Tail Empirical Processes: Limit Theorems and Bootstrap Techniques, with Applications to Risk Measures

Hicham Loukrati

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

Au cours des dernières années, des changements importants dans le domaine des assurances et des finances attirent de plus en plus l'attention sur la nécessité d'élaborer un cadre normalisé pour la mesure des risques. Récemment, il y a eu un intérêt croissant de la part des experts en assurance sur l'utilisation de l'espérance conditionnelle des pertes (CTE) parce qu'elle partage des propriétés considérées comme souhaitables et applicables dans diverses situations. En particulier, il répond aux exigences d'une mesure de risque "cohérente", selon Artzner [2].

Cette thèse représente des contributions à l'inférence statistique en développant des outils, basés sur la convergence des intégrales fonctionnelles, pour l'estimation de la CTE qui présentent un intérêt considérable pour la science actuarielle. Tout d'abord, nous développons un outil permettant l'estimation de la moyenne conditionnelle $\mathbb{E}[X|X > x]$, ensuite nous construisons des estimateurs de la CTE, développons la théorie asymptotique nécessaire pour ces estimateurs, puis utilisons la théorie pour construire des intervalles de confiance.

Pour la première fois, l'approche de bootstrap non paramétrique est explorée dans cette thèse en développant des nouveaux résultats applicables à la valeur à risque (VaR) et à la CTE. Des études de simulation illustrent la performance de la technique de bootstrap.

Acknowledgements

En préambule à cette thèse, je remercie Dieu le tout puissant de m'avoir accordé désir et patience dans la réalisation de ce travail.

Je tiens tout d'abord à remercier mes directeurs de thèse, Docteur Gail Ivanoff et Docteur Rafal Kulik, pour leur précieuse collaboration, leur disponibilité, leurs recommandations judicieuses et leur patience pendant presque six ans. Cette thèse n'aurait jamais pu être accompli sans leur encadrement supérieur.

Je voudrais également remercier les membres du comité: le professeur Hao Yu, La professeure Raluca Balan, le professeur Aaron Smith, le professeur Mohamedou Ould Haye pour avoir siégé en tant que membres de comité examinateurs. Je tiens aussi à vous remercier d'avance de faire de ma défense un moment agréable et de vos commentaires, questions et suggestions constructifs.

J'exprime ma gratitude à toute personne qui a d'une façon ou d'une autre prêté son concours à l'accomplissement de ce projet et à toute personne qui, spontanément ou à ma demande, m'a fait part de ses suggestions et de ses commentaires constructifs visant à rendre ce mémoire conforme aux exigences.

Je voudrais témoigner toute ma reconnaissance aux professeurs qui m'ont fourni les outils nécessaires à la réussite de mes études universitaires, plus particulièrement la Dr Raluca Balan et le Dr. Mohamedou Ould Haye.

Merci à toutes et à tous

Dedication

Á mes parents: Ahmed et Zahra. Je vous rends hommage par ce modeste travail en guise de ma reconnaissance éternelle et de mon infini amour.

Á ma femme, Marina. Tous les mots du monde ne sauraient exprimer l'immense amour que je te porte, ni la profonde gratitude que je te témoigne pour tous les efforts et les sacrifices que tu n'as jamais cessé de consentir pour cette thèse.

Á mes beaux-parents, Anatoly et Galina pour leur soutien tout au long de mes études.

Á ma famille, à mes amis et à tous ceux qui me sont chers.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Très différente de la théorie classique basée sur le théorème central limite, la théorie des valeurs extrêmes (extreme value theory, EVT), après le travail pionnier de Fréchet [22] et de Fisher et Tippett [21] sur les lois limites possibles du maximum, était quelque chose de plutôt spécial. En effet, le besoin urgent de décrire et de résoudre certains des problèmes liés aux phénomènes extrêmes dans divers domaines d'applications a eu un effet considérable sur le développement fondamental de la théorie des valeurs extrêmes. Depuis 1958, Gumbel et Weibull ont démontré de façon impressionnante l'importance des concepts mathématiques du comportement extrême dans plusieurs applications. La théorie des valeurs extrêmes se repose principalement sur des distributions limites des extrêmes et leurs domaines d'attraction. Cependant, on y retrouve deux modèles : loi généralisée des extrêmes et loi de Pareto généralisée. En 1943, Gnedenko [23], Fisher et Tippett ont énoncé un théorème fondamental avec la création de trois domaines d'attraction: domaine d'attraction de Fréchet, Gumbel et Weibull. Ce théorème intéressant fait référence à un paramètre appelé l'indice de queue, qui contrôle la forme de la queue de distribution. Si l'indice de queue est positif nous sommes en présence du domaine d'attraction de Fréchet; si il est négatif, domaine d'attraction de Weibull par contre si l'indice est nul alors domaine de Gumbel. Pour une lecture assez complète sur les principaux résultats théoriques, l'ouvrage de Embrechts et al., [17] est une bonne référence, quant à Reiss et Thomas [37] nous proposent un certain nombre d'exemples pratiques, en finance, en assurance et en sciences environnementales. En 1970, pour la première fois, des propriétés stochastiques des valeurs extrêmes ont été développées dans la thèse écrite par Laurens De Haan intituléeă: "On regular variation and its applications to the weak convergence of sample extremes" [15]. Cela semble être le début du développement théorique de la théorie des valeurs extrêmes. On y montre que la classe des distributions à variations régulières est étroitement liée à la théorie des valeurs extrêmes. En effet, une distribution appartient au domaine d'attraction de Fréchet si et seulement si elle est à variation régulière [17]. Cette notion de variation régulière fournit ainsi un lien entre EVT et des outils probabilistes puissants tels que les processus ponctuels et leur convergence faible. Quant à, l'aspect statistique a dû attendre longtemps avant qu'il n'ait reçu l'attention nécessaire. Depuis quelques années, en raison de la richesse d'idées mathématiques, la théorie des valeurs extrêmes est devenue de plus en plus d'intérêt pour plusieurs chercheurs aussi bien sur le plan théorique que sur le plan pratique. En particulier, on note une attention croissante pour l'application de cette théorie pour la modélisation des événements dits rares dont la probabilité d'apparition est trop faible, c'est-à-dire se trouve dans les queues des distributions. En général, ces événements apparaissent dans de nombreux contextes physiques variés, en particulier les catastrophes naturelles : la prévision des crues en hydrologie [12] et [29], l'étude de la vitesse du vent en météorologie [10], les grands sinistres en assurance [32] et [40], en biologie, en ingénierie, en gestion de l'environnement, en finance, sciences sociales, etc. Von Mises [41] et Jenkinson [28] ont rassemblé les distributions des trois domaines d'attraction en une seule écriture. C'est en ce moment que plusieurs auteurs se sont focalisés aux estimations de l'indice des valeurs extrêmes. Dans le cas où l'indice est positif, on peut citer l'estimateur de Hill, [26] l'estimateur de t-Hill [19] et la classe des estimateurs du moment harmonique [25]. Notre principale motivation est inspirée par les travaux de Holger Rootzen [39] afin de prouver la convergence faible du processus empirique de queue (tail empirical process, TEP) lié à des distributions à queue à variation régulière dans l'espace $\mathscr{D}(0,\infty)$ pour des variables aléatoires indépendantes et identiquement distribuées. En fait, les processus empiriques de queue sont les objets de choix si l'on veut étudier le comportement extrême de la fonction de répartition. Outre que l'intérêt théorique, la motivation vient des méthodes semi-paramétriques appliquées aux extrêmes.

1.2 Contributions and plan of thesis

Le thème principal de cette thèse étant d'étudier la convergence en loi du processus empirique de queue lié à des distributions à queue lourde et s'articule autour de trois axes principaux: estimation fonctionnelle de l'indice des valeurs extrêmes, étude des mesures de risque actuariel pour des pertes extrêmes, en particulier la valeur à risque (Value-at-Risk, VaR) et l'espérance conditionnelle des pertes (Conditional Tail Expectation, CTE) et application de la technique de sous-échantillonnage (bootstrap) à l'approximation des distributions limites de quelques estimateurs. Passant à travers l'introduction et la conclusion, nous commençons par le deuxième chapitre.

1.2.1 Chapter 2

In this chapter we present concepts and tools used throughout the thesis. Indeed, almost all results are based on existing literature. There, we recall the notion of univariate regular variation, second order regular variation and weak convergence. We cite some important results that we will use later in this work (cf. [38], [20], [33], [8], [9], [14], [3], [16], [30], [6], [27], [5] and [42]).

However, Lemma 2.5.6 is new; it presents a mild extension of criteria for convergence in distribution from [5].

1.2.2 Chapter 3

Here, we consider tail empirical processes $G_n(\cdot)$ with deterministic levels and $\widehat{G}_n(\cdot)$ with random levels defined in equations (3.1.5) and (3.1.24), respectively. We prove the weak convergence for $G_n(\cdot)$ (Theorem 3.1.1) and $\widehat{G}_n(\cdot)$ (Theorem 3.1.9). The results in these theorems are not new, they are special cases of results in [39]; see also [14]. For completeness, detailed proofs are provided and a comparison between our results and those in [39] and [14] is included in Section 3.4. Furthermore, the proof of Theorem 3.1.1 provides the basis for the new results on bootstrap procedures presented in Chapter 6.

In Section 3.2 we present new results (Theorem 3.2.1 and Theorem 3.2.2) on weak convergence of integral functionals of the TEP,

$$\int_1^\infty \frac{G_n(s)}{s^m}\,ds \qquad \text{and} \qquad \int_1^\infty \frac{\widehat{G}_n(s)}{s^m}\,ds$$

for integer values of $m \geq 0$. These theorems are keys to the subsequent results on weak convergence of estimators of the tail index and various risk measures. In particular, Theorem 3.2.2 is applied to recover existing results on weak convergence of the Hill estimator of the tail index (cf. Corollary 3.3.1; see Theorem 3.2.5 in [14]), Harmonic Mean estimators of the tail index (cf. Corollary 3.3.2; [19]) and to obtain a new result on estimation of the conditional mean (Corollary 3.3.7). Again, a precise comparison between our results and those in [14] can be found in Section 3.4.

1.2.3 Chapter 4

La littérature sur le risque et les mesures de risque est vaste. Le quatrième chapitre se concentre sur le risque financier et sur quelques mesures de ce risque. La gestion des risques s'est développée à travers les besoins des entreprises d'une part et grâce à la recherche scientifique d'autre part développant ainsi des méthodes modernes de gestion. Ces méthodes de gestion s'appuient principalement sur l'utilisation de modèles mathématiques et statistiques. Aujourd'hui, la maitrise et la pratique de ces méthodes sont nécessaires à la bonne gestion de l'entreprise. En commençant par définir la mesure de risque et les propriétés actuelles que toute mesure de risque acceptable devrait avoir [2], ce chapitre présente l'aspect pratique de la thèse. Nous y montrons l'application des résultats trouvés dans le chapitre 3. En particulier, nous y étudions la normalité asymptotique de la VaR et de la CTE et nous y développons une nouvelle approche pour estimer la mesure de risque CTE (Théorème 4.2.4).

1.2.4 Chapter 5

Dans ce chapitre, nous faisons appel à la méthode de bootstrap non paramétrique (nonparametric bootstrap, Efron [18]). Les résultats développés dans ce chapitre sont complètement nouveaux. Tout d'abord, à la section 5.1, nous établissons la convergence en loi pour le schéma triangulaire des processus empiriques de queue. Ensuite, à la section 5.2, après avoir introduit une bref description de la technique de bootstrap, nous développons de nouveaux résultats en ce qui concerne la version bootstrap de TEP à niveaux déterministes $G_n^{(b)}(\cdot)$ et de celui avec des statistiques d'ordre $\widehat{G}_n^{(b)}(\cdot)$, définies dans (5.2.2) et (5.2.12), respectivement. La technique de bootstrap est applicable si l'on souhaite évaluer la variance de certains estimateurs, plus particulièrement l'estimateur de CTE. Cette application et des applications aux estimateurs de Hill, de t-Hill et de la VaR sont montrées à la section 5.3.

1.2.5 Chapter 6

En dehors de l'étude théorique qui sera menée dans cette thèse, dans ce dernier chapitre on s'attardera sur des simulations dont le but est d'apprécier la qualité de la méthode de bootstrap appliquée aux estimateurs étudiés.

Enfin, certains théorèmes et expressions de nature plutôt technique sont relégués aux annexes. Dans l'annexe A, nous avons regroupé quelques inégalités mathématiques utils dans nos démonstrations. L'annexe B présente le code source des simulations effectuées dans le langage R.

Summary

- The results in Chapter 3 are not new. Proofs are provided for completeness in order to:
  - present a unified framework to deal with different estimators such as the Hill estimator, the Harmonic Mean estimator or an estimator of the conditional mean (see Section 3.3.3). As such we recover both classical and newer results with one single technique, using integral functionals.
  - provide the theoretical basis for the new results on bootstrap procedures.
- The results in Chapter 4 are applications of the delta method. Some results on estimation of the Value-at-Risk can be found in [43] and [14]. The results for the Conditional Tail Expectation are new (to the best of our knowledge). The feature of our approach is that we can cover many risk measures with one technique.
- The results in Chapter 5 are completely new and they are obtained with the help of the tools developed in Chapter 3.
- The theoretical results are illustrated in Chapter 6.

Chapter 2

Preliminaries

2.1 Univariate regular variation

The notion of regular variation introduced by J. Karamata in 1930 proved fruitful in many areas, and finds an ever increasing number of applications in probability theory and statistics. In fact, this notion appears in a natural way in various fields of applied probability, including queuing theory, extreme value theory and others. The theory of regularly varying functions is an essential analytical tool for dealing with heavy tails. Roughly speaking, regularly varying functions are those functions which behave asymptotically like power functions (cf. [38]). This concept of regular variation is used extensively in the theory of extreme values (EVT). This chapter presents the main definitions and results that will be used later.

Definition 2.1.1 *([14], Definition B.1.1) A Lebesgue measurable function $f : \mathbb{R}^+ \mapsto \mathbb{R}$ that is eventually positive is regularly varying at infinity,written $f \in \mathbf{RV}_\alpha$, if for some $\alpha \in \mathbb{R}$,*

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\alpha, \qquad x > 0\,. \tag{2.1.1}$$

We call the real number α the **index of regular variation**.

If we put

$$f(x) = x^{\alpha}\ell(x)\,, \tag{2.1.2}$$

the ratio $f(tx)/f(t)$ will approach x^{α} if and only if

$$\lim_{t\to\infty} \frac{\ell(tx)}{\ell(t)} = 1\,, \tag{2.1.3}$$

for every $x > 0$. Functions with this property vary *slowly*, $\ell(x) \in \mathbf{RV}_0$, and the transformation (2.1.2) reduces regular variation to slow variation. Formally, we use this fact for a definition of regular variation. That is,

- A positive function ℓ defined on $\mathbb{R}_+$ varies slowly at infinity if and only if (2.1.3) holds.
- A function f varies regularly with index $\alpha \in \mathbb{R}$ if and only if it is the form (2.1.2) with ℓ slowly varying.

This definition extends to regular variation at the origin. We say that f varies regularly at 0 if and only if $f(x^{-1})$ varies regularly at infinity. The property of regular variation depends only on the behavior at infinity and it is therefore not necessary that $\ell(x)$ be positive, or even defined, for all $x > 0$ [20].

Example 2.1.2 Typical examples of slowly varying functions are positive constants or functions converging to a positive constant, logarithms and iterated logarithms. For α, $\beta \in \mathbb{R}$ the functions x^{α}, $x^{\alpha}(\log x)^{\beta}$, $x^{\alpha}(\log\log x)^{\beta}$ are $\mathbf{RV}_{\alpha}$. The functions $2{+}\sin(\log\log x)$, $\exp((\log x)^{\alpha})$, with $0 < \alpha < 1$, $\sum_{k\le x} 1/k$, $(\log x)^{\sin(\log\log x)}$ are slowly varying . The functions $2 + \sin x$, $\exp[\log x]$, $2 + \sin(\log x)$, $x\exp(\sin(\log x))$ are not regularly varying where $[\cdot]$ stands for integer part.
All powers of $|\log(x)|$ vary slowly at infinity. Similarly, any function f such that $\lim\limits_{x\to\infty} f(x) = f(\infty)$ exists, is positive and finite, is slowly varying.

The function $f(x) = (1 + x^2)^p$ varies regularly at infinity with index $2p$. The function e^x does not vary regularly at infinity, but it satisfies the conditions of Definition 2.1.1 with $\alpha = \infty$.

The next result gives a full characterization of regular variation.

Theorem 2.1.3 *([14], Theorem B.1.3) Suppose $f : \mathbb{R}^+ \mapsto \mathbb{R}$ is measurable, eventually positive, and*

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} \tag{2.1.4}$$

exists, and is finite and positive for all x in a set of positive Lebesgue measure. Then $f \in \mathbf{RV}_\alpha$ for some $\alpha \in \mathbb{R}$.

If we denote by λ the limiting function of (2.1.4), then λ satisfies $\lambda(xy) = \lambda(x)\lambda(y)$ which implies that λ is a power function.

We introduce some of the most important properties of regularly varying functions.

Theorem 2.1.4 **(Representation Theorem)***[33]*
A positive measurable function ℓ on $[x_0, \infty)$ is slowly varying if and only if it can be written in the form

$$\ell(x) = c(x) \exp\left\{ \int_{x_0}^{x} \frac{\varepsilon(y)}{y} dy \right\}, \tag{2.1.5}$$

where $c(\cdot)$ is a measurable non-negative function such that $\lim_{x\to\infty} c(x) = c_0 \in (0, \infty)$ and $\varepsilon(x) \to 0$ as $x \to \infty$.

We remark that it is clear that a regularly varying function f with index α has representation

$$f(x) = x^\alpha c(x) \exp\left\{ \int_{x_0}^{x} \frac{\varepsilon(y)}{y} dy \right\}, \tag{2.1.6}$$

where $c(\cdot)$ and $\varepsilon(\cdot)$ are as above. Also we may conclude that if $\alpha \neq 0$, as $x \to \infty$,

$$f(x) \to \begin{cases} \infty & \text{if} \quad \alpha > 0, \\ 0 & \text{if} \quad \alpha < 0. \end{cases}$$

Moreover, if ℓ is slowly varying then for every $\epsilon > 0$, as $x \to \infty$

$$x^{-\epsilon}\ell(x) \longrightarrow 0 \qquad \text{and} \qquad x^{\epsilon}\ell(x) \longrightarrow \infty\,.$$

An important result is the fact that convergence in (2.1.1) is uniform on each compact subset of $(0, \infty)$.

Theorem 2.1.5 **(Uniform Convergence Theorem for regular variation)***[33]* *If f varies regularly with index α (in the case $\alpha > 0$, assuming f bounded on each interval $(0, x]$, $x > 0$), then for $0 < a \leq b < \infty$*

$$\begin{aligned} \frac{f(tx)}{f(t)} \quad & \longrightarrow x^{\alpha} \qquad (t \longrightarrow \infty) \qquad \textit{uniformly in } x \\ & (i.) \quad \textit{on each } [a, b] \quad \textit{if} \;\; \alpha = 0, \\ & (ii.) \quad \textit{on each } [0, b] \quad \textit{if} \;\; \alpha > 0, \\ & (iii.) \; \textit{on each } [a, \infty] \quad \textit{if} \;\; \alpha < 0. \end{aligned}$$

In what follows, for any positive functions f and g, $f(x) \sim g(x)$ as $x \longrightarrow x_1$ means

$$\lim_{x \to x_1} \frac{f(x)}{g(x)} = 1\,.$$

The following result of Karamata is often applicable. It essentially says that integrals of regularly varying functions are again regularly varying, or more precisely, one can take the slowly varying function out of the integral.

Theorem 2.1.6 (Karamata's Theorem)[33]

Suppose $\ell \in \mathbf{RV}_0$ and locally bounded in $[x_0, \infty)$ for some $x \geq x_0$. Then,

$$\text{for } \alpha > -1, \qquad \int_{x_0}^{x} s^{\alpha}\ell(s)ds \sim (\alpha+1)^{-1}x^{\alpha+1}\ell(x), \qquad x \to \infty\,, \tag{2.1.7}$$

$$\text{for } \alpha < -1, \qquad \int_{x}^{\infty} s^{\alpha}\ell(s)ds \sim -(\alpha+1)^{-1}x^{\alpha+1}\ell(x), \qquad x \to \infty\,. \tag{2.1.8}$$

Conversely, if (2.1.7) holds with $-1 < \alpha < \infty$, then $\ell \in \mathbf{RV}_0$; if (2.1.8) holds with $-\infty < \alpha < -1$, then $\ell \in \mathbf{RV}_0$.

The result remains true for $\alpha = -1$ in the sense that if $\int_{x_0}^{x}(\ell(s)/s)ds \in \mathbf{RV}_0$,

$$\frac{1}{\ell(x)}\int_{x_0}^{x}\frac{\ell(s)}{s}ds \longrightarrow \infty, \qquad x \to \infty\,,$$

and if $\int_{x_0}^{\infty}(\ell(s)/s)ds < \infty$ and $\int_{x_0}^{x}(\ell(s)/s)ds \in \mathbf{RV}_0$.

$$\frac{1}{\ell(x)}\int_{x}^{\infty}\frac{\ell(s)}{s}ds \longrightarrow \infty, \qquad x \to \infty\,,$$

The following result is crucial for the differentiation of regularly varying functions.

Theorem 2.1.7 (Monotone density theorem)[33]

Assume that $V(x) = \int_0^x v(y)dy$ (or $\int_x^{\infty} v(y)dy$) where v is ultimately monotone. If

$$V(x) \sim c\,x^{\alpha}\ell(x), \qquad x \to \infty\,, \tag{2.1.9}$$

with $c \geq 0$, $\alpha \in \mathbb{R}$ and ℓ is slowly varying, then

$$v(x) \sim c\,\alpha\,x^{\alpha-1}\ell(x), \qquad x \to \infty\,. \tag{2.1.10}$$

For the case where $c = 0$ (2.1.9) and (2.1.10) are interpreted as $V(x) = o(x^{\alpha}\ell(x))$ and $v(x) = o(x^{\alpha-1}\ell(x))$.

The next result concerns global bounds for $f(y)/f(x)$, extending Potter [36].

Theorem 2.1.8 (Potter's Theorem) *[8]*

(i) *If ℓ is slowly varying then for any chosen constants $A > 1$, $\delta > 0$ there exists $B = B(A, \delta)$ such that*

$$\ell(y)/\ell(x) \leq A \max\{(y/x)^{\delta}, (y/x)^{-\delta}\}, \quad (x \geq B, y \geq B)\,. \tag{2.1.11}$$

(ii) *If, further, ℓ is bounded away from 0 and ∞ on every compact subset of $[0, \infty)$, then for every $\delta > 0$ there exists $A' = A'(\delta) > 1$ such that*

$$\ell(y)/\ell(x) \leq A' \max\{(y/x)^{\delta}, (y/x)^{-\delta}\}, \quad (x > 0, y > 0)\,. \tag{2.1.12}$$

(iii) *If $f \in \mathbf{RV}_\rho$ then for any chosen $A > 1$, $\delta > 0$ there exists $B = B(A, \delta)$ such that*

$$f(y)/f(x) \leq A \max\{(y/x)^{\rho+\delta}, (y/x)^{\rho-\delta}\}, \quad (x \geq B, y \geq B)\,. \tag{2.1.13}$$

The following Corollary is a useful property of regularly varying functions.

Corollary 2.1.9 (Potter bounds) *[38]*

Suppose $U \in \mathbf{RV}_\rho$, $\rho \in \mathbb{R}$. Take $\epsilon > 0$. Then there exists t_0 such that for $x > 1$ and $t \geq t_0$,

$$(1-\epsilon)x^{\rho-\epsilon} < \frac{U(tx)}{U(t)} < (1+\epsilon)x^{\rho+\epsilon}\,. \tag{2.1.14}$$

2.1.1 Regularly varying sequences

The following results were developed in [9].

Definition 2.1.10 *A sequence $\{c_n\}_{n=0}^{\infty}$ of positive terms is regularly varying if,*

$$\frac{c_{[\lambda n]}}{c_n} \longrightarrow \psi(\lambda) \in (0,\infty), \quad \textit{as } n \longrightarrow \infty, \quad \forall \lambda > 0\,. \tag{2.1.15}$$

Theorem 2.1.11 *If $\{c_n\}$ is a regularly varying sequence, the above limit function $\psi(\lambda)$ has the form $\psi(\lambda) = \lambda^{\alpha}$ for some finite $\alpha \in \mathbb{R}$ and all λ.*

We call α the index of $\{c_n\}$, and write $c_n \in \mathbf{RV}_{\alpha}$. When $\alpha = 0$, a regularly varying sequence is called a slowly varying. The following results are useful in applications.

Corollary 2.1.12 *Let $\big\{\ell(n)\big\}$ be a slowly varying sequence. A sequence of positive numbers c_n is a regularly varying sequence of index α if and only if, for $n \geq 1$,*

$$c_n = n^{\alpha}\,\ell(n)\,. \tag{2.1.16}$$

Theorem 2.1.13 *If $\{c_n\}$ is a regularly varying sequence with index α, the function $f(x) = c_{[x]}$ defined on $[0,\infty)$ is regularly varying with index α.*

Thus every regularly varying sequence is embeddable as the integer values of a regularly varying function and since we may apply all the properties possessed by regularly varying functions to f, it is easy to deduce the analogous properties for the sequence c_n.

2.2 Properties of inverse functions

Suppose $f : \mathbb{R} \mapsto (a, b)$ is a nondecreasing function on $\mathbb{R}$ with range (a, b), where $-\infty \leq a < b \leq \infty$. With the convention that the infimum of an empty set is ∞, we define the left-continuous inverse $f^{\leftarrow} : (a, b) \mapsto \mathbb{R}$ of f as

$$f^{\leftarrow}(y) = \inf\{x : f(x) \geq y\}\,. \tag{2.2.1}$$

When f is strictly increasing the inverse is defined unambiguously. In case the function f is right continuous, we have the following desirable properties [38]:

i. $A(y) = \{x : f(x) \geq y\}$ is closed,

ii. $f(f^{\leftarrow}(y)) \geq y$,

iii. $f^{\leftarrow}(y) \leq t$ if and only if $y \leq f(t)$.

A useful technique is to translate convergence of a family of monotone functions to a corresponding family of inverses using the following simple result.

Lemma 2.2.1 *[14] Suppose f_n is a sequence of nondecreasing functions and g is a nondecreasing function. Suppose that for each x in some open interval (a, b) that is a continuity point of g,*

$$\lim_{n\to\infty} f_n(x) = g(x)\,. \tag{2.2.2}$$

Let $f_n^{\leftarrow}$, $g^{\leftarrow}$ be the left-continuous inverses of f_n and g. Then, for each x in the interval $(g(a), g(b))$ that is a continuity point of $g^{\leftarrow}$ we have

$$\lim_{n\to\infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x)\,. \tag{2.2.3}$$

Theorem 2.2.2 *([8], Theorem 1.5.12) Let $\alpha > 0$. Let $f \in \mathbf{RV}_\alpha$ at infinity be ultimately monotone. Then there exist a function $g \in \mathbf{RV}_{1/\alpha}$ such that*

$$(f \circ g)(x) \sim (g \circ f)(x) \sim x \,. \tag{2.2.4}$$

Moreover g can be chosen as the left-continuous or the right-continuous inverse of f and satisfying that $g(x) \sim x^{1/\alpha} \ell^\star \left(x^{1/\alpha}\right)$, where $\ell^\star$ is a slowly varying function.

Corollary 2.2.3 (De Bruyn conjugate)*[3]*
If $\ell(x)$ is slowly varying, then there exists a slowly varying function $\ell^\star(x)$, the de Bruyn conjugate of ℓ, such that

$$\ell(x)\ell^\star\left(x\ell(x)\right) \longrightarrow 1, \qquad \text{as } x \to \infty \,. \tag{2.2.5}$$

The de Bruyn conjugate is asymptotically unique in the sense that if also $\tilde{\ell}$ is slowly varying and $\ell(x)\tilde{\ell}\left(x\ell(x)\right) \longrightarrow 1$, then $\ell^\star \sim \tilde{\ell}$.

Example 2.2.4 Let α, β be real numbers such that $\alpha\beta > 0$. Let $f(x) = x^{\alpha\beta}\ell^\alpha(x^\beta)$. Then if f is ultimately monotone and g is its left or right continuous inverse, then

$$g(x) \sim x^{\frac{1}{\alpha\beta}} (\ell^\star)^{\frac{1}{\beta}} (x^{\frac{1}{\alpha}}) \,.$$

2.3 Regularly varying random variables

As we have already said, the notion of regular variation appears in many fields of applied probability. In those areas, random variables (or random vectors) with regularly varying tails have plenty of applications. Let $F(x)$ be the distribution function of any random variable X and $\overline{F}(x) = 1 - F(x)$ its right tail.

Definition 2.3.1 *A non-negative random variable X and its distribution are said to be regularly varying with index $\alpha \geq 0$ if the right tail distribution $\overline{F}$ is regularly varying with index $-\alpha$.*

Example 2.3.2

1. Standard Pareto distribution : $1 - F(x) = x^{-\alpha}$, $x \geq 1$, $\alpha > 0$.
2. Extreme-value distribution : $\Phi_\alpha(x) = exp\left(-x^{-\alpha}\right), x > 0$, where $\Phi_\alpha(x)$ has the property : $1 - \Phi_\alpha(x) \sim x^{-\alpha}$, as $x \longrightarrow \infty$.
3. A stable law with index α, $0 < \alpha < 2$, has the property, $1 - G(x) \sim cx^{-\alpha}$, when $x \longrightarrow \infty$, $c > 0$.

The concept of the tail quantile function is important to deal with extremes. Let X be a random variable with distribution function F. By $x_L = \inf\{x : F(x) > 0\}$ we denote the left-end boundary of F and similarly $x_R = \sup\{x : F(x) < 1\}$. We define the tail quantile function Q on $[1, \infty)$ by

$$Q(t) = \inf\{x : F(x) \geq 1 - 1/t\}\,. \tag{2.3.1}$$

We note that the tail quantile function Q and the inverse function $F^{\leftarrow}$ are linked via the relation $Q(t) = F^{\leftarrow}(1 - 1/t)$. The quantity $Q(t)$ is just a large quantile; it is the high level such that there is only probability $1/t$ that X exceeds the level.

From (2.3.1), we get the following inequalities:

i. If $y < Q(t)$, then $1 - F(y) > 1/t$;

ii. for all $t > 0$, $1 - F(Q(t)) \leq 1/t$;

iii. for all $x < x_R$, $Q(1/(1 - F(x))) \leq x$.

Note that $Q(1) = x_L$ while $Q(\infty) = x_R$.

We can easily prove that if F is continuous, then $1 - F(Q(t)) = 1/t$ and if Q is continuous, then $Q(1/(1 - F(x))) = x$. In any case, Q will be left-continuous while F is right-continuous. The easy transition from F to Q and back is the main reason for always assuming that F and Q are continuous. Here we cite an important relation between regular variation of $\overline{F}$ and Q. If $\overline{F} \in \mathbf{RV}_{-\alpha}$ is continuous, then

$$Q(t) := F^{\leftarrow}(1 - 1/t) \in \mathbf{RV}_{1/\alpha} . \tag{2.3.2}$$

Lemma 2.3.3 *Let X be a nonnegative random variable with continuous distribution function F which has a regularly varying right tail with index $-\alpha$, $\alpha > 0$. Let Q be the tail quantile function of F. Assume that x_n, y_n are nonnegative sequences converging to ∞. If there exist a, b, $(0 < a < b < \infty)$, such that $a < \liminf(x_n/y_n) \leq \limsup(x_n/y_n) < b$, then*

$$\frac{Q(x_n)}{Q(y_n)} \sim \left(\frac{x_n}{y_n}\right)^{1/\alpha} . \tag{2.3.3}$$

Proof: Recall that $Q(x) \in \mathbf{RV}_{1/\alpha}$. Then, by uniform convergence Theorem 2.1.5

$$\frac{Q(x_n)}{Q(y_n)} = \frac{Q\left(y_n \cdot \dfrac{x_n}{y_n}\right)}{Q(y_n)} \sim \left(\frac{x_n}{y_n}\right)^{1/\alpha} .$$

∎

Lemma 2.3.4 *Let X be a nonnegative random variable such that $X \in \mathbf{RV}_\alpha$ with $\alpha > 1$. Then*

$$\lim_{x\to\infty} \mathbb{E}\left[\frac{X}{x}\middle| X > x\right] = \frac{\alpha}{\alpha - 1} . \tag{2.3.4}$$

Proof: $X \in \mathbf{RV}_\alpha$ means that $\overline{F}(x) = x^{-\alpha}\,\ell(x)$, where ℓ is slowly varying at infinity. Using Karamata Theorem 2.1.6 and since $\lim_{s\to\infty} s\overline{F}(s) = 0$, for $\alpha > 1$ and $x \geq 1$, and using integration by parts, we get

$$
\begin{aligned}
\mathbb{E}\left[\frac{X}{x}\middle| X > x\right] &= \frac{1}{x\,\overline{F}(x)}\left(-\int_x^\infty s\,\overline{F}(ds)\right)\\
&= \frac{1}{x\,\overline{F}(x)}\left(-\,s\,\overline{F}(s)\Big|_x^\infty + \int_x^\infty \overline{F}(s)ds\right)\\
&= \frac{1}{x\,\overline{F}(x)}\left(x\,\overline{F}(x) + \int_x^\infty s^{-\alpha}\ell(s)ds\right)\\
&\sim 1 - \frac{\ell(s)}{x\,\overline{F}(x)}\frac{x^{-\alpha+1}}{-\alpha+1} \quad \text{as } x \to \infty,\\
&= 1 + \frac{1}{\alpha-1} = \frac{\alpha}{\alpha-1}\,.
\end{aligned}
$$

∎

2.4 Second order regular variation

Let X be a nonnegative random variable with continuous distribution function F such that $\overline{F} \in \mathbf{RV}_{-\alpha}$ with $\alpha > 0$. This means that, for $t > 0$,

$$\frac{\overline{F}(tx)}{\overline{F}(x)} \longrightarrow t^{-\alpha}, \qquad \text{as } x \longrightarrow \infty\,. \tag{2.4.1}$$

In many applications, in particular statistical applications, it is necessary to know the rate of convergence in (2.4.1). One can write $\overline{F}$ as

$$\overline{F}(x) = c_0\,x^{-\alpha}\exp\left\{\int_1^x \frac{\eta(y)}{y}\mathrm{d}y\right\}\,, \tag{2.4.2}$$

where $\eta(x) \to 0$ as $x \to \infty$ and $c_0 \in (0,\infty)$.

We strengthen (2.4.2) by assuming that $\overline{F}$ is second order regularly varying, as defined in [16].

Definition 2.4.1 *If (2.4.2) holds and if there exists a function η^* bounded, non increasing on $[0,\infty)$ and regularly varying at infinity with index $-\rho$, $\rho \geq 0$, such that $\lim_{t\to\infty} \eta^*(t) = 0$ and there exists $A > 0$ such that for all $s \geq 0$,*

$$|\eta(s)| \leq A\ \eta^*(s)\ , \tag{2.4.3}$$

then the function $\overline{F}$ is said to be of second-order regular variation with index $-\alpha$ and rate function η^*, written $\overline{F} \in 2RV(-\alpha, -\rho)$.

With the help of second order variation, the rate of convergence in (2.4.1) can be controlled.

Lemma 2.4.2 *([30], Lemma 4.1)*
If (2.4.2) and (2.4.3) hold and if η^* is regularly varying at infinity with index $-\rho$, for some $\rho \geq 0$, then for any $\epsilon > 0$, there exists a constant C such that

$$\forall t \geq 0\ , \quad \forall x > 1\ , \quad \left| \frac{\overline{F}(xt)}{\overline{F}(x)} - t^{-\alpha} \right| \leq C\ \eta^*(x) t^{-\alpha-\rho} (t \vee t^{-1})^{\epsilon}\ . \tag{2.4.4}$$

The following example illustrates the meaning of second order regular variation.

Example 2.4.3 Let $\eta^*(s) = s^{-\alpha(\beta-1)}$ for some $(\beta - 1) > 0$. Assume that η is continuous. Then, for some constant $c_0 > 0$

$$\overline{F}(x) = c_0 x^{-\alpha} \left(1 + O(x^{-\alpha(\beta-1)})\right) \quad \text{as } x \to \infty\ , \tag{2.4.5}$$

Indeed, let $\ell \in \mathbf{RV}_0$ at infinity, defined on $[0,\infty)$, such that $\overline{F}(x) = x^{-\alpha}\ \ell(x)$.
By (2.4.2), we can write

$$\ell(t) = \ell(1) \exp\left\{ \int_1^t \eta(s) \frac{\mathrm{d}s}{s} \right\}\ . \tag{2.4.6}$$

This implies that the function ℓ is the solution of the equation

$$\ell(t) = \ell(1) + \int_1^t \eta(s)\ell(s)\,\frac{\mathrm{d}s}{s}\,. \tag{2.4.7}$$

Then, by (2.4.3) and for $C > 0$, we get

$$\begin{aligned}
\overline{F}(x) &= x^{-\alpha}\ell(1) + x^{-\alpha}\int_1^x \eta(s)\ell(s)\,\frac{\mathrm{d}s}{s} \\
&\le x^{-\alpha}\ell(1) + C\,x^{-\alpha}\int_1^x \eta^*(s)\ell(s)\,\frac{\mathrm{d}s}{s} \\
&= x^{-\alpha}\ell(1) + C\,x^{-\alpha}\int_1^x s^{-\alpha(\beta-1)-1}\ell(s)\,\mathrm{d}s \\
&= c_0 x^{-\alpha}\left(1 + O\left(x^{-\alpha(\beta-1)}\right)\right) \quad \text{as } x \to \infty\,.
\end{aligned} \tag{2.4.8}$$

2.5 Weak convergence

2.5.1 Weak convergence in metric spaces

We recall some classical definitions.

Let $(\mathbb{S}, d)$ be a complete, separable metric space, endowed with its Borel σ-algebra, $\mathscr{B}$, of subsets generated by the open sets. Let $\{\mu_n, n \ge 0\}$ be a sequence of probability measures on $\mathbb{S}$. The sequence $\{\mu_n\}$ is said to be tight if for every $\epsilon > 0$ there exists a compact set K such that $\mu_n(K) > 1 - \epsilon$, for all n.

Definition 2.5.1 (Weak convergence)

We say that the sequence $\{\mu_n\}$ of probability measures converges weakly to the non-negative measure μ, written $\mu_n \xrightarrow{\mathscr{W}} \mu$ if for every continuous bounded function $f : \mathbb{S} \to \mathbb{R}$, one has $\mu_n(f) \longrightarrow \mu(f)$.

In this case, μ is automatically a probability measure since $\mu(1) = 1$, and the definition still makes sense if we suppose that μ_n, μ are just finite non-negative measures on $\mathbb{S}$.

Definition 2.5.2 (Convergence in distribution)
If $(X_n, n \geq 0)$ is a sequence of random variables with values in a metric space $(\mathbb{S}, d)$, and defined on possibly different probability spaces $(\Omega_n, \mathscr{F}_n, P_n)$, we say that X_n converges in distribution to a random variable X on $(\Omega, \mathscr{F}, P)$ if the law of X_n converges weakly to that of X. Equivalently, X_n converges in distribution to X, $X_n \xrightarrow{\mathscr{D}} X$, if for every continuous bounded function f, $\mathbb{E}[f(X_n)]$ converges to $\mathbb{E}[f(X)]$.

The following theorem is a good way to think about weak convergence since, it allows one to replace convergence in distribution with almost sure convergence. In a theory that relies heavily on the concept of continuity, this is a big advantage. Almost sure convergence, being pointwise, is very well suited to continuity arguments.

Theorem 2.5.3 (Skorokhod representation theorem)
Let $\xi, \xi_1, \xi_2, \ldots$ be random elements in a separable metric space $(\mathbb{S}, d)$ such that $\xi_n \xrightarrow{\mathscr{D}} \xi$. Then on a suitable probability space, there exist some random elements $\eta \stackrel{d}{=} \xi$ and $\eta_n \stackrel{d}{=} \xi_n$, $n \in \mathbb{N}$, with $\eta_n \xrightarrow{a.s.} \eta$.

2.5.2 Weak convergence in space $\mathscr{D}[0,1]$, $\mathscr{D}[t_0, \infty)$ and $\mathscr{D}(0, \infty)$

Let $\mathscr{C}[0,1]$ be the space of continuous functions on $[0,1]$ endowed with the topology of uniform convergence. Consider $\mathscr{D}[0,1]$ the space of functions x on $[0,1]$ that are right-continuous and have left-hand limits. That is

i. For $0 \leq t < 1$, $x(t+) = \lim_{s \downarrow t} x(s)$ exists and $x(t+) = x(t)$;

ii. For $0 < t \leq 1$, $x(t-) = \lim_{s \uparrow t} x(s)$ exists.

$\mathscr{D}[0,1]$ is endowed with the Skorohod J_1-topology and both $\mathscr{C}[0,1]$ and $\mathscr{D}[0,1]$ are metrizable. Weak convergence on both spaces is discussed in [5] and [6].

In what follows, for $0 < \delta \leq 1$, $x \in \mathscr{D}[0,1]$ and $A \subset [0,1]$, we use the following moduli of continuity :

$$\omega_x(A) = \sup_{s,t\in A}\left|x(s) - x(t)\right| \ , \tag{2.5.1}$$

$$\omega''_x(\delta) = \sup\min\left\{\left|x(t) - x(t_1)\right| \ , \left|x(t_2) - x(t)\right|\right\}, \tag{2.5.2}$$

where the supremum extends over t_1, t, and t_2 satisfying

$$t_1 \le t \le t_2, \qquad t_2 - t_1 \le \delta\,.$$

The following theorem is a special case of Theorem 15.4 of [5]. We shall write $\omega''(X,\delta)$ in place of $\omega''_X(\delta)$, where $\omega''_x(\cdot)$ is defined in (2.5.2).

Theorem 2.5.4 *(cf. [5], Theorem 15.4) A sequence $\{X_n\}$ of random elements in $\mathscr{D}[0,1]$ converges weakly to X in $\mathscr{C}[0,1]$ if*

$$\big(X_n(t_1),\dots,X_n(t_k)\big) \xrightarrow{\mathscr{D}} \big(X(t_1),...,X(t_k)\big) \ , \tag{2.5.3}$$

whenever $t_1,\dots,t_k$ all lie in $[0,1]$, and for each positive ϵ and η, there exist a δ, $0<\delta<1$, and an integer n_0 such that

$$P(\omega''(X_n,\delta)\ge\epsilon)\le\eta, \quad n\ge n_0\,. \tag{2.5.4}$$

Theorem 2.5.5 *([5], Theorem 15.6)*
Let X_n be random elements of $\mathscr{D}[0,1]$ and let $X\in\mathscr{C}[0,1]$. Suppose that (2.5.3) holds and that

$$\mathbb{P}\left\{|X_n(t_2)-X_n(t)|\ge\lambda,\ |X_n(t)-X_n(t_1)|\ge\lambda\right\}\le\frac{1}{\lambda^{2\gamma}}\big[F(t_2)-F(t_1)\big]^{2\alpha}, \tag{2.5.5}$$

for $t_1<t<t_2$ and $n>1$, where $\gamma>0$, $\alpha>1/2$, and F is a nondecreasing, continuous function on $[0,1]$. Then $X_n \xrightarrow{\mathscr{D}} X$ in $\mathscr{D}[0,1]$.

There is a more restrictive version of (2.5.5) involving moments,

$$\mathbb{E}\left\{|X_n(t_2)-X_n(t)|^\gamma\,|X_n(t)-X_n(t_1)|^\gamma\right\} \leq \left[F(t_2)-F(t_1)\right]^{2\alpha}. \tag{2.5.6}$$

In subsequent chapters, we will be consider weak convergence on the spaces $\mathscr{D}[t_0,\infty)$ and $\mathscr{D}(0,\infty)$ of functions on $[t_0,\infty)$, respectively on $(0,\infty)$, that are right continuous and have left-hand limits. Our discussion of $\mathscr{D}[0,1]$ adapts readily to $\mathscr{D}[t_0,T]$ for $0 < t_0 < T < \infty$. Theorem 16.8 of [6] ensures that if $X_n \xrightarrow{\mathscr{D}} X$ in $\mathscr{D}[t_0,T]$ for all $T<\infty$, then $X_n \xrightarrow{\mathscr{D}} X$ in $\mathscr{D}[t_0,\infty)$. Similarly, if $X_n \xrightarrow{\mathscr{D}} X$ in $\mathscr{D}[t_0,T]$ for all $0<t_0<T<\infty$, then $X_n \xrightarrow{\mathscr{D}} X$ in $\mathscr{D}(0,\infty)$.

Theorem 2.5.5 can be extended to $\mathscr{D}[t_0,\infty)$ and condition (2.5.5) can be weakened to (2.5.9) below, where the functions F_n are all monotone increasing or all monotone decreasing, and the sequence (F_n) converges (uniformly on compact intervals) to a monotone continuous function F. Likewise, F can be replaced by F_n on the right hand side of (2.5.6). That is,

$$\mathbb{E}\left\{|X_n(t_2)-X_n(t)|^\gamma\,|X_n(t)-X_n(t_1)|^\gamma\right\} \leq \left[F_n(t_2)-F_n(t_1)\right]^{2\alpha}. \tag{2.5.7}$$

Lemma 2.5.6 (Extension of Theorem 2.5.5) *Suppose that*

$$\big(X_n(t_1),\ldots,X_n(t_k)\big) \xrightarrow{\mathscr{D}} \big(X(t_1),...,X(t_k)\big) \tag{2.5.8}$$

holds whenever $t_1,\ldots,t_k>0$ and that

$$\mathbb{P}\left\{|X_n(t_2)-X_n(t)|\geq\lambda, |X_n(t)-X_n(t_1)|\geq\lambda\right\} \leq \frac{1}{\lambda^{2\gamma}}\left[F_n(t_2)-F_n(t_1)\right]^{2\alpha}, \tag{2.5.9}$$

for $t_0<t_1<t<t_2<\infty$ and $n>1$, where $\gamma>0$, $\alpha>1/2$, and F_n is a sequence of monotone (all increasing or decreasing) functions converging to a continuous function F on $(0,\infty)$. Then, $X_n \xrightarrow{\mathscr{D}} X$ in $\mathscr{D}(0,\infty)$.

Comment: If F_n is increasing (respectively, decreasing) for every n and the sequence (F_n) converges pointwise to a continuous function F (which must be increasing or decreasing, respectively), then the convergence is uniform on compact sets.

Proof: By the discussion preceding Lemma 2.5.6 and the comment above, it is enough to restrict our attention to $\mathscr{D}[0,1]$.
Without loss of generality, by Theorem 15.4 in [6], it suffices to show that for each positive ϵ and η there exists δ, $0<\delta<1$ and n_0 such that

$$P(\omega''(X_n,\delta)\geq\epsilon)\leq\eta,\quad n\geq n_0\,. \tag{2.5.10}$$

Follow exactly the proof in [5] pp. 129-130, replacing F with F_n in (15.24), (15.26), (15.29) and (15.30). The equation (15.30) of [5] now becomes

$$P\{\omega''(X_n,\delta)\geq\epsilon\}\leq\frac{2K}{\epsilon^{2\gamma}}\left[|F_n(1)-F_n(0)|\right]\left[\omega_{F_n}(2\delta)\right]^{2\alpha-1}\,,$$

where K depends on ϵ and α, but not on n (see [5], Theorem 12.5). Note that for any $t_1,t_2\in[0,1]$, $|F_n(t_2)-F_n(t_1)|\leq|F(t_2)-F(t_1)|+2\sup_{t\in[0,1]}|F_n(t)-F(t)|$. Therefore, letting $\rho_n=2\sup_{t\in[0,1]}|F_n(t)-F(t)|$, we have

$$\omega_{F_n}(\delta)\leq\omega_F(\delta)+\rho_n\,, \tag{2.5.11}$$

and

$$P\{\omega''(X_n,\delta)\geq\epsilon\}\leq\frac{2K}{\epsilon^{2\gamma}}\left[|F(1)-F(0)|+\rho_n\right]\left[\omega_F(2\delta)+\rho_n\right]^{2\alpha-1}\,.$$

Since $\rho_n\to0$ as $n\to\infty$ (by the aforementioned uniform convergence) and $\omega_F(\delta)\to0$ as $\delta\to0$ (by continuity of F), given $\eta>0$ we can choose n_0 and δ so that the right hand side of (2.5.11) is less than η for all $n\geq n_0$. This proves (2.5.10) under the assumption (2.5.9). ∎

2.5.3 Inverses and composition

Lemma 2.5.7 *([14], Lemma A.0.2)* **(Vervaat lemma - convergence of inverses)** *Suppose y is a continuous function and $x_1, x_2, \ldots$ are nondecreasing functions on some interval $[a, b]$. Further, let g be a function on $[a, b]$ with a positive derivative g'. Let δ_n be a sequence of positive numbers such that $\delta_n \longrightarrow 0$, $n \longrightarrow \infty$, and*

$$\lim_{n\to\infty} \frac{x_n(s) - g(s)}{\delta_n} = y(s)\,, \tag{2.5.12}$$

uniformly on $[a, b]$. Then

$$\lim_{n\to\infty} \frac{x_n^{\leftarrow}(s) - g^{\leftarrow}(s)}{\delta_n} = -(g^{\leftarrow})'(s)\; y(g^{\leftarrow}(s))\,, \tag{2.5.13}$$

uniformly on $[g(a), g(b)]$, where $g^{\leftarrow}, x_1^{\leftarrow}, x_2^{\leftarrow}, \ldots$ are the inverse functions (right- or left-continuous or defined in any way consistent with monotonicity).

Lemma 2.5.8 *([5], Theorem 4.4)*
If $X'_n \xrightarrow{\mathscr{D}} X'$ and $X''_n \xrightarrow{P} a''$, then $(X'_n, X''_n) \xrightarrow{\mathscr{D}} (X', a'')$.

Let T be a subinterval of the real line. Let S be a complete separable metric space with metric m. Let $\mathscr{D} = \mathscr{D}(T) = \mathscr{D}(T, S)$ be the set of all right-continuous S-valued functions on T with limits from the left. Let $\mathscr{C} = \mathscr{C}(T, S)$ be the subset of continuous functions in $\mathscr{D}$. Now, let T_1, T_2, T_3 be subintervals of $\mathbb{R}$ and suppose $T_2 \subseteq T_3$. Let $\mathscr{D}_0 = \mathscr{D}_0(T_1, T_2)$ be the subset of nondecreasing T_2-valued functions in $\mathscr{D}$; let $\mathscr{C}_0 = \mathscr{C}_0(T_1, T_2)$ be the subset of strictly-increasing T_2-valued functions in $\mathscr{C}$.

Theorem 2.5.9 *([42], Theorem 3.1)*
The composition mapping on $\mathscr{D}(T_3, S) \times \mathscr{D}_0(T_1, T_2)$ is continuous at each $(x, y) \in (\mathscr{C} \times \mathscr{D}_0) \cup (\mathscr{D} \times \mathscr{C}_0)$.

The following Lemma is a version of Theorem 2.5.9 for nonincreasing functions.

Lemma 2.5.10 (Continuity of composition map)
Let $\mathscr{D}_1(0,\infty)$ be the set of non-increasing functions in $\mathscr{D}(0,\infty)$ and $\mathscr{C}_1(0,\infty)$ be the set of continuous non-increasing and positive functions in $\mathscr{C}(0,\infty)$. Then the composition map from $(\mathscr{D}(0,\infty)\times\mathscr{D}_1(0,\infty))$ is continuous at $(\mathscr{C}(0,\infty)\times\mathscr{C}_1(0,\infty))$, where $\mathscr{C}(0,\infty)$ is endowed with the topology of uniform convergence on compact sets.

Proof: Suppose that $\gamma\in\mathscr{C}(0,\infty)$ and $\varphi\in\mathscr{C}_1(0,\infty)$. Then $\gamma\circ\varphi$ lies in $\mathscr{C}(0,\infty)$. Also, suppose that

$$\gamma_n \longrightarrow \gamma \qquad \text{and} \qquad \varphi_n \longrightarrow \varphi$$

uniformly on compact sets. Now, for any compact set $[a,b]$, $0<a<b<\infty$, since $\varphi_n\longrightarrow\varphi$ uniformly in $[a,b]$ and by continuity of φ, there exists a compact set $[c,d]$ in $(0,\infty)$ such that for any $t\in[a,b]$, $\varphi_n(t)$ and $\varphi(t)$ lie in $[c,d]$, for all n sufficiently large. We have

$$\sup_{a\le t\le b}\big|\gamma_n(\varphi_n(t))-\gamma(\varphi(t))\big| \le \sup_{a\le t\le b}\big|\gamma_n(\varphi_n(t))-\gamma(\varphi_n(t))\big| + \sup_{a\le t\le b}\big|\gamma(\varphi_n(t))-\gamma(\varphi(t))\big|\,.$$

Fix $\epsilon>0$, then since γ is uniformly continuous on $[c,d]$, there exists $\delta_\epsilon>0$ such that for $u,v\in[c,d]$, $|u-v|<\delta_\epsilon$ implies that $\big|\gamma(u)-\gamma(v)\big|<\epsilon/2$.
Choose n_ϵ large enough that for all $n\ge n_\epsilon$,

- $\varphi_n(t),\varphi(t)$ lie in $[c,d]$ for all $t\in[a,b]$,
- $\sup_{a\le t\le b}\big|\varphi_n(t)-\varphi(t)\big|<\delta_\epsilon$,
- $\big|\gamma(u)-\gamma(v)\big|<\epsilon/2$, for all $u,v\in[c,d]$, $|u-v|<\delta_\epsilon$,
- $\sup_{c\le s\le d}\big|\gamma_n(s)-\gamma(s)\big|<\epsilon/2$.

Hence, for $n\ge n_\epsilon$,

$$\sup_{a\le t\le b}\big|\gamma_n(\varphi_n(t))-\gamma(\varphi_n(t))\big| \le \sup_{c\le s\le d}\big|\gamma_n(s)-\gamma(s)\big| < \frac{\epsilon}{2}\,.$$

Furthermore,

$$\sup_{a\le t\le b} \left|\gamma(\varphi_n(t)) - \gamma(\varphi(t))\right| \le \sup_{\substack{u,v\in[c,d] \\ |u-v|<\delta_\epsilon}} \left|\gamma(u) - \gamma(v)\right| < \frac{\epsilon}{2}\,.$$

Therefore, $\gamma_n \circ \varphi_n \longrightarrow \gamma \circ \varphi$, uniformly on $[a, b]$. Since $[a, b]$ is arbitrary,

$$\gamma_n \circ \varphi_n \longrightarrow \gamma \circ \varphi\,,$$

in $\mathscr{D}(0, \infty)$. ∎

Lemma 2.5.10 implies the following corollary.

Corollary 2.5.11 *For each n, let Γ_n be random elements of $\mathscr{D}(0, \infty)$ and Φ_n random elements of $\mathscr{D}_1(0, \infty)$. Suppose that*

$$(\Gamma_n, \Phi_n) \xrightarrow{\mathscr{D}} (\Gamma, \Phi)\,,$$

and that

$$\mathbb{P}\{\Gamma \in C(0, \infty)\} = \mathbb{P}\{\Phi \in C_1(0, \infty)\} = 1\,.$$

Then

$$\Gamma_n \circ \Phi_n \xrightarrow{\mathscr{D}} \Gamma \circ \Phi\,,$$

in $\mathscr{D}(0, \infty)$.

∎

Chapter 3

Weak Convergence of the Tail Empirical Process

Based on the results of Rootzén [39], this chapter represents the cornerstone of the thesis, detailing the theoretical properties and applications of the tail empirical process (TEP). In Section 3.1, we prove the convergence in distribution of the TEP with deterministic levels, $G_n(\cdot)$ (Theorem 3.1.1), and the TEP with random levels, $\widehat{G}_n(\cdot)$ (Theorem 3.1.9). These theorems are particular cases of results in [39].

These theorems are used in Section 3.2 to show convergence in distribution of the integral functionals $\int_1^\infty (G_n(s)/s^m)\,\mathrm{d}s$ (Theorem 3.2.1) and $\int_1^\infty (\widehat{G}_n(s)/s^m)\,\mathrm{d}s$ (Theorem 3.2.2). These results are new and are key to all the applications presented in Section 3.3; in particular, the Hill estimator, the harmonic mean estimators and estimators of the conditional mean.

3.1 Tail empirical process

3.1.1 Tail Empirical Process with deterministic levels

Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid sequence of random variables with continuous marginal distribution F which has a regularly varying right tail with index $\alpha > 0$, that is, for $t > 0$,

$$\frac{\overline{F}(tx)}{\overline{F}(x)} \longrightarrow T(t) = t^{-\alpha}, \qquad \text{as } x \longrightarrow \infty . \tag{3.1.1}$$

Let $u_n > 0$ be a non decreasing sequence such that

$$u_n \longrightarrow \infty , \qquad n\overline{F}(u_n) \longrightarrow \infty . \tag{3.1.2}$$

Define the **tail empirical distribution function** by

$$\widetilde{T}_n(t) = \frac{1}{n\overline{F}(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n t\}}, \quad t > 0 . \tag{3.1.3}$$

We note that the dependence of $\widetilde{T}_n$ on the sequence u_n is suppressed in the notation. Then, as $n \longrightarrow \infty$, we have

$$T_n(t) = \mathbb{E}[\widetilde{T}_n(t)] = \frac{\overline{F}(u_n t)}{\overline{F}(u_n)} \longrightarrow t^{-\alpha} . \tag{3.1.4}$$

Define the **tail empirical process** by

$$G_n(t) = \sqrt{n\overline{F}(u_n)} \left\{ \widetilde{T}_n(t) - T_n(t) \right\}, \quad t > 0 . \tag{3.1.5}$$

Our first goal is to obtain a weak convergence theorem for the process G_n.

Theorem 3.1.1 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution function F. Suppose the conditions (3.1.1) and (3.1.2) hold. Then, as $n \longrightarrow \infty$*

$$G_n(\cdot) \xrightarrow{\mathscr{D}} G(\cdot)\,, \tag{3.1.6}$$

in $\mathscr{D}(0,\infty)$, where $G(\cdot)$ is a centered Gaussian process with covariance function

$$\mathbb{C}ov\left(G(t_1), G(t_2)\right) = (t_1 \vee t_2)^{-\alpha}\ .$$

Note that Theorem 3.1.1 is a special case of Theorem 2.1 in [39].

In order to prove the above theorem, we verify finite dimensional convergence and tightness. The first task is achieved by Lemmas 3.1.2 and 3.1.3, while tightness is proven in Lemma 3.1.4.

Lemma 3.1.2 *Under the conditions of Theorem 3.1.1, for each $t > 0$,*

$$G_n(t) \xrightarrow{\mathscr{D}} G(t)\,.$$

Proof: We check the central limit theorem by using Lindeberg's conditions. Let $Z_{j,n}(\cdot)$ be a random process defined by

$$Z_{j,n}(t) = \frac{1}{\sqrt{n\overline{F}(u_n)}}\left(\mathbb{1}_{\{X_j > u_n t\}} - \overline{F}(u_n t)\right),\quad t > 0\,, \tag{3.1.7}$$

so that

$$G_n(t) = \sum_{j=1}^{n} Z_{j,n}(t)\,. \tag{3.1.8}$$

Since $\{Z_{j,n}(t), j \in \mathbb{Z}\}$ are centred and independent, we write

$$
\begin{aligned}
\mathbb{C}ov\left(G_n(t_1), G_n(t_2)\right) &= \mathbb{C}ov\left(\sum_{j=1}^{n} Z_{j,n}(t_1), \sum_{j=1}^{n} Z_{j,n}(t_2)\right) \\
&= \sum_{j=1}^{n} \mathbb{C}ov\left(Z_{j,n}(t_1), Z_{j,n}(t_2)\right) \\
&= \frac{1}{\overline{F}(u_n)} \mathbb{C}ov\left\{\left(\mathbb{1}_{\{X_j > u_n t_1\}} - \overline{F}(u_n t_1)\right), \left(\mathbb{1}_{\{X_j > u_n t_2\}} - \overline{F}(u_n t_2)\right)\right\} \\
&= \frac{1}{\overline{F}(u_n)} \left(\mathbb{E}[\mathbb{1}_{\{X_j > u_n t_1\}} \mathbb{1}_{\{X_j > u_n t_2\}}] - \mathbb{E}[\mathbb{1}_{\{X_j > u_n t_1\}}]\, \mathbb{E}[\mathbb{1}_{\{X_j > u_n t_2\}}]\right) \\
&= \frac{\overline{F}(u_n(t_1 \vee t_2))}{\overline{F}(u_n)} - \frac{\overline{F}(u_n t_1)}{\overline{F}(u_n)} \overline{F}(u_n t_2) \longrightarrow (t_1 \vee t_2)^{-\alpha}, \qquad \text{as } n \longrightarrow \infty\,. \quad (3.1.9)
\end{aligned}
$$

Therefore,

$$
\mathbb{V}ar[G_n(t)] \longrightarrow t^{-\alpha}, \qquad \text{as } n \longrightarrow \infty\,. \tag{3.1.10}
$$

Now, we evaluate the second Lindeberg condition. By the Cauchy-Schwartz inequality,

$$
\sum_{j=1}^{n} \mathbb{E}\left[Z_{j,n}^2(t) \mathbb{1}_{\{|Z_{j,n}(t)| > \epsilon\}}\right] \leq \sum_{j=1}^{n} \left(\mathbb{E}[Z_{j,n}^4(t)]\right)^{\frac{1}{2}} \left(\mathbb{P}(|Z_{j,n}(t)| > \epsilon)\right)^{\frac{1}{2}}\,. \tag{3.1.11}
$$

On the other hand, we have

$$
\begin{aligned}
\left(\mathbb{E}[Z_{j,n}^4(t)]\right)^{\frac{1}{2}} &= \frac{1}{n\overline{F}(u_n)} \left(\mathbb{E}\left[\left(\mathbb{1}_{\{X_j > u_n t\}} - \overline{F}(u_n t)\right)^4\right]\right)^{\frac{1}{2}} \\
&= \frac{1}{n\overline{F}(u_n)} \left(\overline{F}(u_n t) - 4(\overline{F}(u_n t))^2 + 6(\overline{F}(u_n t))^3 - 4(\overline{F}(u_n t))^4 + (\overline{F}(u_n t))^4\right)^{\frac{1}{2}} \\
&\leq \frac{1}{n\overline{F}(u_n)} \left(16\overline{F}(u_n t)\right)^{\frac{1}{2}} \\
&= 4\frac{\left(\overline{F}(u_n t)\right)^{\frac{1}{2}}}{n\overline{F}(u_n)}\,. \qquad (3.1.12)
\end{aligned}
$$

Using Markov's inequality, we obtain

$$\left(\mathbb{P}\left(\left|Z_{j,n}(t)\right| > \epsilon\right)\right)^{\frac{1}{2}} \le \left(\frac{\mathbb{E}[Z_{j,n}^4(t)]}{\epsilon^4}\right)^{\frac{1}{2}} \le \frac{4}{\epsilon^2}\frac{\left(\overline{F}(u_n t)\right)^{\frac{1}{2}}}{n\overline{F}(u_n)}.$$

Then, by (3.1.1) and (3.1.2), we get

$$\begin{aligned}
\sum_{j=1}^{n}\mathbb{E}\left[Z_{j,n}^2(t)\mathbb{1}_{\{\left|Z_{j,n}(t)\right|>\epsilon\}}\right] &\le \sum_{j=1}^{n}\left(\mathbb{E}[Z_{j,n}^4(t)]\right)^{\frac{1}{2}}\left(\mathbb{P}(\left|Z_{j,n}(t)\right| > \epsilon)\right)^{\frac{1}{2}} \\
&\le \frac{16}{\epsilon^2}\frac{\overline{F}(u_n t)}{\overline{F}(u_n)}\frac{1}{n\overline{F}(u_n)} \longrightarrow 0, \quad \text{as } n \longrightarrow \infty.
\end{aligned}$$

∎

Lemma 3.1.3 *Under the conditions of Theorem 3.1.1, for fixed* $t_1, \dots, t_k$

$$\left(G_n(t_1), \dots, G_n(t_k)\right) \xrightarrow{\mathscr{D}} \left(G(t_1), \dots, G(t_k)\right).$$

Proof: For clarity, we prove the theorem for $k = 2$. Assume that $t_2 > t_1$. We will use the Cramer-Wold device, that is for $a, b \in \mathbb{R}$, we need to verify Lindeberg's conditions for $a\,G_n(t_1) + b\,G_n(t_2)$.

First, we compute the limiting variance. Using (3.1.9)

$$\begin{aligned}
\mathbb{V}ar\left[aG_n(t_1) + bG_n(t_2)\right] &= a^2\mathbb{V}ar\left[G_n(t_1)\right] + b^2\mathbb{V}ar\left[G_n(t_2)\right] \\
&\quad + 2ab\,\mathbb{C}ov\left(G_n(t_1), G_n(t_2)\right) \\
&\longrightarrow a^2 t_1^{-\alpha} + b^2 t_2^{-\alpha} + 2ab(t_1 \vee t_2)^{-\alpha}, \qquad \text{as } n \longrightarrow \infty.
\end{aligned}$$

Second, let $Z_{j,n}(\cdot,\cdot)$ be a random process defined by

$$Z_{j,n}(t_1, t_2) = a\,Z_{j,n}(t_1) + b\,Z_{j,n}(t_2), \quad t_1, t_2 > 0. \tag{3.1.13}$$

Then,

$$a\,G_n(t_1) + b\,G_n(t_2) = \sum_{j=1}^{n} Z_{j,n}(t_1,t_2)\,. \tag{3.1.14}$$

Note that,

$$\left|Z_{j,n}^4(t_1,t_2)\right| \le 16\,a^4\left|Z_{j,n}^4(t_1)\right| + 16\,b^4\left|Z_{j,n}^4(t_2)\right|\,. \tag{3.1.15}$$

By the Cauchy-Schwartz inequality we get

$$\sum_{j=1}^{n}\mathbb{E}\left[Z_{j,n}^2(t_1,t_2)\mathbb{1}_{\{|Z_{j,n}(t_1,t_2)|>\epsilon\}}\right] \le \sum_{j=1}^{n}\left(\mathbb{E}[Z_{j,n}^4(t_1,t_2)]\right)^{\frac{1}{2}}\left(\mathbb{P}(|Z_{j,n}(t_1,t_2)|>\epsilon)\right)^{\frac{1}{2}}\,.$$

Now, since $t_2 > t_1$ and using (3.1.12) and (3.1.15), we write

$$\begin{aligned}
\left(\mathbb{E}[Z_{j,n}^4(t_1,t_2)]\right)^{\frac{1}{2}} &\le 4\,a^2\left(\mathbb{E}[Z_{j,n}^4(t_1)]\right)^{\frac{1}{2}} + 4\,b^2\left(\mathbb{E}[Z_{j,n}^4(t_2)]\right)^{\frac{1}{2}} \\
&\le 16\,a^2\frac{\left(\overline{F}(u_nt_1)\right)^{\frac{1}{2}}}{n\overline{F}(u_n)} + 16\,b^2\frac{\left(\overline{F}(u_nt_2)\right)^{\frac{1}{2}}}{n\overline{F}(u_n)} \\
&\le C_1\,\frac{\left(\overline{F}(u_nt_1)\right)^{\frac{1}{2}}}{n\overline{F}(u_n)}\,.
\end{aligned} \tag{3.1.16}$$

Also, by Markov's inequality and (3.1.16)

$$\left(\mathbb{P}\left(\left|Z_{j,n}(t_1,t_2)\right|>\epsilon\right)\right)^{\frac{1}{2}} \le \left(\frac{\mathbb{E}[Z_{j,n}^4(t_1,t_2)]}{\epsilon^4}\right)^{\frac{1}{2}} \le \frac{C_1}{\epsilon^2}\,\frac{\left(\overline{F}(u_nt_1)\right)^{\frac{1}{2}}}{n\overline{F}(u_n)}\,.$$

Therefore, by (3.1.1) and (3.1.2), one can conclude that as $n \longrightarrow \infty$,

$$\sum_{j=1}^{n}\mathbb{E}\left[Z_{j,n}^2(t_1,t_2)\mathbb{1}_{\{|Z_{j,n}(t_1,t_2)|>\epsilon\}}\right] \le C\,\frac{\overline{F}(u_nt_1)}{\overline{F}(u_n)}\,\frac{1}{n\overline{F}(u_n)} \longrightarrow 0\,.$$

∎

Lemma 3.1.4 *Under the conditions of Theorem 3.1.1, the tail empirical process $G_n(\cdot)$ defined in (3.1.5) is tight.*

Proof: Since convergence in $\mathscr{D}(0,\infty)$ is implied by convergence in $\mathscr{D}[a,b]$ for any $0<a<b<\infty$, it suffices to show that for $a \le t_1 < s < t_2 \le b$,

$$\mathbb{E}[|G_n(t_1)-G_n(s)|^2\,|G_n(s)-G_n(t_2)|^2] \le C\big|\xi_n(t_2)-\xi_n(t_1)\big|^2\,, \tag{3.1.17}$$

where C is constant and ξ_n is a sequence of deterministic functions that converge uniformly on $[a,b]$ to a monotone continuous function ξ, cf. Lemma 2.5.6 . We have

$$\begin{aligned}
|G_n(t_1)-G_n(s)|^2 &= \frac{1}{n\overline{F}(u_n)}\left\{\sum_{j=1}^n\Big(\mathbb{1}_{\{X_j>u_nt_1\}}-\mathbb{1}_{\{X_j>u_ns\}}-\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\}\Big)\right\}^2\\
&= \frac{1}{n\overline{F}(u_n)}\left\{\sum_{j=1}^n\Big(\mathbb{1}_{\{u_nt_1<X_j<u_ns\}}-\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\}\Big)\right\}^2 .
\end{aligned}$$

Put

$$U_j = \mathbb{1}_{\{u_nt_1<X_j<u_ns\}}-\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\}\,, \tag{3.1.18}$$

$$V_j = \mathbb{1}_{\{u_ns<X_j<u_nt_2\}}-\{\overline{F}(u_ns)-\overline{F}(u_nt_2)\}\,. \tag{3.1.19}$$

Let $p_1=\overline{F}(u_nt_1)-\overline{F}(u_ns)$, $p_2=\overline{F}(u_ns)-\overline{F}(u_nt_2)$ and $p=\overline{F}(u_nt_1)-\overline{F}(u_nt_2)$, then $p_1\le p$ and $p_2\le p$. Hence,

$$\begin{aligned}
&\mathbb{E}[|G_n(t_1)-G_n(s)|^2\,|G_n(s)-G_n(t_2)|^2]\\
&\qquad= \frac{1}{(n\overline{F}(u_n))^2}\left(\mathbb{E}\left[\left(\sum_{j=1}^n U_j\right)^2\left(\sum_{j=1}^n V_j\right)^2\right]\right)\\
&\qquad= \frac{1}{(n\overline{F}(u_n))^2}\sum_{i=1}^n\mathbb{E}[(U_iV_i)^2]+\frac{1}{(n\overline{F}(u_n))^2}\sum_{\substack{i,j=1\\ i\ne j}}^n\mathbb{E}[U_i^2]\mathbb{E}[V_j^2]\,.
\end{aligned}$$

We have that $\mathbb{E}[U_i^2] = p_1 - p_1^2 \leq p$, $\mathbb{E}[V_i^2] = p_2 - p_2^2 \leq p$ and

$$
\begin{aligned}
(U_iV_i)^2 &= \mathbb{1}_{\{u_nt_1<X_j<u_ns\}}\{\overline{F}(u_ns)-\overline{F}(u_nt_2)\}^2 + \mathbb{1}_{\{u_ns<X_j<u_nt_2\}}\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\}^2 \\
&\quad +\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\}^2\{\overline{F}(u_ns)-\overline{F}(u_nt_2)\}^2 \\
&\quad -2\cdot\mathbb{1}_{\{u_nt_1<X_j<u_ns\}}\{\overline{F}(u_ns)-\overline{F}(u_nt_2)\}^2\cdot\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\} \\
&\quad -2\cdot\mathbb{1}_{\{u_ns<X_j<u_nt_2\}}\{\overline{F}(u_nt_1)-\overline{F}(u_ns)\}^2\cdot\{\overline{F}(u_ns)-\overline{F}(u_nt_2)\} \\
&= \mathbb{1}_{\{u_nt_1<X_j<u_ns\}}p_2^2 + \mathbb{1}_{\{u_ns<X_j<u_nt_2\}}p_1^2 + p_1^2p_2^2 \\
&\quad -2\cdot\mathbb{1}_{\{u_nt_1<X_j<u_ns\}}p_1p_2^2 - 2\cdot\mathbb{1}_{\{u_ns<X_j<u_nt_2\}}p_2p_1^2\,.
\end{aligned}
$$

So,

$$
\mathbb{E}[(U_iV_i)^2] = p_1p_2^2 + p_2p_1^2 + p_1^2p_2^2 - 2p_1^2p_2^2 - 2p_1^2p_2^2 \leq 3p^3 \leq 3p^2\,.
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}[|G_n(t_1)-G_n(s)|^2\,|G_n(s)-G_n(t_2)|^2] \\
&\qquad = \frac{1}{(n\overline{F}(u_n))^2}\sum_{i=1}^{n}\mathbb{E}[(U_iV_i)^2] + \frac{1}{(n\overline{F}(u_n))^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\mathbb{E}[U_i^2]\mathbb{E}[V_j^2] \\
&\qquad \leq \frac{3np^2}{(n\overline{F}(u_n))^2} + \frac{n(n-1)p^2}{(n\overline{F}(u_n))^2} \\
&\qquad = \frac{3}{n}\left|\frac{\overline{F}(u_nt_1)-\overline{F}(u_nt_2)}{\overline{F}(u_n)}\right|^2 + \frac{n^2-n}{n^2}\left|\frac{\overline{F}(u_nt_1)-\overline{F}(u_nt_2)}{\overline{F}(u_n)}\right|^2 \\
&\qquad \leq 4\left|\frac{\overline{F}(u_nt_1)-\overline{F}(u_nt_2)}{\overline{F}(u_n)}\right|^2 = 4\big|T_n(t_2)-T_n(t_1)\big|^2\,.
\end{aligned}
$$

Since the T_n's are monotone and T is continuous, Lemma 2.5.6 implies tightness of $G_n(\cdot)$. ∎

Proof of Theorem 3.1.1. Lemma 3.1.3 yields finite-dimensional convergence. Lemma 3.1.4 yields tightness. This conclude the proof. ∎

3.1.2 Tail Empirical Process with random levels

The tail empirical process considered in (3.1.5) uses deterministic levels u_n that need to fulfill the condition $n\overline{F}(u_n) \to \infty$. In other words, we need to know F in order to choose the appropriate level. Secondly, the centering $T_n(t)$ depends on n and we would like to replace it with its limit $T(t)$. Throughout the sequel, we assume that F is continuous.

To address the first issue, let $X_{n:n} \geq X_{n:n-1} \geq \cdots \geq X_{n:1}$ be the order statistics from the sample $X_1, \ldots, X_n$. Choose a sequence of integers $k = k_n \longrightarrow \infty$ such that $k = o(n)$, that is $k/n \longrightarrow 0$ as $n \to \infty$ and choose u_n so that $k = n\overline{F}(u_n)$.
Define

$$\widehat{T}_n(t) = \frac{1}{k} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{n:n-k} t\}}\,, \quad t > 0\,. \tag{3.1.20}$$

To deal with the second issue, we will require the following condition that can be expressed in terms of u_n or k : for all $\epsilon > 0$,

$$\lim_{n\to\infty} \sup_{t \geq \epsilon} \sqrt{n\overline{F}(u_n)}\,\left|T_n(t) - T(t)\right| = 0\,. \tag{3.1.21}$$

Although (3.1.21) appears to be an ad hoc condition, it is implied by second order regular variation (see Section 2.4). To see this, we use the following lemma.

Lemma 3.1.5 *([30], Proposition 2.8)*
If $\overline{F} \in 2RV(-\alpha, -\rho)$, then for $\epsilon > 0$

$$\sup_{t \geq \epsilon} \left| \frac{\overline{F}(u_n t)}{\overline{F}(u_n)} - t^{-\alpha} \right| = O\left(\eta^*(u_n)\right)\ . \tag{3.1.22}$$

Corollary 3.1.6 *If $\overline{F} \in 2RV(-\alpha, -\rho)$ and $\sqrt{n\overline{F}(u_n)}\ \eta^*(u_n) \to 0$, then condition (3.1.21) holds.*

This motivates the introduction of the following condition, that will be referred to as the *no-bias condition*

$$\sqrt{n\overline{F}(u_n)}\;\eta^*(u_n) = \sqrt{k}\;\eta^*(u_n) = o(1)\,, \qquad \text{as } n \to \infty\,. \tag{3.1.23}$$

Example 3.1.7 This is a continuation of Example 2.4.3. One can deduce that if we put $k = n\overline{F}(u_n)$ and $F^{\leftarrow}(x) = (1-x)^{-1/\alpha}$, we have $u_n = (k/n)^{-1/\alpha}$. Then if $\eta^*(s) = s^{-\alpha(\beta-1)}$, the right hand side of (3.1.22) becomes $O((k/n)^{(\beta-1)})$, and hence the no-bias condition (3.1.23) is satisfied if k is chosen such that $k^{1/2+(\beta-1)}/n^{\beta-1} \to 0$.

Example 3.1.8 If $\overline{F}(x) = x^{-\alpha}\log(x)$ and $\eta^*(x) = 1/\log(x)$ then $\overline{F} \in 2RV(-\alpha, 0)$ and the right hand side of (3.1.22) becomes $O(-\log^{-1}(k/n))$. Hence, the no-bias condition (3.1.23) is satisfied if k is chosen such that $k = \log^{\delta}(n)$ with $\delta < 2$.

Now, we consider the following process

$$\widehat{G}_n(t) = \sqrt{k}\left\{\widehat{T}_n(t) - T(t)\right\}\,, \quad t > 0\,. \tag{3.1.24}$$

Our goal is to prove weak convergence of $\widehat{G}_n$. In order to prove it, we need the assumptions of Theorem 3.1.1 to be fulfilled, together with the additional assumptions (3.1.23) and the following technical condition:

$$T_n^{'}(t) \longrightarrow T^{'}(t) = -\alpha t^{-\alpha-1}\,, \tag{3.1.25}$$

uniformly in a neighborhood of 1.

Theorem 3.1.9 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution F. Suppose the conditions (3.1.1), (3.1.2), (3.1.23) and (3.1.25) hold. Then, in $\mathscr{D}(0,\infty)$ and as $n \longrightarrow \infty$, we have*

$$\widehat{G}_n(\cdot) \xrightarrow{\mathscr{D}} G(\cdot) - G(1)\,T(\cdot)\,. \tag{3.1.26}$$

Note that the proof of Theorem 3.1.9 is given after we introduce some technical tools needed in the proof.

Definition 3.1.10 **(Inverse functions)**
Let $\widetilde{T}_n^{\leftarrow}$ be the right continuous inverse of the tail empirical distribution $\widetilde{T}_n$ defined by

$$\widetilde{T}_n^{\leftarrow}(y) = \sup\{x : \widetilde{T}_n(x) > y\} . \tag{3.1.27}$$

Let $T_n^{\leftarrow}$ be the right continuous inverse of T_n defined by,

$$T_n^{\leftarrow}(y) = \sup\{x : T_n(x) > y\} . \tag{3.1.28}$$

Note that discontinuities of T_n become converted into flat stretches of $T_n^{\leftarrow}$ and flat stretches of T_n into discontinuities of $T_n^{\leftarrow}$.

Convergence of order statistics: We obtain the following result on convergence of order statistics. For this, we require that the conditions of Theorem 3.1.1 are satisfied, together with the differentiability condition (3.1.25). The no-bias condition (3.1.23) is not needed.

Lemma 3.1.11 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution F. Suppose the conditions (3.1.1), (3.1.2) and (3.1.25) hold. Then*

$$\left(G_n, \frac{X_{n:n-k}}{u_n}\right) \xrightarrow{\mathscr{D}} (G, 1) , \tag{3.1.29}$$

in $\mathscr{D}(0,\infty) \times \mathbb{R}$, and

$$\sqrt{k}\left\{\frac{X_{n:n-k}}{u_n} - 1\right\} \xrightarrow{\mathscr{D}} \frac{-G(1)}{T'(1)} = \frac{1}{\alpha}\, G(1) . \tag{3.1.30}$$

Proof: We use the notation introduced in Section 2.5.3. Note that $G_n \in \mathscr{D}(0,\infty)$, $T_n^{\leftarrow} \in \mathscr{D}_1(0,\infty)$, $G \in \mathscr{C}(0,\infty)$, $T^{\leftarrow} \in \mathscr{C}_1(0,\infty)$. Since $T_n^{\leftarrow} \longrightarrow T^{\leftarrow}$ pointwise and hence uniformly on compact sets (by monotonicity of $T_n^{\leftarrow}$ and continuity of $T^{\leftarrow}$) and using Corollary 2.5.11, we have

$$(G_n \circ T_n^{\leftarrow})(\cdot) \xrightarrow{\mathscr{D}} (G \circ T^{\leftarrow})(\cdot)\,, \tag{3.1.31}$$

in $\mathscr{D}(0,\infty)$. Recall that

$$G_n(t) = \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \left(\mathbb{1}_{\{X_j > u_n t\}} - \overline{F}(u_n t) \right) = \sqrt{k} \left\{ \widetilde{T}_n(t) - T_n(t) \right\} .$$

Then,

$$(G_n \circ T_n^{\leftarrow})(t) = \sqrt{k} \left\{ (\widetilde{T}_n \circ T_n^{\leftarrow})(t) - t \right\} .$$

Hence,

$$\sqrt{k} \left\{ (\widetilde{T}_n \circ T_n^{\leftarrow})(\cdot) - (\cdot) \right\} \xrightarrow{\mathscr{D}} (G \circ T^{\leftarrow})(\cdot)\,. \tag{3.1.32}$$

Applying Lemma 2.5.3, one can argue that on some probability space,

$$\sqrt{k} \left\{ (\widetilde{T}_n \circ T_n^{\leftarrow})(\cdot) - (\cdot) \right\} \xrightarrow{a.s.} (G \circ T^{\leftarrow})(\cdot)\,, \tag{3.1.33}$$

uniformly on compact subsets of $(0,\infty)$.

Note that $(\widetilde{T}_n \circ T_n^{\leftarrow})(t)$ is a sequence of nondecreasing functions. We apply Vervaat's Lemma 2.5.7 with $\sqrt{k} = 1/\delta_n$, $(\widetilde{T}_n \circ T_n^{\leftarrow})(t) = x_n(t)$, $g(t) = t$, and $y(t) = (G \circ T^{\leftarrow})(t)$, to (3.1.33), and we conclude that as $n \longrightarrow \infty$,

$$\sqrt{k} \left\{ (\widetilde{T}_n \circ T_n^{\leftarrow})^{\leftarrow}(t) - t \right\} \xrightarrow{a.s.} -(G \circ T^{\leftarrow})(t), \tag{3.1.34}$$

uniformly on compact sets.

Since $(\widetilde{T}_n \circ T_n^{\leftarrow})^{\leftarrow}(t) = (T_n \circ \widetilde{T}_n^{\leftarrow})(t)$, we argue that as $n \longrightarrow \infty$

$$\sqrt{k}\left\{(T_n \circ \widetilde{T}_n^{\leftarrow})(t) - t\right\} \xrightarrow{a.s.} -(G \circ T^{\leftarrow})(t)\,, \tag{3.1.35}$$

uniformly on compact sets.

By (3.1.25) we can apply Taylor's expansion to the left hand side of (3.1.35) to obtain

$$\begin{aligned} T_n\left(\widetilde{T}_n^{\leftarrow}(1)\right) - 1 &= T_n\left(\widetilde{T}_n^{\leftarrow}(1)\right) - T_n\left(T_n^{\leftarrow}(1)\right) \\ &= T_n'\left(T_n^{\leftarrow}(1)\right)\left\{\widetilde{T}_n^{\leftarrow}(1) - T_n^{\leftarrow}(1)\right\}\big(1+o(1)\big)\;. \end{aligned} \tag{3.1.36}$$

Since $\widetilde{T}_n^{\leftarrow}(1) = X_{n:n-k}/u_n$ and $T_n^{\leftarrow}(1) = 1$, we have

$$\begin{aligned} \sqrt{k}\left\{T_n\left(\frac{X_{n:n-k}}{u_n}\right) - 1\right\} &= \sqrt{k}\left\{T_n\left(\widetilde{T}_n^{\leftarrow}(1)\right) - T_n\left(T_n^{\leftarrow}(1)\right)\right\} \\ &= \sqrt{k}\;T_n'\left(T_n^{\leftarrow}(1)\right)\left\{\widetilde{T}_n^{\leftarrow}(1) - T_n^{\leftarrow}(1)\right\}\big(1+o(1)\big) \\ &= \sqrt{k}\;T_n'(1)\left\{\frac{X_{n:n-k}}{u_n} - 1\right\}\big(1+o(1)\big)\;. \end{aligned}$$

Then, as $n \longrightarrow \infty$, using (3.1.35) we get

$$\sqrt{k}\left\{T_n\left(\frac{X_{n:n-k}}{u_n}\right) - 1\right\} = \sqrt{k}\;T_n'(1)\left\{\frac{X_{n:n-k}}{u_n} - 1\right\}\big(1+o(1)\big) \xrightarrow{a.s.} -G(1)\,. \tag{3.1.37}$$

(3.1.37) implies that $X_{n:n-k}/u_n$ converges to 1 in probability. Going back to the original probability space, Theorem 3.1.1 and Lemma 2.5.8 yield (3.1.29). Furthermore, by (3.1.37) and since $T'(t) = -\alpha\, t^{-(\alpha+1)}$ we can conclude that

$$\sqrt{k}\left\{\frac{X_{n:n-k}}{u_n} - 1\right\} \xrightarrow{\mathscr{D}} \frac{-G(1)}{T'(1)} = \frac{1}{\alpha}\,G(1)\,.$$

Hence, (3.1.30) has been proven. ∎

Proof of Theorem 3.1.9: Recall that, for $t > 0$,

$$\widetilde{T}_n(t) = \frac{1}{n\overline{F}(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n t\}} \qquad \text{and} \qquad \widehat{T}_n(t) = \frac{1}{k} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{n:n-k} t\}} .$$

Then, $G_n\left(t\frac{X_{n:n-k}}{u_n}\right) = \sqrt{k}\left\{\widehat{T}_n(t) - T_n\left(t\frac{X_{n:n-k}}{u_n}\right)\right\}$, and so

$$\begin{aligned}
\widehat{G}_n(t) &= \sqrt{k}\left\{\widehat{T}_n(t) - T_n\left(t\frac{X_{n:n-k}}{u_n}\right)\right\} + \sqrt{k}\left\{T_n\left(t\frac{X_{n:n-k}}{u_n}\right) - T\left(t\frac{X_{n:n-k}}{u_n}\right)\right\} \\
&\quad + \sqrt{k}\left\{T\left(t\frac{X_{n:n-k}}{u_n}\right) - T(t)\right\} \\
&= G_n\left(t\frac{X_{n:n-k}}{u_n}\right) + \sqrt{k}\left\{T_n\left(t\frac{X_{n:n-k}}{u_n}\right) - T\left(t\frac{X_{n:n-k}}{u_n}\right)\right\} \\
&\quad + \sqrt{k}\left\{T\left(t\frac{X_{n:n-k}}{u_n}\right) - T(t)\right\} = I_1 + I_2 + I_3 .
\end{aligned}$$

By (3.1.29), G_n and $X_{n:n-k}/u_n$ converge jointly and hence the first term I_1 converges weakly in $\mathscr{D}(0,\infty)$ to $G(\cdot)$, by Theorem 3.1.1. Now, by the no-bias condition (3.1.23),

$$\sqrt{k}\left\{T_n\left(t\frac{X_{n:n-k}}{u_n}\right) - T\left(t\frac{X_{n:n-k}}{u_n}\right)\right\} \xrightarrow{P} 0 , \tag{3.1.38}$$

uniformly in $t > \epsilon$. Hence, the second term I_2 is negligible.

Finally, on account of (3.1.30) and the delta method and since $T'(t) = -\alpha\, t^{-(\alpha+1)}$,

$$\sqrt{k}\left\{T\left(t\,\frac{X_{n:n-k}}{u_n}\right) - T(t)\right\} \xrightarrow{\mathscr{D}} \frac{t}{\alpha}\, G(1)\, T'(t) = -G(1)\, T(t) . \tag{3.1.39}$$

We conclude that

$$\widehat{G}_n(\cdot) \xrightarrow{\mathscr{D}} G(\cdot) - G(1)\, T(\cdot) .$$

in $\mathscr{D}(0,\infty)$. ∎

3.2 Convergence of integral functionals

In this section we are interested in weak convergence of integral functionals

$$\int_1^\infty \frac{G_n(s)}{s^m}\,\mathrm{d}s$$

for integer values of $m \geq 0$. Note that the convergence $x_n \to x$ in $\mathscr{D}(0,\infty)$ does not imply convergence of integrals $\int_1^\infty x_n(t)\,\mathrm{d}t \to \int_1^\infty x(t)\,\mathrm{d}t$. However, if x is continuous, we have $\int_1^M x_n(t)\,\mathrm{d}t \to \int_1^M x(t)\,\mathrm{d}t$ for all $M < \infty$. In order to conclude convergence of integrals over $(1,\infty)$, we need to deal with the tail $\int_M^\infty x_n(t)\,\mathrm{d}t$. These functionals provide a unified approach to all the weak convergence results that follow.

Theorem 3.2.1 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution F. Under the assumptions of Theorem 3.1.1 and if $\alpha > 2(1-m)$, then as $n \longrightarrow \infty$*

$$\int_1^\infty \frac{G_n(t)}{t^m}\,\mathrm{d}t \;\xrightarrow{\mathscr{D}}\; \int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t \sim \mathscr{N}\left(0, \frac{2}{(\alpha+m-1)(\alpha+2m-2)}\right).$$

Proof: First, we argue that the limiting integral is finite. We have

$$\begin{aligned}
\mathbb{V}ar\left[\int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\right] &= \mathbb{E}\left[\left(\int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\right)\left(\int_1^\infty \frac{G(s)}{s^m}\,\mathrm{d}s\right)\right] \\
&= 2\int_1^\infty \int_{t=s}^\infty \mathbb{E}\left[G(t)G(s)\right]\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m} \\
&= 2\int_1^\infty \int_{t=s}^\infty t^{-\alpha}\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m} \\
&= \frac{2}{(\alpha+m-1)(\alpha+2m-2)} < \infty
\end{aligned} \tag{3.2.1}$$

whenever $\alpha > 2(1-m)$. Since any linear functional of a Gaussian process is Gaussian, the limiting distribution is normal. Let $1 < M < \infty$. By Theorem 3.1.1 and the continuous mapping theorem, as $n \longrightarrow \infty$ we have

$$\int_1^M \frac{G_n(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^M \frac{G(t)}{t^m}\,\mathrm{d}t\,. \tag{3.2.2}$$

This is true because the convergence of G_n is in the topology of uniform convergence on compact sets. Moveover, by (3.2.1),

$$\lim_{M\to\infty} \mathbb{V}ar\left[\int_M^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\right] = 0\,.$$

Hence, letting $M \longrightarrow \infty$, we get

$$\int_1^M \frac{G(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\,. \tag{3.2.3}$$

To finish the proof, it is enough to show that for $\epsilon > 0$, (cf. [7], Theorem 25.5)

$$\lim_M \limsup_n \mathbb{P}\left[\left|\int_M^\infty \frac{G_n(t)}{t^m}\,\mathrm{d}t\right| \geq \epsilon\right] = 0\,. \tag{3.2.4}$$

$$\begin{aligned}
\mathbb{P}\left[\left|\int_M^\infty \frac{G_n(s)}{s^m}\,\mathrm{d}s\right| \geq \epsilon\right] &\leq \frac{1}{\epsilon^2}\mathbb{V}ar\left[\int_M^\infty \frac{G_n(s)}{s^m}\,\mathrm{d}s\right] \\
&= \frac{1}{\epsilon^2}\mathbb{V}ar\left[\int_M^\infty \sqrt{k}\left[\frac{1}{k}\sum_{j=1}^n \mathbb{1}_{\{X_j > u_n s\}} - \frac{\overline{F}(u_n s)}{\overline{F}(u_n)}\right]\frac{\mathrm{d}s}{s^m}\right] \\
&= \frac{1}{k\epsilon^2}\mathbb{V}ar\left[\sum_{j=1}^n \int_M^\infty \mathbb{1}_{\{X_j > u_n s\}}\frac{\mathrm{d}s}{s^m}\right] \\
&\leq \frac{n}{k\epsilon^2}\mathbb{E}\left\{\left[\int_M^\infty \mathbb{1}_{\{X_1 > u_n s\}}\frac{\mathrm{d}s}{s^m}\right]^2\right\} \\
&= \frac{2n}{k\epsilon^2}\int_M^\infty \int_{t=s}^\infty \mathbb{P}[X_1 > u_n t]\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m} \\
&= \frac{2}{\epsilon^2}\int_M^\infty \int_{t=s}^\infty \frac{\overline{F}(u_n t)}{\overline{F}(u_n)}\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m}\,.
\end{aligned}$$

If $m \geq 1$, the above expression is bounded above by

$$\frac{2}{\epsilon^2}\int_M^\infty \int_{t=s}^\infty \frac{\overline{F}(u_n t)}{\overline{F}(u_n)}\frac{\mathrm{d}t}{t}\frac{\mathrm{d}s}{s}\,.$$

By Potter's bounds (Corollary 2.1.9), since $\overline{F} \in RV_{-\alpha}$, for any $\delta > 0$, there exists C such that for all $n \geq 1$ and $t > 1$,

$$\frac{\overline{F}(u_n t)}{\overline{F}(u_n)} \leq C\, t^{-\alpha+\delta}\,.$$

Then for $\alpha > \delta$, we get

$$\begin{aligned}
\mathbb{P}\left[\left|\int_M^\infty \frac{G_n(s)}{s^m}\,\mathrm{d}s\right| \geq \epsilon\right] &\leq \frac{2}{\epsilon^2}\int_M^\infty \frac{\overline{F}(u_n t)}{\overline{F}(u_n)}\int_M^t \frac{\mathrm{d}s}{s}\,\frac{\mathrm{d}t}{t} \\
&\leq \frac{2C}{\epsilon^2}\int_M^\infty \ln(t)\, t^{\delta-\alpha-1}\mathrm{d}t \\
&= \frac{2C}{\epsilon^2}\left(\frac{\ln(M)\, M^{\delta-\alpha}}{\alpha-\delta} + \frac{M^{\delta-\alpha}}{(\delta-\alpha)^2}\right) \longrightarrow 0 \quad \text{as } M \to \infty\,.
\end{aligned}$$

If $m = 0$, use Potter's bounds with $0 < \delta < \alpha - 2$. Then

$$\begin{aligned}
\mathbb{P}\left[\left|\int_M^\infty G_n(s)\,\mathrm{d}s\right| \geq \frac{\epsilon}{2}\right] &\leq \frac{2}{\epsilon^2}\int_M^\infty \frac{\overline{F}(u_n t)}{\overline{F}(u_n)}\int_M^t \mathrm{d}s\;\mathrm{d}t \\
&\leq \frac{2C}{\epsilon^2}\int_M^\infty t^{\delta-\alpha+1}\mathrm{d}t \\
&= \frac{2C}{\epsilon^2}\left(\frac{M^{\delta-\alpha+2}}{\alpha-2-\delta}\right) \longrightarrow 0 \quad \text{as } M \to \infty
\end{aligned}$$

whenever $\alpha > \delta + 2$.

Thus, (3.2.4) holds, and the proof is complete. ∎

Theorem 3.2.2 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution F. Under the assumptions of Theorem 3.1.9 and if $\alpha > 2(1-m)$, then as $n \longrightarrow \infty$*

$$\begin{aligned} \int_1^\infty \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t \quad &\xrightarrow{\mathscr{D}} \quad \int_1^\infty \frac{G(t) - T(t)G(1)}{t^m}\,\mathrm{d}t \\ &\sim \quad \mathscr{N}\left(0, \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)}\right) . \end{aligned} \tag{3.2.5}$$

Proof: First, we compute the variance of the limiting normal distribution. Since

$$\begin{aligned} \mathbb{V}ar\left[\int_1^\infty \frac{G(t) - G(1)\,T(t)}{t^m}\,\mathrm{d}t\right] &= 2\int_1^\infty \int_{t=s}^\infty t^{-\alpha}\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m} - \frac{2}{\alpha+m-1}\int_1^\infty t^{-\alpha}\frac{\mathrm{d}t}{t^m} \\ &\quad + \frac{1}{(\alpha+m-1)^2} \\ &= \frac{2}{\alpha+m-1}\,\frac{1}{\alpha+2m-2} - \frac{1}{(\alpha+m-1)^2} \\ &= \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)} . \end{aligned}$$

It is finite whenever $\alpha > 2(1-m)$.

Now, for any $M > 1$, we can write

$$\int_1^\infty \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t = \int_1^M \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t + \int_M^\infty \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t\,. \tag{3.2.6}$$

By Theorem 3.1.9 and the continuous mapping theorem, as $n \longrightarrow \infty$ we have

$$\int_1^M \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^M \frac{G(t) - G(1)\,T(t)}{t^m}\,\mathrm{d}t\,, \tag{3.2.7}$$

This is true because the convergence of $\widehat{G}_n$ is in the topology of uniform convergence on compact sets.

Moreover, whenever $\alpha > 2(1-m)$

$$\begin{aligned}
\mathbb{V}ar\left[\int_M^\infty \frac{G(t)-G(1)\,T(t)}{t^m}\,\mathrm{d}t\right] &= \mathbb{V}ar\left[\int_M^\infty \frac{G(t)}{t^m}\,\mathrm{d}t - G(1)\int_M^\infty t^{-\alpha-m}\,\mathrm{d}t\right] \\
&= \mathbb{V}ar\left[\int_M^\infty \frac{G(t)}{t^m}\,\mathrm{d}t - \frac{G(1)\,M^{-\alpha-m+1}}{\alpha+m-1}\right] \\
&= \mathbb{V}ar\left[\int_M^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\right] + \mathbb{V}ar\left[\frac{G(1)\,M^{-\alpha-m+1}}{\alpha+m-1}\right] \\
&\quad - 2\mathbb{E}\left[\left(\int_M^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\right)\left(\frac{G(1)\,M^{-\alpha-m+1}}{\alpha+m-1}\right)\right] \\
&= \frac{2\,M^{-\alpha-2m+2}}{(\alpha+m-1)(\alpha+2m-2)} + \frac{M^{-2\alpha-2m+2}}{(\alpha+m-1)^2} - \frac{2\,M^{-\alpha-m+1}}{\alpha+m-1}\int_M^\infty \mathbb{E}[G(1)G(t)]\frac{\mathrm{d}t}{t^m} \\
&= \frac{2\,M^{-\alpha-2m+2}}{(\alpha+m-1)(\alpha+2m-2)} + \frac{M^{-2\alpha-2m+2}}{(\alpha+m-1)^2} - \frac{2\,M^{-2\alpha-2m+2}}{(\alpha+m-1)^2} \\
&= \frac{2\,M^{-\alpha-2m+2}}{(\alpha+m-1)(\alpha+2m-2)} - \frac{M^{-2\alpha-2m+2}}{(\alpha+m-1)^2} \\
&\longrightarrow 0, \quad \text{as } M \to \infty\,.
\end{aligned}$$

Letting $M \longrightarrow \infty$, we get

$$\int_1^M \frac{G(t)-G(1)\,T(t)}{t^m}\,\mathrm{d}t \quad \stackrel{\mathscr{D}}{\longrightarrow} \quad \int_1^\infty \frac{G(t)-G(1)\,T(t)}{t^m}\,\mathrm{d}t\,. \tag{3.2.8}$$

Then to prove (3.2.5), we show that (cf. [7], Theorem 25.5)

$$\lim_M \limsup_n \mathbb{P}\left[\left|\int_M^\infty \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t\right| \geq \epsilon\right] = 0\,, \tag{3.2.9}$$

for positive ϵ. To do this, we follow the steps below.

Let $\xi_n = X_{n:n-k}/u_n$ and recall that

$$G_n(t) = \sqrt{k}\left\{\widetilde{T}_n(t) - T_n(t)\right\} \quad \text{where} \quad \widetilde{T}_n(t) = \frac{1}{k}\sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n t\}} \quad \text{and} \quad T_n(t) = \frac{\overline{F}(u_n t)}{\overline{F}(u_n)} .$$

Then $\widehat{T}_n(t) = \widetilde{T}_n(\xi_n t)$. Thus, we get

$$\begin{aligned}
\widehat{G}_n(t) &= \sqrt{k}\left\{\widehat{T}_n(t) - T(t)\right\} = \sqrt{k}\left\{\widetilde{T}_n(\xi_n t) - T_n(\xi_n t)\right\} + \sqrt{k}\left\{T_n(\xi_n t) - T(t)\right\} \\
&= G_n(\xi_n t) + \sqrt{k}\left\{T_n(\xi_n t) - T(t)\right\} \ ,
\end{aligned}$$

and then, in order to prove (3.2.9), it suffices to show that for all $\epsilon > 0$,

$$\lim_M \limsup_n \mathbb{P}\left[\left|\int_M^\infty \frac{G_n(\xi_n t)}{t^m}\,\mathrm{d}t\right| \geq \epsilon\right] = 0 \ , \tag{3.2.10}$$

and that

$$\lim_M \limsup_n \mathbb{P}\left[\sqrt{k}\left|\int_M^\infty \frac{T_n(\xi_n t) - T(t)}{t^m}\mathrm{d}t\right| \geq \epsilon\right] = 0 \ . \tag{3.2.11}$$

Let $s = \xi_n t$, then the expression (3.2.10) becomes

$$\begin{aligned}
\lim_M \limsup_n \mathbb{P}\left[\left|\int_M^\infty \frac{G_n(\xi_n t)}{t^m}\,\mathrm{d}t\right| \geq \epsilon\right] &= \lim_M \limsup_n \mathbb{P}\left[\left|\xi_n^{m-1}\int_{\xi_n M}^\infty \frac{G_n(s)}{s^m}\,\mathrm{d}s\right| \geq \epsilon\right] \\
&\leq \lim_M \limsup_n \mathbb{P}\left[\left|\xi_n^{m-1}\int_{M}^\infty \frac{G_n(s)}{s^m}\,\mathrm{d}s\right| \geq \frac{\epsilon}{2}\right] \\
&\quad + \lim_M \limsup_n \mathbb{P}\left[\left|\xi_n^{m-1}\int_{M}^{\xi_n M} \frac{G_n(s)}{s^m}\,\mathrm{d}s\right| \geq \frac{\epsilon}{2}\right] \\
&= RHS1 + RHS2 \ .
\end{aligned}$$

Since $\xi_n \xrightarrow{P} 1$, by (3.2.4), the first term (RHS1) converges to 0.

Now, it remains to show that (RHS2) is negligible.

In fact, fixing M, and for any $\delta > 0$ we have

$$\limsup_n \mathbb{P}\left[\left|\xi_n^{m-1}\int_M^{\xi_n M}\frac{G_n(s)}{s^m}\,\mathrm{d}s\right| \geq \frac{\epsilon}{2}\right] \leq \limsup_n \mathbb{P}\left[C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)}\left|\frac{G_n(s)}{s^m}\right|\,\mathrm{d}s \geq \frac{\epsilon}{2}\right] + \limsup_n \mathbb{P}\big[|\xi_n - 1| \geq \delta\big]\ .$$

where $C_{m,\delta} = (1+\delta)^{m-1} \vee (1-\delta)^{m-1}$.

Since $\xi_n \xrightarrow{P} 1$, then as $n \longrightarrow \infty$ the second term converges to 0. Furthermore, by continuity of the limit in distribution, we have uniform convergence on the compact set $\big[M(1-\delta),\, M(1+\delta)\big]$. Therefore, by Theorem 3.1.1 and since $G(t)$ is centered Gaussian with variance $t^{-\alpha}$ and applying Markov's inequality, for $\epsilon > 0$, we get that for $\alpha > 2(1-m)$

$$\begin{aligned}
\limsup_n \mathbb{P}&\left[C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)}\left|\frac{G_n(s)}{s^m}\right|\,\mathrm{d}s \geq \frac{\epsilon}{2}\right] \\
&= \mathbb{P}\left[C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)}\big|G(s)\big|\,\frac{\mathrm{d}s}{s^m} \geq \frac{\epsilon}{2}\right] \\
&\leq \frac{2}{\epsilon}\,\mathbb{E}\left[C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)}\big|G(s)\big|\,\frac{\mathrm{d}s}{s^m}\right] \\
&= \frac{2\sqrt{2}}{\epsilon\sqrt{\pi}}\,C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)} s^{-\alpha/2}\,\frac{\mathrm{d}s}{s^m} \\
&= \frac{2\sqrt{2}}{\epsilon\sqrt{\pi}}\frac{2C_{m,\delta}}{\alpha+2m-2}\left\{\big[(1-\delta)M\big]^{-\alpha/2-m+1} - \big[(1+\delta)M\big]^{-\alpha/2-m+1}\right\} \\
&\leq \frac{4\sqrt{2}C_{m,\delta}}{\epsilon\sqrt{\pi}(\alpha+2m-2)}\,\big[(1-\delta)M\big]^{-\alpha/2-m+1} \\
&\longrightarrow 0 \quad \text{as } M \to \infty\ .
\end{aligned}$$

Therefore, (3.2.10) has been proven.

Now, in order to prove (3.2.11), since $T(t) = t^{-\alpha}$, one can write that

$$\begin{aligned}
\sqrt{k}\int_M^\infty \frac{T_n(\xi_n t) - T(t)}{t^m}\mathrm{d}t &= \sqrt{k}\int_M^\infty \frac{T_n(\xi_n t) - T(\xi_n t)}{t^m}\mathrm{d}t + \sqrt{k}\int_M^\infty \frac{T(\xi_n t) - T(t)}{t^m}\mathrm{d}t \\
&= \sqrt{k}\int_M^\infty \frac{T_n(\xi_n t) - T(\xi_n t)}{t^m}\mathrm{d}t + \sqrt{k}\left(\xi_n^{-\alpha} - 1\right)\frac{M^{-\alpha-m+1}}{\alpha+m-1} \\
&= L_{n,m,M} + R_{n,m,M} \ .
\end{aligned}$$

By (3.1.30) and the delta method, one can argue that $\sqrt{k}\left(\xi_n^{-\alpha} - 1\right)$ converges in distribution to a normal law and then $R_{n,m,M} \xrightarrow{P} 0$ as first $n \longrightarrow \infty$ and then $M \longrightarrow \infty$.

Now, let $s = \xi_n t$, then by Lemma 2.4.2 and the second order condition (3.1.23), for any $\epsilon > 0$ there exists a constant C such that

$$\begin{aligned}
L_{n,m,M} &= \xi_n^{m-1}\sqrt{k}\int_{\xi_n M}^\infty \left|\frac{T_n(s) - T(s)}{s^m}\right|\mathrm{d}s = \xi_n^{m-1}\sqrt{k}\int_{\xi_n M}^\infty \left|\frac{\overline{F}(u_n s)}{\overline{F}(u_n)} - s^{-\alpha}\right|\frac{\mathrm{d}s}{s^m} \\
&\leq C\,\xi_n^{m-1}\sqrt{k}\;\eta^*(u_n)\int_{\xi_n M}^\infty \left\{s^{-\alpha+\rho-m}\left(s \vee s^{-1}\right)^\epsilon\right\}\mathrm{d}s \\
&= o_p(1)\int_{\xi_n M}^\infty \left\{s^{-\alpha+\rho-m}\left(s \vee s^{-1}\right)^\epsilon\right\}\mathrm{d}s\,. \qquad (3.2.12)
\end{aligned}$$

Recall that $\rho < 0$. The integral in (3.2.12) can be written as follows:

$$\begin{aligned}
I &= \left\{\int_{\xi_n M}^\infty s^{-\alpha+\rho-m+\epsilon}\,\mathrm{d}s\right\}\mathbb{1}(\xi_n > 1/M) \\
&\quad + \left\{\int_{\xi_n M}^1 s^{-\alpha+\rho-m-\epsilon}\,\mathrm{d}s + \int_1^\infty s^{-\alpha+\rho-m+\epsilon}\,\mathrm{d}s\right\}\mathbb{1}(\xi_n \leq 1/M) \\
&= \frac{(\xi_n M)^{-\alpha+\rho-m+\epsilon+1}}{\alpha-\rho+m-\epsilon-1}\,\mathbb{1}(\xi_n > 1/M) \\
&\quad + \left\{\frac{(\xi_n M)^{-\alpha+\rho-m-\epsilon+1}}{\alpha-\rho+m+\epsilon-1} - \frac{1}{\alpha-\rho+m+\epsilon-1} + \frac{1}{\alpha-\rho+m-\epsilon-1}\right\}\mathbb{1}(\xi_n \leq 1/M) \\
&= R_1 + R_2\ .
\end{aligned}$$

Since $\xi_n \xrightarrow{P} 1$ as $n \to \infty$, for all $M > 1$, $\mathbb{1}(\xi_n \le 1/M) \xrightarrow{P} 0$ and so $R_2 \xrightarrow{P} 0$ as $n \to \infty$. On the other hand, $\mathbb{1}(\xi_n > 1/M) \xrightarrow{P} 1$ as $n \to \infty$ and so

$$R_1 = O_p(1) \, \frac{M^{-\alpha+\rho-m+\epsilon-1}}{\alpha - \rho + m - \epsilon + 1} \xrightarrow{P} 0 \, , \qquad \text{as } M \to \infty \, .$$

Thus, $L_{n,m,M} \xrightarrow{P} 0$ as first $n \to \infty$ and then $M \to \infty$.
This completes the proof. ∎

Remark 3.2.3 We note that the limiting variance in Theorem 3.2.2 is always smaller than the one in Theorem 3.2.1. This is the result of replacing the unknown quantile u_n with the empirical quantile.

3.3 Estimation of tail index and conditional mean

Suppose that $X_1, \dots, X_n$ are non-negative i.i.d. random variables with distribution function $F(x)$ that has a regularly varying right tail with index $-\alpha$, that is

$$\mathbb{P}(X > x) = x^{-\alpha} \ell(x) \, , \qquad x > 0 \, , \tag{3.3.1}$$

where $\ell(x)$ is slowly varying at infinity.
Equivalently, for $t > 0$,

$$\lim_{x \to \infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha} \, . \tag{3.3.2}$$

The problem of how to estimate α is crucial in many applications. There have been many tail index estimators proposed since the early work of Hill [26] and Pickands [35]. Here we will consider estimators of $1/\alpha$: the Hill estimator that will be denoted by $\widehat{\alpha}_1^{-1}$ and the harmonic moment estimators that will be denoted by $\widehat{\alpha}_m^{-1}$ for $m \neq 1$. These estimators are also used to estimate the risk measures discussed in Chapter 4.

Several ad-hoc methods have been proposed to deal specifically with the asymptotic normality of these estimators. We will use a unified approach which is based on weak convergence of integral functionals of the tail empirical process with random levels.

3.3.1 The Hill estimator

Among the statistical estimators for the parameter α of a regularly varying tail $\overline{F}(x) = x^{-\alpha}\ell(x)$, the Hill estimator defined below has become particularly popular. Let $X_{n:n} \geq X_{n:n-1} \geq \cdots \geq X_{n:1}$ be the order statistics. The Hill estimator of α^{-1} is defined by

$$\widehat{\alpha}_1^{-1} = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right) . \tag{3.3.3}$$

We give a heuristic, nonrigorous explanation of the definition. Note that (3.3.2) can be interpreted as, for $t > 0$

$$\mathbb{P}(X/x > t \,|\, X > x) \longrightarrow t^{-\alpha} , \qquad \text{as } x \longrightarrow \infty . \tag{3.3.4}$$

Let $N_t = \sum_{j=1}^{n} \mathbb{1}_{\{X_j/x>t\}}$. Define $j_1 = \inf\{i : X_i/x > t\}$, $j_l = \inf\{i > j_{l-1} : X_i/x > t\}$ for $l = 2, \ldots, N_t$ and $Y_i = X_{j_i}$, for $i = 1, \ldots, N_t$. Then the limiting log-likelihood function based on conditional excesses $(Y_1, \ldots, Y_{N_t})$ is

$$\log L(Y_1, \ldots, Y_{N_t}) \approx N_t \log\alpha - (\alpha+1)\sum_{j=1}^{N_t} \log Y_j .$$

Hence, we get $\displaystyle\frac{d}{d\alpha}[\log L(Y_1, \ldots, Y_{N_t})] = 0 \Longleftrightarrow \frac{1}{\alpha} = \frac{1}{N_t}\sum_{j=1}^{N_t} \log Y_j .$

If we choose $x = X_{n:n-k}$ for $1 \leq k < n$ where $k \longrightarrow \infty$ and put $N_t = k$, we obtain the Hill estimator of α^{-1}.

Although the Hill estimator is motivated by likelihood methods, it is not a true MLE, so a different approach is needed to study its asymptotic behaviour.
The asymptotic normality of the Hill estimator can be obtained from weak convergence of the tail empirical process with random levels; see [30]. Indeed, recall that for $t > 0$,
$\widehat{G}_n(t) = \sqrt{k}\left\{\widehat{T}_n(t) - T(t)\right\}$, where $\widehat{T}_n(t) = \frac{1}{k}\sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{n:n-k}t\}}$ and $T(t) = t^{-\alpha}$.
Then,

$$\begin{aligned}
\int_1^\infty \frac{\widehat{G}_n(t)}{t}\,\mathrm{d}t &= \sqrt{k}\left\{\frac{1}{k}\int_1^\infty \sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{n:n-k}t\}}\frac{1}{t}\,\mathrm{d}t - \int_1^\infty t^{-\alpha-1}\,\mathrm{d}t\right\} \\
&= \sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^{k}\int_1^\infty \mathbb{1}_{\{X_{n:n-j+1} > X_{n:n-k}t\}}\frac{1}{t}\,\mathrm{d}t - \frac{1}{\alpha}\right\} \\
&= \sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^{k}\int_1^{\frac{X_{n:n-j+1}}{X_{n:n-k}}} \frac{1}{t}\,\mathrm{d}t - \frac{1}{\alpha}\right\} \\
&= \sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right) - \frac{1}{\alpha}\right\} \\
&= \sqrt{k}\left(\widehat{\alpha}_1^{-1} - \alpha^{-1}\right).
\end{aligned}$$

Then the following corollary is an immediate consequence of Theorem 3.2.2 with $m = 1$.

Corollary 3.3.1 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution F. Then under the assumptions of Theorem 3.1.9,*

$$\sqrt{k}\left(\widehat{\alpha}_1^{-1} - \alpha^{-1}\right) \xrightarrow{\mathscr{D}} \int_1^\infty \frac{G(t) - G(1)\,T(t)}{t}\,\mathrm{d}t \sim \mathscr{N}\left(0, \alpha^{-2}\right). \tag{3.3.5}$$

3.3.2 The harmonic moment estimators (HME$_m$)

Let $m \geq 0$. If moreover $m > 1-\alpha$, we have that

$$\zeta_m = \int_1^\infty T(t)\,\frac{\mathrm{d}t}{t^m} = \frac{1}{\alpha+m-1}\ . \tag{3.3.6}$$

Hence, we can estimate the right hand side of (3.3.6) by $\int_1^\infty (\widehat{T}_n(t)/t^m)\,\mathrm{d}t$.

For $m \neq 1$, one can write that

$$\begin{aligned}
\widehat{\zeta}_m &= \int_1^\infty \frac{\widehat{T}_n(t)}{t^m}\,\mathrm{d}t = \frac{1}{k}\int_1^\infty \sum_{j=1}^n \mathbb{1}_{\{X_j > X_{n:n-k}t\}}\frac{1}{t^m}\,\mathrm{d}t \\
&= \frac{1}{k}\sum_{j=1}^k \int_1^{\frac{X_{n:n-j+1}}{X_{n:n-k}}} \frac{1}{t^m}\,\mathrm{d}t \\
&= \frac{1}{k}\sum_{j=1}^k \left(\frac{1-\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right)^{1-m}}{m-1}\right) \\
&= \frac{1}{m-1} - \frac{1}{(m-1)k}\sum_{j=1}^k \left(\frac{X_{n:n-k}}{X_{n:n-j+1}}\right)^{m-1}\ .
\end{aligned}$$

This leads to the following harmonic moment estimator of $1/\alpha$ considered in [25] and defined by

$$\widehat{\alpha}_m^{-1} = \frac{1}{m-1}\left[\frac{1}{\frac{1}{k}\sum_{j=1}^k \left(\frac{X_{n:n-k}}{X_{n:n-j+1}}\right)^{m-1}} - 1\right]\ . \tag{3.3.7}$$

Note that for $m=1$, the harmonic moment estimator is interpreted as a limit for $m \to 1$, that is

$$\widehat{\alpha}_1^{-1} = \lim_{m\to 1}\widehat{\alpha}_m^{-1} = \frac{1}{k}\sum_{j=1}^k \log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right)\ ,$$

which coincides with the Hill estimator.

Corollary 3.3.2 *Let $m > 1 - \alpha/2$ and $m \neq 1$. Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with continuous marginal distribution F. Then under the assumptions of Theorem 3.1.9,*

$$\begin{aligned}\sqrt{k}\left\{\widehat{\zeta}_m - \zeta_m\right\} &\xrightarrow{\mathscr{D}} \int_1^\infty \frac{G(t) - G(1)\,T(t)}{t^m}\,\mathrm{d}t \\ &\sim \mathscr{N}\left(0, \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)}\right).\end{aligned} \tag{3.3.8}$$

Proof: Since

$$\sqrt{k}\left\{\widehat{\zeta}_m - \zeta_m\right\} = \int_1^\infty \frac{\widehat{G}_n(t)}{t^m}\,\mathrm{d}t,$$

the proof is a direct consequence of Theorem 3.2.2. ∎

Applying the delta method to (3.3.8) we obtain the following result.

Corollary 3.3.3 *Under the assumptions of Corollary 3.3.1, and if $m > 1 - \alpha/2$ and $m \neq 1$, then*

$$\sqrt{k}\left\{\widehat{\alpha}_m^{-1} - \alpha^{-1}\right\} \xrightarrow{\mathscr{D}} \mathscr{N}\left(0, \frac{(\alpha+m-1)^2}{\alpha^3(\alpha+2m-2)}\right). \tag{3.3.9}$$

Proof: Let $g(x) = \dfrac{x}{1 - mx + x}$, then $g'(x) = \dfrac{1}{(1-mx+x)^2}$. Therefore, we get

$$\begin{aligned}\sqrt{k}\left\{\widehat{\alpha}_m^{-1} - \alpha^{-1}\right\} &= \sqrt{k}\left\{g(\widehat{\zeta}_m) - g(\zeta_m)\right\} \\ &\xrightarrow{\mathscr{D}} \mathscr{N}\left(0, \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)}\left[g'\left(\frac{1}{\alpha+m-1}\right)\right]^2\right) \\ &= \mathscr{N}\left(0, \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)}\left(\frac{\alpha+m-1}{\alpha}\right)^4\right) \\ &= \mathscr{N}\left(0, \frac{(\alpha+m-1)^2}{\alpha^3(\alpha+2m-2)}\right).\end{aligned}$$

∎

Remark 3.3.4 The condition $m > 1 - \alpha/2$ imposes some restrictions on m and α. If $\alpha > 2$, then Corollaries 3.3.2 and 3.3.3 are applicable to all values of $m \geq 0$.

Remark 3.3.5 It is clear from Theorem 3.2.2 that the variance of the limiting random variable $\int_1^\infty (G(t)/t^m)\mathrm{d}t$ decreases with m. For $m = 1$ we obtain directly the Hill estimator. However, the Harmonic Mean Estimators are obtained by transforming the integral functionals resulting in the Hill estimator having the smallest limiting variance among all HMEs.

Remark 3.3.6 For $m = 2$, we obtain the t-Hill estimator introduced in [19],

$$\widehat{\alpha}_2^{-1} = \left(\frac{1}{k} \sum_{j=1}^{k} \frac{X_{n:n-k}}{X_{n:n-j+1}} \right)^{-1} - 1 \ . \tag{3.3.10}$$

As a result,

$$\sqrt{k} \left\{ \widehat{\alpha}_2^{-1} - \alpha^{-1} \right\} \xrightarrow{\mathscr{D}} \mathscr{N}\left(0, \frac{(\alpha+1)^2}{\alpha^3(\alpha+2)} \right) \ . \tag{3.3.11}$$

As indicated above, the Hill estimator has a smaller limiting variance than any HME. However, the HME can be asymptotically more efficient than the Hill estimator if the no-bias condition is not satisfied. Indeed, if MSE_m denotes the Mean Squared Error of the HME_m estimator, then it was calculated in [4] that it is possible that $\mathrm{MSE}_2 < \mathrm{MSE}_1$. Even though in this thesis we work only under the no-bias condition, the theoretical result of [4] indicates that e.g. the t-Hill estimator may perform better in finite samples, where a bias is always present - see Section 6.1.5.

3.3.3 Application to conditional mean

Our goal is to approximate $\mathbb{E}[X|X > x]$ when x is large. The results in this section will be applied to the risk measures to be discussed in Section 4.2. To do this, let

$$\vartheta_n = \mathbb{E}\left[\frac{X}{u_n}\middle| X > u_n\right] = \frac{1}{\overline{F}(u_n)}\frac{1}{u_n}\mathbb{E}[X\mathbb{1}_{\{X>u_n\}}], \tag{3.3.12}$$

where $u_n \to \infty$ and $n\overline{F}(u_n) \to \infty$.

Regular variation implies that for $\alpha > 1$, $\vartheta_n \longrightarrow \vartheta = \alpha/(\alpha-1)$, (see Lemma 2.3.4). By replacing u_n with $X_{n:n-k}$ and $n\overline{F}(u_n)$ with k, we can approximate ϑ_n by

$$\widehat{\vartheta}_{n,0} = \frac{1}{k}\frac{1}{X_{n:n-k}}\sum_{j=1}^{n} X_j \mathbb{1}_{\{X_j > X_{n:n-k}\}} = \frac{1}{k}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}} \tag{3.3.13}$$

Recall the notation from (3.3.6). Since $\widehat{\zeta}_0 - \zeta_0 = \widehat{\vartheta}_{n,0} - \vartheta$, the following corollary is equivalent to Corollary 3.3.2 with $m = 0$.

Corollary 3.3.7 *Under the assumptions of Corollary 3.3.1 and if $\alpha > 2$, then*

$$\begin{aligned}\sqrt{k}(\widehat{\vartheta}_{n,0} - \vartheta) &\xrightarrow{\mathscr{D}} \int_1^\infty (G(t) - G(1)\,T(t))\,\mathrm{d}t \\ &\sim \mathscr{N}\left(0, \frac{\alpha}{(\alpha-1)^2(\alpha-2)}\right).\end{aligned} \tag{3.3.14}$$

We note that $\widehat{\vartheta}_{n,0}$ can be viewed as the plug-in estimator of $\vartheta = \alpha/(\alpha-1)$, where α is estimated using HME_0. We can consider the family $\widehat{\vartheta}_{n,m}$ of plug-in estimators of ϑ, where ϑ is estimated using HME_m :

$$\widehat{\vartheta}_{n,m} = \frac{1}{1 - \widehat{\alpha}_m^{-1}}\,, \tag{3.3.15}$$

where $\widehat{\alpha}_m^{-1}$ is defined in (3.3.7).

The following corollary is obvious using Corollary 3.3.3 and the delta method.

Corollary 3.3.8 *Under the assumptions of Corollary 3.3.1 and if $m > 1 - \alpha/2$, then*

$$\sqrt{k}(\widehat{\vartheta}_{n,m} - \vartheta) \xrightarrow{\mathscr{D}} \mathscr{N}\left(0, \frac{\alpha(\alpha + m - 1)^2}{(\alpha - 1)^4(\alpha + 2m - 2)}\right). \tag{3.3.16}$$

3.3.4 Summary

In Sections 3.3.1, 3.3.2 and 3.3.3, we were able to apply Theorem 3.2.2 in order to obtain the asymptotic behaviour of the class of the harmonic moment estimators in a unified manner. This includes the Hill estimator, the HMEs of all orders and estimators of the conditional mean.

3.4 Discussion and Examples

As indicated at the beginning of the chapter, the results of Section 3.1 are special cases of those in [39]. Since our results are stated for i.i.d. random variables, we compare our result to the existing literature in this case, in particular to [14].

3.4.1 No-bias condition

Consider the function

$$U(t) = (1/\overline{F})^{\leftarrow}(t)\,. \tag{3.4.1}$$

If $\overline{F}$ is regularly varying with index $-\alpha$, then U is regularly varying with index $1/\alpha$. This can be obviously seen if $\overline{F}(t) = t^{-\alpha}$, since then $U(t) = t^{1/\alpha}$. Moreover, this is easy to see that if $F(F^{\leftarrow}(t)) = F^{\leftarrow}(F(t)) = t$, then the functions U and Q defined by (cf. (2.3.1) and (4.1.3))

$$Q(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad t > 1$$

coincide.

The no-bias condition in this chapter (see (3.1.23)) is linked to the second order regular variation for the tail function. The second order regular variation used here is in the spirit of [16] and essentially says that

$$\overline{F}(x) = c_0\; x^{-\alpha} \exp\left\{\int_1^x \frac{\eta^*(y)}{y}\mathrm{d}y\right\},$$

with $\eta^*(x)$ being regularly varying with index $-\rho$ and $c_0 \in (0,\infty)$. Under this condition we can control speed of convergence via Lemma 2.4.2:

$$\forall t \geq 0 \; , \quad \forall x > 1 \; , \quad \left| \frac{\overline{F}(xt)}{\overline{F}(x)} - t^{-\alpha} \right| \leq C \; \eta^*(x) \; t^{-\alpha-\rho} (t \vee t^{-1})^{\epsilon} \; .$$

Hence, via Corollary 3.1.6, the no-bias condition is expressed as

$$\sqrt{n\overline{F}(u_n)} \; \eta^*(u_n) \approx \sqrt{k} \; \eta^* \left(Q(n/k) \right) \longrightarrow 0 \; . \tag{3.4.2}$$

In [14] the approach to the bias control is different. First, one assumes existence of function $a(t)$ defined through

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} =: D_\gamma(x) \; . \tag{3.4.3}$$

The function a is regularly varying with index $\gamma = 1/\alpha$ and in fact for $\overline{F}(t) = t^{-\alpha}$ we have $a(t) = \gamma \, t^\gamma$. Next, the second order condition (Definition 2.3.1 in [14]) is defined as follows: there exists a function A such that $\lim_{t \to \infty} A(t) = 0$ and the limit

$$\lim_{t \to \infty} \frac{1}{A(t)} \left\{ \frac{U(tx) - U(t)}{a(t)} - D_\gamma(x) \right\} = H(x) \tag{3.4.4}$$

exists for some function H. Theorem 2.3.3 in [14] implies that A has to be regularly varying with index $-\tilde{\rho}$, $\tilde{\rho} > 0$. Theorem 2.3.9 in [14] states that

$$\lim_{t \to \infty} \frac{\frac{U(tx)}{U(x)} - x^\gamma}{A(t)} = x^\gamma \frac{x^{-\tilde{\rho}} - 1}{-\tilde{\rho}} \tag{3.4.5}$$

and

$$\left| \frac{\frac{U(tx)}{U(x)} - x^\gamma}{A(t)} - x^\gamma \frac{x^{-\tilde{\rho}} - 1}{-\tilde{\rho}} \right| \leq C x^{\gamma - \tilde{\rho}} \{ x^\delta \vee x^{-\delta} \}$$

for arbitrary C, δ and t, tx sufficiently large.

The no-bias condition in [14] becomes

$$\lim_{k\to\infty} \sqrt{k}A(n/k) = 0\ . \tag{3.4.6}$$

The link between (3.4.2) and (3.4.6) is established through Remark 2.3.10 in [14]. Indeed, (3.4.5) is equivalent to

$$\lim_{t\to\infty} \frac{\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-\alpha}}{\tilde{A}(t)} = x^{-\alpha}\frac{x^{-\alpha\tilde{\rho}} - 1}{-\tilde{\rho}/\alpha}$$

with $\tilde{A}(t) = A\big(1/\overline{F}(t)\big) = A\big(U^{\leftarrow}(t)\big)$, where the latter equality holds if F is continuous. In other words, the function $A(n/k)$ plays the same role for quantile-based no-bias condition as the function $\eta^*\big(Q(n/k)\big)$ for the tail function-based no-bias condition. As a consequence, (3.4.2) and (3.4.6) are essentially equivalent.

3.4.2 Comparison with [14]

The function F fulfills the Von-Mises condition (Eq. (1.1.29) in [14]) if

$$\lim_{t\to\infty} \left(\frac{\overline{F}}{F'}\right)'(t) = \gamma \in (0,\infty)\ .$$

If the Von-Mises condition holds then (Theorem 2.2.1 in [14])

$$\sqrt{k}\frac{X_{n:n-k} - U(n/k)}{\frac{n}{k}U'(n/k)} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1)\ . \tag{3.4.7}$$

Instead of imposing the differentiability assumption that appears implicitly in the Von-Mises condition, we can consider the second order condition (Definition 2.3.1 in [14]). As a consequence (Theorem 2.4.1 in [14]) we have

$$\sqrt{k}\frac{X_{n:n-k} - U(n/k)}{a(n/k)} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1)$$

as long as (3.4.6) holds. Furthermore, the result is strengthened to the functional convergence of the quantile process $X_{n:n-[ns]}$, $s \in [s_0, 1)$ for some $s_0 > 0$.

In their proofs in [14], the authors work directly with the quantile processes and order statistics. This is the main difference between the approach in [14] (as well as in [13] in case of dependent random variables) and our approach here: we work with the empirical processes and relevant results for quantiles and quantile-based statistics are obtained via inverse functionals and integral transformations. This way we avoid some technical problems that arise while working with quantile processes in a neighbourhood of 0 and 1.

As such, the result in (3.4.7) corresponds to our Lemma 3.1.11, with the Von-Mises differentiability condition replaced with (3.1.25) and normalization expressed in terms of quantile function U instead of u_n. Likewise, the no-bias condition (3.1.23), expressed in terms of the tail distribution function, corresponds to (3.4.6), expressed in terms of the quantile function.

As for the Hill estimator, Theorem 3.2.5 in [14] states that under second order regular variation and the no-bias condition (3.4.6), the appropriately centered and scaled Hill estimator is approximately normal with mean zero.

3.4.3 Examples

Recall that the function η^* is regularly varying with index $-\rho$, $\rho > 0$. On the other hand, the function A is regularly varying with index $-\tilde{\rho}$, $\tilde{\rho} > 0$. We also noticed that $A(t)$ plays the same role as $\eta^*\big(Q(t)\big)$ as $t \to \infty$. Since $Q(t)$ is regularly varying with index $1/\alpha$, we conclude that

$$\tilde{\rho} = \rho/\alpha \, . \tag{3.4.8}$$

In what follows we discuss the no-bias assumption for two commonly used classes of distributions.

Example 3.4.1 (Burr distribution) (cf. [27, Example 2]) The Burr distribution with parameters a and b has the survival function $\overline{F}(x) = (1 + x^b)^{-a}$, $x, a, b > 0$. Clearly,

$$\overline{F}(x) = x^{-ab}\left\{1 - ax^{-b} + o(x^{-b})\right\} \ , \quad \text{as} \quad x \to \infty \ .$$

This implies that the tail index is $\alpha = ab$, while the second order indices are, respectively, $\tilde{\rho} = b$ and $\rho = \tilde{\rho}\alpha = ab^2$. The no bias conditions (3.4.2) and (3.4.6) are fulfilled whenever

$$\sqrt{k}(k/n)^b \to 0 \ .$$

Of course, one has to point out that even if we know the tail index (or we estimate it), we do not know b unless we estimate the parameter a.

Example 3.4.2 (Student t-distribution) Assume that a random variable X has a Student-t density with ν degrees of freedom, that is the density is given by

$$f(x) = c_\nu \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \ , \quad x \in R \ ,$$

for some explicit constant c_ν. Then (cf. [14, Exercise 2.15])

$$\overline{F}(x) = c_\nu \nu^{\nu/2} x^{-\nu}\left\{1 + d_\nu x^{-2} + o(x^{-4})\right\} \ , \quad \text{as} \quad x \to \infty \ ,$$

where d_ν is another explicit constant. This implies that the tail index is $\alpha = \nu$, while the second order indices are, respectively, $\tilde{\rho} = 2$ and $\rho = \tilde{\rho}\alpha = 2\nu$. The no bias conditions (3.4.2) and (3.4.6) are fulfilled whenever

$$\sqrt{k}(k/n)^2 \longrightarrow 0 \ .$$

That is, $k^{5/2}/n^2 \to 0$. For example, we can take $k = \sqrt{n}$. See also [27, Example 3].

This is one of the very few examples where the second order parameter can be explicitly specified. Another example is the Inverse Gamma distribution, see [27, Example 4].

Example 3.4.3 (Mixture distribution) Consider the tail distribution function

$$\overline{F}(x) = \frac{1}{2}\{x^{-\alpha} + x^{-\alpha\beta}\} \ , \quad x > 1, \ \beta > 1 \ .$$

Then

$$\overline{F}(x) = \frac{1}{2}x^{-\alpha}\{1 + x^{-\alpha(\beta-1)}\} \ .$$

That is, $\tilde{\rho} = \alpha(\beta - 1)$ and $\rho = \beta - 1$. Cf. Example 2.4.3.

■

Chapter 4

The risk measures

The literature on financial risk and risk measures is vast. This chapter concentrates on financial risks and on financial risk measures. We start by defining the risk measure, present properties that any acceptable risk measure should have, investigate some risk measures used in the financial literature, in particular, the Value-at-Risk (VaR) and the conditional tail expectation (CTE), and we establish some estimates of these risk measures.

The word "risk" is closely connected with phenomena generated by uncertainty. In modern risk management, a central question is the measurement of risk. For example, a financial institution needs to quantify risk to determine the amount of capital that the institution has to hold as a buffer against unexpected losses.

Consider a given financial portfolio. For a given time horizon, Δ_t, the performance of the portfolio (like gain or loss) is often evaluated by a real-valued random variable X. For our purposes, X will denote the negative return (i.e. if $X > 0$, there is a loss of X; if $X < 0$, there is a gain of $-X$). Measuring risk is equivalent to establishing a correspondence ϱ between the space of random variables X and a non-negative real number R. In particular, in this sense, variance or standard deviation could be measures of risk.

The definition of a risk measure is very general since any positive real function of a random variable can be regarded as being a measure of risk. In practice, such measures should have certain mathematical properties.

Artzner et al. [2] go back to “first principles" and ask what basic features a risk measure should have. They wrote down a list of properties that a good risk measure should have. Measures satisfying these axioms are called coherent.

Consider a random variable X viewed as an element of a linear space $\mathscr{L}$ of measurable functions, defined on an appropriate probability space. A functional $\varrho : \mathscr{L} \to \mathbb{R}_+ \cup \{+\infty\}$ is said to be a coherent risk measure for $\mathscr{L}$ in the sense of [2] if it satisfies the following properties for $X, Y \in \mathscr{L}$:

1. *Normalized* : $\varrho(0) = 0$. The risk of holding no assets is zero.

2. *Positive homogeneity* : $\varrho(aX) = a\varrho(X)$, for all positive $a \in \mathbb{R}$. An increase in the size of a portfolio leads to an increase in the risk of this portfolio.

3. *Subadditivity* : $\varrho(X+Y) \leq \varrho(X) + \varrho(Y)$. The aggregate portfolio risk must be less than or equal to the sum of the individual risks of the assets that compose this portfolio.

4. *Monotonicity* : $X \leq Y \implies \varrho(X) \leq \varrho(Y)$. The lower the potential loss of a portfolio, the lower the risk.

5. *Transitional invariance* : $\varrho(X-b) = \varrho(X) - b$, for all $\varrho(X) \geq b \geq 0$. This can be interpreted as adding cash b to the portfolio X.

Once a risk has been defined mathematically, we must be able to measure it in relation to a real financial environment.

The standard deviation and variance were the first risk measures to be used. However, variance does not satisfy properties 2 – 5 and standard deviation does not satisfy 3 – 5.

Value-at-Risk (VaR) is probably the most widely used risk measure by financial institutions. This concept, which originated in the insurance sector, was imported in the late 1980s in the financial markets in the United States. This risk measure is based on a simple concept and easily explainable : VaR is the value of the amount that will cover the loss X with a probability $1-p$. However, VaR in general turns out not to be a coherent risk measure, since it violates the property of sub-additivity that reasonable risk measures should have. An immediate consequence is that value at risk might discourage diversification.

A coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. It can be shown that for continuous F (losses), CTE is the average of VaR over all confidence levels greater than p, focusing more than VaR does on extremal losses. Thus, CTE is more conservative than VaR at the same level of confidence (i.e., CTE $\geq$ VaR) and provides an effective tool for analyzing tail risks.

Nevertheless, VaR is still being used extensively in the financial industry. Therefore, we analyse an estimator of VaR. Furthermore, we will repeat the exercise for CTE. To the best of our knowledge, Theorem 4.2.4 on the asymptotic behaviour of the estimator of the CTE is new.

4.1 Value-at-Risk (VaR)

Let $X = -Y\mathbb{1}_{\{Y\le 0\}}$, where Y is the return on an investment. Hence $X > 0$ if there is a loss and $X = 0$, if there is a gain. The VaR (or more precisely, p-VaR) represents the maximum loss that, with probability $1-p$, will not be exceeded. More precisely, assuming that X is continuous, $\mathrm{VaR}_p(X)$ is defined by

$$\mathbb{P}\big(X \le \mathrm{VaR}_p(X)\big) = 1 - p \,. \tag{4.1.1}$$

Thus

$$\mathrm{VaR}_p(X) = F_X^{\leftarrow}(1-p) \,, \tag{4.1.2}$$

where F_X is the distribution function of X. Define

$$Q_X(t) = F_X^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad t > 1 \,. \tag{4.1.3}$$

Then

$$Q_X(1/p) = \mathrm{VaR}_p(X) \,. \tag{4.1.4}$$

Let $\{X_j,\ j = 1, \dots, n\}$ be a random sample from X. We are interested in the case when p is very small, that is the expected number of observations above $Q_X(1/p)$, np, is also very small. Since we are dealing with asymptotic analysis, we are obliged to assume that p depends on n, $p = p_n$, and that $\lim_{n\to\infty} p_n = 0$. That is, suppressing dependence of p on n in the notation, we want to estimate $Q_X(1/p)$ with $p \longrightarrow 0$, as $n \longrightarrow \infty$. We shall assume that the random variable X is regularly varying with the tail index α. In other words,

$$\overline{F}_X(x) = 1 - F_X(x) = x^{-\alpha}\ell(x) \,, \tag{4.1.5}$$

where $\ell(x)$ is a slowly varying function at infinity. For the moment, we assume that $\ell(x) = 1$. A natural nonparametric method to estimate the distribution function is the empirical distribution $F_n(x) = (1/n)\sum_{j=1}^n \mathbb{1}(X_j \le x)$. Then, the empirical estimate of $Q_X(1/p)$ is given by

$$F_n^{\leftarrow}(1-p) = X_{n:n-[np]} \ . \tag{4.1.6}$$

However, for a very small value of p, this procedure is not very reliable since for $p_1 \neq p_2$ such that $[np_1] = [np_2]$, the values $Q_X(1/p_1)$ and $Q_X(1/p_2)$ may differ significantly, but both values will be estimated by the same $X_{n:n-[np_1]}$. In particular, for all $p < 1/n$, $Q_X(1/p)$ will be always estimated by $X_{n:n}$.

Weissman [43] proposed an alternative approach by combining the semi-parametric form of Q_X with a nonparametric estimator. Indeed, let k be an intermediate sequence, that is $k = k_n \longrightarrow \infty$ and $k/n \longrightarrow 0$. From (4.1.5) and keeping in mind that $\ell(x) = 1$, it is easy to see that $F_X^{\leftarrow}(1-p) = p^{-1/\alpha}$. Then we have

$$Q_X(1/p) = p^{-1/\alpha} \ , \tag{4.1.7}$$

and

$$Q_X(n/k) = \left(\frac{k}{n}\right)^{-1/\alpha} \ . \tag{4.1.8}$$

Thus we get

$$\frac{Q_X(n/k)}{Q_X(1/p)} = \left(\frac{k}{np}\right)^{-1/\alpha} \ . \tag{4.1.9}$$

In general, when $\overline{F} \in \mathbf{RV}_{-\alpha}$, in view of Lemma 2.3.3, we have

$$\frac{Q_X(n/k)}{Q_X(1/p)} = \frac{Q_X(\frac{1}{p}\frac{np}{k})}{Q_X(\frac{1}{p})} \sim \left(\frac{k}{np}\right)^{-1/\alpha} \tag{4.1.10}$$

as long as (np/k) is bounded away from 0 and ∞. Henceforth, we make this assumption. By (4.1.6), $Q_X(n/k)$ can be estimated by $X_{n:n-k}$. This suggests the following estimator of $Q_X(1/p)$,

$$\widetilde{Q}_X(1/p) = X_{n:n-k}\left(\frac{k}{np}\right)^{1/\alpha}, \tag{4.1.11}$$

if the parameter α is known, otherwise

$$\widehat{Q}_X(1/p) = X_{n:n-k}\left(\frac{k}{np}\right)^{\widehat{\alpha}^{-1}}, \tag{4.1.12}$$

where $\widehat{\alpha}^{-1}$ is an estimate of α^{-1}.

This procedure will now be shown to be valid for general regularly varying random variables X. In what follows, we will prove several asymptotic results, which will follow from weak convergence of the tail empirical process:

1. Lemma 4.1.1 deals with consistency of the estimator $\widetilde{Q}_X(1/p)$ defined in (4.1.11).
2. We extend the consistency to asymptotic normality in Lemma 4.1.2.

The first result deals with consistency of the Value-at-Risk estimator $\widetilde{Q}_X(1/p)$.

Lemma 4.1.1 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with strictly increasing marginal distribution F_X continuous on $(0,\infty)$. Suppose that np/k is bounded away from zero and ∞ and that the conditions (3.1.1), (3.1.2) and (3.1.25) hold. Then as $n \longrightarrow \infty$,*

$$\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)} - 1 \xrightarrow{P} 0\,. \tag{4.1.13}$$

Proof: Recall that $k = n\overline{F}_X(u_n)$, then $u_n = F_X^{\leftarrow}(1-k/n)$ and by (4.1.3) we get $Q_X(n/k) = u_n$. So, by (4.1.11)

$$\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)} = \frac{X_{n:n-k}}{u_n}\,\frac{\left(\frac{k}{np}\right)^{1/\alpha} Q_X(n/k)}{Q_X(1/p)}\,. \tag{4.1.14}$$

By (4.1.10), the second ratio on the right hand side converges to 1. Furthermore, Lemma 3.1.11 implies that $X_{n:n-k}/u_n$ converges in probability to 1. This completes the proof. ∎

To prove asymptotic normality of the Value-at-Risk estimator $\widetilde{Q}_X(1/p)$, we need the following no-bias condition:

$$\limsup_{n\to\infty} \sqrt{k} \left| \frac{\left(\frac{k}{np}\right)^{-1/\alpha} Q_X(1/p)}{Q_X(n/k)} - 1 \right| = 0\,. \tag{4.1.15}$$

Note that if $\ell(x) = 1$, then the no-bias condition (4.1.15) is always valid. In general, the condition is valid for some sequences $\{k_n\}$, see the discussion in Section 4.3.

Lemma 4.1.2 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with strictly increasing marginal distribution F_X continuous on $(0,\infty)$. Suppose that np/k is bounded away from zero and ∞ and that the conditions (3.1.1), (3.1.2), (3.1.25) and (4.1.15) hold. Then, as $n \to \infty$,*

$$\sqrt{k}\left\{\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)} - 1\right\} \xrightarrow{\mathscr{D}} \frac{1}{\alpha} G(1)\,, \tag{4.1.16}$$

where $G(\cdot)$ is the centered Gaussian process that is defined in Theorem 3.1.1.

Proof: By (4.1.14), we can write

$$\sqrt{k}\left\{\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)} - 1\right\} = \sqrt{k}\left\{\frac{X_{n:n-k}}{u_n}\,\frac{\left(\frac{k}{np}\right)^{1/\alpha} Q_X(n/k)}{Q_X(1/p)} - 1\right\}$$

$$= \sqrt{k}\left\{\frac{X_{n:n-k}}{u_n}-1\right\}+\frac{\left(\frac{k}{np}\right)^{1/\alpha}Q_X(n/k)}{Q_X(1/p)}\frac{X_{n:n-k}}{u_n}\sqrt{k}\left\{\frac{\left(\frac{k}{np}\right)^{-1/\alpha}Q_X(1/p)}{Q_X(n/k)}-1\right\}.$$

Lemma 3.1.11, (4.1.10) and the condition (4.1.15) finish the proof of (4.1.16). ∎

We continue with the case when α is unknown.

Lemma 4.1.3 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with strictly increasing marginal distribution F_X continuous on $(0,\infty)$. Suppose that k/np is bounded away from zero and ∞ and that the conditions (3.1.1), (3.1.2) and (3.1.25) hold. Assume moreover that $\widehat{\alpha}$ is a consistent estimator of α. Then, as $n \longrightarrow \infty$*

$$\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)}-1 \xrightarrow{P} 0\,. \tag{4.1.17}$$

Proof: Write

$$\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)}-1=\left\{\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)}-1\right\}+\frac{\widehat{Q}_X(1/p)-\widetilde{Q}_X(1/p)}{Q_X(1/p)}\,. \tag{4.1.18}$$

Lemma 4.1.1 implies that the first term on the right hand side converges in probability to zero. To deal with the second term, we apply the Taylor expansion of the function $\gamma \to (k/(np))^\gamma$ around $1/\alpha$ to get

$$\left(\frac{k}{np}\right)^{\widehat{\alpha}^{-1}} = \left(\frac{k}{np}\right)^{1/\alpha}+\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)\left(\frac{k}{np}\right)^{1/\alpha}\log\left(\frac{k}{np}\right)+R_n\,, \tag{4.1.19}$$

where $R_n = \frac{1}{2}\left(\frac{k}{np}\right)^{1/\alpha}\log^2\left(\frac{k}{np}\right)O_p\left(\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)^2\right)$.

The Taylor expansion in (4.1.19) yields

$$\begin{aligned} X_{n:n-k}\left(\frac{k}{np}\right)^{\widehat{\alpha}^{-1}} &= X_{n:n-k}\left(\frac{k}{np}\right)^{1/\alpha}+X_{n:n-k}\,\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)\,\left(\frac{k}{np}\right)^{1/\alpha}\log\left(\frac{k}{np}\right)\\ &\quad +X_{n:n-k}\,R_n\;, \end{aligned}$$

so that

$$\widehat{Q}_X(1/p)=\widetilde{Q}_X(1/p)+X_{n:n-k}\,\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)\,\left(\frac{k}{np}\right)^{1/\alpha}\log\left(\frac{k}{np}\right)+X_{n:n-k}\,R_n\;.$$

Thus, bearing in mind $u_n=Q_X(n/k)$ and by (4.1.10) we write

$$\begin{aligned} \frac{\widehat{Q}_X(1/p)-\widetilde{Q}_X(1/p)}{Q_X(1/p)} &= \frac{X_{n:n-k}}{u_n}\,\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)\left(\frac{k}{np}\right)^{1/\alpha}\frac{Q_X(n/k)}{Q_X(1/p)}\,\log\left(\frac{k}{np}\right)\\ &\quad +\frac{X_{n:n-k}}{Q_X(1/p)}\,R_n\\ &= \frac{X_{n:n-k}}{u_n}\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)\log\left(\frac{k}{np}\right)\big(1+o(1)\big)\\ &\quad +\frac{X_{n:n-k}}{Q_X(1/p)}\,R_n\,. \end{aligned} \tag{4.1.20}$$

Consistency of $\widehat{\alpha}^{-1}$, Lemma 3.1.11 and boundedness of the log imply that the first term on the right hand side converges in probability to zero. Also, we show that the second term on the right hand side converges in probability to 0. Indeed, by (4.1.10) we have

$$\begin{aligned} \frac{X_{n:n-k}}{Q_X(1/p)}\,R_n &= \frac{1}{2}\frac{X_{n:n-k}}{u_n}\,\left(\frac{k}{np}\right)^{1/\alpha}\frac{Q_X(n/k)}{Q_X(1/p)}\,\log^2\left(\frac{k}{np}\right)O_p\left(\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)^2\right)\\ &= \frac{1}{2}\frac{X_{n:n-k}}{u_n}(1+o(1))\log^2\left(\frac{k}{np}\right)O_p\left(\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)^2\right)\;. \end{aligned} \tag{4.1.21}$$

Consistency of $1/\widehat{\alpha}$ and Lemma 3.1.11 finish the proof. ∎

Theorem 4.1.4 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with strictly increasing marginal distribution F_X continuous on $(0,\infty)$. Suppose that k/np is bounded away from zero and ∞ and that the conditions (3.1.1), (3.1.2), (3.1.25) and (4.1.15) hold. Assume that*

$$\sqrt{\delta_n}\left(\widehat{\alpha}^{-1} - \frac{1}{\alpha}\right) \xrightarrow{\mathscr{D}} \Delta\,, \tag{4.1.22}$$

where Δ is a normal random variable, δ_n is a sequence such that $\delta_n \longrightarrow \infty$ with $\delta_n = o(n)$ as $n \longrightarrow \infty$ and that

$$\lim_{n\to\infty} \frac{\sqrt{k}}{\sqrt{\delta_n}} \log\left(\frac{k}{np}\right) = r \in (-\infty,\infty)\,. \tag{4.1.23}$$

Then

$$\sqrt{k}\left\{\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right\} \xrightarrow{\mathscr{D}} \frac{1}{\alpha}\,G(1) + r\Delta\,. \tag{4.1.24}$$

where $G(\cdot)$ is the centered Gaussian process defined in Theorem 3.1.1.

Proof: By (4.1.18) and (4.1.20), one can argue that

$$\begin{aligned}
\sqrt{k}\left\{\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right\} &= \sqrt{k}\left\{\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)} - 1\right\} \\
&\quad + \frac{X_{n:n-k}}{u_n}\,\sqrt{\delta_n}\left(\widehat{\alpha}^{-1} - \frac{1}{\alpha}\right)\frac{\sqrt{k}}{\sqrt{\delta_n}}\log\left(\frac{k}{np}\right)(1+o(1)) \\
&\quad + \sqrt{k}\,\frac{X_{n:n-k}}{Q_X(1/p)}\,R_n\,.
\end{aligned} \tag{4.1.25}$$

By Lemma 4.1.2, the first term on the right hand side converges in distribution to $(1/\alpha)\,G(1)$. Lemma 3.1.11 and assumptions (4.1.22), (4.1.23) imply that the second term converges in distribution to $r\Delta$. It remains to show that the third term converges in probability to 0.

Indeed, by (4.1.21) we have

$$\sqrt{k}\,\frac{X_{n:n-k}}{Q_X(1/p)}\,R_n = \frac{1}{2}\,\frac{X_{n:n-k}}{u_n}\,\sqrt{\delta_n}\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right)\frac{\sqrt{k}}{\sqrt{\delta_n}}\log\left(\frac{k}{np}\right)\log\left(\frac{k}{np}\right) \times O_p\left(\widehat{\alpha}^{-1}-\frac{1}{\alpha}\right).$$

Lemma 3.1.11, assumptions (4.1.22) and (4.1.23) complete the proof. ∎

Corollary 4.1.5 *Under the assumptions of Theorem 4.1.4, if $\delta_n = k_n$ for all n, $\widehat{\alpha}^{-1} = \widehat{\alpha}_1^{-1}$ (Hill estimator), and $k/(np)$ is bounded away from zero and ∞ then*

$$\sqrt{k}\left\{\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)}-1\right\} \xrightarrow{\mathscr{D}} \frac{1}{\alpha}\,G(1) + r\int_1^\infty \frac{G(t)-G(1)T(t)}{t}\,dt \sim \mathscr{N}\left(0,\frac{1+r^2}{\alpha^2}\right). \tag{4.1.26}$$

Proof:

By Corollary 3.3.1, the limiting distribution of $\widehat{\alpha}_1^{-1}$ is $\int_1^\infty\{G(t)-G(1)\,T(t)\}/t\,dt$ with limiting variance α^{-2} and the variance $(1+r^2)/\alpha^2$ is an immediate consequence of the fact that $\mathbb{E}\big[G(1)\{G(t)-G(1)T(t)\}\big]=0$. ∎

Remark 4.1.6 The Hill estimator could be replaced by any HME, $\widehat{\alpha}_m$, in the preceding Corollary. The limiting variance is easily found by observing that $G(1)$ and $\int_1^\infty\{G(t)-G(1)\,T(t)\}/t^m\,dt$ are uncorrelated.

4.2 Conditional Tail Expectation (CTE)

The CTE is an important risk measure and useful tool in financial risk assessment. In risk analysis, the CTE describes the expected amount of risk that can be experienced given that a potential risk exceeds a threshold value. Assume that $X = -R\mathbb{1}_{\{R\leq 0\}}$ is the loss random variable defined in Section 4.1, with regularly varying tail distribution function $\overline{F}_X$ continuous on $(0,\infty)$. Our goal is to estimate the CTE defined as follows:

$$\theta(p) = \mathbb{E}\left[X|X > Q_X(1/p)\right] \ , \tag{4.2.1}$$

when p is small.

If X is the Pareto distribution with parameter $\alpha > 1$, then we can easily check that

$$\theta(p) = \mathbb{E}\left[X|X > p^{-1/\alpha}\right] = \vartheta\, p^{-1/\alpha} \ , \tag{4.2.2}$$

where $\vartheta = \alpha/(\alpha - 1)$.

More generally, if $\overline{F}_X \in \mathbf{RV}_{-\alpha}$, then, when $p \longrightarrow 0$ and as long as $np = O(k)$, Lemma 2.3.4 and (4.1.10) lead to

$$\begin{aligned} \theta(p) &= \vartheta\, Q_X(1/p)\big(1 + o(1)\big) && (4.2.3)\\ &= \vartheta\, Q_X(n/k)\left(\frac{k}{np}\right)^{1/\alpha}\big(1 + o(1)\big) \ . && (4.2.4) \end{aligned}$$

Since $Q_X(n/k)$ can be estimated by $X_{n:n-k}$ and by (4.1.11), if α is known, we can estimate $\theta(p)$ by

$$\begin{aligned} \widetilde{\theta}(p) &= \vartheta\, X_{n:n-k}\left(\frac{k}{np}\right)^{1/\alpha} && (4.2.5)\\ &= \vartheta\, \widetilde{Q}_X(1/p) \ . && (4.2.6) \end{aligned}$$

If α is unknown, we estimate $\theta(p)$ by

$$\begin{aligned}\widehat{\theta}(p) &= \widehat{\vartheta}_n\, X_{n:n-k}\left(\frac{k}{np}\right)^{\widehat{\alpha}^{-1}} &(4.2.7)\\ &= \widehat{\vartheta}_n\, \widehat{Q}_X(1/p)\,, &(4.2.8)\end{aligned}$$

where $\widehat{\alpha}^{-1}$ is an estimator of α^{-1} and $\widehat{\vartheta}_n$ is an estimator of ϑ. We will use the Hill estimator (3.3.3) for α^{-1}. The estimator for ϑ will be chosen as either

$$\widehat{\vartheta}_{n,0} = \frac{1}{k}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}} \quad \text{(cf. (3.3.13))}\,,$$

or

$$\widehat{\vartheta}_{n,1} = \frac{1}{1-\widehat{\alpha}^{-1}}\,.$$

Define

$$\widehat{\theta}_0(p) := \widehat{\vartheta}_{n,0}\, \widehat{Q}_X(1/p)\,, \quad \widehat{\theta}_1(p) := \widehat{\vartheta}_{n,1}\, \widehat{Q}_X(1/p)\,. \tag{4.2.9}$$

We note that since $Q_X(n/k) = u_n$ and from (4.2.3) and (4.2.6) we get

$$\frac{\widetilde{\theta}(p)}{\theta(p)} = \frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)}\big(1+o(1)\big)\,. \tag{4.2.10}$$

Hence, consistency of the estimator $\widetilde{Q}_X(1/p)$ in Lemma 4.1.1 allows us to state the following obvious corollary.

Corollary 4.2.1 *Under the conditions of Lemma 4.1.1 and if $\alpha > 1$, then as $n \to \infty$*

$$\frac{\widetilde{\theta}(p)}{\theta(p)} - 1 \xrightarrow{P} 0\,. \tag{4.2.11}$$

To prove the asymptotic normality of $\widetilde{\theta}(p)$, we need the following no-bias condition :

$$\limsup_{n\to\infty} \sqrt{k}\left|\frac{\vartheta\, Q_X(1/p)}{\theta(p)} - 1\right| = 0\,. \tag{4.2.12}$$

We note that when X is Pareto with parameter $\alpha > 1$, then the no-bias condition holds for any k, cf. (4.2.2).

Corollary 4.2.2 *Assume that the conditions of Lemma 4.1.2 and the no-bias condition (4.2.12) hold. If $\alpha > 1$ then as $n \to \infty$*

$$\sqrt{k}\left\{\frac{\widetilde{\theta}(p)}{\theta(p)} - 1\right\} \xrightarrow{\mathscr{D}} \frac{1}{\alpha}\, G(1)\ , \tag{4.2.13}$$

where $G(\cdot)$ is the centered Gaussian process defined in Theorem 3.1.1.

Proof: By (4.2.6), we have

$$\begin{aligned}
\sqrt{k}\left\{\frac{\widetilde{\theta}(p)}{\theta(p)} - 1\right\} &= \sqrt{k}\left\{\frac{\vartheta\, \widetilde{Q}_X(1/p)}{\theta(p)} - 1\right\} \\
&= \sqrt{k}\left\{\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)} - 1\right\} + \sqrt{k}\left\{\frac{\vartheta\, Q_X(1/p)}{\theta(p)} - 1\right\}\frac{\widetilde{Q}_X(1/p)}{Q_X(1/p)}\,.
\end{aligned}$$

By Lemma 4.1.1 and (4.2.12) the second term on the right hand side converges in probability to zero. By Lemma 4.1.2, the first term on the right hand side converges in distribution to $(1/\alpha)\, G(1)$. This completes the proof. ∎

Now, we prove consistency and asymptotic normality of the estimators $\widehat{\theta}_0(p)$ and $\widehat{\theta}_1(p)$ when the parameter α is unknown.

By (4.2.3) and (4.2.8) one can argue that, for $i = 0, 1$,

$$\frac{\widehat{\theta}_i(p)}{\theta(p)} = \frac{\widehat{\vartheta}_{n,i}}{\vartheta}\,\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)}\big(1 + o(1)\big)\,. \tag{4.2.14}$$

This allows us to state the following corollary.

Corollary 4.2.3 *Under the conditions of Lemma 4.1.3 and Corollary 3.3.8, we have for $i = 0, 1$,*

$$\frac{\widehat{\theta}_i(p)}{\theta(p)} - 1 \xrightarrow{P} 0\,, \qquad \text{as } n \to \infty\,. \tag{4.2.15}$$

Proof: From (4.2.14) we get

$$\frac{\widehat{\theta}_i(p)}{\theta(p)} - 1 \;=\; \left[\left\{\frac{\widehat{\vartheta}_{n,i}}{\vartheta} - 1\right\} + \left\{\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right\}\frac{\widehat{\vartheta}_{n,i}}{\vartheta}\right]\big(1 + o(1)\big)\;. \tag{4.2.16}$$

Corollary 3.3.8 implies that both $\widehat{\vartheta}_{n,0}$ and $\widehat{\vartheta}_{n,1}$ are consistent estimators of ϑ. So the first term on the right hand side converges in probability to 0 and Lemma 4.1.3 finishes the proof. ∎

The next result is the main theorem for the estimation of CTE. For this, we need the conditions of Corollary 3.3.8 and Theorem 4.1.4 to hold.

Theorem 4.2.4 *Assume that $\{X_j, j \in \mathbb{Z}\}$ is an iid regularly varying sequence of random variables with strictly increasing marginal distribution F_X continuous on $(0, \infty)$. Assume that $\overline{F}_X \in 2RV(-\alpha, -\rho)$ and (3.1.23) holds.*
Suppose that the conditions (3.1.25), (4.1.15) and (4.2.12) hold. Assume that k/np is bounded away from zero and ∞ and that

$$\lim_{n\to\infty} \log\left(\frac{k}{np}\right) = r \in (-\infty, \infty)\,.$$

Then, as $n \to \infty$, we have the following:

1. If $\alpha > 2$,

$$\sqrt{k}\left\{\frac{\widehat{\theta}_0(p)}{\theta(p)} - 1\right\}$$

$$\xrightarrow{\mathscr{D}} \frac{1}{\alpha}\, G(1) + r\int_1^\infty \frac{G(t) - G(1)T(t)}{t}\, dt + \frac{1}{\vartheta}\int_1^\infty \big(G(t) - G(1)\, T(t)\big)\ dt \tag{4.2.17}$$

$$\sim \mathscr{N}\left(0, \frac{(r^2+2r+2)\alpha^2 - (3r^2+4r+4)\alpha + 2(1+r^2)}{\alpha^2(\alpha-1)(\alpha-2)}\right). \tag{4.2.18}$$

2. If $\alpha > 1$,

$$\sqrt{k}\left\{\frac{\widehat{\theta}_1(p)}{\theta(p)} - 1\right\}$$

$$\xrightarrow{\mathscr{D}} \frac{1}{\alpha}\, G(1) + (r+\vartheta)\int_1^\infty \frac{G(t) - G(1)T(t)}{t}\, dt \tag{4.2.19}$$

$$\sim \mathscr{N}\left(0, \frac{(r^2+2r+2)\alpha^2 - 2(r^2+r+1)\alpha + (1+r^2)}{\alpha^2(\alpha-1)^2}\right). \tag{4.2.20}$$

Proof: By (4.2.9), we have for $i = 0, 1$,

$$\begin{aligned}
\sqrt{k}\left\{\frac{\widehat{\theta}_i(p)}{\theta(p)} - 1\right\} &= \sqrt{k}\left\{\frac{\widehat{\vartheta}_{n,i}\ \widehat{Q}_X(1/p)}{\theta(p)} - 1\right\} \\
&= \sqrt{k}\left\{\frac{\widehat{\vartheta}_{n,i}}{\vartheta} - 1\right\} + \sqrt{k}\left\{\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right\}\frac{\widehat{\vartheta}_{n,i}}{\vartheta} \qquad (4.2.21) \\
&\quad + \sqrt{k}\left\{\frac{\vartheta\ Q_X(1/p)}{\theta(p)} - 1\right\}\frac{\widehat{\vartheta}_{n,i}}{\vartheta}\ \frac{\widehat{Q}_X(1/p)}{Q_X(1/p)}\,. \qquad (4.2.22)
\end{aligned}$$

From Corollary 3.3.7 we have, for $\alpha > 2$,

$$\sqrt{k}\left(\frac{\widehat{\vartheta}_{n,0}}{\vartheta} - 1\right) \xrightarrow{\mathscr{D}} \frac{1}{\vartheta}\int_1^\infty \big(G(t) - G(1)\, T(t)\big)\ dt\,.$$

On the other hand, if $\alpha > 1$, by Corollary 3.3.1,

$$\begin{aligned}\sqrt{k}\left(\frac{\widehat{\vartheta}_{n,1}}{\vartheta}-1\right) &= \sqrt{k}\left(\widehat{\alpha}_1^{-1}-\alpha^{-1}\right)\widehat{\vartheta}_{n,1} \\ &\xrightarrow{\mathscr{D}} \vartheta\int_1^\infty \frac{(G(t)-G(1)\,T(t))}{t}\,dt\,.\end{aligned}$$

Consistency of $\widehat{\vartheta}_{n,i}$ implies that the ratio $\widehat{\vartheta}_{n,i}/\vartheta$ converges in probability to 1 for $i = 0, 1$, hence by Corollary 4.1.5, the second term in (4.2.21) converges in distribution to

$$\alpha^{-1}\,G(1) + r\int_1^\infty \frac{G(t)-G(1)T(t)}{t}\,dt\,.$$

Consistency of $\widehat{\vartheta}_{n,i}$, Lemma 4.1.3 and the no-bias condition (4.2.12) imply that the term in (4.2.22) converges in probability to zero. Putting this together yields the asymptotic representations in (4.2.17) and (4.2.19).

It remains to calculate the asymptotic variances in (4.2.18) and (4.2.20), which are justified in Lemma 4.2.5. ∎

Lemma 4.2.5 *The limiting variances in (4.2.18) and (4.2.20) are respectively*

$$\frac{(r^2+2r+2)\alpha^2-(3r^2+4r+4)\alpha+2(1+r^2)}{\alpha^2(\alpha-1)(\alpha-2)}\,,$$
$$\frac{(r^2+2r+2)\alpha^2-2(r^2+r+1)\alpha+(1+r^2)}{\alpha^2(\alpha-1)^2}\,.$$

Proof:

$$\mathbb{V}ar[G(1)] = 1\,,$$
$$\mathbb{V}ar\left[\int_1^\infty \frac{G(t)-G(1)\,T(t)}{t}\,dt\right] = \frac{1}{\alpha^2}\,,$$
$$\mathbb{V}ar\left[\int_1^\infty (G(t)-G(1)\,T(t))\,dt\right] = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}\,,$$

We note that $G(1)$ and $\int_1^\infty\{G(t)-G(1)\,T(t)\}/t^m\,dt$ are uncorrelated for $m \geq 0$. Hence, it remains to calculate the covariance between the two integrals.

Recall that $T(t) = t^{-\alpha}$ and $\mathbb{E}[G(t)G(s)] = (t \vee s)^{-\alpha}$.

Since

$$\int_1^\infty T(t)\,dt = \frac{1}{\alpha - 1} \quad \text{and} \quad \int_1^\infty \frac{T(t)}{t}\,dt = 1/\alpha\ ,$$

we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_1^\infty \frac{G(t)}{t}\,dt - \frac{G(1)}{\alpha}\right)\left(\int_1^\infty G(s)\,ds - \frac{G(1)}{\alpha-1}\right)\right] \\
&\quad = \mathbb{E}\left[\int_1^\infty \frac{G(t)}{t}\,dt \int_1^\infty G(s)\,ds\right] - \frac{1}{\alpha-1}\int_1^\infty s^{-\alpha}\,\frac{ds}{s} - \frac{1}{\alpha}\int_1^\infty t^{-\alpha}\,dt + \frac{1}{\alpha(\alpha-1)} \\
&\quad = \int_1^\infty t^{-\alpha-1}\int_1^t ds\,dt + \int_1^\infty t^{-1}\int_t^\infty s^{-\alpha}\,ds\,dt - \frac{1}{\alpha(\alpha-1)} \\
&\quad = \int_1^\infty t^{-\alpha-1}(t-1)\,dt + \int_1^\infty t^{-1}\frac{t^{-\alpha+1}}{\alpha-1}\,dt - \frac{1}{\alpha(\alpha-1)} \\
&\quad = \int_1^\infty t^{-\alpha}\,dt - \int_1^\infty t^{-\alpha-1}\,dt + \frac{1}{\alpha-1}\int_1^\infty t^{-\alpha}\,dt - \frac{1}{\alpha(\alpha-1)} \\
&\quad = \frac{1}{\alpha-1} - \frac{1}{\alpha} + \frac{1}{(\alpha-1)^2} - \frac{1}{\alpha(\alpha-1)} \\
&\quad = \frac{1}{(\alpha-1)^2}\ .
\end{aligned}$$

Therefore, the limiting variances are

$$\begin{aligned}
\frac{1}{\alpha^2} + \left(r + \frac{\alpha}{\alpha-1}\right)^2 \frac{1}{\alpha^2} &= \frac{1}{\alpha^2}\left(1 + r^2 + \frac{2r\alpha}{\alpha-1} + \frac{\alpha^2}{(\alpha-1)^2}\right) \\
&= \frac{\alpha^2 - 2\alpha + 1 + r^2\alpha^2 - 2r^2\alpha + r^2 + 2r\alpha^2 - 2r\alpha + \alpha^2}{\alpha^2(\alpha-1)^2} \\
&= \frac{(r^2+2r+2)\alpha^2 - 2(r^2+r+1)\alpha + (1+r^2)}{\alpha^2(\alpha-1)^2}\ .
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\alpha^2}+\frac{r^2}{\alpha^2}+\left(\frac{\alpha-1}{\alpha}\right)^2\frac{\alpha}{(\alpha-1)^2(\alpha-2)}+2r\frac{\alpha-1}{\alpha}\frac{1}{(\alpha-1)^2}\\
&\qquad=\frac{1+r^2}{\alpha^2}+\frac{1}{\alpha(\alpha-2)}+\frac{2r}{\alpha(\alpha-1)}\\
&\qquad=\frac{(1+r^2)(\alpha^2-3\alpha+2)+\alpha(\alpha-1)+2r\alpha(\alpha-2)}{\alpha^2(\alpha-1)(\alpha-2)}\\
&\qquad=\frac{\alpha^2-3\alpha+2+\alpha^2r^2-3\alpha r^2+2r^2+\alpha^2-\alpha+2r\alpha^2-4r\alpha}{\alpha^2(\alpha-1)(\alpha-2)}\\
&\qquad=\frac{(r^2+2r+2)\alpha^2-(3r^2+4r+4)\alpha+2(1+r^2)}{\alpha^2(\alpha-1)(\alpha-2)}\,,
\end{aligned}$$

∎

Comment 4.2.6 It is tedious but straightforward to show that the variance in (4.2.18) is greater than that in (4.2.20); for all r, the difference is $[\alpha(\alpha-1)^2(\alpha-2)]^{-1}$. For $\alpha \gg 2$, this difference is insignificant, but we see that as $\alpha \to 2^+$, the estimator $\widehat{\theta}_1(p)$ is much more stable.

4.3 Discussion

4.3.1 No-bias condition

In this chapter we consider two no-bias conditions: (4.1.15) and (4.2.12). The former is

$$\limsup_{n\to\infty} \sqrt{k} \left| \frac{\left(\frac{k}{np}\right)^{-1/\alpha} Q_X(1/p)}{Q_X(n/k)} - 1 \right| = 0\,. \tag{4.3.1}$$

The latter one is used for the CTE. The results on CTE seem to be new and as such we cannot compare the no-bias condition (4.2.12) to an existing one. As for (4.3.1), we recall that in case of continuous F we have $U(x) = Q(x)$ and hence (4.3.1) is implied by (3.4.5) and (3.4.6). Bearing in mind again (cf. Section 3.4) that (3.4.2) and (3.4.6) are essentially equivalent, we conclude that the no-bias condition (4.3.1) used in this chapter is essentially the same as the no-bias condition used in Chapter 3. We do not claim, however, that these conditions are equivalent for all distribution functions F.

4.3.2 Comparison with existing results

Estimation of the value-at-risk, using a slightly different approach, is tackled in Theorem 4.3.8 in [14]. There, the Value-At-Risk is defined via

$$\mathrm{VaR}_p(X) = U(1/p)\,, \tag{4.3.2}$$

where U is defined in (3.4.1). In [14] the following approximation is used:

$$U(1/p) \approx U(n/k) + a(n/k)\frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma}\,,$$

where a is defined in (3.4.3). Noting that $a(t) \approx tU'(t) \approx \gamma U(t)$ and that $U(n/k)$ can be approximated by $X_{n:n-k}$, we obtain (4.1.12). We do not claim any novelty of Theorem 4.1.4, but it should be pointed out that in [14] only the case of $r = \infty$ was considered with a different normalization, see Theorem 4.3.8 of [14]. We also note that the authors used the assumptions (3.4.5) and (3.4.6).

Estimation of CTE is certainly connected to the behaviour of $\sum_{j=1}^{k} X_{n:n-j+1}$. In [11] the authors studied the sums

$$a_n \sum_{j=1}^{k} X_{n:n-j+1} - c_n$$

with $a_n = (nk^{-1})^{1/2-1/\alpha} n^{-1/2}$ and a centering sequence c_n and obtained asymptotic normality.

■

Chapter 5

Functional Central Limit Theorem for the Bootstrapped TEP

The asymptotic results of Chapter 3 all involve functionals of the Gaussian process $G(\cdot)$ of Theorem 3.1.1, which in turn depends on the unknown tail index α. Furthermore, the first limiting variance for the CTE in Lemma 4.2.5 is finite only when $\alpha > 2$, while the second variance gets large for small values of $\alpha > 1$. Because the estimators of the tail index α tend to be very unstable, the theoretical formulas for the limiting variance may be of limited use. Therefore, we turn to bootstrap techniques in order to approximate the limiting distributions of the various estimators that are functionals of the tail empirical process.

In this chapter, we shall see that Efron's non parametric bootstrap (cf. [18]) can be applied in a straightforward way to approximate the limiting Gaussian distribution of Theorem 3.2.2. To the best of our knowledge this is the first result of this nature for the integral of the TEP with random levels.

5.1 TEP with deterministic levels for arrays

We begin by extending Theorem 3.1.1 to one for arrays of random variables. Recall that Theorem 3.1.1 states that the tail empirical process

$$\begin{aligned} G_n(\cdot) &= \sqrt{n\overline{F}(u_n)}\left\{\widetilde{T}_n(\cdot) - T_n(\cdot)\right\} \\ &= \sqrt{n\overline{F}(u_n)}\left\{\frac{\overline{F}_n(u_n\cdot) - \overline{F}(u_n\cdot)}{\overline{F}(u_n)}\right\}, \end{aligned}$$

where $\overline{F}_n(x) = (1/n)\sum_{j=1}^n \mathbb{1}_{\{X_j > x\}}$, converges in distribution to the Gaussian process $G(\cdot)$ on $\mathscr{D}(0,\infty)$ if the following conditions hold : $\overline{F}(u_n t)/\overline{F}(u_n) \longrightarrow t^{-\alpha}$, $u_n \longrightarrow \infty$ and $n\overline{F}(u_n) \longrightarrow \infty$ as $n \longrightarrow \infty$, where F is a continuous marginal distribution.

Now, let $\{X_{1n}, \ldots, X_{nn},\ n = 1, 2, \ldots\}$ be an array of random variables such that $(X_{1n}, \ldots, X_{nn})$ are i.i.d. with distribution function $F^{(n)}$ such that for $\alpha > 0$ and $t > 0$,

$$\frac{\overline{F}^{(n)}(u_n t)}{\overline{F}^{(n)}(u_n)} \longrightarrow t^{-\alpha}, \quad \text{as } n \longrightarrow \infty\,, \tag{5.1.1}$$

where u_n is a non decreasing sequence such that

$$u_n \longrightarrow \infty\,, \qquad n\overline{F}^{(n)}(u_n) \longrightarrow \infty\,. \tag{5.1.2}$$

Let

$$\overline{F}_n^{(n)}(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_{jn} > x\}}$$

denote the **tail empirical distribution** generated by $(X_{1n}, \ldots, X_{nn})$ and consider the **tail empirical process for arrays**

$$G_n^{(n)}(t) = \sqrt{n\overline{F}^{(n)}(u_n)}\left\{\frac{\overline{F}_n^{(n)}(u_n t) - \overline{F}^{(n)}(u_n t)}{\overline{F}^{(n)}(u_n)}\right\}, \quad t > 0\,. \tag{5.1.3}$$

Theorem 3.1.1 holds if we replace G_n by $G_n^{(n)}$, since the proofs of Lemmas 3.1.2, 3.1.3 and 3.1.4 remain identical with F replaced by $F^{(n)}$. This allows us to state the following corollary to Theorem 3.1.1 without proof.

Corollary 5.1.1 *Assume that $\{X_{1n}, \ldots, X_{nn},\ n = 1, 2, \ldots\}$ is an array of random variables such that $(X_{1n}, \ldots, X_{nn})$ are i.i.d. with distribution function $F^{(n)}$. Suppose the conditions (5.1.1)-(5.1.2) hold. Then, as $n \longrightarrow \infty$*

$$G_n^{(n)}(\cdot) \xrightarrow{\mathscr{D}} G(\cdot)\,, \tag{5.1.4}$$

in $\mathscr{D}(0, \infty)$, where $G(\cdot)$ is the centered Gaussian process defined in Theorem 3.1.1.

5.2 Bootstrap for TEP

5.2.1 Description of bootstrap method

Let $\mathscr{X}_n = \{X_1, \ldots, X_n\}$ be a sample from an unknown distribution function F and let

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}\,, \tag{5.2.1}$$

denote the empirical distribution generated by our original sequence $\{X_1, \ldots, X_n\}$. The bootstrap procedure starts with drawing a sample $\left\{X_1^{(b)}, X_2^{(b)}, \ldots, X_n^{(b)}\right\}$ with replacement, n times, from the data $\mathscr{X}_n$. Conditionally on $\mathscr{X}_n$, the random variables $\left\{X_1^{(b)}, X_2^{(b)}, \ldots, X_n^{(b)}\right\}$ are i.i.d. Moreover we have

$$\mathbb{P}_{(b)}(X_1^{(b)} = X_i) := \mathbb{P}(X_1^{(b)} = X_i | \mathscr{X}_n) = 1/n, \qquad i = 1\,, \ldots, n\,,$$

where $\mathbb{P}_{(b)}$ denotes the conditional probability given the sample $\{X_1, X_2, \ldots, X_n\}$. This means that given the sample $\mathscr{X}_n$, the common distribution function of $X_i^{(b)}$ $(i = 1, 2, ..., n)$ is equal to the empirical distribution function F_n.

Example 5.2.1 Consider estimation of the expected value $\mu = \mathbb{E}(X_1)$ by the sample mean $\overline{X}_n$. Denote by $\overline{X}_n^{(b)} = (1/n)\sum_{i=1}^n X_i^{(b)}$ the bootstrap sample mean. Let $\mathbb{E}_{(b)}$ and $\mathbb{V}ar_{(b)}$ be the expectation and the variance with respect to $\mathbb{P}_{(b)}$ respectively. We have the following moment properties:

$$\mathbb{E}_{(b)}\left[X_1^{(b)}\right] = \mathbb{E}\left[X_1|\mathscr{X}_n\right] = \int x dF_n(x) = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n \ ,$$

$$\mathbb{E}_{(b)}\left[\overline{X}_n^{(b)}\right] = \mathbb{E}\left[\overline{X}_n^{(b)}|\mathscr{X}_n\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_i|\mathscr{X}_n\right] = \frac{1}{n} n\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \overline{X}_n \ ,$$

$$\mathbb{E}\left[\overline{X}_n^{(b)}\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_i^{(b)}\right] = \mathbb{E}\left[\mathbb{E}\left[X_1^{(b)}|\mathscr{X}_n\right]\right] = \mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}[X] \ ,$$

$$\mathbb{V}ar_{(b)}\left[X_1^{(b)}\right] = \int x^2 dF_n(x) - \left(\int x dF_n(x)\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \ .$$

5.2.2 Bootstrap for TEP with deterministic levels

The purpose of this section is to obtain weak convergence for the **bootstrapped tail empirical process** $G_n^{(b)}$ defined hereafter by

$$G_n^{(b)}(t) = \sqrt{n\overline{F}_n(u_n)}\left\{\frac{\overline{F}_n^{(b)}(u_n t) - \overline{F}_n(u_n t)}{\overline{F}_n(u_n)}\right\}, \quad t > 0 \,. \tag{5.2.2}$$

where $\overline{F}_n$ denotes the tail empirical distribution function generated by the i.i.d. random variables $\{X_1, \ldots, X_n\}$ and $\overline{F}_n^{(b)}$ denotes the bootstrapped tail empirical distribution defined as follows :

$$\overline{F}_n^{(b)}(x) = \frac{1}{n}\sum_{i=1}^n 1_{\left\{X_i^{(b)} > x\right\}} \,. \tag{5.2.3}$$

Lemma 5.2.2 *Under the conditions of Theorem 3.1.1 and as $n \longrightarrow \infty$, for $t > 0$, we have*

$$\frac{\overline{F}_n(u_n t)}{\overline{F}_n(u_n)} \longrightarrow t^{-\alpha} \quad \textit{and} \quad n\overline{F}_n(u_n) \longrightarrow \infty\,, \tag{5.2.4}$$

in probability.

Proof: From Theorem 3.1.1, for every $t > 0$ and as $n \longrightarrow \infty$, we have

$$\frac{\left|\overline{F}_n(u_n t) - \overline{F}(u_n t)\right|}{\overline{F}(u_n)} \xrightarrow{P} 0\,. \tag{5.2.5}$$

Therefore, since $\overline{F}$ is regularly varying with index $\alpha > 0$,

$$\left|\frac{\overline{F}_n(u_n t)}{\overline{F}(u_n)} - t^{-\alpha}\right| \leq \left|\frac{\overline{F}_n(u_n t) - \overline{F}(u_n t)}{\overline{F}(u_n)}\right| + \left|\frac{\overline{F}(u_n t)}{\overline{F}(u_n)} - t^{-\alpha}\right| \xrightarrow{P} 0\,. \tag{5.2.6}$$

Applying (5.2.6) with $t = 1$ yields

$$\frac{\overline{F}_n(u_n)}{\overline{F}(u_n)} \xrightarrow{P} 1, \qquad \text{as } n \longrightarrow \infty\,. \tag{5.2.7}$$

Combining (5.2.6) and (5.2.7) finishes the proof. ∎

Theorem 5.2.3 *Assume that $\left\{X_1^{(b)}, \ldots, X_n^{(b)}, n = 1, 2, \ldots\right\}$ is a bootstrap array of random variables from the empirical distribution functions F_n. Under the assumptions of Theorem 3.1.1, as $n \longrightarrow \infty$, we have that on $\mathscr{D}(0,\infty)$*

$$G_n^{(b)}(\cdot) \xrightarrow{\mathscr{D}} G(\cdot)\,, \tag{5.2.8}$$

in probability, where $G(\cdot)$ is the centered Gaussian process defined in Theorem 3.1.1. (i.e. for every subsequence $n' \longrightarrow \infty$, there is a further subsequence $n'' \longrightarrow \infty$ such that $G_{n''}^{(b)} \xrightarrow{\mathscr{D}} G$ with probability 1).

Proof: From Lemma 5.2.2 and by the monotonicity of the functions $\overline{F}_n(u_n\cdot)/\overline{F}_n(u_n)$ and the continuity of the limit $t^{-\alpha}$, we observe that for any $0 < a < b < \infty$,

$$\sup_{a \le t \le b} \left| \frac{\overline{F}_n(u_n t)}{\overline{F}_n(u_n)} - t^{-\alpha} \right| \xrightarrow{P} 0 ,$$

(see, for example, Lemma 1 in [31]). This implies that for any subinterval $[a_k, b_k]$, for every subsequence $n' \to \infty$, there is a further subsequence $n'' \to \infty$ such that with probability 1,

$$\sup_{a_k \le t \le b_k} \left| \frac{\overline{F}_{n''}(u_{n''} t)}{\overline{F}_{n''}(u_{n''})} - t^{-\alpha} \right| \to 0 . \tag{5.2.9}$$

Now, letting $a_k \to 0$ and $b_k \to \infty$, a diagonalization argument can be used to define a single subsequence $n'' \to \infty$ such that with probability 1, (5.2.9) holds for all k. Since $\cup_k [a_k, b_k] = (0, \infty)$ it follows that, for every $0 < t < \infty$, with probability 1

$$\frac{\overline{F}_{n''}(u_{n''} t)}{\overline{F}_{n''}(u_{n''})} \longrightarrow t^{-\alpha} . \tag{5.2.10}$$

By Lemma 5.2.2, we can also choose n'' so that with probability 1

$$n'' \overline{F}_{n''}(u_{n''}) \longrightarrow \infty . \tag{5.2.11}$$

Now, to finish the proof, we apply Corollary 5.1.1. ∎

5.2.3 Bootstrap for TEP with order statistics

Let $X_{n:n} \ge X_{n:n-1} \ge \cdots \ge X_{n:1}$ be the order statistics from the sample $X_1, \ldots, X_n$. Set $k = n\overline{F}(u_n)$ such that $k = k_n \longrightarrow \infty$ and $k/n \longrightarrow 0$ as $n \longrightarrow \infty$.
Consider the bootstrapped process

$$\widehat{G}_n^{(b)}(t) = \frac{n}{\sqrt{k}} \left\{ \overline{F}_n^{(b)}(X_{n:n-k}\, t) - \overline{F}_n(X_{n:n-k}\, t) \right\} , \quad t > 0 , \tag{5.2.12}$$

where $\overline{F}_n^{(b)}$ is the tail of the bootstrapped empirical distribution defined in (5.2.3). Recall that

$$\widehat{T}_n(t) = (1/k)\sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{n:n-k}t\}} \quad \text{and} \quad \overline{F}_n(x) = (1/n)\sum_{j=1}^{n} \mathbb{1}_{\{X_j > x\}}\,.$$

Then, one can deduce that

$$\widehat{T}_n(t) = \frac{\overline{F}_n(X_{n:n-k}t)}{k/n}\,. \tag{5.2.13}$$

Theorem 5.2.4 *Assume that* $\left\{X_1^{(b)},\ldots,X_n^{(b)}, n=1,2,\ldots\right\}$ *is a bootstrap array of random variables from the empirical distribution functions* F_n*. Suppose that the conditions (3.1.1), (3.1.2) and (3.1.25) hold. Then as* $n \longrightarrow \infty$*, we have that on* $\mathscr{D}(0,\infty)$

$$\widehat{G}_n^{(b)}(\cdot) \quad \overset{\mathscr{D}}{\longrightarrow} \quad G(\cdot) \tag{5.2.14}$$

and

$$\widehat{G}_n^{(b)}(\cdot) - \widehat{G}_n^{(b)}(1)\,\widehat{T}_n(\cdot) \quad \overset{\mathscr{D}}{\longrightarrow} \quad G(\cdot) - G(1)\,T(\cdot)\,, \tag{5.2.15}$$

in probability, where $G(\cdot)$ *is the centered Gaussian process defined in Theorem 3.1.1.*

Proof: Recall that, for $t > 0$,

$$G_n^{(b)}(t) = \sqrt{n\overline{F}_n(u_n)}\left\{\frac{\overline{F}_n^{(b)}(u_n t) - \overline{F}_n(u_n t)}{\overline{F}_n(u_n)}\right\}\,. \tag{5.2.16}$$

Then, by (5.2.12) and since $k/n = \overline{F}(u_n)$, we have

$$\begin{aligned} G_n^{(b)}\left(t\,\frac{X_{n:n-k}}{u_n}\right) &= \sqrt{\frac{n}{\overline{F}_n(u_n)}}\,\left\{\overline{F}_n^{(b)}(X_{n:n-k}\;t) - \overline{F}_n(X_{n:n-k}\;t)\right\} \\ &= \sqrt{\frac{k/n}{\overline{F}_n(u_n)}}\,\widehat{G}_n^{(b)}(t) = \sqrt{\frac{\overline{F}(u_n)}{\overline{F}_n(u_n)}}\,\widehat{G}_n^{(b)}(t)\,. \end{aligned}$$

So, since $\overline{F}_n(u_n)/\overline{F}(u_n) \xrightarrow{P} 1$, as $n \longrightarrow \infty$, we conclude that as $n \longrightarrow \infty$

$$\widehat{G}_n^{(b)}(t) = G_n^{(b)}\left(t\,\frac{X_{n:n-k}}{u_n}\right)\big(1 + o_p(1)\big)\ , \tag{5.2.17}$$

where the $o_p(1)$ term does not depend on t.

Lemma 3.1.11 implies that $X_{n:n-k}/u_n$ converges in probability to 1, so we have

$$t\,\frac{X_{n:n-k}}{u_n} \xrightarrow{P} t\ , \tag{5.2.18}$$

in $\mathscr{C}_0(0,\infty)$. Hence, Theorem 5.2.3 and continuity of the composition map (Theorem 2.5.9) can be applied to the right hand side of (5.2.17) to prove (5.2.14).

Now, from Theorem 3.1.9, we have

$$\widehat{T}_n(t) = \frac{\overline{F}_n(X_{n:n-k}\ t)}{k/n} \xrightarrow{P} t^{-\alpha}\,, \tag{5.2.19}$$

in $\mathscr{C}_0(0,\infty)$. Hence, (5.2.15) has been proven. ∎

Remark 5.2.5 We note that Theorem 3.1.9 requires regular variation of the tail distribution $\overline{F}$ (3.1.1), convergence of the sequence $u_n \to \infty$ so that $n\overline{F}_n(u_n) \to \infty$ (3.1.2), the "no-bias" condition (3.1.23), as well as the differentiability condition (3.1.25) for T_n. These conditions are required to determine the asymptotic behaviour of each of the three terms in the decomposition of the process $\widehat{G}_n$ given on page 41. As a result, we have that $\quad \widehat{G}_n(\cdot) \xrightarrow{\mathscr{D}} G(\cdot) - G(1)T(\cdot)$.

The extra term $G(1)T(\cdot)$ arises as a result of replacing the deterministic (but unknown) upper (k/n)th quantile u_n with the upper order statistic $X_{n:n-k}$ (recall that $X_{n:n-k}/u_n \to 1$ in probability).

On the other hand, Theorem 5.2.4 follows from Theorem 3.1.1 and Lemma 3.1.11 and does not require the no-bias condition (3.1.23). $\widehat{G}_n^{(b)}(\cdot) \xrightarrow{\mathscr{D}} G(\cdot)$ in probability because $X_{n:n-k}$ plays the same role for the process $\widehat{G}_n^{(b)}$ as u_n does for the original process G_n (they are the upper (k/n)th quantiles of F_n and F, respectively). Indeed, since the analogues of (3.1.1) and (3.1.2),

$$\widehat{T}_n(t) = \frac{\overline{F}_n(X_{n:n-k}t)}{\overline{F}_n(X_{n:n-k})} = \frac{\overline{F}_n(X_{n:n-k}t)}{k/n} \longrightarrow t^{-\alpha} \quad \text{and} \quad n\overline{F}_n(X_{n:n-k}) \longrightarrow \infty,$$

are satisfied in probability, conditional weak convergence to $G(\cdot)$ follows in probability. Then, in order to achieve the same limiting distribution as in Theorem 3.1.9, we use the fact that $\widehat{T}_n(t) \xrightarrow{P} t^{-\alpha}$, which leads to (5.2.15).

The question that naturally arises is: why not use random levels in the bootstrapped CTE? In other words, can we replace $X_{n:n-k}$ with its bootstrap counterpart $X^{(b)}_{n:n-k}$ in (5.2.12) and achieve the limiting distribution of Theorem 3.1.9 directly? We conjecture that this may be the case, but have been unable to provide a rigorous proof. The difficulty is in proving a counterpart for the differentiability condition (3.1.25). Since the empirical distribution F_n is not differentiable, such a condition would involve convergence in probability of

$$\frac{\overline{F}_n(X_{n:n-k}(t+\delta_n)) - \overline{F}_n(X_{n:n-k}t)}{\delta_n\overline{F}_n(X_{n:n-k})}$$

to $T^{'}(t) = -\alpha t^{-\alpha-1}$ uniformly in a neighbourhood of 1 for any sequence $\delta_n \to 0$. This question remains open.

5.3 Bootstrapping tail index and risk measures

5.3.1 Hill and t-Hill estimators using bootstrap technique

Let $m \geq 0$. The Theorem 5.3.1 below, which is the bootstrap counterpart of Theorem 3.2.2, allows us to approximate the distribution of $\sqrt{k}\left(\widehat{\alpha}_1^{-1} - \alpha^{-1}\right)$, where $\widehat{\alpha}_1^{-1}$ denotes the Hill estimator defined by

$$\widehat{\alpha}_1^{-1} = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right) = \int_1^{\infty} \widehat{T}_n(t)\frac{\mathrm{d}t}{t} \tag{5.3.1}$$

and as well $\sqrt{k}\left\{\widehat{\zeta}_m - \zeta_m\right\}$ for $m \neq 1$, where

$$\zeta_m = \int_1^{\infty} \frac{T(t)}{t^m}\,\mathrm{d}t = \frac{1}{\alpha + m - 1} \qquad \text{and} \qquad \widehat{\zeta}_m = \int_1^{\infty} \frac{\widehat{T}_n(t)}{t^m}\,\mathrm{d}t\,.$$

Theorem 5.3.1 *Under the assumptions of Theorem 3.1.9 and if $\alpha > 2(1-m)$, then as $n \longrightarrow \infty$*

$$\begin{aligned}\int_1^{\infty} \frac{\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\,\widehat{T}_n(t)}{t^m}\,\mathrm{d}t \quad &\xrightarrow{\mathscr{D}} \quad \int_1^{\infty} \frac{G(t) - G(1)\,T(t)}{t^m}\,\mathrm{d}t \\ &\sim \quad \mathscr{N}\left(0, \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)}\right)\,.\end{aligned} \tag{5.3.2}$$

in probability, where $G(\cdot)$ is a centered Gaussian process with covariance function

$$\mathbb{C}ov\left(G(t_1), G(t_2)\right) = (t_1 \vee t_2)^{-\alpha}\,.$$

We present two lemmas that will provide the proof of Theorem 5.3.1.

Lemma 5.3.2 *Assume that (3.1.1), (3.1.2) and (3.1.25) hold. If $\alpha > 2(1-m)$, then we have that on $\mathscr{D}(0,\infty)$,*

$$\int_1^\infty \frac{\widehat{G}_n^{(b)}(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\,, \qquad as\ n \longrightarrow \infty\,, \tag{5.3.3}$$

in probability.

Proof: We will show that for any subsequence n', there is a further subsequence n'' such that

$$\int_1^\infty \frac{\widehat{G}_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t, \quad \text{w.p.1.} \tag{5.3.4}$$

Assume that $\alpha > 2(1-m)$. Begin by choosing a subsequence n'' of n' such that

(i) $\widehat{G}_{n''}^{(b)}(\cdot) \xrightarrow{\mathscr{D}} G(\cdot)$ w.p.1 (by Theorem 5.2.4);

(ii) $\xi_{n''} = \dfrac{X_{n'':n''-k}}{u_{n''}} \to 1$ w.p.1 (by Lemma 3.1.11);

(iii) $\dfrac{\overline{F}_{n''}(u_{n''})}{\overline{F}(u_{n''})} \to 1$ w.p.1 $\big($by (5.2.10) with $t = 1\big)$;

For any $M > 1$, we can write that

$$\int_1^\infty \frac{\widehat{G}_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t = \int_1^M \frac{\widehat{G}_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t + \int_M^\infty \frac{\widehat{G}_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t\,. \tag{5.3.5}$$

By (i) and the continuous mapping theorem,

$$\int_1^M \frac{\widehat{G}_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^M \frac{G(t)}{t^m}\,\mathrm{d}t\,, \tag{5.3.6}$$

with probability 1 as $n'' \longrightarrow \infty$.

Letting $M \longrightarrow \infty$, we get that

$$\int_1^M \frac{G(t)}{t^m}\,\mathrm{d}t \longrightarrow \int_1^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\,. \tag{5.3.7}$$

This is true since

$$\mathbb{V}ar\left[\int_M^\infty \frac{G(t)}{t^m}\,\mathrm{d}t\right] = \frac{2\,M^{-\alpha-2m+2}}{(\alpha+m-1)(\alpha+2m-2)} \longrightarrow 0\,, \quad \text{as } M \longrightarrow \infty\,.$$

Now, let $\xi_{n''} = X_{n'':n''-k}/u_{n''}$. Then, by (5.2.17), **(ii)** and **(iii)** we have

$$\int_M^\infty \frac{\widehat{G}^{(b)}_{n''}(t)}{t^m}\,\mathrm{d}t = \int_M^\infty \frac{G^{(b)}_{n''}(\xi_{n''}t)}{t^m}\big(1+o(1)\big)\,\mathrm{d}t, \quad \text{w.p.1}, \tag{5.3.8}$$

where the $o(1)$ term is uniform in t.

Hence, by Theorem 4.2 of [5], to finish the proof, it suffices to show that for all $\epsilon > 0$,

$$\lim_M \limsup_{n''} \mathbb{P}_{(b)}\left[\left|\int_M^\infty \frac{G^{(b)}_{n''}(\xi_{n''}t)}{t^m}\,\mathrm{d}t\right| \geq \epsilon\right] = 0, \quad \text{w.p.1.} \tag{5.3.9}$$

We put $s = \xi_{n''}t$, then

$$\begin{aligned}
\mathbb{P}_{(b)}\left[\left|\int_M^\infty \frac{G^{(b)}_{n''}(\xi_{n''}t)}{t^m}\,\mathrm{d}t\right| \geq \epsilon\right] &= \mathbb{P}_{(b)}\left[\left|\xi_{n''}^{m-1}\int_{\xi_{n''}M}^\infty \frac{G^{(b)}_{n''}(s)}{s^m}\,\mathrm{d}s\right| \geq \epsilon\right] \\
&\leq \mathbb{P}_{(b)}\left[\left|\xi_{n''}^{m-1}\int_{M}^\infty \frac{G^{(b)}_{n''}(s)}{s^m}\,\mathrm{d}s\right| \geq \frac{\epsilon}{2}\right] \\
&\quad + \mathbb{P}_{(b)}\left[\left|\xi_{n''}^{m-1}\int_{\xi_{n''}M}^M \frac{G^{(b)}_{n''}(s)}{s^m}\,\mathrm{d}s\right| \geq \frac{\epsilon}{2}\right] \\
&= I_1 + I_2\,.
\end{aligned}$$

First, consider I_2 and note that $\xi_{n''}$ is $\sigma(X_1,\ldots,X_n)$ measurable. Fixing M, we choose $0<\delta<1$ such that for $C_{m,\delta} = (1+\delta)^{m-1} \vee (1-\delta)^{m-1}$

$$I_2 \leq \mathbb{P}_{(b)}\left[C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)} \left|\frac{G^{(b)}_{n''}(s)}{s^m}\right|\,\mathrm{d}s \geq \frac{\epsilon}{2}\right] + \mathbb{1}_{\left\{|\xi_{n''}-1|\geq\delta\right\}}\,. \tag{5.3.10}$$

Therefore, along subsequence n'', by (i), (ii), the continuous mapping theorem and Markov's inequality, since $\alpha > 2(1-m)$, we have with probability 1

$$
\begin{aligned}
\limsup_{n''} I_2 &\leq \mathbb{P}\left[C_{m,\delta}\int_{M(1-\delta)}^{M(1+\delta)} \left|G(s)\right| \frac{\mathrm{d}s}{s^m} \geq \frac{\epsilon}{2}\right] \\
&\leq \frac{2\,C_{m,\delta}}{\epsilon}\,\mathbb{E}\left[\int_{M(1-\delta)}^{M(1+\delta)} \left|G(s)\right| \frac{\mathrm{d}s}{s^m}\right] \\
&= \frac{2\sqrt{2}\,C_{m,\delta}}{\epsilon\sqrt{\pi}} \int_{M(1-\delta)}^{M(1+\delta)} s^{-\alpha/2-m}\,\mathrm{d}s \\
&= \frac{2\sqrt{2}\,C_{m,\delta}}{\epsilon\sqrt{\pi}}\frac{2}{\alpha+2(m-1)}\left\{\big[(1-\delta)M\big]^{\frac{2(1-m)-\alpha}{2}} - \big[(1+\delta)M\big]^{\frac{2(1-m)-\alpha}{2}}\right\} \\
&\longrightarrow 0 \quad \text{as } M \longrightarrow \infty\,.
\end{aligned}
$$

Next we consider I_1 and recall that

$$
G_n^{(b)}(t) = \sqrt{n\overline{F}_n(u_n)}\left\{\frac{\overline{F}_n^{(b)}(u_n t) - \overline{F}_n(u_n t)}{\overline{F}_n(u_n)}\right\} \quad \text{and} \quad \overline{F}_n^{(b)}(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{\left\{X_j^{(b)}>x\right\}}\,.
$$

Then, since $\xi_{n''}^{m-1} \longrightarrow 1$ w.p.1 by (ii), it suffices to show that

$$
\lim_{M}\limsup_{n''}\mathbb{P}_{(b)}\left[\left|\int_M^\infty \frac{G_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t\right| \geq \frac{\epsilon}{2}\right] = 0\,.
$$

In fact, by Chebyshev's inequality, we can write

$$
\begin{aligned}
\mathbb{P}_{(b)}\left[\left|\int_M^\infty \frac{G_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t\right| \geq \frac{\epsilon}{2}\right] &\leq \frac{4}{\epsilon^2}\mathbb{V}ar_{(b)}\left[\int_M^\infty \frac{G_{n''}^{(b)}(t)}{t^m}\,\mathrm{d}t\right] \\
&= \frac{4}{\epsilon^2}\mathbb{V}ar_{(b)}\left[\int_M^\infty \frac{1}{\sqrt{n\overline{F}_{n''}(u_{n''})}}\sum_{j=1}^{n}\left[\mathbb{1}_{\left\{X_j^{(b)}>u_{n''}t\right\}} - \overline{F}_{n''}(u_{n''}t)\right]\frac{\mathrm{d}t}{t^m}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{4}{\epsilon^2}\mathbb{V}ar_{(b)}\left[\int_M^\infty \frac{1}{\sqrt{n\overline{F}_{n''}(u_{n''})}}\sum_{j=1}^n \mathbb{1}_{\left\{X_j^{(b)}>u_{n''}t\right\}}\frac{\mathrm{d}t}{t^m}\right] \\
&= \frac{4}{n\overline{F}_{n''}(u_{n''})\,\epsilon^2}\mathbb{V}ar_{(b)}\left[\sum_{j=1}^n\int_M^\infty \mathbb{1}_{\left\{X_j^{(b)}>u_{n''}t\right\}}\frac{\mathrm{d}t}{t^m}\right] \\
&= \frac{4}{\overline{F}_{n''}(u_{n''})\,\epsilon^2}\mathbb{V}ar_{(b)}\left[\int_M^\infty \mathbb{1}_{\left\{X_1^{(b)}>u_{n''}t\right\}}\frac{\mathrm{d}t}{t^m}\right] \\
&\le \frac{4}{\overline{F}_{n''}(u_{n''})\,\epsilon^2}\mathbb{E}_{(b)}\left\{\left[\int_M^\infty \mathbb{1}_{\left\{X_1^{(b)}>u_{n''}t\right\}}\frac{\mathrm{d}t}{t^m}\right]^2\right\} \\
&= \frac{8}{\overline{F}_{n''}(u_{n''})\,\epsilon^2}\int_M^\infty\int_{t=s}^\infty \mathbb{P}_{(b)}\left[X_1^{(b)}>u_{n''}t\right]\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m} \\
&= \frac{8}{\epsilon^2}\int_M^\infty\int_{t=s}^\infty \frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})}\frac{\overline{F}(u_{n''})}{\overline{F}_{n''}(u_{n''})}\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m}\ .
\end{aligned}
$$

By **(iii)**, we get that with probability 1

$$
\mathbb{P}_{(b)}\left[\left|\int_M^\infty \frac{G_{n''}^{(b)}(s)}{s^m}\,\mathrm{d}s\right|\ge\frac{\epsilon}{2}\right]\le\left(\frac{8}{\epsilon^2}\int_M^\infty\int_{t=s}^\infty\frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})}\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m}\right)\big(1+o(1)\big).
$$

To finish the proof, it is sufficient to show that with probability 1

$$
\lim_M\limsup_{n''}\int_M^\infty\int_{t=s}^\infty\frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})}\frac{\mathrm{d}t}{t^m}\frac{\mathrm{d}s}{s^m}=0\ . \tag{5.3.11}
$$

Here, we will consider two cases: the first case when $m\ge 1$ and $\alpha>0$ and the second case when $m=0$ and $\alpha>2$. In both cases we show that $\overline{F}_{n''}(u_{n''}t)$ can be replaced with $\overline{F}(u_{n''}t)$ in (5.3.11).

Case 1: $m \geq 1$ and $\alpha > 0$.

Since for $M > 1$ and $m \geq 1$,

$$\left[\int_M^\infty \int_s^\infty \frac{\overline{F}_n(u_n t)}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t^m} \frac{\mathrm{d}s}{s^m}\right] \leq \left[\int_M^\infty \int_s^\infty \frac{\overline{F}_n(u_n t)}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s}\right],$$

it is enough in this case to prove (5.3.11) for $m = 1$ and $\alpha > 0$.

Let

$$W_{n'',M} := \int_M^\infty \int_{t=s}^\infty \frac{\overline{F}_{n''}(u_{n''}t) - \overline{F}(u_{n''}t)}{\overline{F}(u_{n''})} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s}\ . \tag{5.3.12}$$

Using Potter's bounds, for any chosen $C > 1$ and $\delta > 0$ such that $\alpha > \delta$, we get for all n,

$$\begin{aligned}
&\mathbb{V}ar\left[W_{n,M}\right] = \mathbb{V}ar\left[\int_M^\infty \int_{t=s}^\infty \frac{\sum_{j=1}^n \mathbb{1}_{\{X_j > u_n t\}}}{n\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s}\right] \\
&= n\mathbb{V}ar\left[\int_M^\infty \int_{t=s}^\infty \frac{\mathbb{1}_{\{X_1 > u_n t\}}}{n\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s}\right] \\
&= \frac{1}{n}\mathbb{V}ar\left[\int_M^\infty \int_{t=s}^\infty \frac{\mathbb{1}_{\{X_1 > u_n t\}}}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s}\right] \\
&\leq \frac{1}{n}\mathbb{E}\left\{\left[\int_M^\infty \int_{t=s}^\infty \frac{\mathbb{1}_{\{X_1 > u_n t\}}}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s}\right]^2\right\} \\
&= \frac{4}{n}\mathbb{E}\left\{\int_M^\infty \int_{t=s}^\infty \frac{\mathbb{1}_{\{X_1 > u_n t\}}}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \cdot \int_M^\infty \int_{v=u}^\infty \frac{\mathbb{1}_{\{X_1 > u_n v\}}}{\overline{F}(u_n)} \frac{\mathrm{d}v}{v} \frac{\mathrm{d}u}{u}\right\} \\
&= \frac{8}{k_n}\left\{\int_{u=M}^\infty \int_{s=u}^\infty \int_{t=s}^\infty \int_{v=u}^t \frac{\overline{F}(u_n t)}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}v}{v} \frac{\mathrm{d}u}{u} + \int_{u=M}^\infty \int_{s=u}^\infty \int_{t=s}^\infty \int_{v=t}^\infty \frac{\overline{F}(u_n v)}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}v}{v} \frac{\mathrm{d}u}{u}\right\} \\
&= \frac{8}{k_n}\left\{\int_{u=M}^\infty \int_{s=u}^\infty \int_{t=s}^\infty \frac{\overline{F}(u_n t)}{\overline{F}(u_n)} \int_{v=u}^t \frac{\mathrm{d}v}{v} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}u}{u} + \int_{u=M}^\infty \int_{s=u}^\infty \int_{t=s}^\infty \int_{v=t}^\infty \frac{\overline{F}(u_n v)}{\overline{F}(u_n)} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}v}{v} \frac{\mathrm{d}u}{u}\right\} \\
&\leq \frac{8C}{k_n}\left\{\int_{u=M}^\infty \int_{s=u}^\infty \int_{t=s}^\infty t^{\delta-\alpha}\left(\ln t - \ln u\right) \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}u}{u} + \int_{u=M}^\infty \int_{s=u}^\infty \int_{t=s}^\infty \int_{v=t}^\infty v^{\delta-\alpha-1}\mathrm{d}v\ \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}u}{u}\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{8C}{k_n}\left\{\int_{u=M}^{\infty}\int_{s=u}^{\infty}\int_{t=s}^{\infty}\ln(t)t^{\delta-\alpha-1}\mathrm{d}t\ \frac{\mathrm{d}s}{s}\frac{\mathrm{d}u}{u}+\int_{u=M}^{\infty}\int_{s=u}^{\infty}\int_{t=s}^{\infty}\frac{t^{\delta-\alpha-1}}{(\alpha-\delta)}\ \mathrm{d}t\ \frac{\mathrm{d}s}{s}\frac{\mathrm{d}u}{u}\right\}\\
&=\frac{8C}{k_n}\left\{\int_{u=M}^{\infty}\int_{s=u}^{\infty}\left(\frac{\ln(s)s^{\delta-\alpha}}{(\alpha-\delta)}+\frac{s^{\delta-\alpha}}{(\alpha-\delta)^2}\right)\frac{\mathrm{d}s}{s}\frac{\mathrm{d}u}{u}+\int_{u=M}^{\infty}\int_{s=u}^{\infty}\frac{s^{\delta-\alpha-1}}{(\alpha-\delta)^2}\ \mathrm{d}s\ \frac{\mathrm{d}u}{u}\right\}\\
&=\frac{8C}{k_n}\left\{\int_{u=M}^{\infty}\int_{s=u}^{\infty}\frac{\ln(s)s^{\delta-\alpha-1}}{(\alpha-\delta)}\ \mathrm{d}s+2\int_{u=M}^{\infty}\int_{s=u}^{\infty}\frac{s^{\delta-\alpha-1}}{(\alpha-\delta)^2}\ \mathrm{d}s\ \frac{\mathrm{d}u}{u}\right\}\\
&=\frac{8C}{k_n}\left\{\int_{M}^{\infty}\left(\frac{\ln(u)u^{\delta-\alpha}}{(\alpha-\delta)^2}-\frac{u^{\delta-\alpha}}{(\delta-\alpha)^3}\right)\frac{\mathrm{d}u}{u}+2\int_{M}^{\infty}\frac{u^{\delta-\alpha-1}}{(\alpha-\delta)^3}\ \mathrm{d}u\right\}\\
&=\frac{8C}{k_n}\left\{\int_{M}^{\infty}\frac{\ln(u)u^{\delta-\alpha-1}}{(\alpha-\delta)^2}\mathrm{d}u+3\int_{M}^{\infty}\frac{u^{\delta-\alpha-1}}{(\alpha-\delta)^3}\ \mathrm{d}u\right\}\\
&=\frac{8C}{k_n}\left\{\frac{\ln(M)M^{\delta-\alpha}}{(\alpha-\delta)^3}+4\frac{M^{\delta-\alpha}}{(\alpha-\delta)^4}\right\}\ .
\end{aligned}
$$

It follows that for all n'' sufficiently large and $\epsilon > 0$,

$$
\mathbb{P}\left[\left|W_{n'',M}\right| > \epsilon\right] \leq \frac{8C}{\epsilon^2\, k_{n''}}\left\{\frac{\ln(M)M^{\delta-\alpha}}{(\alpha-\delta)^3}+4\frac{M^{\delta-\alpha}}{(\alpha-\delta)^4}\right\}\ .
$$

The subsequence (n'') can be chosen so that $\sum_{n''}\frac{1}{k_{n''}} < \infty$. In this case, by the first Borel-Cantelli lemma, for every $\epsilon > 0$

$$
\mathbb{P}\left[\left|W_{n'',M}\right| > \epsilon \text{ i.o.}\right] = 0\,.
$$

Hence,

$$
\mathbb{P}\left[\limsup_{n''}\left|W_{n'',M}\right| = 0\right] = \mathbb{P}\left(\left[\bigcup_{r=1}^{\infty}\left\{\left|W_{n'',M}\right| > 1/r \ \ \text{i.o.}\right\}\right]^c\right) = 1\,,
$$

and so with probability 1, since $\overline{F}(u_{n''}t)/\overline{F}(u_{n''}) \longrightarrow t^{-\alpha}$ and by Potter's bounds and dominated convergence, for all $M > 1$ we get

$$\begin{aligned}\limsup_{n''} \int_M^\infty \int_s^\infty \frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})} \frac{\mathrm{d}t}{t^m} \frac{\mathrm{d}s}{s^m} &\le \limsup_{n''} \int_M^\infty \int_s^\infty \frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \\ &= \lim_{n''} \int_M^\infty \int_s^\infty \frac{\overline{F}(u_{n''}t)}{\overline{F}(u_{n''})} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} \\ &= \int_M^\infty \int_s^\infty t^{-\alpha} \frac{\mathrm{d}t}{t} \frac{\mathrm{d}s}{s} = \frac{M^{-\alpha}}{\alpha^2} \ .\end{aligned}$$

Therefore, with probability 1,

$$\lim_M \limsup_{n''} \int_M^\infty \int_s^\infty \frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})} \frac{\mathrm{d}t}{t^m} \frac{\mathrm{d}s}{s^m} \quad \le \quad \lim_M \frac{M^{-\alpha}}{\alpha^2} = 0 \ .$$

This proves (5.3.11) in Case 1.

Case 2: $m = 0$ and $\alpha > 2$.

As noted before, to prove (5.3.11) we will show that

$$\int_M^\infty \int_{t=s}^\infty \frac{\overline{F}_{n''}(u_{n''}t)}{\overline{F}(u_{n''})} \mathrm{d}t\mathrm{d}s$$

can be replaced with

$$\int_M^\infty \int_{t=s}^\infty \frac{\overline{F}(u_{n''}t)}{\overline{F}(u_{n''})} \mathrm{d}t\mathrm{d}s \ .$$

However, the method of Case 1 is not applicable, since the variance of the first double integral is infinite. Instead, we appeal to the law of large numbers of [1].

Fix $M \ge 1$. For $j = 1, ..., n$, let

$$V_{j,n} = \int_M^\infty \int_s^\infty \frac{\mathbb{1}_{\{X_j > u_n t\}}}{\overline{F}(u_n)} \, \mathrm{d}t \, \mathrm{d}s \ ,$$

so that

$$\frac{1}{n} \sum_{j=1}^n V_{j,n} = \int_M^\infty \int_s^\infty \frac{\overline{F}_n(u_n t)}{\overline{F}(u_n)} \, \mathrm{d}t \, \mathrm{d}s \, .$$

For each n, $V_{1,n}, ..., V_{n,n}$ are i.i.d., nonnegative random variables with expectation

$$\mathbb{E}[V_{j,n}] = \int_M^\infty \int_s^\infty \frac{\overline{F}(u_n t)}{\overline{F}(u_n)} \,\mathrm{d}t \,\mathrm{d}s \,.$$

By Potter's bounds, $\mathbb{E}[V_{j,n}]$ is finite if $\alpha > 2$.

We will apply Theorem 5.3.3 below to prove that as $n \to \infty$

$$\int_M^\infty \int_s^\infty \frac{\overline{F}_n(u_n t) - \overline{F}(u_n t)}{\overline{F}(u_n)} \,\mathrm{d}t \,\mathrm{d}s = \frac{1}{n}\sum_{j=1}^n \big(V_{j,n} - \mathbb{E}[V_{j,n}]\big) \xrightarrow{P} 0 \,. \tag{5.3.13}$$

Theorem 5.3.3 (**special version of Theorem 3.1 of [1]**)
Let $\big(V_{j,n}, 1 \le j \le n\big)$, be an array of nonnegative independent and identically distributed random variables. Assume that as $n \longrightarrow \infty$

$$\int_0^1 z\, f_n(z)\, dz \longrightarrow 0 \,, \tag{5.3.14}$$

where $f_n(z) = \sum_{j=1}^n \mathbb{P}[V_{j,n} > nz] = n\,\mathbb{P}[V_{1,n} > nz]$. Then,

$$\frac{1}{n}\sum_{j=1}^n \Big(V_{j,n} - \mathbb{E}\big[V_{j,n}\mathbb{1}_{\{V_{j,n}\le n\}}\big]\Big) \xrightarrow{P} 0 \,. \tag{5.3.15}$$

For (5.3.15) to hold, we verify the condition (5.3.14). In fact, we have

$$\begin{aligned}
V_{j,n} &= \int_M^\infty \int_s^\infty \frac{\mathbb{1}_{\{X_j > u_n t\}}}{\overline{F}(u_n)} \,\mathrm{d}t \,\mathrm{d}s = \frac{1}{\overline{F}(u_n)} \int_M^{X_j/u_n} \int_s^{X_j/u_n} \mathbb{1}_{\{X_j > u_n s\}} \,\mathrm{d}t \,\mathrm{d}s \\
&= \frac{1}{\overline{F}(u_n)} \int_M^{X_j/u_n} \left(\frac{X_j}{u_n} - s\right) \mathbb{1}_{\{X_j > u_n M\}} \,\mathrm{d}s \\
&= \frac{1}{2\overline{F}(u_n)} \left(2\frac{X_j^2}{u_n^2} - \frac{X_j^2}{u_n^2} - 2\frac{X_j}{u_n}M + M^2\right) \\
&= \frac{1}{2\overline{F}(u_n)} \left(\frac{X_j}{u_n} - M\right)^2 .
\end{aligned}$$

Recalling that $k_n = n\overline{F}(u_n)$, we have

$$\begin{aligned} V_{j,n} > nz \quad \iff \quad X_j \; &> \; u_n\left(M + \sqrt{2zn\overline{F}(u_n)}\right) \\ &= \; u_n\left(M + \sqrt{2zk_n}\right) . \end{aligned} \tag{5.3.16}$$

Therefore,

$$f_n(z) = n\overline{F}\left(u_n\left(M + \sqrt{2zk_n}\right)\right) = k_n \frac{\overline{F}\left(u_n\left(M + \sqrt{2zk_n}\right)\right)}{\overline{F}(u_n)} . \tag{5.3.17}$$

Apply Potter's bounds to (5.3.17). For any δ such that $(\alpha-4)\vee 0 < \delta < \alpha - 2$, there exists $C < \infty$ such that for all $M \geq 1$

$$f_n(z) \leq C\, k_n \left(M + \sqrt{2zk_n}\right)^{-(\alpha-\delta)} .$$

Let $\epsilon = \frac{\alpha-\delta-2}{2}$. Note that $\epsilon \in (0,1)$. Therefore,

$$\begin{aligned} \int_0^1 z f_n(z)\mathrm{d}z \; &\leq \; Ck_n \int_0^1 z\left(M + (2zk_n)^{1/2}\right)^{-(\alpha-\delta)} \mathrm{d}z \\ &\leq \; 2^{-(\alpha-\delta)/2} C k_n^{-\epsilon} \int_0^1 z^{-\epsilon}\mathrm{d}z \\ &\leq \; \frac{2^{-(\alpha-\delta)/2} C k_n^{-\epsilon}}{1-\epsilon} \longrightarrow 0 \qquad \text{as } n \to \infty , \end{aligned}$$

and (5.3.15) holds.

To see that (5.3.15) implies (5.3.13), we must show that

$$\mathbb{E}\left[V_{j,n}\mathbb{1}_{\{V_{j,n}>n\}}\right] \longrightarrow 0 .$$

Indeed, by (5.3.16) and integration by parts we have

$$
\begin{aligned}
E[V_{j,n}\mathbf{1}_{\{V_{j,n}>n\}}] &= \frac{1}{2\overline{F}(u_n)} E\left[\left(\frac{X_j}{u_n}-M\right)^2 \mathbf{1}_{\{X_j>u_n(M+\sqrt{2k_n})\}}\right] \\
&\leq \frac{1}{2\overline{F}(u_n)} E\left[\left(\frac{X_j}{u_n}\right)^2 \mathbf{1}_{\{X_j>u_n(M+\sqrt{2k_n})\}}\right] \\
&= -\frac{1}{2\overline{F}(u_n)u_n^2}\int_{u_n(M+\sqrt{2k_n})}^{\infty} x^2 d\overline{F}(x) \\
&= \frac{1}{2\overline{F}(u_n)u_n^2}\left[-x^2\overline{F}(x)|_{u_n(M+\sqrt{2k_n})}^{\infty} + 2\int_{u_n(M+\sqrt{2k_n})}^{\infty} x\overline{F}(x)\mathrm{d}x\right] \\
&= \frac{(M+\sqrt{2k_n})^2}{2}\,\frac{\overline{F}\left(u_n(M+\sqrt{2k_n})\right)}{\overline{F}(u_n)} + \frac{1}{u_n^2}\int_{u_n(M+\sqrt{2k_n})}^{\infty} x\frac{\overline{F}(x)}{\overline{F}(u_n)}\mathrm{d}x\ .
\end{aligned}
$$

Apply Potter's bounds, that is for any chosen $\delta>0$ such that $\alpha>2+\delta$, $M\geq 1$ and n large enough we get

$$
\begin{aligned}
&E[V_{j,n}\mathbf{1}_{\{V_{j,n}>n\}}] \\
&\leq \frac{1+\delta}{2}\left(M+\sqrt{2k_n}\right)^{\delta-\alpha+2} + \frac{1+\delta}{u_n^2}\int_{u_n(M+\sqrt{2k_n})}^{\infty} x\left(\frac{x}{u_n}\right)^{\delta-\alpha}\mathrm{d}x \\
&= \frac{1+\delta}{2}\left(M+\sqrt{2k_n}\right)^{\delta-\alpha+2} + (1+\delta)\,u_n^{-\delta+\alpha-2}\int_{u_n(M+\sqrt{2k_n})}^{\infty} x^{\delta-\alpha+1}\mathrm{d}x \\
&= \frac{1+\delta}{2}\left(M+\sqrt{2k_n}\right)^{\delta-\alpha+2} - \frac{1+\delta}{\delta-\alpha+2}\,u_n^{-\delta+\alpha-2}\left(u_n\left(M+\sqrt{2k_n}\right)\right)^{\delta-\alpha+2} \\
&= C\left(M+\sqrt{2k_n}\right)^{\delta-\alpha+2}\ ,
\end{aligned}
$$

where C is a constant that does not depend on n or M.

Since $k_n \longrightarrow \infty$ as $n\longrightarrow\infty$, for any M, $E[V_{j,n}\mathbf{1}_{\{V_{j,n}>n\}}]\longrightarrow 0$ as $n\longrightarrow\infty$.

Therefore, as $n \longrightarrow \infty$

$$\int_M^\infty \int_s^\infty \frac{\overline{F}_n(u_n t) - \overline{F}(u_n t)}{\overline{F}(u_n)} \, \mathrm{d}t \, \mathrm{d}s \xrightarrow{P} 0 \, .$$

In this case, the subsequence (n'') can be chosen so that with probability 1, for $\alpha > 2$ and by Potter's bounds and dominated convergence

$$\begin{aligned} \lim_M \limsup_{n''} \int_M^\infty \int_s^\infty \frac{\overline{F}_{n''}(u_{n''} t)}{\overline{F}(u_{n''})} \, \mathrm{d}t \, \mathrm{d}s &= \lim_M \limsup_{n''} \int_M^\infty \int_s^\infty \frac{\overline{F}(u_{n''} t)}{\overline{F}(u_{n''})} \, \mathrm{d}t \, \mathrm{d}s \\ &= \lim_M \frac{M^{-(\alpha-2)}}{(\alpha-1)(\alpha-2)} = 0 \, . \end{aligned}$$

This completes the proof of Case 2 and (5.3.11) follows. ∎

Lemma 5.3.4 *Under the assumptions of Theorem 3.1.9 and if $\alpha > 2(1-m)$, as $n \longrightarrow \infty$, we have that on $\mathscr{D}(0,\infty)$*

$$\int_1^\infty \widehat{G}_n^{(b)}(1) \, \widehat{T}_n(t) \, \frac{\mathrm{d}t}{t^m} \xrightarrow{\mathscr{D}} \int_1^\infty G(1) \, t^{-\alpha-m} \, \mathrm{d}t = G(1) \, \frac{1}{\alpha+m-1} \, , \tag{5.3.18}$$

in probability.

Proof: By Theorem 5.2.4, we know that $\widehat{G}_n^{(b)}(1) \xrightarrow{\mathscr{D}} G(1)$ in probability. So, it remains to show that

$$\int_1^\infty \widehat{T}_n(t) \frac{\mathrm{d}t}{t^m} \xrightarrow{\mathscr{D}} \int_1^\infty t^{-\alpha-m} \, \mathrm{d}t = \frac{1}{\alpha+m-1} \, , \tag{5.3.19}$$

in probability.

This is an immediate consequence of Corollaries 3.3.1 and 3.3.2. ∎

Proof of Theorem 5.3.1 : From Corollaries 3.3.1 and 3.3.2, we know that

$$\begin{aligned} \mathbb{E}\left[\int_1^\infty \frac{G(t)-G(1)\,t^{-\alpha}}{t^m}\,\mathrm{d}t\right] &= 0\ , \\ \mathbb{V}ar\left[\int_1^\infty \frac{G(t)-G(1)\,t^{-\alpha}}{t^m}\,\mathrm{d}t\right] &= \frac{\alpha}{(\alpha+m-1)^2(\alpha+2m-2)}\ . \end{aligned}$$

It remains to show that

$$\int_1^\infty \frac{\widehat{G}_n^{(b)}(t)-\widehat{G}_n^{(b)}(1)\,\widehat{T}_n(t)}{t^m}\,\mathrm{d}t \xrightarrow{\mathscr{D}} \int_1^\infty \frac{G(t)-G(1)\ t^{-\alpha}}{t^m}\,\mathrm{d}t\ ,$$

in probability. In fact, Lemmas 5.3.2 and 5.3.4 finish the proof. ∎

5.3.2 VaR and CTE using the bootstrap technique

The goal is to get the bootstrap versions of Corollary 4.1.5 and Theorem 4.2.4. From the point of view of applications, Proposition 5.3.5 may not be relevant, since the limiting variance in Corollary 4.1.5 can be easily estimated. However, as pointed out previously, Proposition 5.3.6 is of particular importance whenever α is close to 2. In this case, due to instability of estimates of α, the resulting estimate of the first limiting variance in Lemma 4.2.5 can be negative and hence of no use and the second limiting variance can get large. In these cases, the bootstrap comes to the rescue.

Proposition 5.3.5 *Suppose the conditions (3.1.1), (3.1.2), (3.1.23) and (3.1.25) hold. Then as $n \longrightarrow \infty$*

$$\widehat{\alpha}_1^{-1}\widehat{G}_n^{(b)}(1) + r\int_1^\infty \frac{\widehat{G}_n^{(b)}(t)-\widehat{G}_n^{(b)}(1)\,\widehat{T}_n(t)}{t}\,\mathrm{d}t \tag{5.3.20}$$

$$\xrightarrow{\mathscr{D}} \alpha^{-1}G(1) + r\int_1^\infty \frac{G(t)-G(1)T(t)}{t}\,\mathrm{d}t \sim \mathscr{N}\left(0, \frac{1+r^2}{\alpha^2}\right) \tag{5.3.21}$$

in probability, where $r = \lim_{n\longrightarrow\infty}\log\big(k/(np)\big)$ and $\widehat{\alpha}_1^{-1}$ is the Hill estimator.

Proof: Recall that $\widehat{\alpha}_1^{-1} = \int_1^\infty \widehat{T}(t)\mathrm{d}t/t$ and $\alpha^{-1} = \int_1^\infty T(t)\mathrm{d}t/t$. Hence, (5.3.21) can be written as

$$r\int_1^\infty \frac{G(t)}{t}\mathrm{d}t + (1-r)G(1)\int_1^\infty \frac{T(t)}{t}\mathrm{d}t\,.$$

Furthermore, (5.3.20) can be written as

$$r\int_1^\infty \frac{\widehat{G}_n^{(b)}(t)}{t}\,\mathrm{d}t + (1-r)\,\widehat{G}_n^{(b)}(1)\int_1^\infty \frac{\widehat{T}_n(t)}{t}\,\mathrm{d}t\,.$$

Then, we let $m = 1$ and we apply Lemmas 5.3.2 and 5.3.4 to get the desired result. ∎

Proposition 5.3.6 *Suppose the conditions (3.1.1), (3.1.2), (3.1.23) and (3.1.25) hold. As $n \longrightarrow \infty$ we have*

1. *If $\alpha > 2$, then*

$$\begin{aligned}
&\widehat{\alpha}_1^{-1}\widehat{G}_n^{(b)}(1) + r\int_1^\infty \frac{\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\,\widehat{T}_n(t)}{t}\,\mathrm{d}t + \left(1-\widehat{\alpha}_1^{-1}\right)\int_1^\infty \left(\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\widehat{T}_n(t)\right)\mathrm{d}t\\
&\xrightarrow{\mathscr{D}} \alpha^{-1}G(1) + r\int_1^\infty \frac{G(t)-G(1)T(t)}{t}\,\mathrm{d}t + \left(1-\alpha^{-1}\right)\int_1^\infty \left(G(t)-G(1)T(t)\right)\mathrm{d}t\\
&\sim \mathscr{N}\left(0, \frac{(r^2+2r+2)\alpha^2 - (3r^2+4r+4)\alpha + 2(1+r^2)}{\alpha^2(\alpha-1)(\alpha-2)}\right).
\end{aligned} \tag{5.3.22}$$

2. *If $\alpha > 1$, then*

$$\begin{aligned}
&\widehat{\alpha}_1^{-1}\widehat{G}_n^{(b)}(1) + \left(r + \frac{1}{1-\widehat{\alpha}_1^{-1}}\right)\int_1^\infty \frac{\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\,\widehat{T}_n(t)}{t}\,\mathrm{d}t\\
&\xrightarrow{\mathscr{D}} \quad \alpha^{-1}G(1) + \left(r + \frac{1}{1-\alpha^{-1}}\right)\int_1^\infty \frac{G(t)-G(1)T(t)}{t}\,\mathrm{d}t\\
&\sim \quad \mathscr{N}\left(0, \frac{(r^2+2r+2)\alpha^2 - 2(r^2+r+4)\alpha + (1+r^2)}{\alpha^2(\alpha-1)^2}\right),
\end{aligned} \tag{5.3.23}$$

in probability, where $r = \lim_{n\longrightarrow\infty} \log\left(k/(np)\right)$ and $\widehat{\alpha}_1^{-1}$ is the Hill estimator.

Proof: First, keep in mind that by Lemma 5.3.4,

$$\widehat{\alpha}_1^{-1} = \int_1^{\infty} \widehat{T}(t) \frac{\mathrm{d}t}{t} \longrightarrow \alpha^{-1}$$

in probability. Proposition 5.3.5 deals with the first and second terms of the left hand sides of (5.3.22) and (5.3.23). For the last term of (5.3.22), we apply Theorem 5.3.1 with $m = 0$. ∎

■

Chapter 6

Simulations

We provide simulation studies to illustrate the theoretical results derived in the previous chapters.

6.1 Comparison of the Hill and t-Hill estimators

This section presents some simulations of the Hill and t-Hill estimators for the Pareto distribution with different choices of the tail index $\alpha > 0$ or equivalently $\gamma = 1/\alpha$, which is the notation that we will follow.

Let $X_{n:n} \geq X_{n:n-1} \geq \cdots \geq X_{n:1}$ be the order statistics from Pareto distribution $P(X > x) = x^{-\alpha}$. We pick $k < n$ (the number of upper order statistics used in the estimation) and consider the Hill estimator

$$\widehat{\gamma}_1 = \frac{1}{k} \sum_{j=1}^{k} \log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right) \ . \tag{6.1.1}$$

Also, we consider the harmonic moment estimator

$$\widehat{\gamma}_m = \frac{1}{m-1}\left[\frac{1}{\frac{1}{k}\sum_{j=1}^{k}\left(\frac{X_{n:n-k}}{X_{n:n-j+1}}\right)^{m-1}} - 1\right] , \tag{6.1.2}$$

where $m \geq 0$. We note that for $m \to 1$, the HME coincides with the Hill estimator and when $m = 2$, the HME coincides with t-Hill estimator.

We recall that among all HME estimators, the Hill estimator has the lowest variance: $\sigma_1^2 = \gamma^2$.

6.1.1 Histogram from Pareto(4)

In order to evaluate the quality of the Gaussian approximation of the asymptotic distributions of the standardized Hill and t-Hill estimators of the tail index $\alpha = 4$ ($\gamma = 0.25$), we plot the histograms based on $N = 100$ samples of size $n = 1000$ with $k = 0.1n$ (on the left) and $k = 0.05n$ (on the right). That is, the histograms display the empirical distribution of

$$\sqrt{k}\,\{\hat{\gamma} - \gamma\} \ ,$$

where $\hat{\gamma}$ is either $\hat{\gamma}_1$ (Hill estimator) or $\hat{\gamma}_2$ (t-Hill estimator). Figure 6.1 shows the empirical distribution of the Hill estimator versus its theoretical asymptotic distribution, which is given by the centered normal distribution with variance $\sigma_1^2 = 1/\alpha^2 = 0.0625$, and Figure 6.2 shows the empirical distribution of the t-Hill estimator versus its theoretical asymptotic distribution, given by a centered normal distribution with variance $\sigma_2^2 = \frac{(\alpha+1)^2}{\alpha^3(\alpha+2)} = 0.06510417$.

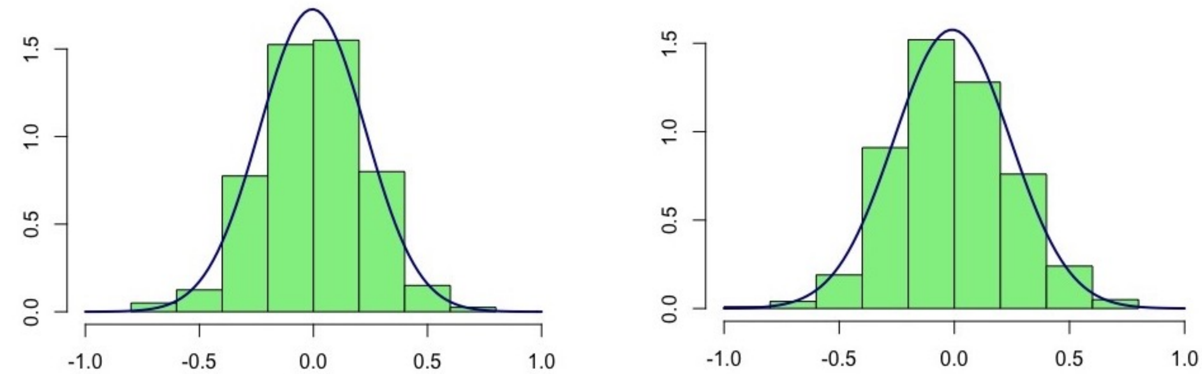

Figure 6.1: Histogram of Hill estimator ($\alpha = 4$)

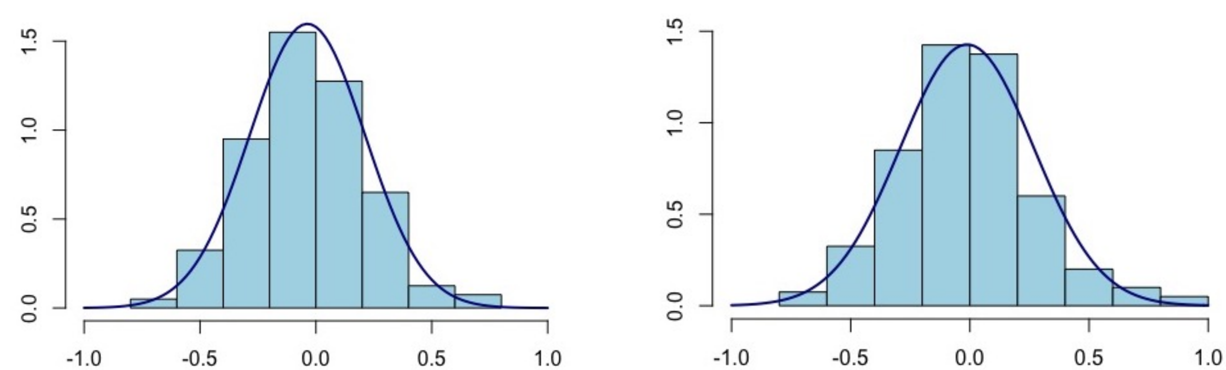

Figure 6.2: Histogram of t-Hill estimator ($\alpha = 4$)

6.1.2 The Hill and t-Hill plots from Pareto(4)

In choosing an appropriate value of k, the basic instrument is the so-called Hill plot, graphing

$$\left\{(k, \widehat{\gamma}),\ 5 \leq k < n\right\},$$

where $n = 2000$ is the sample size and $\widehat{\gamma}$ is an estimator of γ. The main idea is to choose k from a region, where the Hill plot appears to be stable.

Figures 6.3 and 6.4 show that the Hill plots of both estimators ($\widehat{\gamma}_1$, $\widehat{\gamma}_2$) stabilize at height roughly γ. For the Hill estimator the stable region is about $k \in [30; 100]$, whilst for t-Hill, it is roughly $k \in [40; 100]$. Also, one can see that the graphs are volatile when very few order statistics are used.

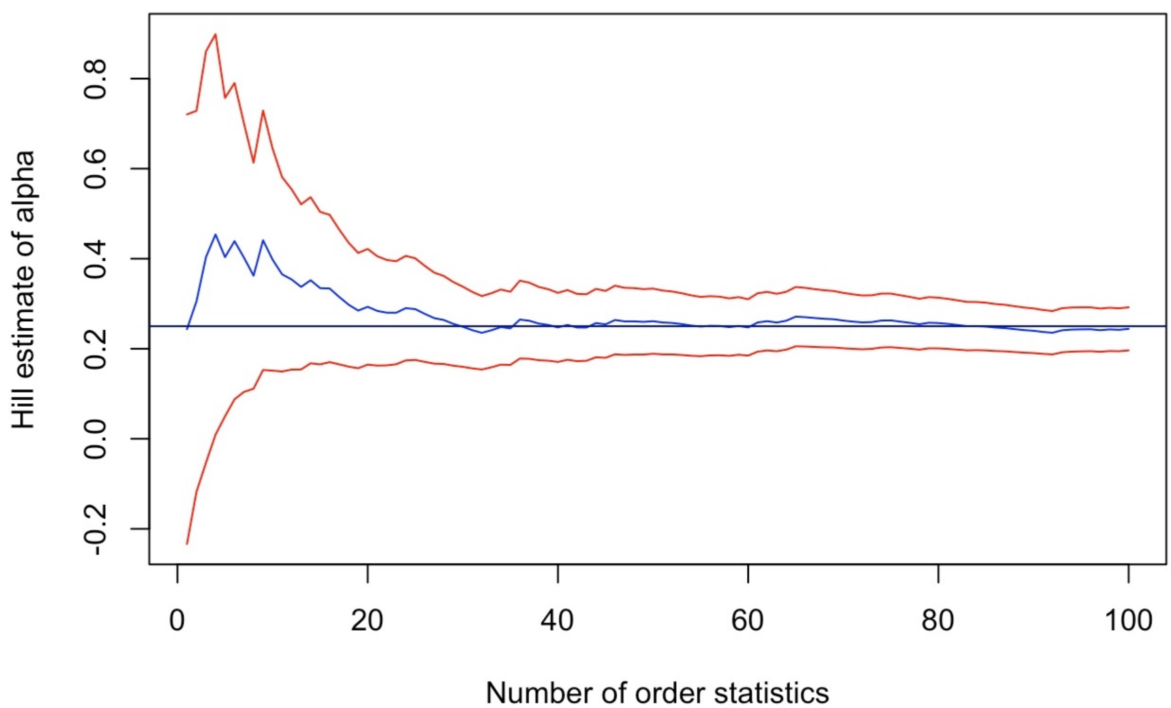

Figure 6.3: Hill plot ($\alpha = 4$). Blue lines - estimators, Red lines - confidence intervals.

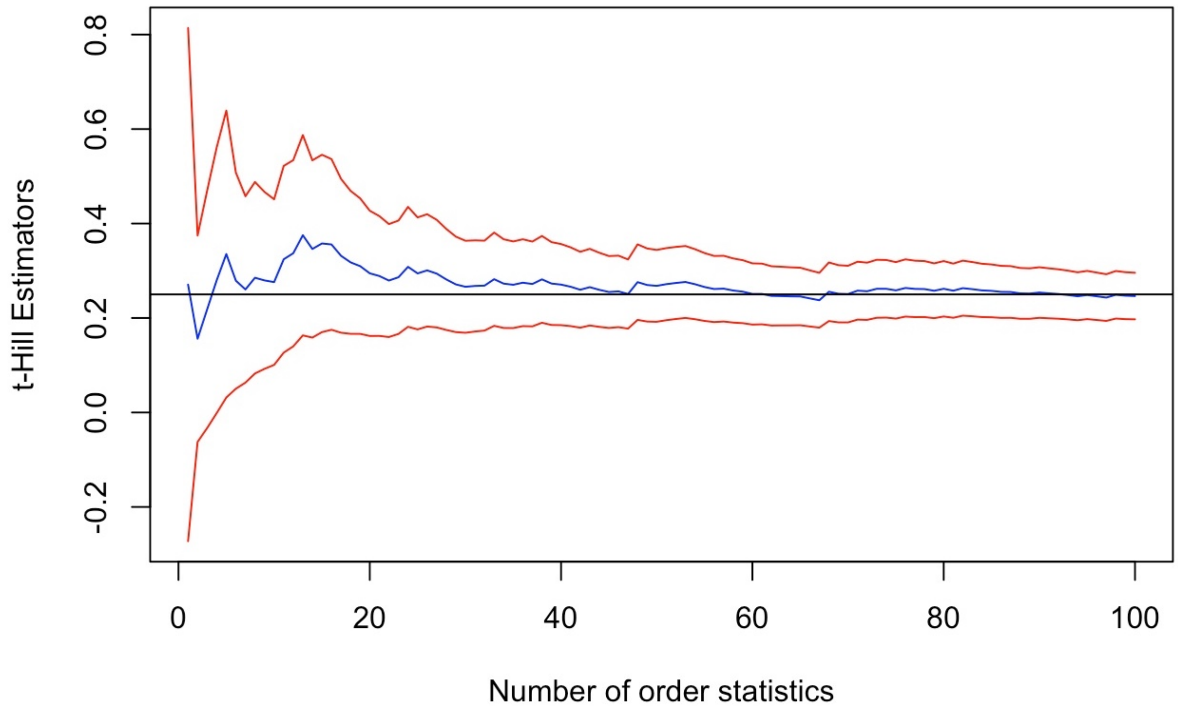

Figure 6.4: t-Hill plot ($\alpha = 4$). Blue lines - estimators, Red lines - confidence intervals.

6.1.3 Plot of variances

Figure 6.5 shows the estimated variances of the Hill and t-Hill estimators for 100 Pareto experiments with $\alpha = 4$, $n = 2000$ and $k = 100$. The averages of the sample variances are shown by blue lines ($\widehat{\sigma}^2_{\text{Hill}} \approx 0.06265415$ and $\widehat{\sigma}^2_{\text{t-Hill}} \approx 0.06417094$). The theoretical variances are shown by red lines ($\sigma^2_{\text{Hill}} = 0.0625$ and $\sigma^2_{\text{t-Hill}} \approx 0.06510417$).

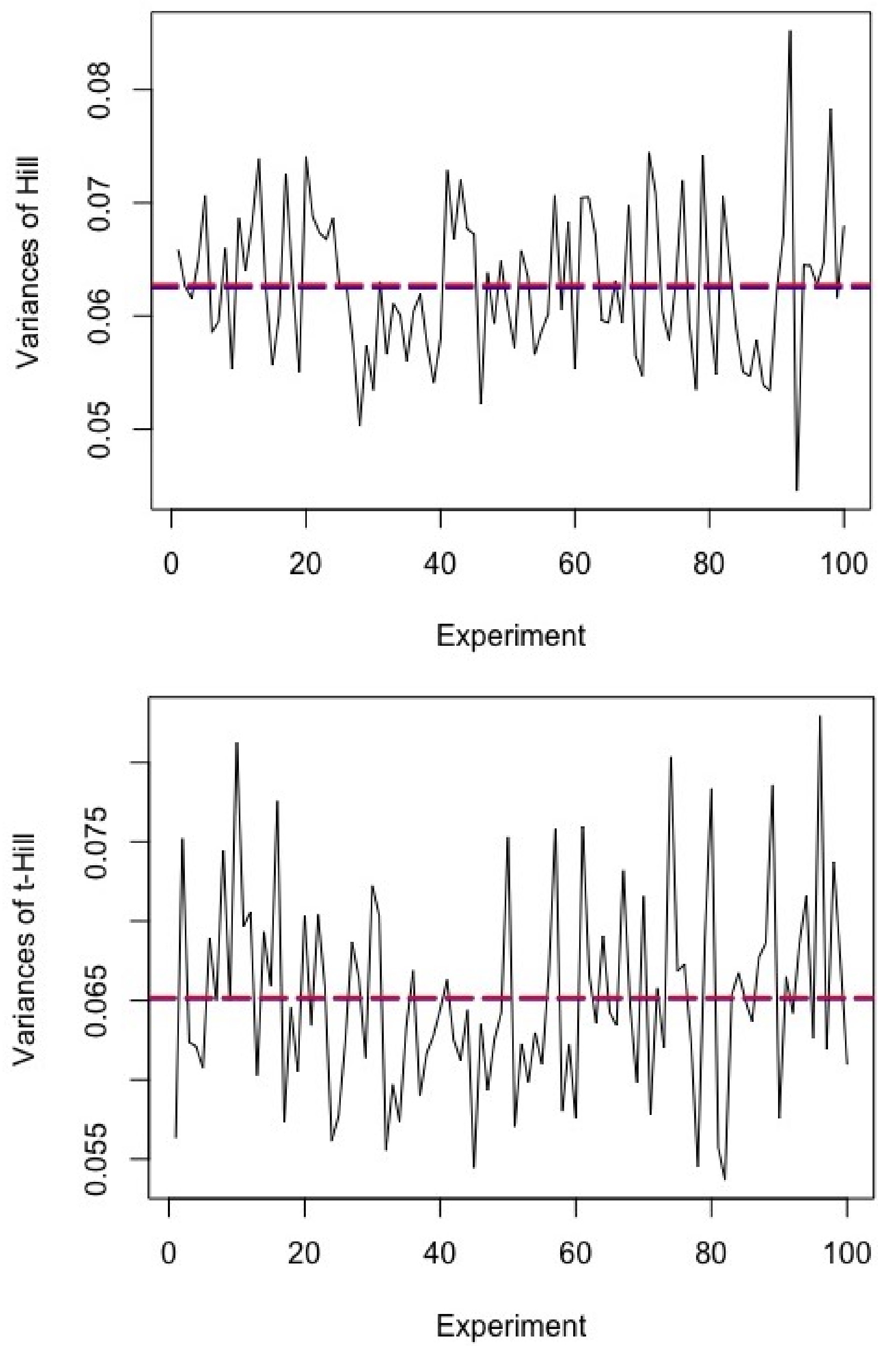

Figure 6.5: Plots of Hill and t-Hill variances ($\alpha = 4$)

6.1.4 Boxplots from Pareto(4)

Let $\alpha = 4$ and $n = 2000$. We produce three boxplots for three different choices of k; $k_1 = 0.05n$, $k_2 = 0.1n$ and $k_3 = 0.2n$.

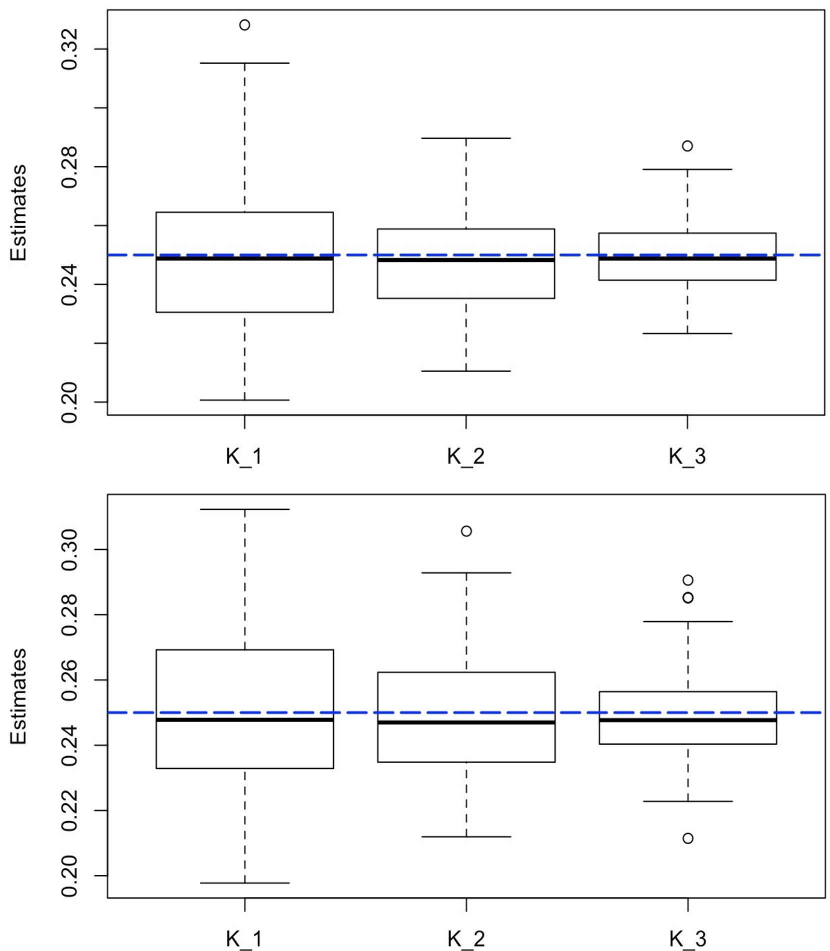

Figure 6.6: Boxplot of Hill and t-Hill estimators with different thresholds

For both estimators, the median appears near the center of each of the three boxplots, indicating a symmetric distribution. The interquartile range and variability of the estimators decrease as k increases, although we also observe outliers with larger values of k.

6.1.5 The Hill and the t-Hill plots from Student-t , $\alpha = 4$

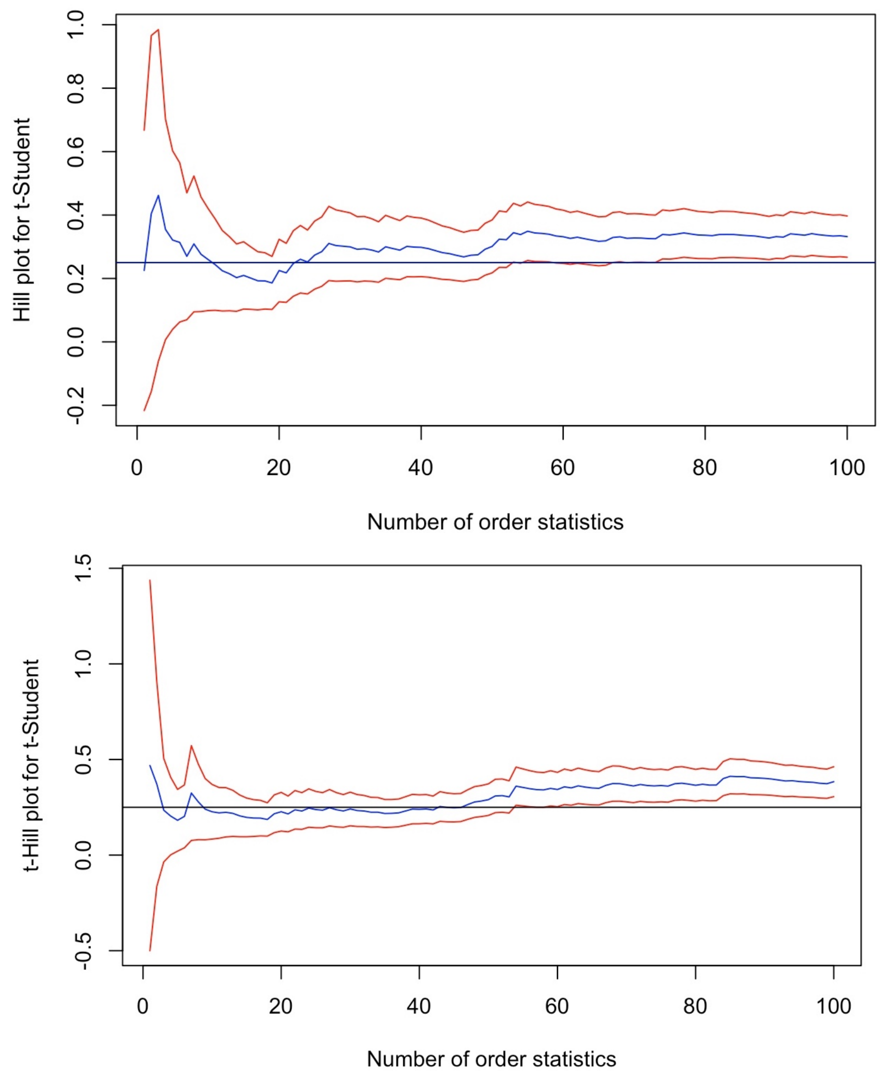

Figure 6.7: Hill and t-Hill plots for Student ($\alpha = 4$)

It is well-known that the Hill estimator may not work well for non-Pareto distributions. We can see that the stable region for the Hill estimator ($k \in [50, 100]$) would overestimate γ, while the stable region for the t-Hill ($k \in [20, 50]$) would return a less biased estimate.

6.1.6 Boxplots from Student-t distribution

We consider $\alpha = 4$, $n = 2000$ and boxplots for different choices of k : $k_1 = 0.025n$, $k_2 = 0.05n$ and $k_3 = 0.1n$.

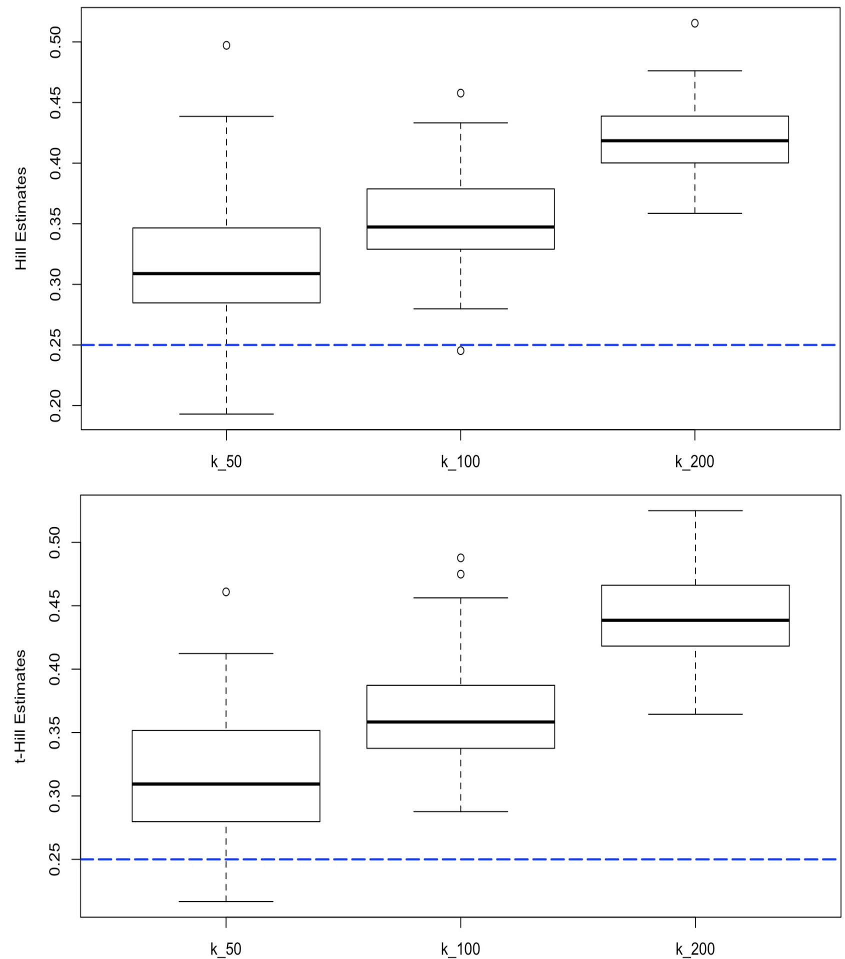

Figure 6.8: Boxplot of Hill and t-Hill estimators with different thresholds

Both estimators exhibit an upward bias that increases with the number of order statistics used in the estimator, while the variability decreases.

6.2 Performance of estimators of CTE

We simulate $n = 2000$ observations from three models: Pareto, Student-t and mixture distribution defined by $\overline{F}_m(x) = {}^1\!/_2(x^{-\alpha} + x^{-\alpha\beta}),\ \beta > 1$.

Recall that from Theorem 4.2.4, the CTE, $\theta(p)$, can be estimated by

1. $\widehat{\theta}_0(p) = \left(\frac{k}{np}\right)^{\widehat{\alpha}^{-1}} \frac{1}{k} \sum_{j=1}^{k} X_{n:n-j+1}\,, \qquad \text{if} \quad \alpha > 2\,;$

2. $\widehat{\theta}_1(p) = \frac{1}{1-\widehat{\alpha}^{-1}} X_{n:n-k} \left(\frac{k}{np}\right)^{\widehat{\alpha}^{-1}}, \qquad \text{if} \quad \alpha > 1\,.$

We note that for $\alpha > 2$ either estimator can be used, but the limiting variance of $\widehat{\theta}_1(p)$ is smaller than that of $\widehat{\theta}_0(p)$. Although our estimators are asymptotically unbiased, for finite samples the effect of bias may have a significant effect on the mean squared error (MSE). In practice, the estimator with the smallest MSE is the one that is most efficient. The tables below show us the estimated MSE for various values of α and k (upper order statistics) for the three models mentioned above with $p = 0.05$.

Upper order statistics	**Parameter α**	**Mean squared Error**	
		$MSE(\widehat{\theta}_0)$	$MSE(\widehat{\theta}_1)$
$k = 0.05n$	$\alpha = 3$	0.07547705	0.0716851
	$\alpha = 4$	0.01342647	0.0120190
$k = 0.1n$	$\alpha = 3$	0.06169946	0.06512302
	$\alpha = 4$	0.01018146	0.01006419

Table 6.1: Estimated mean squared error for Pareto distribution

Since we observe $MSE(\widehat{\theta}_0) \approx MSE(\widehat{\theta}_1)$, either estimator is appropriate when the underlying distribution is Pareto and $\alpha \gg 2$.

Upper order statistics	**Parameter α**	**Mean squared Error**	
		$MSE(\widehat{\theta}_0)$	$MSE(\widehat{\theta}_1)$
$k = 0.025n$	$\alpha = 3$	0.1255236	0.09653583
	$\alpha = 4$	0.1884831	0.2317549
$k = 0.05n$	$\alpha = 3$	0.09830069	0.07814961
	$\alpha = 4$	0.1757575	0.2421394
$k = 0.1n$	$\alpha = 3$	0.08800889	0.1739078
	$\alpha = 4$	0.2726713	0.4907442

Table 6.2: Estimated Mean squared error for Student-t distribution

For Student-t with $\alpha \gg 2$, one would choose $\widehat{\theta}_0$, since its MSE is either close to or much smaller than that of $\widehat{\theta}_1$. One possible explanation for this behaviour is that the Hill estimator $\widehat{\alpha}^{-1}$ is unstable. Since the estimator $\widehat{\alpha}^{-1}$ appears only once in the formula for $\widehat{\theta}_0$ and twice in that for $\widehat{\theta}_1$, this instability seems to have a greater effect on the MSE of $\widehat{\theta}_1$.

For the mixture distribution $\overline{F}_m(x) = (1/2)(x^{-\alpha} + x^{-\alpha\beta})$, $\beta > 1$, simulations (not displayed here) showed that a stable value of k for the Hill and t-Hill estimators was approximately half that of the Pareto distribution. Therefore, we estimate the MSE of $\widehat{\theta}_0$ and $\widehat{\theta}_1$ for $k = .025n$ and $k = .05n$. Consider $n = 2000$, $\alpha = 3$ or 4, $\beta = 10$ and $p = 0.05$.

Upper order statistics	Parameter α	Mean squared Error	
		$MSE(\widehat{\theta}_0)$	$MSE(\widehat{\theta}_1)$
$k = 50$	$\alpha = 3$	0.1447182	0.1441957
	$\alpha = 4$	0.03622246	0.03692423
$k = 100$	$\alpha = 3$	0.1365983	0.137092
	$\alpha = 4$	0.02895653	0.02811501

Table 6.3: Mean squared error for Mixture distribution $\overline{F}_m$

We remark that as in Pareto case, one can chose either $\widehat{\theta}_0$ or $\widehat{\theta}_1$ to estimate $\theta(.05)$, since their mean squared errors are very similar.

Summary: These simulations suggest the following practical steps for estimating $\widehat{\theta}(p)$:

- First, $\gamma = 1/\alpha$ should be estimated by $\widehat{\alpha}^{-1}$, calculated using a value of k chosen from a stable region of the Hill plot.

- If $1/\widehat{\alpha}^{-1} \gg 2$, then either estimator can be used if it is reasonable to assume that $\overline{F}(t) \approx t^{-\alpha}$ for large values of t. If this assumption cannot be made, then $\widehat{\theta}_0(p)$ might produce a less biased estimate. We have observed that this is the case for the t-distribution with a judicious choice of k.

- Since the variance of $\widehat{\theta}_0(p)$ diverges as $\alpha \to 2^+$, then if $1/\widehat{\alpha}^{-1}$ is close to or less than 2, the CTE $\theta(p)$ should be estimated using $\widehat{\theta}_1(p)$. Neither estimator should be used for values of α near 1. These points will be illustrated in the next section.

6.2.1 Comparison of confidence intervals for $\theta(p)$ using estimators $\widehat{\theta}_0(p)$ and $\widehat{\theta}_1(p)$

In this section, we use Theorem 4.2.4 to construct confidence intervals for $\theta(p)$ and then illustrate the effects of decreasing values of α on the confidence bounds when the underlying distribution is Pareto.

Recall that from Theorem 4.2.4, as $n \to \infty$,

$$\text{If } \alpha > 2, \qquad \sqrt{k}\left\{\frac{\widehat{\theta}_0(p)}{\theta(p)} - 1\right\} \sim \mathcal{N}\left(0, var_0\right),$$

$$\text{If } \alpha > 1, \qquad \sqrt{k}\left\{\frac{\widehat{\theta}_1(p)}{\theta(p)} - 1\right\} \sim \mathcal{N}\left(0, var_1\right),$$

where

$$var_0 = \frac{(r^2+2r+2)\alpha^2 - (3r^2+4r+4)\alpha + 2(1+r^2)}{\alpha^2(\alpha-1)(\alpha-2)},$$

$$var_1 = \frac{(r^2+2r+2)\alpha^2 - 2(r^2+r+1)\alpha + (1+r^2)}{\alpha^2(\alpha-1)^2},$$

and $r = \lim_{n\to\infty} \log\left(\frac{k}{np}\right)$.

We recall that $var_0 > var_1$ for all α, r. We estimate var_i by replacing α with $\widehat{\alpha}$ via Hill estimator in the formulas above. Denote the estimates by $\widehat{var}_i,\ i = 0, 1$. Since $\theta(p)/\widehat{\theta}_i(p) > 0$, we have

$$\mathbb{P}\left(\left(0 \vee \left[1 - 1.96\sqrt{\frac{\widehat{var}_i}{k}}\right]\right) < \widehat{\theta}_i(p)/\theta(p) < 1 + 1.96\sqrt{\frac{\widehat{var}_i}{k}}\right) \approx 0.95,$$

so a 95% confidence interval for $\theta(p)$ is given by:

$$\theta(p) \in \left[\widehat{\theta}_i(p)\left(1 + 1.96\sqrt{\frac{\widehat{var}_i}{k}}\right)^{-1} ;\ \widehat{\theta}_i(p)\left(0 \vee \left[1 - 1.96\sqrt{\frac{\widehat{var}_i}{k}}\right]\right)^{-1}\right]. \quad (6.2.1)$$

Comment 6.2.1 Before continuing, we observe two potential problems with the confidence interval above. First, we see that the upper confidence bound is infinite if $1.96\sqrt{\widehat{var}_i/k} > 1$. We will see in our simulations that this can occur with smaller values of α. A more serious problem occurs when $\widehat{var}_i$ becomes negative: this can occur when α is close to 2 in the case of $\widehat{var}_0$ and when α is close to 1 in the case of $\widehat{var}_1$. In these cases, confidence bounds must be constructed using an alternative approach.

Note: In all of our graphs in this and subsequent sections, infinite upper confidence bounds will be illustrated by black dots, and cases of negative variances will be shown by discontinuities since no bounds can be given.

We now restrict our attention to the Pareto distribution. Continuing with our simulations, consider $n = 2000$, $p = 0.05$ and $k = 0.05n$. The figures below compare the two confidence intervals for $\theta(.05)$ for different values of α.

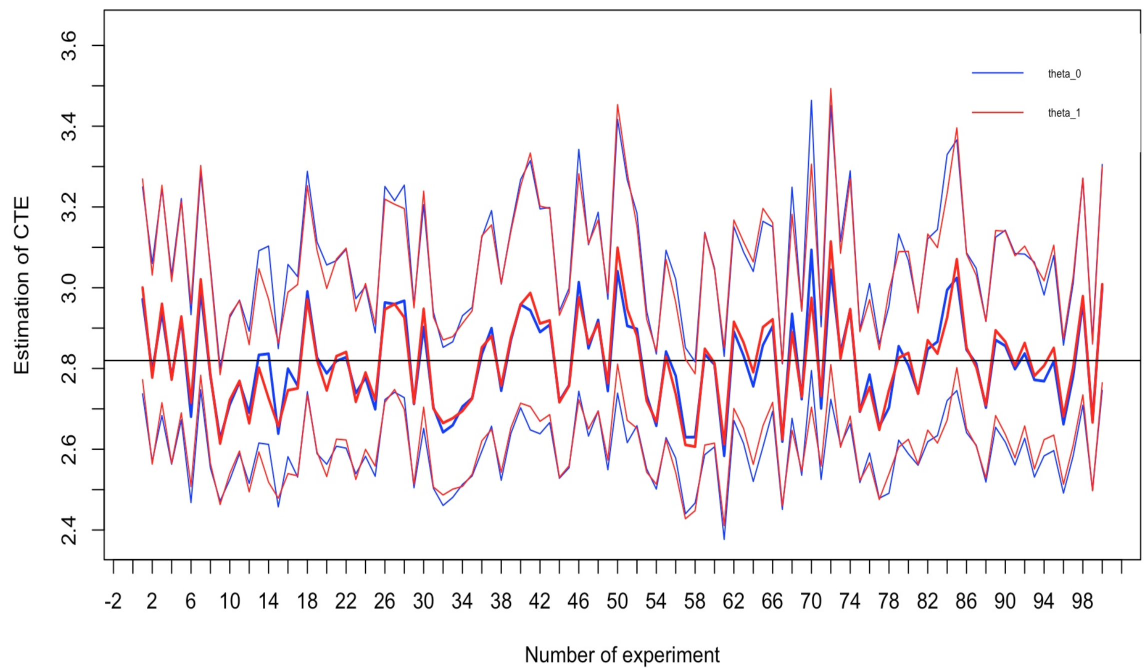

Figure 6.9: Estimates and 95% confidence intervals for $\theta(.05)$ using $\widehat{\theta}_0$ and $\widehat{\theta}_1$ with $\alpha = 4$

Here, the blue lines indicate the estimate and confidence bounds for $\widehat{\theta}_0$ and the red lines indicate the estimate and confidence bounds for $\widehat{\theta}_1$.

In Figure 6.9 we observe that when $\alpha = 4$, both confidence intervals are almost superimposed and they cover the true value of the CTE 97% of the time.

Fgure 6.10 clearly illustrates the effect of a smaller value of α. When $\alpha = 3$, the confidence intervals for $\theta(.05)$ using $\widehat{\theta}_1$ are narrower. This effect is even more evident in Figure 6.11 when $\alpha = 2.7$.

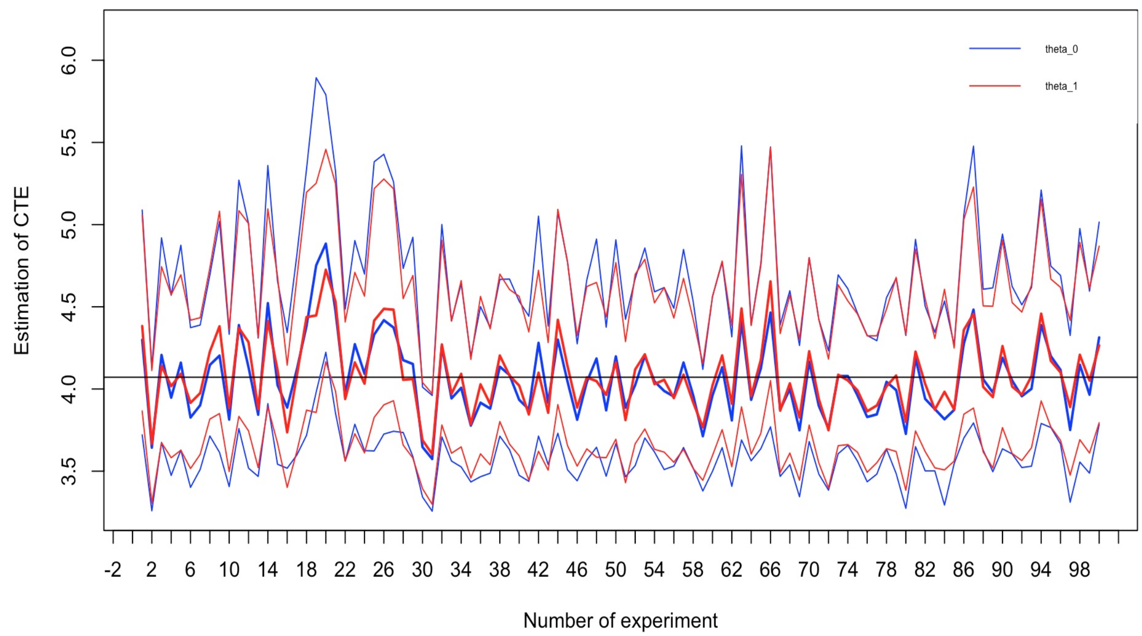

Figure 6.10: Estimates and 95% confidence intervals for $\theta(.05)$ using $\widehat{\theta}_0$ and $\widehat{\theta}_1$ with $\alpha = 3$

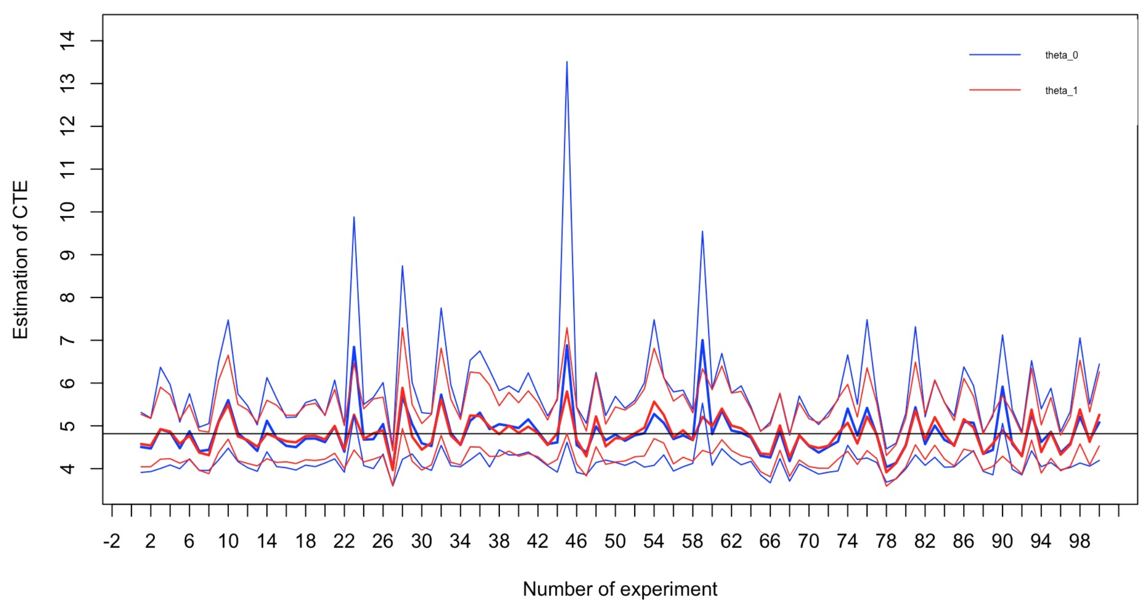

Figure 6.11: Estimates and 95% confidence intervals for $\theta(.05)$ using $\widehat{\theta}_0$ and $\widehat{\theta}_1$ with $\alpha = 2.7$

If α is close to 2, $\widehat{\theta}_0(p)$ becomes highly unstable as $\widehat{var}_0$ tends to ∞. Indeed, as noted before the Hill estimator might yield values less than 2, in which case $\widehat{var}_0$ becomes negative. In this case we estimate the CTE $\theta(.05)$ using $\widehat{\theta}_1(.05)$. However, a similar problem arises with $\widehat{\theta}_1(p)$ when $\alpha \to 1^+$, as shown in Figure 6.12. When $1/\widehat{\alpha}^{-1} < 1$, confidence bounds cannot be constructed and we see a discontinuity (the 9th experiment). When $1/\widehat{\alpha}^{-1}$ is greater than but close to 1 and $1.96\sqrt{\widehat{var}_1/k} > 1$, the upper confidence bound is infinite, illustrated by black dots.

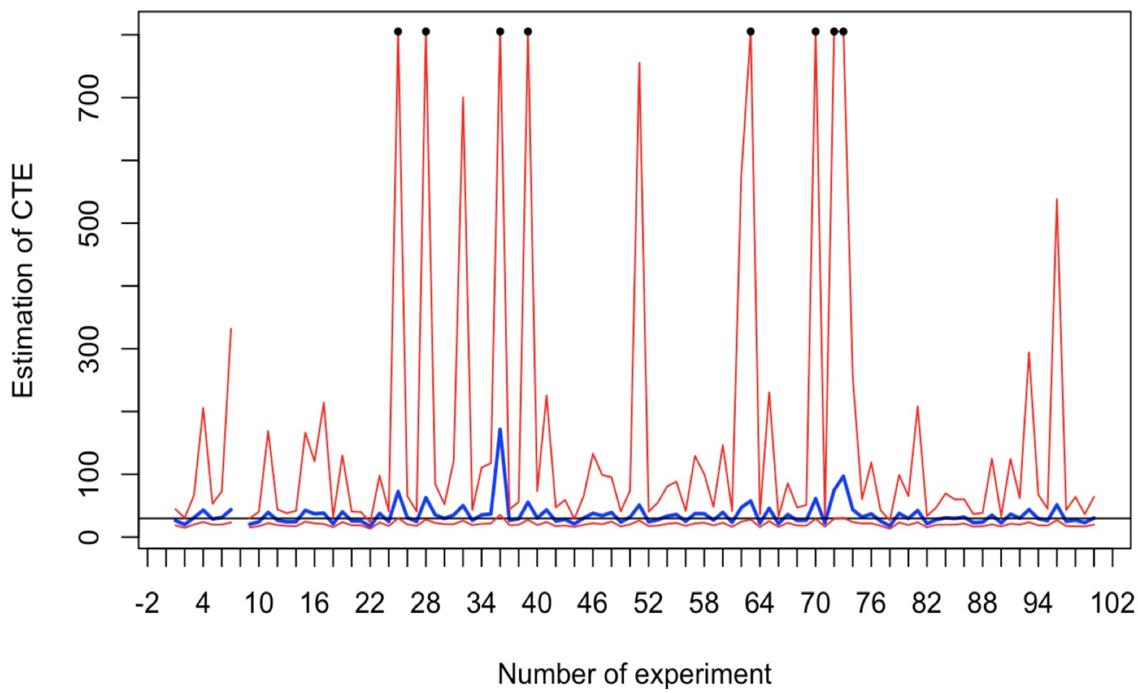

Figure 6.12: Estimates and 95% confidence intervals for $\theta(.05)$ using $\widehat{\theta}_1$ with $\alpha = 1.4$

In these situations, we shall see in the next section that the bootstrap technique can come to the rescue.

6.2.2 Bootstrapped confidence intervals for $\theta(p)$

We have seen that $\widehat{var}_i$ can become negative due to the instability of the Hill estimator, in which case the confidence interval (6.2.1) cannot be used. Instead, we can use Proposition 5.3.6 to find a bootstrapped confidence interval for $\theta(p)$. Using the bootstrap we can simulate data from either the $\mathscr{N}(0, var_1)$ or the $\mathscr{N}(0, var_0)$ distribution. A confidence interval is then given by

$$\theta(p) \in \left[\widehat{\theta}_i(p)\left(1+\frac{q_{0.975}}{\sqrt{k}}\right)^{-1} \;;\; \widehat{\theta}_i(p)\left(0 \vee \left[1+\frac{q_{0.025}}{\sqrt{k}}\right]\right)^{-1} \right]. \tag{6.2.2}$$

where $i = 0, 1$ and $q_{0.025}$, $q_{0.975}$ are the 0.025 and 0.975 quantiles of the corresponding bootstrapped data. (Alternatively, similar results are obtained by replacing the quantiles $q_{0.025}$, $q_{0.975}$ with $-1.96 s_b$ and $1.96 s_b$, respectively, where s_b^2 is the sample variance of the bootstrapped sample.) The bootstrap data are constructed as follows.

Recall that

$$\int_1^\infty (\widehat{T}_n(t)/t) = \widehat{\alpha}_1^{-1} \text{ and } \widehat{G}_n^{(b)}(t) = \frac{n}{\sqrt{k}}\left\{\overline{F}_n^{(b)}(X_{n:n-k}\, t) - \overline{F}_n(X_{n:n-k}\, t)\right\},$$

where $\overline{F}_n^{(b)}(x) = (1/n)\sum_{i=1}^n \mathbb{1}_{\left\{X_i^{(b)} > x\right\}}$ and $\overline{F}_n(x) = (1/n)\sum_{j=1}^n \mathbb{1}_{\{X_j > x\}}$. Then:

$$\begin{aligned}
\int_1^\infty \widehat{G}_n^{(b)}(t)\,\mathrm{d}t &= \frac{1}{\sqrt{k}}\sum_{j=1}^n \int_1^\infty \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k} t\}}\,\mathrm{d}t - \frac{1}{\sqrt{k}}\sum_{j=1}^n \int_1^\infty \mathbb{1}_{\{X_j > X_{n:n-k} t\}}\,\mathrm{d}t \\
&= \frac{1}{\sqrt{k}}\sum_{j=1}^n \int_1^{\frac{X_j^{(b)}}{X_{n:n-k}}} \mathrm{d}t \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1}{\sqrt{k}}\sum_{j=1}^k \int_1^{\frac{X_{n:n-j+1}}{X_{n:n-k}}} \mathrm{d}t \\
&= \frac{1}{\sqrt{k}}\sum_{j=1}^n \left(\frac{X_j^{(b)}}{X_{n:n-k}} - 1\right)\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1}{\sqrt{k}}\sum_{j=1}^k \left(\frac{X_{n:n-j+1}}{X_{n:n-k}} - 1\right)
\end{aligned}$$

$$= \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \left(\frac{X_j^{(b)}}{X_{n:n-k}} \right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{X_{n:n-j+1}}{X_{n:n-k}}$$

$$- \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} + \sqrt{k} \,, \tag{6.2.3}$$

$$\int_1^\infty \frac{\widehat{G}_n^{(b)}(t)}{t} \,\mathrm{d}t = \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \int_1^\infty \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k} t\}} \frac{\mathrm{d}t}{t} - \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \int_1^\infty \mathbb{1}_{\{X_j > X_{n:n-k} t\}} \frac{\mathrm{d}t}{t}$$

$$= \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \log \left(\frac{X_j^{(b)}}{X_{n:n-k}} \right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{\sqrt{k}}{k} \sum_{j=1}^{k} \log \left(\frac{X_{n:n-j+1}}{X_{n:n-k}} \right)$$

$$= \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \log \left(\frac{X_j^{(b)}}{X_{n:n-k}} \right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \sqrt{k}\, \widehat{\alpha}_1^{-1} \,, \tag{6.2.4}$$

$$\int_1^\infty \widehat{T}_n(t) \,\mathrm{d}t = \frac{1}{k} \int_1^\infty \sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{n:n-k} t\}} \,\mathrm{d}t = \frac{1}{k} \sum_{j=1}^{k} \frac{X_{n:n-j+1}}{X_{n:n-k}} - 1 \tag{6.2.5}$$

$$\widehat{G}_n^{(b)}(1) = \frac{1}{\sqrt{k}} \left\{ \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i > X_{n:n-k}\right\}} \right\}$$

$$= \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \sqrt{k} \,. \tag{6.2.6}$$

We evaluate the formulas in Proposition 5.3.6 using the expressions above. First,

$$\widehat{\alpha}_1^{-1} \widehat{G}_n^{(b)}(1) + \left(r + \frac{1}{1 - \widehat{\alpha}_1^{-1}} \right) \int_1^\infty \frac{\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\, \widehat{T}_n(t)}{t} \,\mathrm{d}t$$

$$= (1 - r - \widehat{\vartheta}_{n,1}) \widehat{\alpha}_1^{-1} \widehat{G}_n^{(b)}(1) + (r + \widehat{\vartheta}_{n,1}) \int_1^\infty \frac{\widehat{G}_n^{(b)}(t)}{t} \,\mathrm{d}t$$

$$= (1 - r - \widehat{\vartheta}_{n,1}) \widehat{\alpha}_1^{-1} \left(\frac{1}{\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \sqrt{k} \right)$$

$$+ (r + \widehat{\vartheta}_{n,1}) \left(\frac{1}{\sqrt{k}} \sum_{j=1}^{n} \log \left(\frac{X_j^{(b)}}{X_{n:n-k}} \right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \sqrt{k}\, \widehat{\alpha}_1^{-1} \right)$$

$$= \frac{1 - r - \widehat{\vartheta}_{n,1}}{\sqrt{k}} \widehat{\alpha}_1^{-1} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} + \frac{r + \widehat{\vartheta}_{n,1}}{\sqrt{k}} \sum_{j=1}^{n} \log\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}}$$
$$- \sqrt{k}\widehat{\alpha}_1^{-1} .$$

Next,

$$\widehat{\alpha}_1^{-1}\widehat{G}_n^{(b)}(1) + r\int_1^\infty \frac{\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\,\widehat{T}_n(t)}{t}\,\mathrm{d}t + \left(1 - \widehat{\alpha}_1^{-1}\right)\int_1^\infty \left(\widehat{G}_n^{(b)}(t) - \widehat{G}_n^{(b)}(1)\widehat{T}_n(t)\right)\,\mathrm{d}t$$
$$= (1-r)\widehat{\alpha}_1^{-1}\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}\mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \sqrt{k}\right) + \frac{r}{\sqrt{k}}\sum_{j=1}^{n}\log\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right)\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}}$$
$$- r\sqrt{k}\,\widehat{\alpha}_1^{-1} + \left(1 - \widehat{\alpha}_1^{-1}\right)\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{n}\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right)\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}}\right.$$
$$\left. - \frac{1}{\sqrt{k}}\sum_{j=1}^{n}\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} + \sqrt{k}\right)$$
$$- \left(1 - \widehat{\alpha}_1^{-1}\right)\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}\mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \sqrt{k}\right)\left(\frac{1}{k}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}} - 1\right)$$

$$= \frac{(1-r)\widehat{\alpha}_1^{-1}}{\sqrt{k}}\sum_{i=1}^{n}\mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \widehat{\alpha}_1^{-1}\sqrt{k} + \frac{r}{\sqrt{k}}\sum_{j=1}^{n}\log\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right)\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}}$$
$$+ \left(1 - \widehat{\alpha}_1^{-1}\right)\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{n}\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right)\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}}\right.$$
$$\left. - \frac{1}{\sqrt{k}}\sum_{j=1}^{n}\mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} + \sqrt{k}\right)$$
$$- \left(1 - \widehat{\alpha}_1^{-1}\right)\left(\frac{1}{k\sqrt{k}}\sum_{i=1}^{n}\mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}}\right.$$
$$\left. - \frac{1}{\sqrt{k}}\sum_{i=1}^{n}\mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} - \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{X_{n:n-j+1}}{X_{n:n-k}} + \sqrt{k}\right)$$

$$
\begin{aligned}
&= \frac{(1-r)\widehat{\alpha}_1^{-1}}{\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} + \frac{r}{\sqrt{k}} \sum_{j=1}^{n} \log\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \widehat{\alpha}_1^{-1}\sqrt{k} \\
&\quad + \left(1-\widehat{\alpha}_1^{-1}\right)\left(\frac{1}{\sqrt{k}} \sum_{j=1}^{n} \left(\frac{X_j^{(b)}}{X_{n:n-k}}\right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{X_{n:n-j+1}}{X_{n:n-k}}\right. \\
&\qquad - \frac{1}{\sqrt{k}} \sum_{j=1}^{n} \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} + \sqrt{k} - \frac{1}{k\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} \sum_{j=1}^{k} \frac{X_{n:n-j+1}}{X_{n:n-k}} \\
&\qquad \left. + \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} + \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{X_{n:n-j+1}}{X_{n:n-k}} - \sqrt{k}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{(1-r)\widehat{\alpha}_1^{-1}}{\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} + \frac{r}{\sqrt{k}} \sum_{j=1}^{n} \log\left(\frac{X_j^{(b)}}{X_{n:n-k}}\right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \widehat{\alpha}_1^{-1}\sqrt{k} \\
&\quad + \frac{1-\widehat{\alpha}_1^{-1}}{\sqrt{k}} \sum_{j=1}^{n} \left(\frac{X_j^{(b)}}{X_{n:n-k}}\right) \mathbb{1}_{\{X_j^{(b)} > X_{n:n-k}\}} - \frac{1-\widehat{\alpha}_1^{-1}}{k\sqrt{k}} \sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(b)} > X_{n:n-k}\right\}} \sum_{j=1}^{k} \frac{X_{n:n-j+1}}{X_{n:n-k}}.
\end{aligned}
$$

In what follows, we consider $n = 2000$ and $p = 0.05$. The data is simulated from the Pareto, Student-t and Mixture distributions. For different choices of the value of α, we will compare two 95% confidence intervals for $\theta(.05)$: the blue line represents estimates for $\theta(.05)$ using $\widehat{\theta}_0$ or $\widehat{\theta}_1$, the red lines represent the confidence bounds calculated using $\widehat{var}_i$ (we call this the "plugin" interval) and the black lines represent the bounds of the bootstrapped confidence interval given in (6.2.2).

a) **Pareto case** with $k = 0.05n$

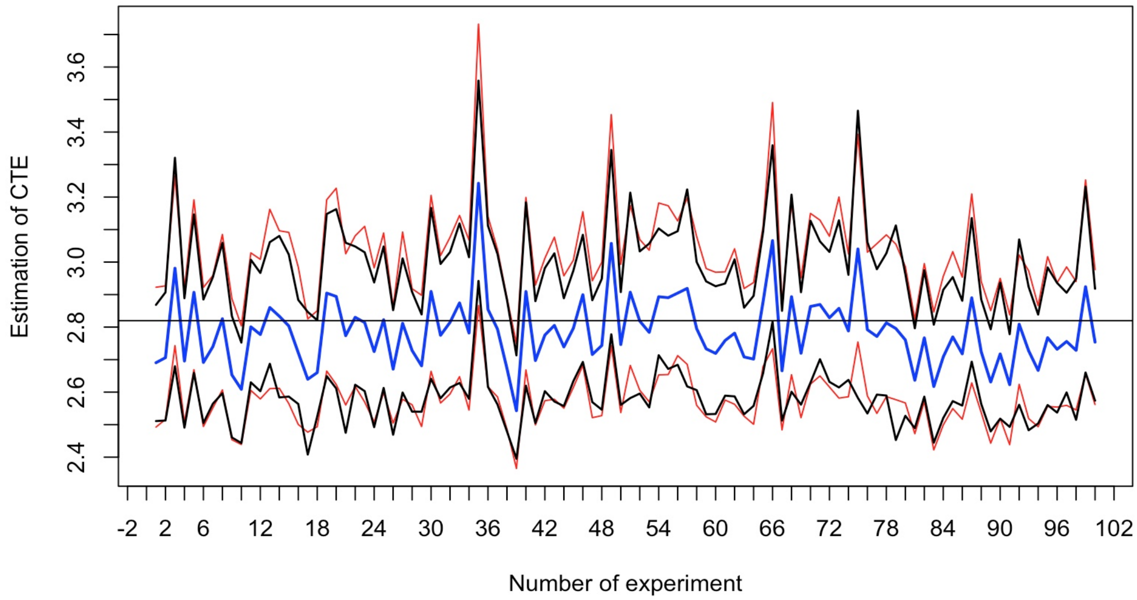

Figure 6.13: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 4$

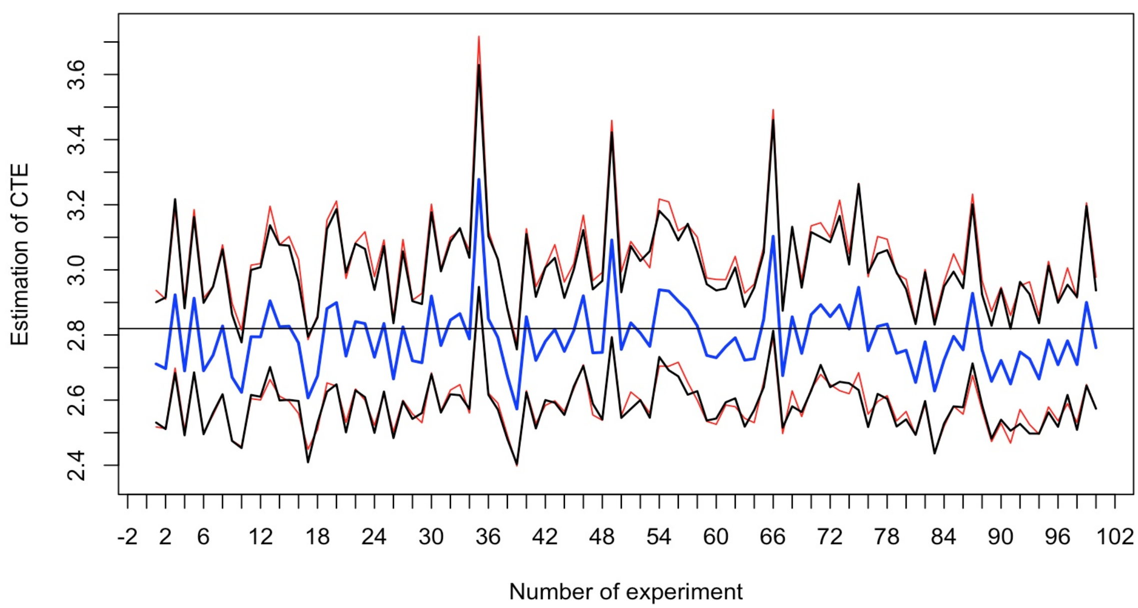

Figure 6.14: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 4$

Figures 6.13 and 6.14 illustrate that for $\alpha \gg 2$, the bootstrap confidence intervals appear to perform slightly better than the plugin intervals. As noted previously, either estimator $\widehat{\theta}_0$ or $\widehat{\theta}_1$ can be used. When α is close to 2, we have seen that $\widehat{\theta}_0$ becomes unstable; $1/\widehat{\alpha}^{-1}$ can yield negative variance or infinite upper confidence bounds. While bootstrapped confidence intervals can always be constructed, infinite upper confidence bounds can still occur. This instability is illustrated in Figure 6.15 for $\alpha = 2.3$.

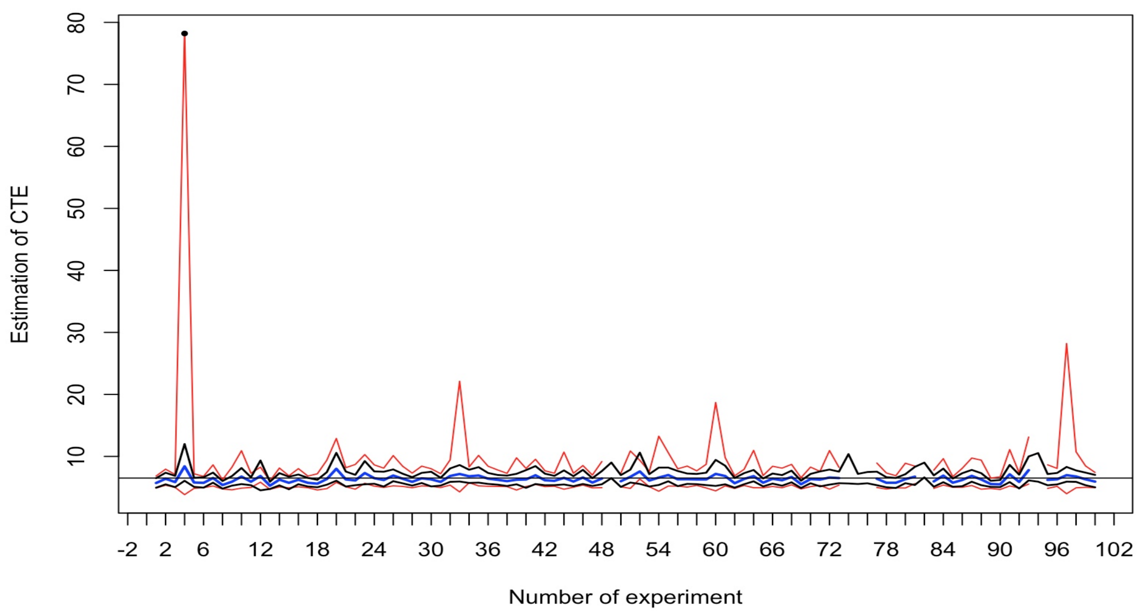

Figure 6.15: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.3$

In this case, it is recommended that $\widehat{\theta}_1$ be used and Figure 6.16 shows that the plugin and bootstrapped bounds are very similar, with the bootstrap performing slightly better.

When $1 < \alpha < 2$, the estimator $\widehat{\theta}_1$ must be used. Figure 6.17 clearly illustrates the instability of the upper confidence bounds, with the bootstrap frequently significantly outperforming the plugin procedure.

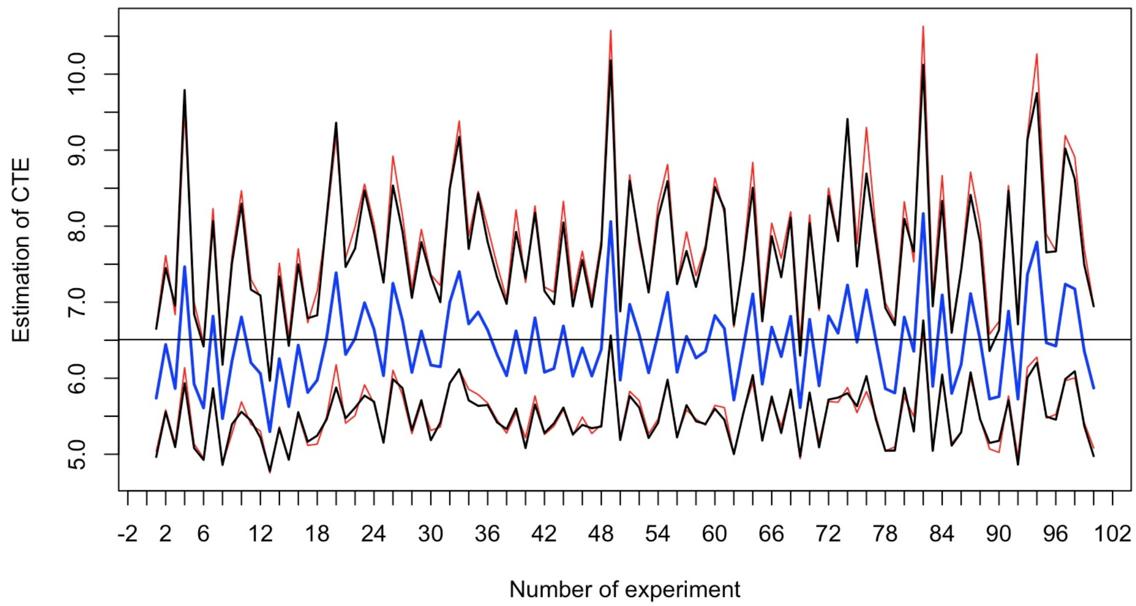

Figure 6.16: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.3$

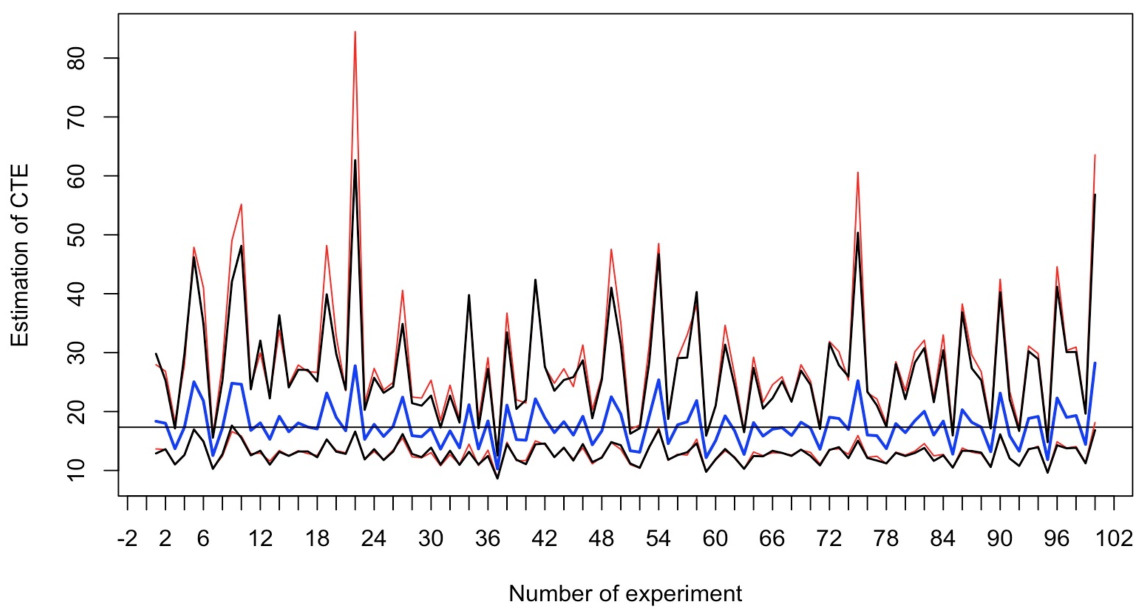

Figure 6.17: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 1.6$

b) Student-t case with $k = 0.025n$.

In the case of the Student-t distribution, for all values of α we find that the estimators of $\theta(.05)$ are biased upwards; however, we see that the bias appears to decrease with α. In the case of $\alpha = 4$, Figures 6.18 and 6.19 illustrate that neither the plugin nor the bootstrap confidence intervals of $\widehat{\theta}_0$ and $\widehat{\theta}_1$ give good coverage of the true value of θ.

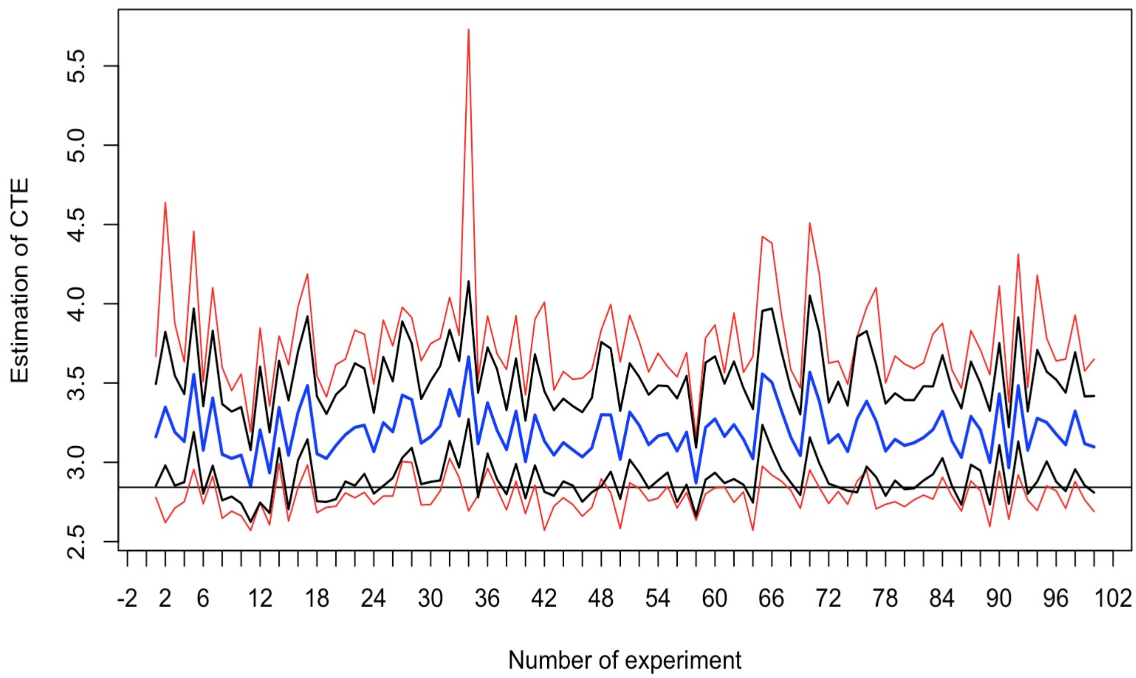

Figure 6.18: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 4$

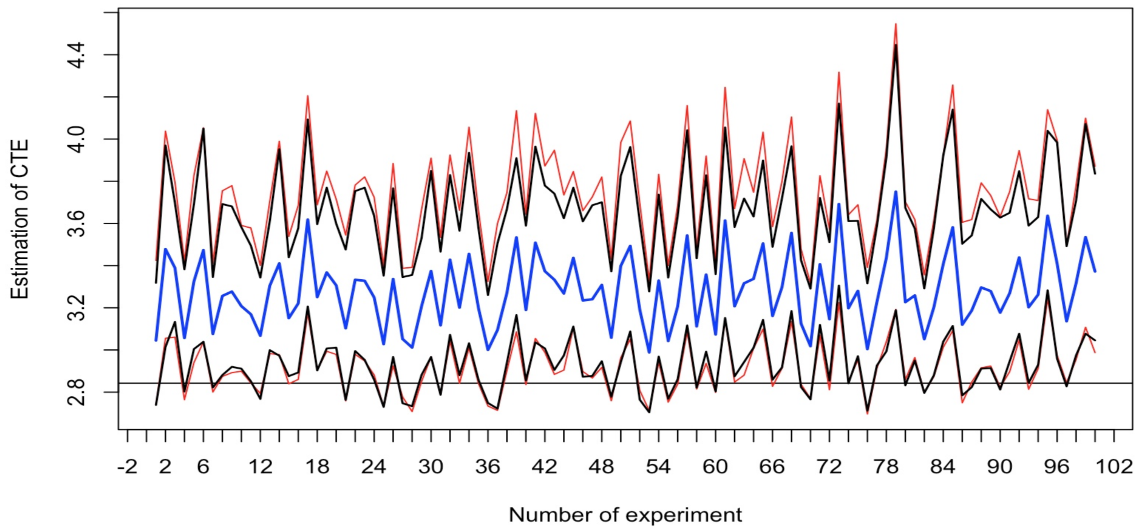

Figure 6.19: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 4$

For values of α closer to 2, Figures 6.20 and 6.21 show both decreasing bias but increasing volatility of the estimator $\widehat{\theta}_0$. The bootstrap gives much better upper confidence bounds than the plugin.

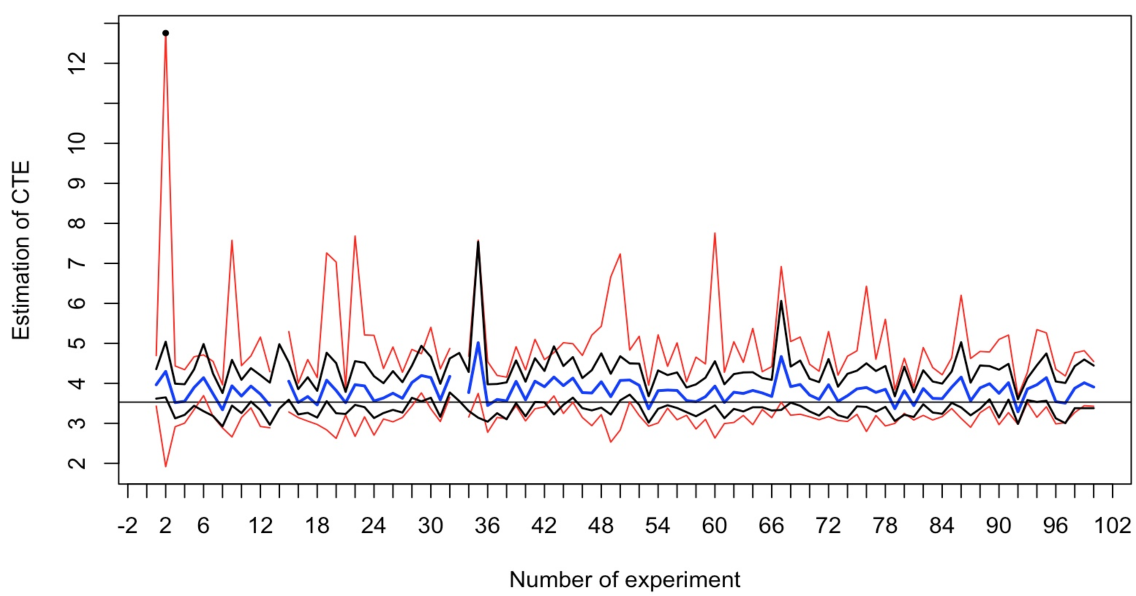

Figure 6.20: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 3$

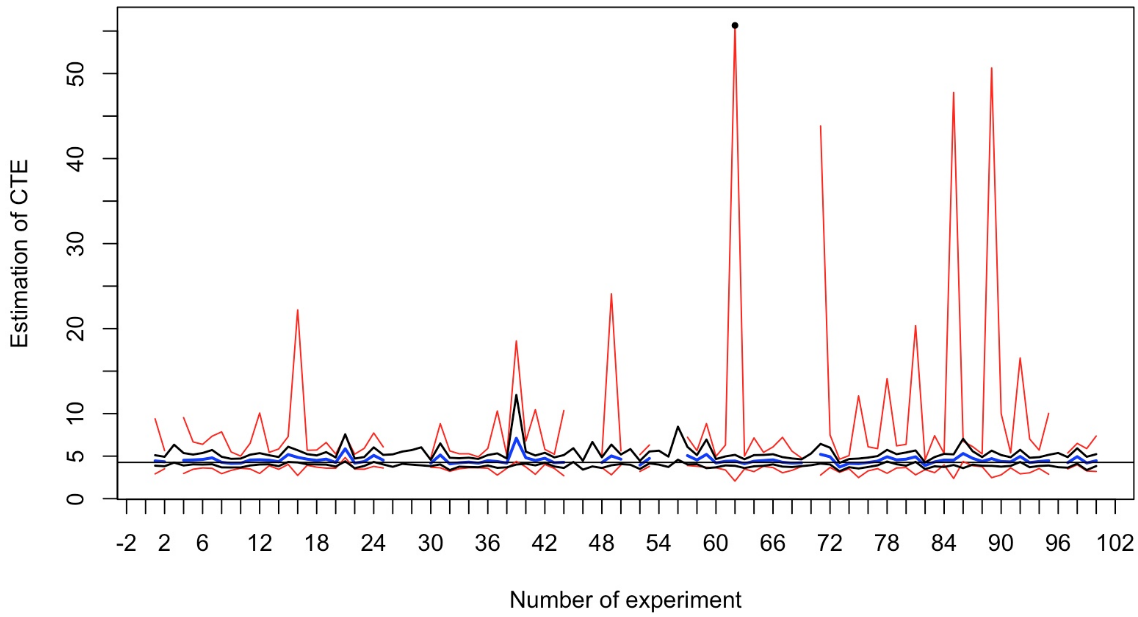

Figure 6.21: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.5$

Figures 6.22 and 6.23 also illustrate decreasing bias and increasing volatility of the estimator $\widehat{\theta}_1$ for smaller values of α. In both cases, the plugin and bootstrap confidence bounds are similar. For $\alpha = 3$, the bounds are close to the bootstrapped bounds obtained using $\widehat{\theta}_0$ seen in Figures 6.20 and 6.21. For $\alpha = 2.5$, it is clear that $\widehat{\theta}_1$ should be used.

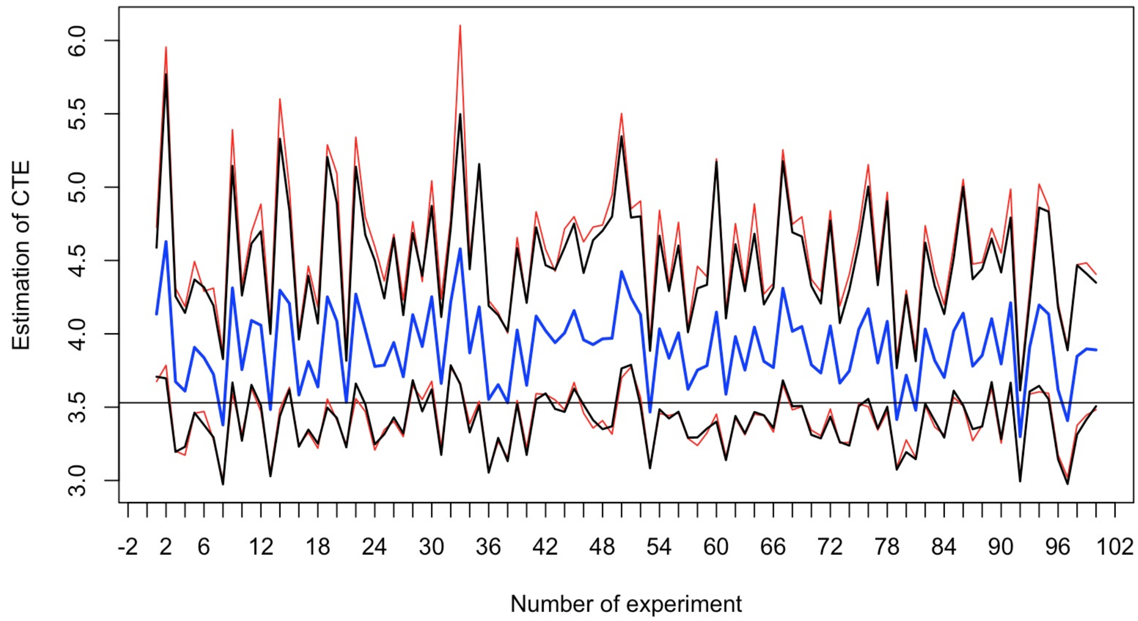

Figure 6.22: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 3$

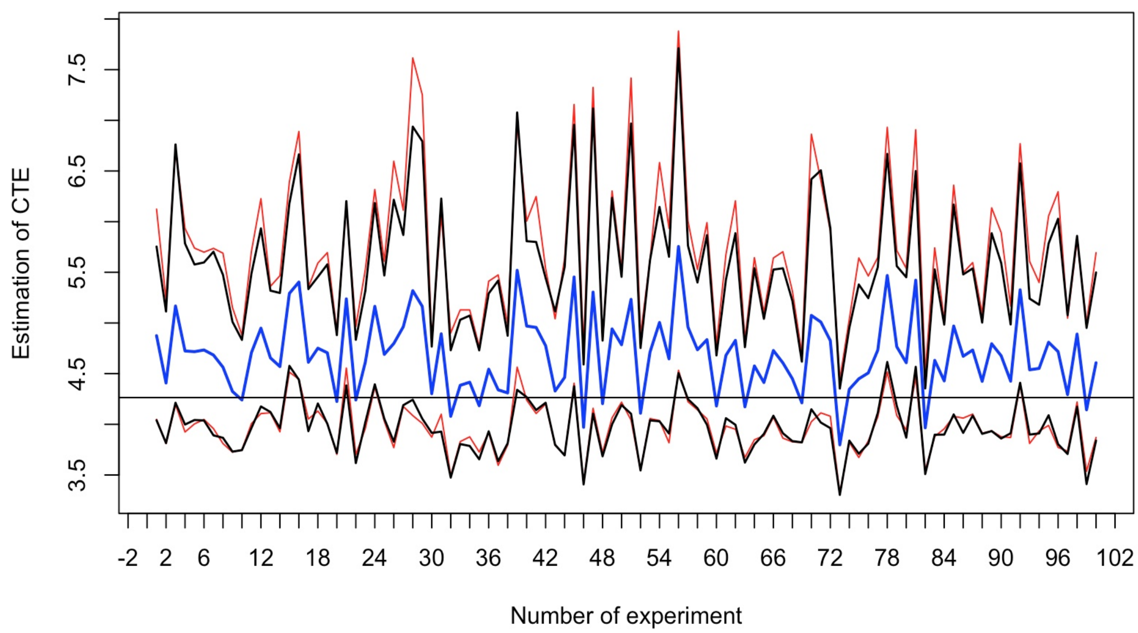

Figure 6.23: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.5$

For $\alpha = 2$, we must use the estimator $\widehat{\theta}_1$. Figure 6.24 show that the bias is much less noticeable and the confidence intervals give good coverage. The upper confidence bounds provided by the bootstrap are noticeably less than those of the plugin.

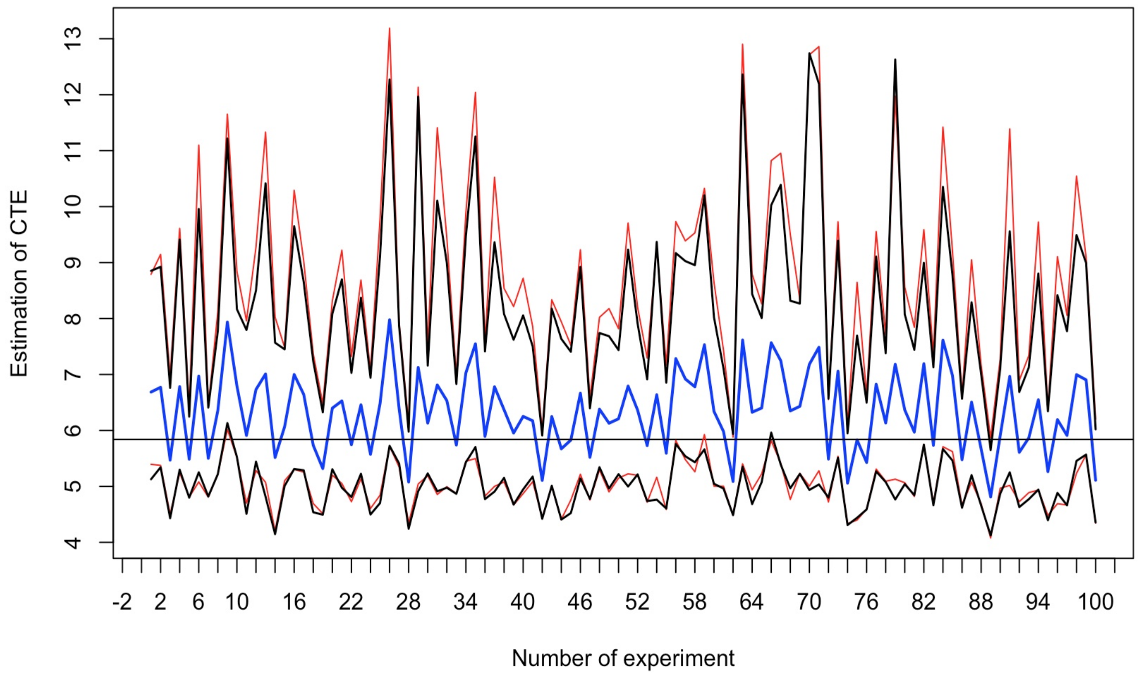

Figure 6.24: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2$

For $\alpha < 1.7$, Figures 6.25 and 6.26 show that there is little bias but both plugin and bootstrap methods become unstable. However, the bootstrap continues to perform significantly better than the plugin as the variance increases.

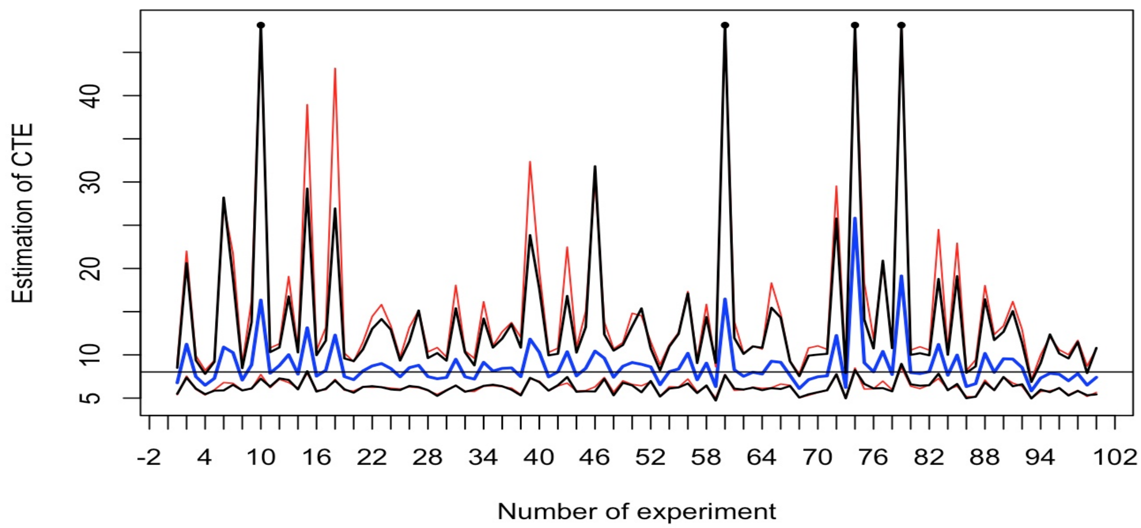

Figure 6.25: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 1.7$

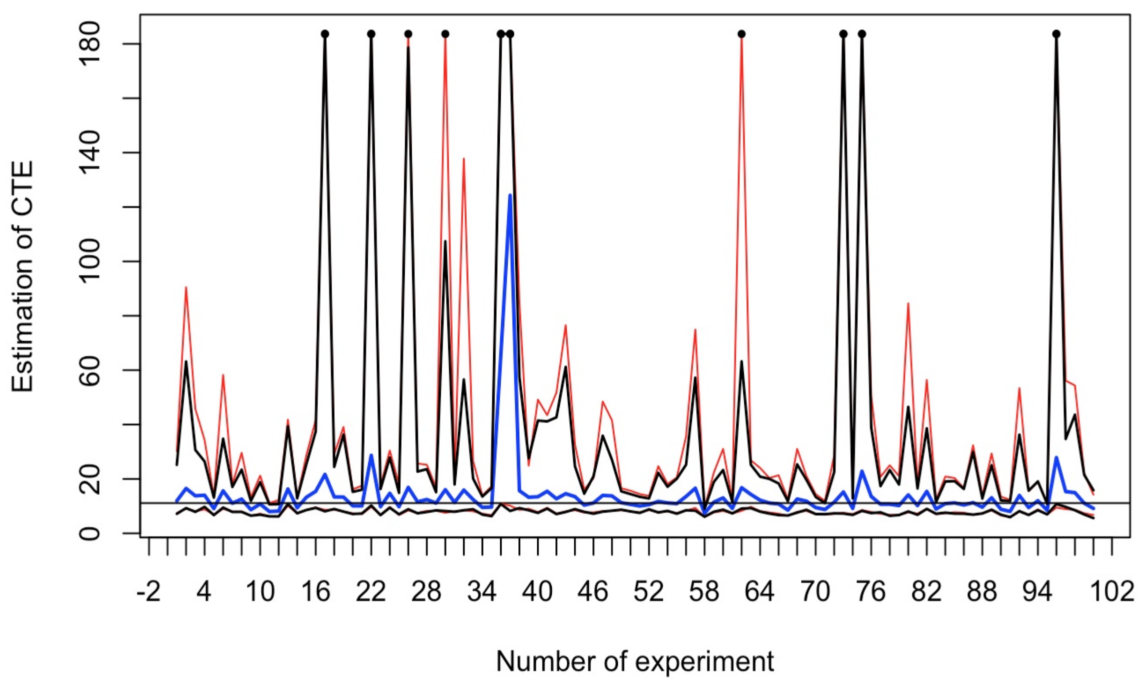

Figure 6.26: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 1.5$

As observed previously, the Hill estimator of $1/\alpha$ does not perform well for the t-distribution, and this likely affects the performance of the estimators of $\theta(p)$. We have seen that other estimators, such as the t-Hill, may be more appropriate and could replace $\widehat{\alpha}_1^{-1}$ in (4.2.7). However, the asymptotic behaviour of the estimator would change and bootstrap techniques would need to be justified.

c) **Mixture case** with $k = 0.025n$.

Recall that

$$\overline{F}_m(x) = {}^1\!/\!_2(x^{-\alpha} + x^{-\alpha\beta}),\ \ \alpha, \beta > 1.$$

Consider $n = 2000$, $p = 0.05$, two choices of parameter β ($\beta = 2$ and $\beta = 10$) and different choices of parameter α.

We will examine the effect of the parameters α, β on the confidence intervals for $\theta(.05)$ based on $\widehat{\theta}_0$, $\widehat{\theta}_1$ and their bootstrapped counterparts.

If $\alpha = 3$ or 4, for both values of β the behaviour of the plugin and the bootstrapped confidence intervals is similar to that of the Pareto distribution and so graphs are not included in the thesis.

When α converges to 2^+, there is a noticeable effect of β on the estimator $\widehat{\theta}_0$ of the CTE. The estimator $\widehat{\theta}_0$ is far less stable when $\beta = 10$ than it is when $\beta = 2$. In both cases, the boostrapped confidence intervals are much better than the plugin, as illustrated in Figures 6.27 and 6.28. However, as seen in the previous examples, $\widehat{\theta}_1$ is a good alternative estimator of the CTE, as shown in Figures 6.29 and 6.30. There is little difference between the plugin and bootstrapped confidence intervals. If $\alpha = 2$ and $\beta = 2,\ 10$, $\widehat{\theta}_1$ provides a stable estimator, but becomes increasingly unstable as α decreases, as illustrated in Figures 6.31 and 6.32.

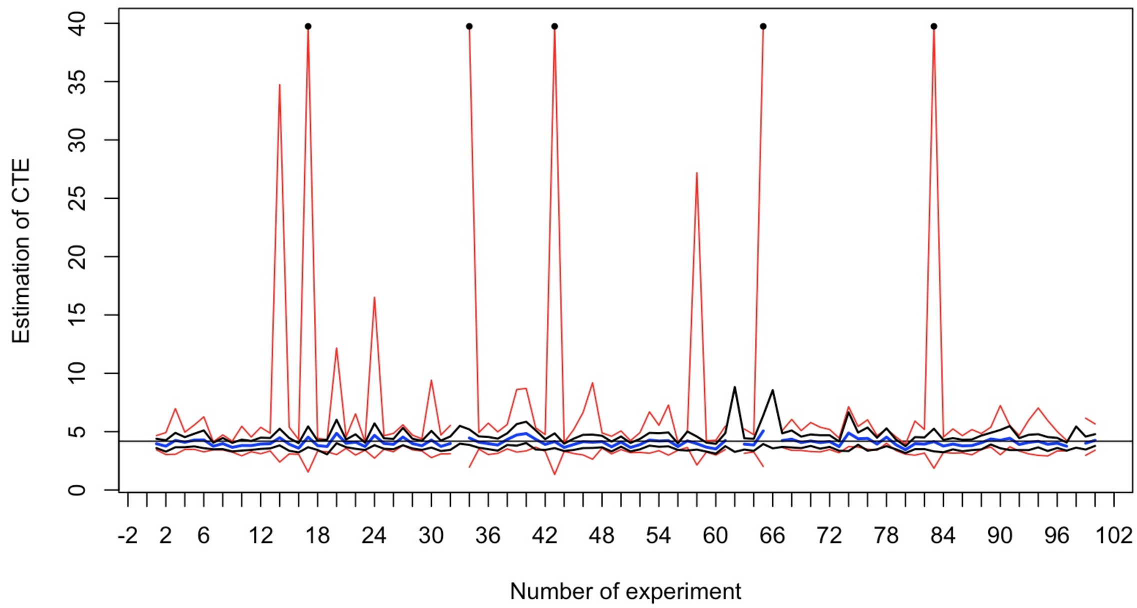

Figure 6.27: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.5$ and $\beta = 10$

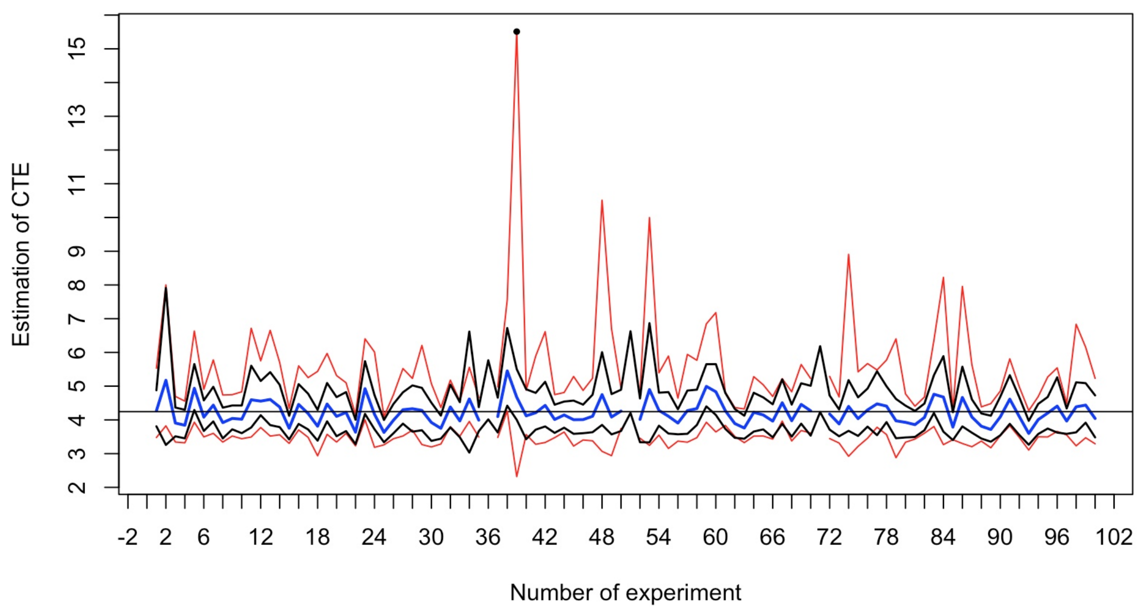

Figure 6.28: Plugin $\widehat{\theta}_0$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.5$ and $\beta = 2$

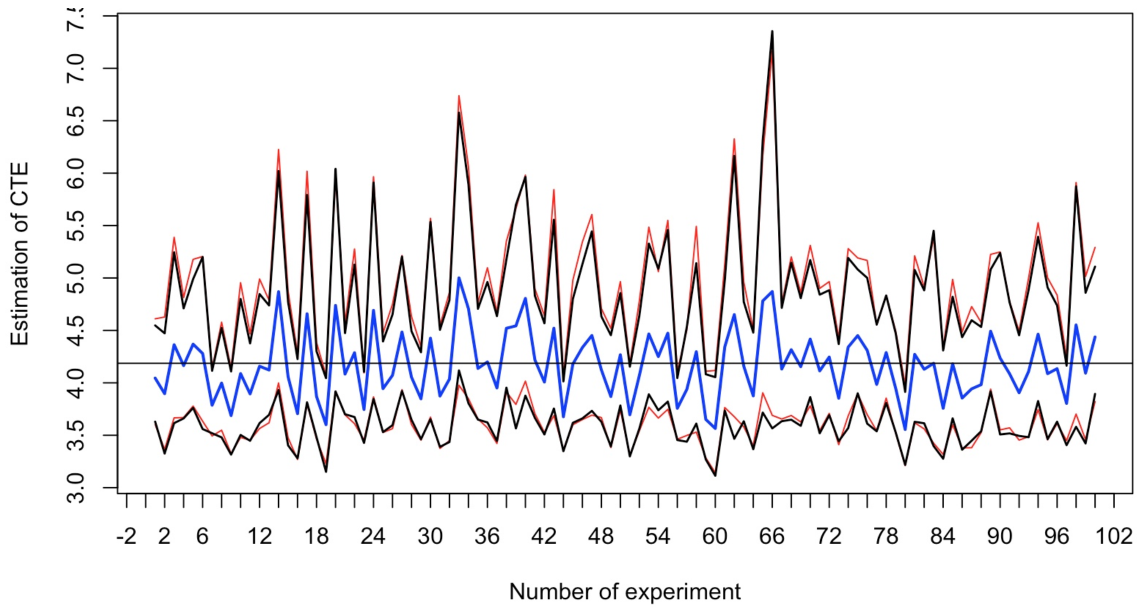

Figure 6.29: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.5$ and $\beta = 10$

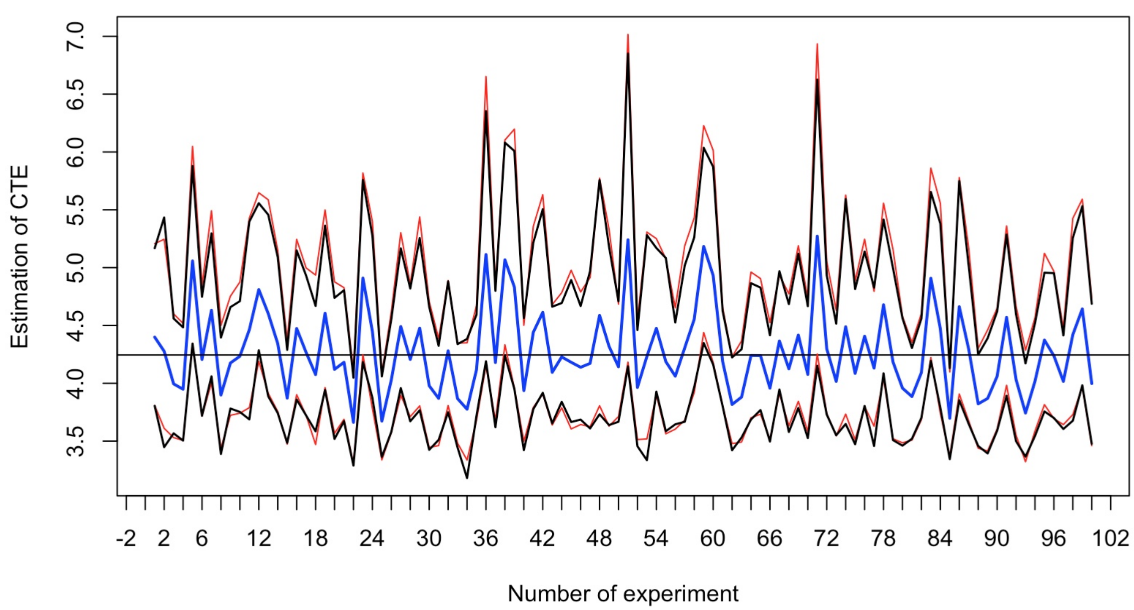

Figure 6.30: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 2.5$ and $\beta = 2$

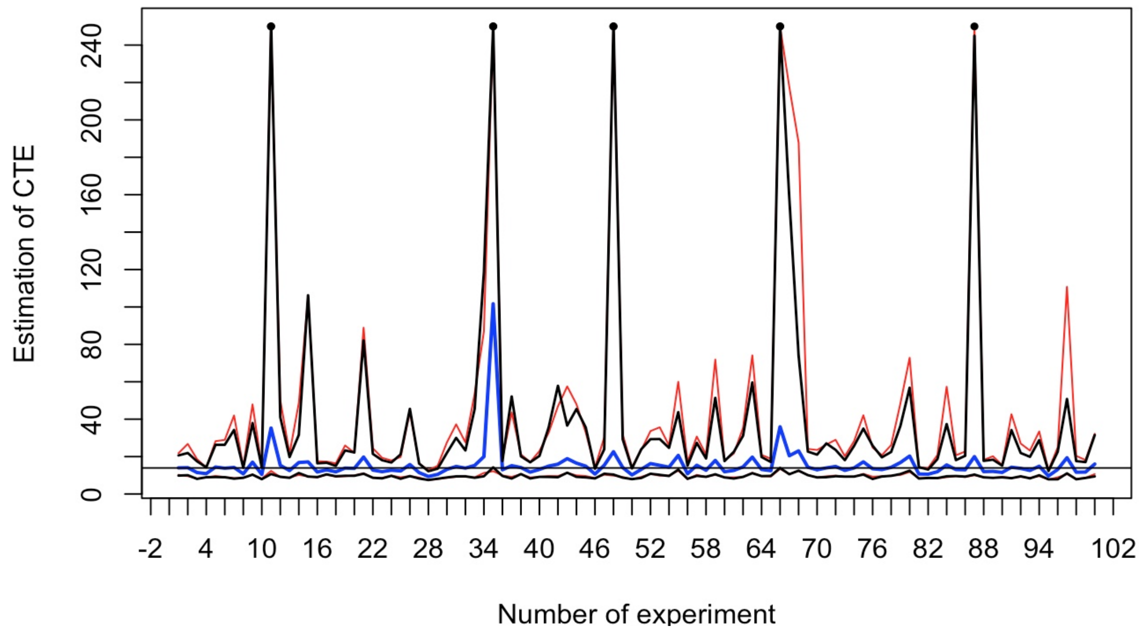

Figure 6.31: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 1.5$ and $\beta = 10$

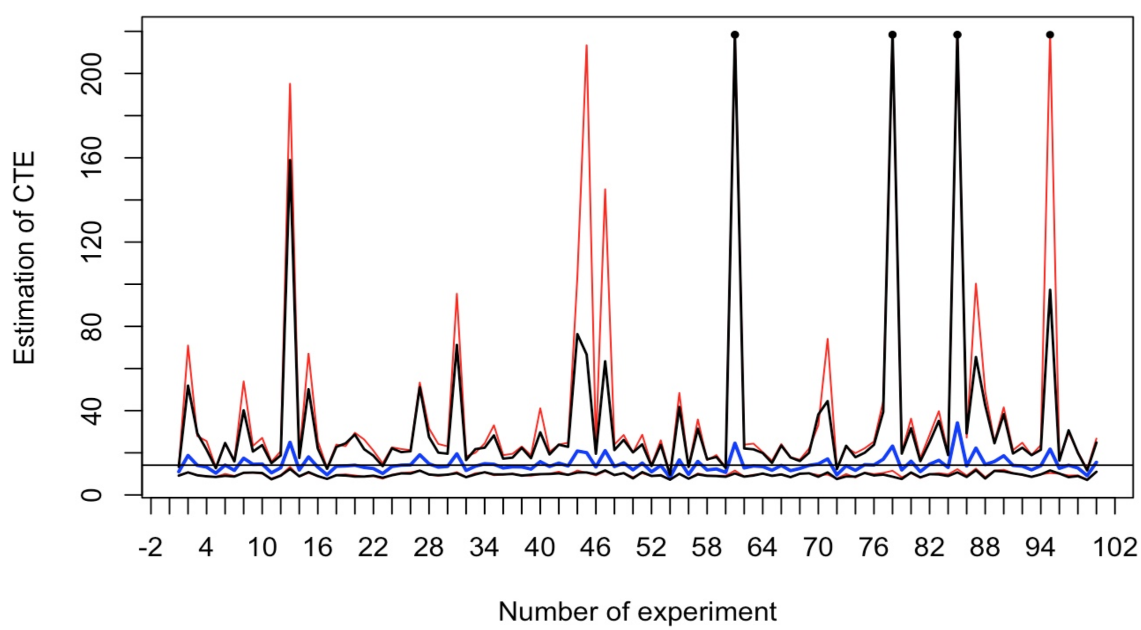

Figure 6.32: Plugin $\widehat{\theta}_1$ and bootstrapped 95% confidence intervals for $\theta(.05)$ with $\alpha = 1.5$ and $\beta = 2$

Summary: Overall, the results of our simulations for estimators of the CTE can be summarized as follows.

- In all cases, the performance of both estimators of $\theta(p)$ is similar for larger values of α. As $\alpha \to 2^+$, $\widehat{\theta}_1(p)$ should be used, and if it is suspected that $1 < \alpha \leq 2$, $\widehat{\theta}_1(p)$ must be used.

- In the case of the Pareto and mixture distributions, the estimators show little bias and perform well for larger values of α. However, the choice of k must be made carefully using a Hill plot.

- The estimators show significant upward bias in the case of the t-distribution. Alternative estimators of $\widehat{\alpha}^{-1}$ and $\widehat{\vartheta}$ should be considered in (4.2.7) for estimating $\theta(p)$.

- Bootstrap confidence intervals appear to perform at least as well as and, in many cases, significantly better than plugin intervals. This is particularly noticeable for smaller values of α. When $\widehat{var}_i$ is negative, the bootstrap must be used to construct a confidence interval for $\theta(p)$.

■

Chapter 7

Conclusion

Dans cette thèse nous avons abordé des concepts essentiels de la théorie univariée de variation régulière développée par Karamata depuis 1903. Une nouvelle lancée au sujet a été fourni, en 1970, par Laurens De Haan dans sa thèse “On Regular Variation and its Applications to the Weak Convergence of Sample Extreme” où il a introduit une généralisation substantielle de cette théorie. De nos jours, cette théorie fait partie des connaissances standards pour les probabilistes et les statisticiens.

Le paramètre d'importance en théorie des valeurs extrêmes est l'indice de queue. Il contrôle le comportement de la queue de distribution du premier ordre et de nombreuses méthodes d'estimation de ce paramètre ont été données dans la littérature.

Après avoir rappelé certaines des notions clés en théorie de variation régulière ainsi que quelques concepts en théorie de convergence faible dans le deuxième chapitre, où nous avons présenté une extension, Lemme 2.5.6, au critère de convergence en loi dans [5], nous nous sommes intéressés dans le troisième chapitre à une méthode unifiée d'estimation, Théorèmes 3.2.1 et 3.2.2, basée sur la convergence en loi des intégrales fonctionnelles

$$\int_1^\infty \frac{G_n(s)}{s^m}\,ds \qquad \text{et} \qquad \int_1^\infty \frac{\widehat{G}_n(s)}{s^m}\,ds$$

où l'on considère les processus de queue empiriques $G_n(\cdot)$ et $\widehat{G}_n(\cdot)$.

Cette nouvelle approche nous a permis d'étudier les propriétés asymptotiques de différents types d'estimateurs de l'indice de queue α, dont l'estimateur de Hill et la classe des estimateurs de la moyenne harmonique. En conséquence, nous avons développé un nouveau résultat, Corollaire 3.3.7, permettant l'estimation de la moyenne conditionnelle $\mathbb{E}[X|X > x]$.

Dans le quatrième chapitre, nous avons abordé les mesures les plus populaires de risque: la valeur à risque, VaR et l'espérance conditionnelle des pertes, CTE. Du point de vue de gestion des risques, l'étude asymptotique du comportement de la queue des distributions est pertinente pour comprendre l'ampleur de grandes pertes et la performance de mesures de risque correspondantes. Dans cette perspective, nous nous sommes intéressés à l'estimation de ces mesures de risque pour des pertes à queue lourde et à variation régulière. Par ailleurs, nous avons proposé dans ce chapitre une nouvelle approche pour pouvoir estimer la mesure de risque CTE. Le Théorème 4.2.4, propose deux estimateurs de la CTE, en étudiant leur normalité asymptotique.

Les propriétés asymptotiques des ces estimateurs sont basées sur le paramètre de queue α supposé inconnu et que l'on estime via l'estimateur de Hill. Dans certains cas, les variances estimées de ces estimateurs peuvent être négatives, aussi la borne supérieure de l'intervalle de confiance peut égaler l'infini. Ces limitations rend l'utilisation de ces estimateurs impossible.

Dans le cinquième chapitre, nous avons développé de nouveaux résultats permettant l'utilisation de la technique de bootstrap afin de rectifier le problème de la variance estimée. La méthode développée nous a permis d'approcher les lois asymptotiques fonctionnelles des estimateurs proposés de la CTE.

En économie, les distributions Pareto et Student-t sont les plus populaires. En particulier, Student-t est couramment utilisée dans la finance et la gestion des risques. Elle est souvent présentée comme étant plus adaptée que la distribution normale, pour caractériser les facteurs de risques du marché, car la variation de son degré de liberté permet un meilleur contrôle du kurtosis. En fait, le libre choix du degré de liberté, qui correspond bien à l'indice de queue α, permettrait de mieux adapter les queues de distributions aux réalités financières.

Partant de ce fait, au sixième chapitre, nous nous sommes attardés sur des simulations afin d'apprécier la qualité de la méthode de bootstrap, développée dans le chapitre 5 et appliquée aux estimateurs proposés de la CTE. Dans nos simulations, nous avons utilisé 3 distributions différentes: Pareto, Student-t et de mélange. Les résultats obtenus sont très encourageants illustrant ainsi l'utilité du bootstrap pour fournir de meilleurs intervalles de confiance, plus particulièrement pour la distribution Student-t.

Enfin, nous dressons ci-après une liste non exhaustive des directions possibles que peuvent prendre des recherches futures:

- ☞ Remplacer l'estimateur de Hill par d'autres estimateurs dans l'estimation de la CTE, notamment l'estimateur t-Hill et la classe HME. Notons que la technique de bootstrap doit être justifier;
- ☞ Développer la technique de bootstrap en considérant la condition de dérivabilité (3.1.25);
- ☞ Étendre les résultats obtenus aux vecteurs aléatoires à variation régulière multivariées;
- ☞ Considérer différents choix du nombre k de la statistique d'ordre supérieur, de r et de la probabilité p dans la partie de simulation;
- ☞ Considérer autre distributions à variation régulière dans les applications.

Appendix A

Useful Inequalities

For x, $y \in \mathbb{R}$ the standard triangular inequality states that $|x + y| \leq |x| + |y|$. The following lemma is the analog for powers.

Lemma A.0.1 *Let $r > 0$, and suppose that x, $y > 0$. Then*

$$(x+y)^r \leq \begin{cases} 2^r \ (x^r + y^r)\,, & \textit{for } r > 0\,, \\ x^r + y^r, & \textit{for } 0 < r \leq 1\,, \\ 2^{r-1} \ (x^r + y^r)\,, & \textit{for } r \geq 1\,. \end{cases}$$

Proof: For $r > 0$,

$$(x+y)^r \leq \big(2 \max\{x, y\}\big)^r = 2^r \left(\max\{x, y\}\right)^r \leq 2^r \left(x^r + y^r\right) .$$

Next, suppose that $0 < r \leq 1$. Then, since $x^{1/r} \leq x$ for any $0 < x < 1$, it follows that

$$\left(\frac{x^r}{x^r + y^r}\right)^{1/r} + \left(\frac{y^r}{x^r + y^r}\right)^{1/r} \leq \frac{x^r}{x^r + y^r} + \frac{y^r}{x^r + y^r} = 1\,,$$

and, hence, that

$$x + y \leq \left(x^r + y^r\right)^{1/r} \ ,$$

which is equivalent to the second inequality.

For $r \geq 1$ we exploit the fact that the function $|x|^r$ is convex, so that, in particular,

$$\left(\frac{x+y}{2}\right)^r \leq \frac{1}{2}x^r + \frac{1}{2}y^r,$$

which is easily reshuffled into the third inequality. ∎

Theorem A.0.2 *[24] Let $r > 0$. Suppose that $E\left[|X|^r\right] < \infty$ and $E\left[|Y|^r\right] < \infty$. Then*

$$E\left[|X+Y|^r\right] \leq 2^r \left(E\left[|X|^r\right] + E\left[|Y|^r\right]\right).$$

Proof: Set $x = X(\omega)$ and $y = Y(\omega)$. The triangle inequality and Lemma A.0.1 together yield

$$E\left[|X+Y|^r\right] \leq E\left[\left(|X|+|Y|\right)^r\right] \leq 2^r \left(E\left[|X|^r\right] + E\left[|Y|^r\right]\right).$$

∎

Theorem A.0.3 *[34] Let $p \geq 1$. Then for every random variable S having finite pth moment,*

$$\mathbb{E}\left[|S - \mathbb{E}[S]|^p\right] \leq C_p \mathbb{E}\left[|S|^p\right],$$

where C_p is positive number.

Appendix B

R code

```
#------------------------------------------------------------------------
#    Estimation of CTE using theta_0, theta_1 and bootstrap
#------------------------------------------------------------------------

a=alpha=2.5; b=beta=2; gama=1/a; n=2000; ns=100; m=400; p=0.05;
k=floor(0.05*n);
#k=floor(0.025*n); # for mixture and Student-t
r=log(k/(n*p)); varteta=a/(a-1);
CTE=varteta*p^(-gama); # for Pareto
#CTE=varteta*qt(.95, df=a); # for Student-t
theta_0=NULL; theta_1=NULL; mse0=NULL; mse1=NULL; Hill=NULL;
ratio_0=NULL; ratio_1=NULL; var_theta_0=NULL; var_theta_1=NULL
CI_up_0=NULL; CI_low_0=NULL; CI_up_1=NULL; CI_low_1=NULL;
CI_up_GI_0=NULL; CI_low_GI_0=NULL; CI_up_GI_1=NULL; CI_low_GI_1=NULL;
index_alpha_0=NULL; index_alpha_1=NULL; index_alpha_11=NULL; index_alpha_00=NULL
ind_CI_up_0=NULL; ind_CI_up_1=NULL; ind_CI_up_GI_0=NULL; ind_CI_up_GI_1=NULL;

# Define mixture function
#------------------------------
#require(GoFKernel)
#mixt=function(x) 1-0.5*(x^(-a)+x^(-a*b));
#inv=inverse(mixt,lower=1,upper=10); # inverse function
#Qx=inv(1-p); # quantile at level p
#CTE_est=a/(a-1)*Qx; # using formula (4.2.3)
#CTE=-(0.5*a/(p-a*p)*Qx^(1-a)+0.5*a*b/(p-a*b*p)*Qx^(1-a*b)); # using formula (4.2.1)

```

```r
for (i in 1:ns) {
  #------ observations from Pareto ------
  data=(1-runif(n))^(-gama)
  #------ observations from Student-t ------
  #data=rt(n,a)
  #------ observations from mixture ------
  #U=(1-runif(n))^(-1/a);  V=(1-runif(n))^(-1/(a*b));
  #coin=rbinom(n,1,0.5);
  #data=coin*U+(1-coin)*V;
  #------------------------------------------
  sorted_data=sort(data, decreasing=TRUE);
  Xn_k=sorted_data[k+1];
  upper_X=sorted_data[1:k];
  log_upper_X=log(upper_X/Xn_k);
  sumupper_X=sum(upper_X);
  #------------------------------------------
  Hill[i]=mean(log_upper_X); # vector of Hill estimator
  hill=mean(log_upper_X); # Hill estimator
  a_est=1/hill;
  #----------------------------------------------
  Q_x=Xn_k*(k/(n*p))^(hill)
  # estimators of conditional mean
  #--------------------------------------
  varteta_0=mean(upper_X)/Xn_k;
  varteta_1=1/(1-hill);
  # Estimators of CTE
  #-------------------------------------------
  theta_0[i]=varteta_0*Q_x;
  theta_1[i]=varteta_1*Q_x;
  # Variances
  #------------------------
  var_theta_true_0=((r^2+2*r+2)*a^2-(3*r^2+4*r+4)*a+2*(1+r^2))/(a^2*(a-1)*(a-2));
  var_theta_true_1=((r^2+2*r+2)*a^2-2*(r^2+r+1)*a+(1+r^2))/(a^2*(a-1)^2);
  var_theta_0[i]=((r^2+2*r+2)*a_est^2-(3*r^2+4*r+4)*a_est+2*(1+r^2))/(a_est^2*(a_est-1)*(a_est-2));
  var_theta_1[i]=((r^2+2*r+2)*a_est^2-2*(r^2+r+1)*a_est+(1+r^2))/(a_est^2*(a_est-1)^2);
  # Mean squared error
  #------------------------------
  mse0[i]=(theta_0[i]-CTE)^2;
  mse1[i]=(theta_1[i]-CTE)^2;

```

```
# If alpha_hat<2, then theta_0 not applicable
#-------------------------------------------------
if (1/Hill[i]<2) {theta_0[i]=NA}
# If alpha_hat<1, then theta_1 not applicable
#-------------------------------------------------
if (1/Hill[i]<1) {theta_1[i]=NA}
# indices when alpha_hat<2 or alpha_hat<1
#-------------------------------------------------
if (1/Hill[i]<2) {index_alpha_0[i]=i}
if (1/Hill[i]<1) {index_alpha_1[i]=i}
# cases when CI_up is infinite:
# ---- for Pareto distribution ----
if (1/Hill[i]<2.019205 & 1/Hill[i]>2) {index_alpha_00[i]=i}
if (1/Hill[i]<1.19867 & 1/Hill[i]>1) {index_alpha_11[i]=i}
# ---- for Student-t and mixture distributions ----
#if (1/Hill[i]<2.0384 & 1/Hill[i]>2) {index_alpha_00[i]=i}
#if (1/Hill[i]<1.28389 & 1/Hill[i]>1) {index_alpha_11[i]=i}
#            Bootstrapped data
#-----------------------------------------------
for (j in 1:m){
  bootX=sample(data,replace=T);
  upper_bootX=subset(bootX, bootX > Xn_k);
  sumupper_bootX=sum(upper_bootX);
  sumLogupper_bootX=sum(log(upper_bootX/Xn_k));
  # Ratio of CTE_hat/CTE
  #--------------------------
  A=(1-r)*hill/sqrt(k)*length(upper_bootX);
  B=r/sqrt(k)*sumLogupper_bootX-hill*sqrt(k);
  C=(1-hill)/sqrt(k)*sumupper_bootX/Xn_k;
  D=(1-hill)/(k*sqrt(k))*length(upper_bootX)*sumupper_X/Xn_k;
  ratio_0[j]=A+B+C-D; # for theta_0
  # for theta_1
  ratio_1[j]=(1-r-varteta_1)/sqrt(k)*hill*length(upper_bootX)
  +(r+varteta_1)/sqrt(k)*sumLogupper_bootX-sqrt(k)*hill;
}
# bootstrapped confidence intervals
#-----------------------------------------
CI_up_GI_0[i]=theta_0[i]*max(0,1+quantile(ratio_0,probs = 0.025)/sqrt(k))^(-1);
CI_low_GI_0[i]=theta_0[i]*(1+quantile(ratio_0,probs = 0.975)/sqrt(k))^(-1);
CI_up_GI_1[i]=theta_1[i]*max(0,1+quantile(ratio_1,probs = 0.025)/sqrt(k))^(-1);
CI_low_GI_1[i]=theta_1[i]*(1+quantile(ratio_1,probs = 0.975)/sqrt(k))^(-1);

```

```
  # If alpha_hat<2 or alpha_hat<1, theta_0 and theta_1 not applicable
  #--------------------------------------------------------------------
  if (1/Hill[i]<2) {theta_0[i]=NA}
  if (1/Hill[i]<1) {theta_1[i]=NA}

  # Plugin confidence intervals theta_0 and theta_1
  #-----------------------------------------------------------------
  CI_up_0[i]=theta_0[i]*max(0,1-1.96*sqrt(var_theta_0[i])/sqrt(k))^(-1);
  CI_low_0[i]=theta_0[i]*(1+1.96*sqrt(var_theta_0[i])/sqrt(k))^(-1);
  CI_up_1[i]=theta_1[i]*max(0,1-1.96*sqrt(var_theta_1[i])/sqrt(k))^(-1);
  CI_low_1[i]=theta_1[i]*(1+1.96*sqrt(var_theta_1[i])/sqrt(k))^(-1);
}

#
index_alpha_0=subset(index_alpha_0, is.finite(index_alpha_0)>0)
index_alpha_00=subset(index_alpha_00, is.finite(index_alpha_00)>0)
index_alpha_1=subset(index_alpha_1, is.finite(index_alpha_1)>0)
index_alpha_11=subset(index_alpha_11, is.finite(index_alpha_11)>0)
#
# indices when CI_up are infinites
#-----------------------------------
ind_CI_up_0=which(CI_up_0 %in% Inf)
ind_CI_up_GI_0=which(CI_up_GI_0 %in% Inf)
ind_CI_up_1=which(CI_up_1 %in% Inf)
ind_CI_up_GI_1=which(CI_up_GI_1 %in% Inf)
#
# When CI_up is infinite, replace with maxCI_up+constante
#-------------------------------------------------------------
CI_up_0=replace(CI_up_0, is.infinite(CI_up_0),max(CI_up_0[is.finite(CI_up_0)])+50);
CI_up_1=replace(CI_up_1, is.infinite(CI_up_1),max(CI_up_1[is.finite(CI_up_1)])+50);
CI_up_GI_0=replace(CI_up_GI_0, is.infinite(CI_up_GI_0),max(CI_up_GI_0[is.finite(
    CI_up_GI_0)])+50);
CI_up_GI_1=replace(CI_up_GI_1, is.infinite(CI_up_GI_1),max(CI_up_GI_1[is.finite(
    CI_up_GI_1)])+50);
#
# indices when CI_up is infinite
#--------------------------------------
X_0=subset(CI_up_0,CI_up_0==max(CI_up_0,na.rm = T))
X_1=subset(CI_up_1,CI_up_1==max(CI_up_1,na.rm = T))
X_GI_0=subset(CI_up_GI_0,CI_up_GI_0==max(CI_up_GI_0,na.rm = T))
X_GI_1=subset(CI_up_GI_1,CI_up_GI_1==max(CI_up_GI_1,na.rm = T))

```

```r
#------------------------------------------------------------
#          plots for theta_0
#------------------------------------------------------------
limsup=max(CI_up_0,CI_up_GI_0,na.rm = T);
liminf=min(CI_low_0,CI_low_GI_0,na.rm = T);
plot(theta_0, type = 'l', col="blue",lwd = 2, ylim=c(liminf,limsup),
      xlab="Number of experiment",ylab="Estimation of CTE",lab=c(50,10,0));
abline(h=CTE, col="black");
points(CI_up_0, type='l', pch = 20, cex=0.5, col="red")
points(CI_low_0, type='l', pch = 21, cex=0.5, col="red")
points(CI_up_GI_0, type='l', lwd = 1.5, pch = 21, cex=0.5, col="black");
points(CI_low_GI_0, type='l', lwd = 1.5, pch = 21, cex=0.5, col="black")
points(ind_CI_up_0,X_0,type='p', pch = 19, cex=0.5, col="black")
points(ind_CI_up_GI_0, X_GI_0, type='p', pch = 19, cex=0.5, col="red")
#
index_alpha_0 # alpha_hat < 2
1/Hill[index_alpha_0]
index_alpha_00 # 2<alpha_hat<2.019205 or 2<alpha_hat<2.0384
1/Hill[index_alpha_00]
#------------------------------------------------------------
#          plots for theta_1
#------------------------------------------------------------
liminf=min(CI_low_1,CI_low_GI_1,na.rm = T);
limsup=max(CI_up_1,CI_up_GI_1,na.rm = T);
plot(theta_1, type = 'l', col="blue",lwd = 2, ylim=c(liminf,limsup),
      xlab="Number of experiment",ylab="Estimation of CTE",lab=c(50,10,0));
abline(h=CTE, col="black")
points(CI_up_1, type='l', pch = 20, cex=0.5, col="red");
points(CI_low_1, type='l', pch = 21, cex=0.5, col="red")
points(CI_up_GI_1, type='l', lwd = 1.5, pch = 21, cex=0.5, col="black");
points(CI_low_GI_1, type='l', lwd = 1.5, pch = 21, cex=0.5, col="black")
points(ind_CI_up_1,X_1,type='p', pch = 19,cex=0.5, col="black")
points(ind_CI_up_GI_1,X_GI_1,type='p', pch = 19, cex=0.5, col="red")
#
index_alpha_1 # alpha_hat < 1
1/Hill[index_alpha_1]
index_alpha_11 # 1<alpha_hat<1.19867 or 1<alpha_hat<1.28389
1/Hill[index_alpha_11]
```

Bibliography

[1] Ankirchner S., Kruse T., Urusov M. (2017). WLLN for arrays of nonnegative random variables. *Statistics and Probability Letters.* **122**, 73-78. (Cited on § 1 on page 101, § 5.3.3 on page 102)

[2] Artzner P., Delbaen F., Eber J.-M., Heath D. (1999). Coherent measures of risk. *Mathematical Finance.* **9 (3)**, 203 - 228. (Cited on § 1.2.3 on page 5, § 4 on page 65)

[3] Beirlant J., Goegebeur Y, Teugels J., Segers J. (2004). Statistics of Extremes : Theory and Applications. John Wiley & Sons, England. (Cited on § 1.2.1 on page 3, § 2.2.3 on page 15)

[4] Beran J., Schell D., Stehlik M. (2013). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Annals of Institute of Statistical Mathematics.* **66 (1)**, 193 - 220. (Cited on § 3.3.2 on page 55)

[5] Billingsley P. (1968). Convergence of Probability Measures. Wiley, New York. (Cited on § 1.2.1 on page 3, § 1.2.1 on page 4, § 2.5.2 on page 21, § 2.5.2 on page 22, § 2.5.2 on page 24, § 2.5.8 on page 25, § 5.3.1 on page 96)

[6] Billingsley P. (1999). Convergence of Probability Measures. Second Edition.Wiley, New York. (Cited on § 1.2.1 on page 3, § 2.5.2 on page 21, § 2.5.2 on page 23, § 2.5.2 on page 24)

[7] Billingsley P. (1995). Probability and Measure. Wiley, New York. (Cited on § 3.2 on page 43, § 3.2 on page 46)

[8] Bingham N. H., Goldie C. M., Teugels J. L. (1987). Regular Variation. Cambridge University Press, Cambridge. (Cited on § 1.2.1 on page 3, § 2.1.8 on page 12)

[9] Bojanic R., Seneta E. (1973). A Unified Theory of Regularly Varying Sequences. *Mathematische Zeitschrift.* **134**, 91-106. (Cited on § 1.2.1 on page 3, § 2.1.1 on page 13)

[10] Coles S. G., Walshaw D. (1994). Directional modelling of extreme wind speeds. *Applied Statistics.* **43**, 139-157. (Cited on § 1.1 on page 2)

[11] Csorgo, S., Mason, D. M. (1986). The asymptotic distribution of sums of extreme values from a regularly varying distribution. Ann. Probab. 11, 974983. MR0841597. (Cited on § 4.3.2 on page 84)

[12] Davison C. A., Smith R. L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society.* **52 (3)**, 393 - 442. (Cited on § 1.1 on page 2)

[13] Drees, H. (2002). Tail empirical processes under mixing conditions. *Empirical process techniques for dependent data*, 325–342. Birkhäuser, Boston, MA. (Cited on § 3.4.2 on page 61)

[14] De Haan L., Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer, New York. (Cited on § 1.2.2 on page 4, § 2.2.1 on page 14, § 3.4.1 on page 59, § 3.4.2 on page 60, § 3.4.2 on page 61, § 4.3.2 on page 83)

[15] De Haan L. (1970). On regular variation and its applications to the weak convergence of sample extremes. *Mathematisch Centrum. Tract 32* Mathematisch Centrum, Amsterdam. (Cited on § 1.1 on page 2)

[16] Drees H. (1998). Optimal rates of convergence for estimates of the extreme value index. *The Annals of Statistics.* **26 (1)**, 434 - 448. (Cited on § 1.2.1 on page 3, § 2.4 on page 18, § 3.4.1 on page 58)

[17] Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag. (Cited on § 1.1 on page 1)

[18] Efron B. (1979). Bootstrap method : Another look at the Jackknife. *The Annals of Statistics.* **7 (1)**, 1-26. (Cited on § 1.2.5 on page 6, § 5 on page 85)

[19] Fabian Z., Stehlik M. (2009). On robust and distribution sensitive Hill like method. *IFAS Research Paper Series.* **43 (4)**, 1 - 9. (Cited on § 1.2.1 on page 3, § 1.2.2 on page 4, § 3.3.6 on page 55)

[20] Feller W. (1971). An Introduction to Probability Theory and Its Applications. Vol. 2. Wiley, New York. (Cited on § 1.2.1 on page 3, § 2.1 on page 8)

[21] Fisher, R., Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society.* **24(2)**, 180-190. (Cited on § 1.1 on page 1)

[22] Fréchet M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société polonaise de mathématiques.* **6 (3)**, 93 - 122. (Cited on § 1.1 on page 1)

[23] Gnedenko B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics, Second Series.* **44 (3)**, 423 - 453. (Cited on § 1.1 on page 1)

[24] Gut A. (2013). Probability: A Graduate Course. Second Edition Springer, New York. (Cited on § A.0.2 on page 148)

[25] Henry III, J.B. (2009). A harmonic moment tail index estimator. *Journal of Statistical Theory and Applications.* **8 (2)**, 141 - 162. (Cited on § 1.2.1 on page 3, § 3.3.2 on page 53)

[26] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics.* **3 (5)**, 1163 - 1174. (Cited on § 1.2.1 on page 3, § 3.3 on page 50)

[27] Hua, L., Joe. H. (2011). Second order regular variation and conditional tail expectation of multiple risks. *Insurance Math. Econom.* **49 (3)**, 537 - 546. (Cited on § 1.2.1 on page 3)

[28] Jenkinson A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *The Quarterly Journal of the Royal Meteorological Society* **81**, 158 - 171. (Cited on § 1.1 on page 2)

[29] Klajnmic H. (2003). Niveaux de retour pour les vitesses extrêmes des vents. XXXV Journés de Statistique, Lyon. (Cited on § 1.1 on page 1)

[30] Kulik R., Soulier P. (2011). The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications.* **121**, 109-134. (Cited on § 1.2.1 on page 3, § 2.4.2 on page 19, § 3.1.2 on page 36, § 3.3.1 on page 52)

[31] McLeish D. L. (1978). An extended martingale invariance principle. *The Annals of Probability.* (Cited on § 5.2.2 on page 90)

[32] McNeil, A. J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance.* **7**, 271-300. (Cited on § 1.1 on page 1)

[33] Mikosch T. (1999). Regular Variation, Subexponentiality and Their Applications in Probability Theory. *Euradom report. Eindhoven University of Technology.* **99**, issue 13. (Cited on § 1.2.1 on page 3, § 2.1.4 on page 9, § 2.1.5 on page 10, § 2.1.7 on page 11)

[34] Móri, T. F. (2009). Sharp inequalities between centered moments. *Journal of inequalities in pure and applied mathematics.* **10 (4)**, p.2. (Cited on § A.0.3 on page 148)

[35] Pickands J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics.* **3 (1)**, 119-131. (Cited on § 3.3 on page 50)

[36] Potter H. S. A. (1942). The mean value of certain Dirichlet series, II. *Proceeding of London Mathematical Society.* **s2 (47)**, 1-19. (Cited on § 2.1 on page 12)

[37] Reiss, R. D., Thomas, M. S. (2007). Statistical analysis of extreme values. From insurance, finance, hydrology and other fields. Birkhauser Verlag, Basel. (Cited on § 1.1 on page 1)

[38] Resnick, S. I. (2007). Heavy-Tail Phenomena. Probabilistic and Statistical Modeling. Springer, New York. (Cited on § 1.2.1 on page 3, § 2.1 on page 7, § 2.1.9 on page 12, § 2.2 on page 14)

[39] Rootzén H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications.* **119 (2)**, 468-490. (Cited on § 1.2.1 on page 3, § 1.2.2 on page 4, § 3 on page 28, § 3.1.1 on page 30, § 3.4 on page 58)

[40] Rootzén H., Tajvidi N. (1997). Extreme value statistics and wind storm losses a case study. *Scand. Actuar. J.* **1**, 70-94. (Cited on § 1.1 on page 1)

[41] Von Mises, R. (1954). La distribution de la plus grande de n valeurs. *Amer. Math. Soc.* **2**, 271 - 294. (Cited on § 1.1 on page 1)

[42] Whitt W. (1980). Some Useful Functions for Functional Limit Theorems. *Mathematics of Operations Research.* **5 (1)**, 67 - 85. (Cited on § 1.2.1 on page 3, § 2.5.9 on page 25)

[43] Weissman I. (1978). Estimation of Parameters and Large Quantiles Based on the K Largest Observations. *Advances in Applied Probability.* **10 (2)**, 331 - 332. (Cited on § 1.2.5 on page 6, § 4.1 on page 68)