Therefore, only the estimators applicable to estimate negative extreme value indices as well were applied to this data set. The results are plotted in Figure 6.3.9. In case of the general kernel type estimator, the parameter α was taken to be 0.6 and the kernel as defined in (6.2).

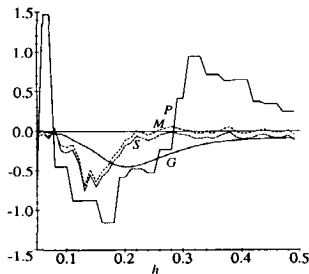


Figure 6.3.9: Estimates of the extreme value index of the Lobith data

6.4 Discussion of the results

With exception of the plot concerning the Lobith data, the plots presented in the previous subsections show for each estimator the average of 1000 estimates. Therefore most estimators, as a function of the parameter h , behave rather smoothly. However, the plot based on the Lobith data clearly shows that only the general kernel type estimator $\hat{\gamma}^G(h)$

is a smooth function of the parameter h when only one sample is considered. This is due to the fact that the other estimators, the moment estimator, the maximum likelihood estimator and Pickands' estimator, all depend on the number of order statistics in a discrete way. Since we are dealing with (intermediate) order statistics, the addition of a single order statistic can change the value of these estimates considerably: each order statistic is equally weighted in these estimators. In case of the general kernel type estimator, the order statistics are weighted using a smooth kernel that decreases to 0 with decreasing rank of the order statistics. I.e., using an additional intermediate order statistic does not change the value of the estimate too much since it enters the estimate with a rather small weight.

The same advantage will be present when comparing the Hill estimator and the kernel type estimator for positive γ on a single sample: the Hill estimator again depends discretely on the number of used order statistics, whereas the kernel type estimator smoothes out the effect of adding an extra order statistic.

As discussed before, not all estimators are applicable to each situation: the Hill estimator and the kernel type estimator can only be used in case of estimating a positive extreme value index. In case the extreme value index is not too close to zero, the kernel type estimator behaves quite well as far as the bias is concerned, see Figures 6.3.1 and 6.3.2. However, as soon as the extreme value index is close to zero, both the Hill estimator and the kernel type estimator get seriously biased as seen in Figures 6.3.5 and 6.3.8.

The generally applicable estimators behave moderately to very well for each of the considered distributions, as far as the bias is concerned. However, it is more interesting to consider the Mean Squared Error. Since we generated 1000 samples for each of the considered distributions, the bias and the variance can be estimated using the estimates for these samples along with the knowledge about the parameter that was estimated. Figures 6.4.10 to 6.4.12 show the estimated Mean Squared Error for the estimators in case of the Cauchy distribution, the Generalized Pareto Distribution with $\gamma = 0.1$ and $\sigma = 2$ and the Uniform (2,5) distribution respectively. Even though Pickands' esti-

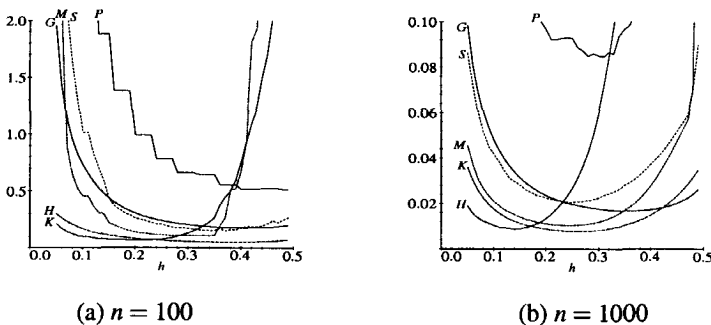


Figure 6.4.10: Estimated Mean Squared Error in case of Cauchy data

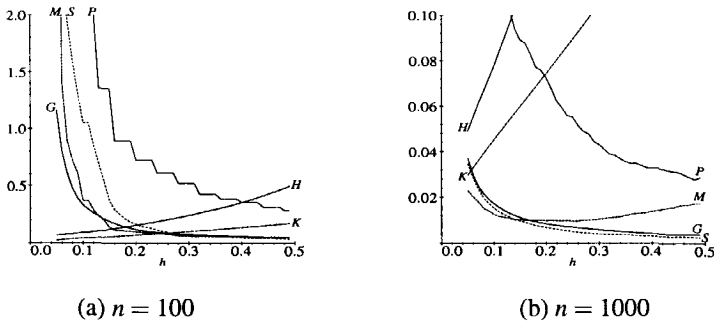
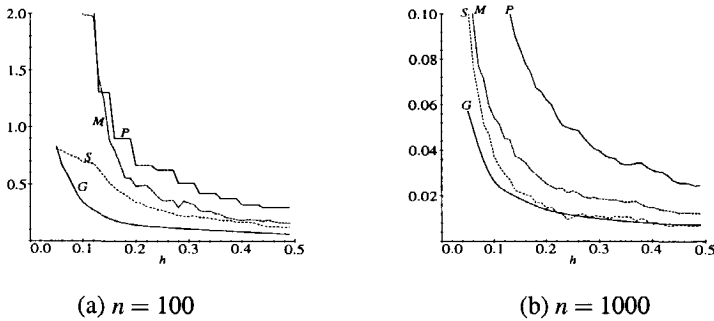


Figure 6.4.11: Estimated Mean Squared Error in case of GPD(0.1, 2) data

Figure 6.4.12: Estimated Mean Squared Error in case of $\mathcal{U}(2, 5)$ data

mator behaves quite well considering its bias, in Mean Squared Error sense it behaves very poorly: the variance of this estimator is quite large. The other estimators behave moderately to very well in this Mean Squared Error sense in most cases.

Figures 6.4.10 and 6.4.11 however show that the Mean Squared Error behaviour of the Hill estimator and the kernel type estimator for positive γ again depends on the value of γ that is to be estimated: in case of the Generalized Pareto data ($\gamma = 0.1$), these estimators behave quite poorly, whereas in case of the Cauchy data they outperform the other estimators in Mean Squared Error sense. This shows that even in case the extreme value index that is to be estimated can be assumed to be positive, the use of generally applicable estimators can yield better estimates, especially if the index is assumed to be close to zero.

Considering all simulations, it seems that the average behaviour of the general kernel type estimator is comparable to the behaviour of the more often used moment estimator, both in bias and in Mean Squared Error sense. The plots of the Mean Squared Error indicate that the use of near-optimal bandwidths would still result in a reasonable estimate. Moreover, the smoothness of the estimator $\hat{\gamma}^G(h)$ itself then yields that the

estimate is likely to be close to the estimate that one would have gotten if one had used the optimal bandwidth. This feature is not apparent in the other generally applicable estimators: even though in Mean Squared Error sense these estimators also seem to behave constantly well in a neighbourhood of the optimal bandwidth, the wobbly behaviour of these estimators as a function of the fraction of used order statistics, does not yield comparable estimates in that same neighbourhood. Indeed, the use of a near-optimal fraction of order statistics for these estimators could lead to an estimate that differs substantially from the estimate using the optimal fraction of order statistics.

Even though the general kernel type estimator as considered in the present simulation study already behaves very well, the effect of choosing different values of α or using different kernels still needs to be investigated more thoroughly. Indeed, the general kernel type estimator as introduced in Chapter 5 actually represents a whole class of new estimators.

Bibliography

- BALKEMA, A.A. AND DE HAAN, L. (1972). On R. von Mises' condition for the domain of attraction of $\exp(-\exp(-x))$. *Annals of Mathematical Statistics*, **43**, 1352–1354.
- BALKEMA, A.A. AND DE HAAN, L. (1974). Residual lifetime at great age. *Annals of Probability*, **2**, 792–804.
- BILLINGSLEY, P. (1968). *Convergence of probability measures*. Wiley, New York.
- BINGHAM, N.H., GOLDIE, C.M. AND TEUGELS, J.L. (1987). *Regular variation*. (Encyclopedia of mathematics and its applications; 27) Cambridge University Press, Cambridge.
- CSÖRGŐ, S. (1984). Adaptive estimation of the parameter of stable laws. *Colloq. Math. Soc. J. Bolyai* **36** (*Limit Theorems in Probability and Statistics, Vol I.*), ed. P. Révész 305–368 North Holland.
- CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. AND MASON, D. (1986). Weighted empirical and quantile processes. *Annals of Probability*, **14**, 31–85.
- CSÖRGŐ, S., DEHEUVELS, P. AND MASON, D. (1985). Kernel estimates of the tail index of a distribution. *Annals of Statistics*, **13**, 1050–1077.
- DANIELSSON, J., DE HAAN, L., PENG, L. AND DE VRIES, C. (1997). *Using a bootstrap method to choose the sample fraction in tail index estimation*. Discussion paper TI 97–016/4, Tinbergen Institute, Rotterdam.
- DEHEUVELS, P., HAEUSLER, E. AND MASON, D.M. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, **104**, 371–381.
- DEKKERS, A.L.M. (1991). *On Extreme-Value Estimation*. Ph.D.-thesis, Erasmus University Rotterdam.
- DEKKERS, A.L.M., EINMAHL, J.H.J. AND DE HAAN, L. (1989). A moment estimator for the index of an extreme value distribution. *Annals of Statistics*, **17**, 1833–1855.
- DEKKERS, A.L.M. AND DE HAAN, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics*, **17**, 1795–1832.
- DEKKERS, A.L.M. AND DE HAAN, L. (1991). Optimal choice of sample fraction in extreme-value estimation. Submitted.

- DONOHU, D.L. AND LIU, R.C. (1987). Geometrizing rates of convergence I. Technical Report 137a, Department of Statistics, University of California, Berkeley.
- DONOHU, D.L. AND LIU, R.C. (1991). Geometrizing rates of convergence II. *Annals of Statistics*, **19**, 633–667.
- DREES, H. (1995). *Optimal rates of convergence for estimates of the extreme value index*. Report no. 9557/A, Econometric Institute, Erasmus University Rotterdam.
- EINMAHL, J.H.J. (1994). *Empirical and related processes with applications*. AIO-Course, Academic Year 1993–1994, Utrecht.
- FISHER, R.A. AND TIPPETT, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.
- FRAGA ALVES, M.I. AND GOMES, M.I. (1996). Statistical choice of extreme value domains of attraction - a comparative analysis. *Communications in Statistics, Part A: Theory and Methods*, **25(4)**, 789–811.
- GELUK, J.L. AND DE HAAN, L. (1987). *Regular variation, extensions and Tauberian theorems*. Mathematical Centre Tract 40, Centre for Mathematics and Computer Science, Amsterdam.
- GNEDENKO, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, **44**, 423–453.
- GROENEBOOM, P. (1993). *Peaks of water discharges at Lobith*. in: Government report 'Evaluation of underlying assumptions for river dike fortifications', January 1993, Appendix A.
- GRÜBEL, R.J. AND DE WOLF, P.P. (1994). *Estimation of the tail index of a distribution*. Technical report 94-79, Faculty of Mathematics, Delft University of Technology.
- GUMBEL, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York.
- DE HAAN, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Mathematical Centre Tract 32, Centre for Mathematics and Computer Science, Amsterdam.
- DE HAAN, L. (1976). Sample extremes: an elementary introduction. *Statistica Neerlandica*, **30**, 161–172.
- DE HAAN, L. (1984). Slow variation and characterization of domains of attraction. *Statistical Extremes and Applications*, Ed. J. Tiago de Oliveira, 31–48 Reidel, Dordrecht.
- DE HAAN, L., PENG, L. AND PEREIRA, T. (1997). *A bootstrap-based method to achieve optimality in estimating the extreme-value index*. Discussion paper TI 97-099/4, Tinbergen Institute, Rotterdam.
- DE HAAN, L. AND PEREIRA, T. (1999). Estimating the index of a stable distribution. *Statistics and Probability Letters*, **41**, 39–55.

- HAEUSLER, E. AND TEUGELS, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *Annals of Statistics*, **13**, 743–756.
- HALL, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameters in nonparametric problems. *Journal of Multivariate Analysis*, **32**, 177–203.
- HALL, P. AND WELSH, A.H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Annals of Statistics*, **12**, 1079–1084.
- HALL, P. AND WELSH, A.H. (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics*, **13**, 331–341.
- HASOFER, A.M. AND WANG, Z. (1992). A test for extreme value domain of attraction. *Journal of the American Statistical Association*, **87**, 171–177.
- HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**, 1163–1174.
- JONGBLOED, G. (1995). *Three Statistical Inverse Problems, estimators-algorithms-asymptotics*. PhD. thesis, Faculty of Mathematics, Delft University of Technology.
- LEADBETTER, M.R., LINDGREN, G. AND ROOTZÉN, H. (1983). *Extremes and related properties of random sequences and processes*. Springer, New York.
- MARSAGLIA, G. AND ZAMAN, A. (1991). A new class of random number generators. *Annals of Applied Probability*, **1**, 462–480.
- MASON, D.M. (1982a). Laws of large numbers for sums of extreme values. *Annals of Probability*, **10**, 754–764.
- MASON, D.M. (1982b). Some characterizations of almost sure bounds for weighted multidimensional empirical distributions and a Glivenko–Cantelli theorem for sample quantiles. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **59**, 505–513.
- VON MISES, R. (1936). La distribution de la plus grande de n valeurs. *Selected Papers II (American Mathematical Society)*, 271–294.
- OMEY, E. AND WILLEKENS, E. (1987). Π -variation with remainder. *Journal of the London Mathematical Society (2)*, **37**, 105–118.
- PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, **3**, 119–131.
- PICKANDS, J. (1986). The continuous and differentiable domains of attraction of the extreme value distributions. *Annals of Probability*, **14**, 996–1004.
- POLLARD, D. (1990). *Empirical processes: theory and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2.
- PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T. AND FLANNERY, B.P. (1992). *Numerical Recipes in C*. Cambridge University Press.

- PYKE, R. (1965). Spacings. *Journal of the Royal Statistical Society*, series B **27**, 395–436.
- RESNICK, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
- RESNICK, S.I. AND STĂRICA, C. (1995). *Smoothing the Hill estimator*. Technical report No. 1114, Cornell University, Ithaca, New York.
- RICE, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics*, **12**, 1215–1230.
- SENETA, E. (1976). *Regularly Varying Functions*. Lecture Notes in Mathematics, 508, Springer, Berlin.
- SMITH, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **72**, 67–92.
- SMITH, R.L. (1987). Estimating tails of probability distributions. *Annals of Statistics*, **15**, 1174–1207.
- SMITH, R.L. (1988). Approximations in extreme value theory. Unpublished manuscript.
- WELLNER, J.A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **45**, 73–88.
- DE WOLF, P.P. (1991). *On the estimation of the index of variation*. Masters thesis, Faculty of Technical Mathematics and Informatics, Delft University of Technology (appeared as STUDREP.91.116, KSEPL, Shell, Rijswijk).

Summary

Estimating the extreme value index

– tales of tails –

For many different phenomena it is interesting to study the asymptotics of the average behaviour. This often leads to an application of some form of the Central Limit Theorem. In other cases, it is the behaviour of extreme events that is of interest. Indeed, in many different areas, there is a dire need to describe extreme situations in order to try to prevent or at least reduce severe damage, financial or otherwise. E.g., accurate description of extreme wave heights can be used to design better strategies that aim to protect inhabited areas against flooding.

Extreme value theory provides, among other things, an analogy of the Central Limit Theorem in case of extreme events. The distribution of the properly scaled and shifted sample maximum will, in most cases, converge to an Extreme Value Distribution. However, contrary to the situation of the Central Limit Theorem, where the Normal distribution always turns up as the limiting distribution, three different limiting types of Extreme Value Distributions exist: the Gumbel, the Fréchet and the (inverse) Weibull distribution. Even though these distributions behave quite differently, they can be combined into one formula containing a single parameter, called the extreme value index, that identifies each type. This thesis focuses on the estimation of that extreme value index. An accurate estimate of the extreme value index is often needed to be able to analyse extreme events in more detail.

The first chapter of this thesis contains additional general information on extreme events, extreme value theory and related topics.

Over the years, several attempts have been made to obtain a measure of the difficulty of estimating the extreme value index. A frequently used approach in assessing the difficulty of estimation procedures is to consider the minimax risk of such a procedure over certain classes of distribution functions. At the start of Chapter 2, we provide an asymptotic lower bound to the minimax risk of estimating the extreme value index over a large class of distribution functions in a rather straightforward way.

Following on the derivation of that lower bound, several existing estimators of the extreme value index are discussed, like the Hill estimator, Pickands' estimator, the Moment estimator and a kernel type estimator. Moreover, considering that kernel type estimator of the extreme value index as a weighted average of Hill estimators, it is argued that the asymptotic behaviour of a recently introduced estimator could more easily be derived using results concerning the asymptotic behaviour of that kernel type estimator.

At the end of the chapter it is shown that the lower bound to the minimax risk can be attained in the sense of convergence in distribution for several classes of distribution functions. This is explicitly shown in case of Pickands' estimator, but other estimators attain the lower bound as well.

The asymptotic behaviour of the just mentioned kernel type estimator is discussed in more detail in Chapter 3. Just like any other estimator of the extreme value index, the calculation of the estimate is done using a certain number of the largest values in the sample (the upper order statistics). That number of used order statistics, relative to the sample size, is in case of the kernel type estimator determined by a parameter called the bandwidth. Considering the estimator as a stochastic process indexed by that parameter, we derive a limiting Gaussian process. As a special case, taking a fixed sequence of bandwidths, a limiting normal distribution is obtained.

Choosing the bandwidth in an optimal way by minimizing the (asymptotic) mean squared error would lead to an optimal estimator. However, the optimal bandwidth depends on characteristics of the underlying distribution function and is hence unknown in practice. A corollary of our theorem on the limiting process yields that the use of a consistent estimate of the optimal bandwidth would lead to the same limiting distribution as would be obtained using the true optimal bandwidth. A first attempt is made to produce a consistent estimator of that optimal bandwidth.

The kernel type estimator discussed in detail in Chapter 3 is applicable only in case the extreme value index is known to be positive. In Chapter 4 this estimator is adjusted in such a way that an estimator is obtained that is consistent for any (real valued) extreme value index. To that end, a second parameter is introduced. Asymptotic normality is derived and, by means of a small simulation study, the dependency of the estimator on the two parameters is discussed.

A disadvantage of this estimator is given by the fact that both parameters need to be chosen appropriately or even optimally. As an alternative generalization of the kernel type estimator of Chapter 3, a whole new class of kernel type estimators is introduced in Chapter 5. Although this class of estimators is characterized by an additional parameter as well, that parameter can (almost arbitrarily) be fixed. Then only the bandwidth needs to be chosen in an optimal way. Hence, the estimators essentially only have one parameter.

Asymptotic properties of these estimators will be derived, including consistency under very weak conditions and asymptotic normality under slightly more restrictive conditions. Moreover, the asymptotic bias is discussed in more detail.

A main advantage of kernel type estimators over the more often used Hill estimator and Moment estimator, is their dependence on the number of used order statistics. Kernel type estimators are smooth functions of the fraction of used order statistics whereas the other estimators depend rather erratically on that fraction. In fact, adding a small fraction of order statistics in the calculation of the estimate will not change the point estimate too much in case of kernel type estimators, whereas in case of the other estimators it can change the point estimate considerably. In several practical situations that are described in the literature, the extreme value index turns out to be close to zero. Since the behaviour of the (limiting) distributions differs substantially for positive and negative index, an estimate of that index that behaves not too erratically is of great importance.

Finally, in Chapter 6, a simulation study is performed to show the advantages and disadvantages of various estimators in case of small and intermediate size samples from several theoretical distributions. Moreover, some of the estimators are applied in case of real data concerning peaks in the water discharges at Lobith, the Netherlands, during the period 1901–1991.

Samenvatting

Het schatten van de extreme waarde index

– verhalen met een staartje –

In veel situaties is het interessant om de asymptotische eigenschappen van het gemiddelde gedrag te bestuderen. Vaak leidt dit tot een toepassing van een of andere vorm van de Centrale Limietstelling. In bepaalde gevallen kan het bestuderen van extreem gedrag minstens zo interessant en nuttig zijn. Zo kan bijvoorbeeld een nauwkeurige beschrijving van extreme golfhoogtes gebruikt worden bij het bepalen van betere strategieën om bewoonde gebieden te beschermen tegen overstromingen.

Eén van de resultaten uit de extreme waarde theorie, is een analogon van de Centrale Limietstelling. De verdeling van het steekproefmaximum, mits op de juiste manier getransformeerd, convergeert in de meeste gevallen naar een Extreme Waarde Verdeling. In tegenstelling tot de Centrale Limietstelling, waar de Normale verdeling altijd als limiet verdeling opduikt, zijn er drie verschillende typen limietverdelingen: de Gumbel, de Fréchet en de (inverse) Weibull verdeling. Hoewel de drie limietverdelingen totaal verschillende eigenschappen hebben, is het mogelijk om die drie typen te combineren tot een functievoorschrift waarin slechts één parameter, de extreme waarde index genaamd, het onderscheid aangeeft tussen de typen. Dit proefschrift richt zich op het schatten van die extreme waarde index. Een goede schatting daarvan vormt vaak de basis voor een verdere analyse van extreme situaties.

In het eerste hoofdstuk van dit proefschrift wordt nog wat extra achtergrond informatie gegeven over extreme gebeurtenissen, extreme waarde theorie en daaraan gerelateerde onderwerpen.

De laatste jaren zijn er verschillende pogingen gedaan om de moeilijkheid van het schatten van de extreme waarde index te kwantificeren. Een veel gebruikte methode voor het bepalen van de moeilijkheid van een schattingsprobleem bestaat uit het bepalen van het minimax risico daarvan over een bepaalde klasse van verdelingen. Aan het begin van Hoofdstuk 2 wordt een asymptotische ondergrens bepaald voor het minimax risico in het geval van het schatten van de extreme waarde index. Die ondergrens wordt op een inzichtelijke manier bepaald voor een vrij grote klasse van verdelingen.

Volgend op die berekeningen worden verschillende al bestaande schatters van de extreme waarde index besproken, waaronder de Hill schatter, de Pickands schatter, de Momenten schatter en een kernschatter. Bovendien wordt aangetoond dat de genoemde kernschatter te beschouwen is als een gewogen gemiddelde van Hill schatters. Daarmee wordt aannemelijk gemaakt dat de asymptotische eigenschappen van een recentelijk geïntroduceerde schatter eenvoudiger zouden kunnen worden afgeleid door gebruik te maken van de asymptotische eigenschappen van die kernschatter.

Aan het einde van het hoofdstuk wordt aangetoond, dat de ondergrens van het minimax risico voor het schatten van de extreme waarde index in de zin van convergentie in verdeling gehaald kan worden voor verschillende klassen van verdelingsfuncties. Dit

wordt expliciet aangetoond voor de Pickands schatter, maar ook andere in dit hoofdstuk besproken schatters halen die ondergrens.

De asymptotische eigenschappen van de zojuist al genoemde kernschatter worden in Hoofdstuk 3 nog wat gedetailleerder bekeken. Zoals bij iedere schatter van de extreme waarde index, wordt ook in dit geval bij de berekening van de schatting gebruik gemaakt van een bepaald aantal grootste waarnemingen uit de steekproef. De fractie van gebruikte grootste waarnemingen wordt in het geval van de kernschatter bepaald door één parameter: de bandbreedte. Door de kernschatter als een stochastisch proces te beschouwen, geïndexeerd door die bandbreedte, wordt een Gaussisch limietproces afgeleid. Door vervolgens een vaste rij van getallen als bandbreedte te nemen wordt als speciaal geval de normale verdeling als limietverdeling van de schatter gevonden.

Door geschikte keuze van de bandbreedte kan de best mogelijke kernschatter worden gedefinieerd. Die optimale bandbreedte, gedefinieerd als de bandbreedte die de verwachte kwadratische fout van de schatter minimaliseert, hangt echter af van eigenschappen van de onderliggende verdeling en is in de praktijk dus niet te bepalen. Als gevolg van onze stelling betreffende het Gaussische limietproces, kan wel worden afgeleid dat de kernschatter, gebaseerd op een consistente schatter van de optimale bandbreedte, dezelfde asymptotische verdeling heeft als de kernschatter gebaseerd op de echte optimale bandbreedte. In Hoofdstuk 3 wordt dan ook een eerste aanzet gegeven tot een consistente schatter voor de optimale bandbreedte.

De kernschatter uit Hoofdstuk 3 is alleen een consistente schatter in het geval dat de te schatten extreme waarde index positief is. In Hoofdstuk 4 wordt deze schatter zodanig aangepast, dat een schatter ontstaat die ook consistent is voor niet positieve waarden van de extreme waarde index. Daartoe wordt, naast de bandbreedte, een extra parameter geïntroduceerd. De asymptotische normaliteit van deze schatter wordt aangetoond en aan het einde van het hoofdstuk wordt met behulp van een simulatiestudie bestudeerd op welke manier de schatter afhangt van de twee parameters.

Een nadeel van deze schatter is dat er nu twee parameters zijn die geschikt, of zelfs optimaal, moeten worden gekozen. Als alternatief wordt in Hoofdstuk 5 een nieuwe klasse van kernschatters geïntroduceerd die ook als een generalisatie van de kernschatter uit Hoofdstuk 3 kan worden gezien. Hoewel deze nieuwe schatters nog steeds van twee parameters afhangen, kan nu één van de parameters (vrij willekeurig) vastgezet worden, waarna de andere parameter (de bandbreedte) optimaal gekozen kan worden.

Ook van deze kernschatters worden asymptotische eigenschappen afgeleid, waaronder de consistentie onder zeer zwakke voorwaarden en de asymptotische normaliteit onder strengere voorwaarden. Tevens wordt nader ingegaan op de mogelijke vormen van de asymptotische onzuiverheid van deze schatters.

Een groot voordeel dat kernschatters hebben boven de meer bekende Hill schatter en Momenten schatter, ligt in de manier waarop die schatters afhangen van het aantal gebruikte grootste waarden uit de steekproef. Kernschatters hangen op een gladde manier af van de fractie gebruikte grootste waarden, terwijl de andere schatters daar op een zeer grillige manier van afhangen. Door een kleine fractie van grootste waarden

toe te voegen aan de berekening, zal bij het gebruik van kernschatters de puntschatting niet veel veranderen, terwijl bij het gebruik van de Hill schatter en de Momenten schatter de puntschatting totaal anders uit kan vallen. In veel praktische situaties die in de literatuur zijn beschreven, blijkt dat de extreme waarde index vaak in de buurt van nul ligt. Gezien het totaal verschillende karakter van de (limiet)verdelingen bij positieve en negatieve index, is een niet al te grillig gedrag van een schatter van die index dan ook zeer gewenst.

In Hoofdstuk 6 is een simulatiestudie opgenomen, waarin naar de voor- en nadelen van de verschillende schatters wordt gekeken voor kleine en middelgrote steekproeven uit verschillende (kunstmatige) verdelingen. Een aantal schatters is ook toegepast op echte waarnemingen, die betrekking hebben op pieken in de waterlozing bij Lobith, gedurende de periode 1901–1991.

Curriculum Vitae

De schrijver van dit proefschrift werd op 10 oktober 1967 in Zwijndrecht geboren. Van 1980 to 1986 bezocht hij daar de Openbare Scholen Gemeenschap Walburg, waar hij zijn VWO diploma haalde met de vakken Nederlands, Engels, Latijn, wiskunde 1, natuurkunde, scheikunde en biologie.

Aansluitend begon hij de studie wiskunde aan de Technische Universiteit Delft (TUD) en behaalde hij in 1987 de propedeuse (cum laude). In 1991 studeerde hij, na een stage van negen maanden bij KSEPL (Shell) in Rijswijk, onder leiding van prof. dr. P. Groeneboom cum laude af op de scriptie 'On the estimation of the index of variation'. Voor die scriptie mocht hij ook de VVS scriptieprijs ontvangen.

Van 1991 tot 1995 verdiepte hij zich als assistent in opleiding aan de Technische Universiteit Delft verder in de extreme waarde theorie en verrichtte hij promotieonderzoek op dit gebied, onder supervisie van prof. dr. P. Groeneboom. Tijdens dat AIO-schap was hij gedurende een zestal maanden in Chapel Hill (North Carolina, USA) om daar aan de 'University of North Carolina' samen met prof. dr. R. Smith en prof. dr. S. Marron onderzoek te doen.

Sinds mei 1996 is hij als statistisch methodoloog werkzaam op het Centraal Bureau voor de Statistiek (CBS) en heeft hij in zijn vrije tijd het promotieonderzoek verder voltooid.

