People 's Democratic Republic of Algeria

Ministry of Higher Education and Scientifc Research

Mohamed Khider University, Biskra

Faculty of Exact Sciences, Natural Sciences and Life Sciences

Department of Mathematics

LMD THIRD CYCLE

A Thesis Presented for the Degree of

DOCTOR in Mathematics (Statistics)

Presented by

Nour Elhouda Guesmia

On Risk Measures and their Estimation

Examination Committee Members :

Abdelhakim Necir	Professor	U. Biskra	President
Djamel Meraghni	Professor	U. Biskra	Supervisor
Fatah Benatia	Professor	U. Biskra	Examiner
Djabrane Yahia	Professor	U. Biskra	Examiner
Hacen Zelaci	MCA	U. El-Oued	Examiner

January 28, 2025

Acknowledgements

I would like to express my deepest gratitude to **Allah** for granting me the strength, knowledge, and ability to undertake this research study and for providing the opportunity to persevere and complete it to the best of my abilities.

I am extremely grateful to my supervisor, **Pr. Djamel Meraghni**, for his invaluable contributions and expert guidance throughout this research project, as well as his relentless support. I truly appreciate his patience, integrity, and nurturing mentorship, without which this work would not have been possible. I am genuinely thankful for the opportunity to learn from him.

I am profoundly grateful to **Pr. Abdelhakim Necir** for his invaluable time, effort, suggestions and providing crucial assistance in difficult situations.

I would like to extend my heartfelt thanks to **Dr. Louiza Soltane** for her invaluable contributions, insightful vision and constructive suggestions.

I would like to thank **the examination committee members** for dedicating their valuable time to reviewing this thesis and for their comments, criticisms and suggestions.

I especially thank **my family** for their unwavering encouragement and support, which has been instrumental in my journey. I also want to acknowledge the significant contributions of **my close friends**, who have been a constant source of inspiration and motivation.

Contents

List of Figures

List of Tables

Publications and presentations

Article

Estimating the conditional tail expectation of randomly right-censored heavy-tailed data. *Journal of Statistical Theory and Practice, 18(3), 30 (joint work with* Meraghni, D., & Soltane, L., 2024).

Nelson-Aalen kernel estimator of the tail index for right-censored Pareto-type data. *In preparation. (joint work with Necir . A* & Meraghni, D).

Divergence-based tail index estimation for randomly censored data. *In preparation. (joint work with Necir . A* & Meraghni, D).

Oral presentations

1. *The conditional tail expectation of a heavy-tailed distribution under random censoring. The first international conference on pure and applied mathematics on May 26-27, 2021, Ouargla, Algeria.*

2. *Construction of the conditional tail expectation estimator for incomplete heavy-tailed data. The national conference of mathematics and applications on December 11, 2021, Mila, Algeria.*

3. *Confidence bounds for the conditional tail expectation of randomly censored heavy-tailed data. The first edition of the national seminar on mathematics and computer science applied to sciences on May 11-12, 2022, Tamanrasset, Algeria.*

4. *Hill-type estimator of the shape parameter for a heavy-tailed distribution under random censoring: Application to the conditional tail expectation.* The *international online conference on applied mathematics on June 01-03, 2022, Fez, Morocco.*

5. *Weissman estimator of extreme quantiles for randomly censored data: Application to the Tail Value at Risk. The second international conference on innovative academic studies on January 28-31, 2023, Konya, Turkey.*

6. *Kaplan-Meier estimator under random censorship : Application to the Conditional Tail Expectation. The first international conference on trends in advanced research on March 4-7, 2023, Konya, Turkey.*

7. *Estimation of the mean for randomly censored heavy-tailed distribution. The first international conference on recent academic studies on May 2-4, 2023, Konya, Turkey.*

8. *On the asymptotic normality of the Tail Value at Risk. The fifth international conference on applied engineering and natural sciences on July 10-12, 2023, Konya, Turkey.*

Introduction

In light of global market volatility, there has been increased scrutiny towards conventional risk assessment methods like standard deviation, due to their limitations in effectively addressing rare but severe events. This criticism has spurred a quest for improved methodologies capable of handling these rare events with significant impacts. A common query arising from this context is: How extreme might adverse outcomes become if they were to materialize. The problem lies in how to model these rare phenomena and accurately estimate their effects. In such a situation it seemsessential to rely on a well founded methodology. Extreme Value Theory (EVT) offers the essential principles for statistically modeling such occurrences and calculating extreme risk measures. Risk measures serve as crucial instruments for investors and financial entities to make well-informed choices regarding investments and risk mitigation tactics. These measures are widely utilized, tailored for various objectives and audiences. The correct risk measures depend on purpose. This is why there are so many good and equally effective risk measures to assess and manage risks.

Risk measures are mathematically derived from the distribution function (df) F or its tail counterpart $\overline{F}$, where

$$F(x) = P(X \leq x) \quad \text{and} \quad \overline{F}(x) = 1 - F(x),\ x \in \mathbb{R}.$$

This formulation enables the quantification of risk by analyzing the behavior of the random variable (r.v) X relative to its distribution.

Graphical tools (boxplots, time series plots, QQ-plots,...) provide evidence that financial and actuarial datasets include extreme values which are due to market fluctuations. The existence of such extremes makes the probability distributions decay very slowly towards the x-axis. In other words, the tails of these datasets are of heavy type. The class of heavy-tailed distributions (also known as Pareto-type or Pareto-like distributions) offers very appropriate statistical models in situations of rare events that could have negative impacts. They have important practical applications in various fields such as insurance, finance, hydrology and telecommunications. The models of Pareto, Burr and Fréchet are perfect examples (of such distributions) that are commonly used in illustrative simulations. For full details on heavy-tailed distributions and their applications, we refer to the excellent textbooks of [Embrechts et al.(1997)], [Beirlant et al.(2004)], [de Haan and Ferreira(2006)], [Reiss and Thomas(2007)] and [Resnick(2007)].
In statistical analyses related to many real world applications the observations are not necessarily fully available. Data is often missing in research in economics, medical sciences, social sciences and other fields of study. In other words, the variable of interest is usually randomly censored or truncated. In this case, we talk of incomplete data whose treatment requires specific techniques as described in, for instance, [Klein and Moeshberger(2003)]. The difference between truncation and censoring lies in the nature of the incomplete data and the way they are handled in the analysis. In truncation, the exact value of the variable of interest is not known for the cases that fall beyond the boundary, no note is recorded when a value exceeds a bound, due to some restriction of the study. In censoring, the censored value is known, but the exact value of the variable of interest is not known for the cases that fall beyond the boundary, a note is recorded documenting which bound had been exceeded and the value of that bound. In this thesis, we are only interested in estimation in the case of random right censored data. Examples of sets of censored data with apparent heavy tails can be found in, for instance, [Einmahl et al.(2008)], [Gomes and Neves(2011)], [Reynkens et al.(2017)] and [Beirlant et al.(2018)]. Finally, it is worthmentioning that,

in the case of complete data, the modelling of distribution tails have got a great deal of interest in the last three decades, while the analysis of extreme values of randomly censored and/or truncated data is a relatively new research topic.
For complete datasets, many different estimation of risk measures has got a great deal of interest by several authors. For instance, [Jones and Zitikis(2003)] and [Jones and Zitikis(2007)] introduced empirical estimators for risk measures and related quantities and [Kaiser and Brazauskas(2006)] discussed confidence interval estimation of various risk measures. For large losses, evaluating and/or estimating risk measures become more crucial for insurance companies. Among the works that were done on this subject, we can cite the papers of [Necir et al.(2007)] and [Necir and Meraghni(2009)] who focused on the estimation of a particular insurance premium, namely the Proportional Hazard Premium (PHP). [Necir et al.(2010)], [Lala Bouali et al.(2021)] and [Goegebeur et al.(2022)] proposed asymptotically normal estimators for the Conditional Tail Expectation (CTE) in the case of Pareto-like distributions. The estimation of risk measures and insurance premiums for randomly right censored data will be the primary focus of Chapter 4.

In this thesis, we present well-known risk measures and their estimations for right censored data, specifically in the context of managing high-risk situations. The following is an outline of the five chapters that comprise this work:

- **Chapter 1** : Risk measures.

In the first section, we review the fundamental concepts and characteristics of risk measures. The second section covers the most common principles for premium calculation, highlighting some of their essential properties. In the third section, we introduce widely used risk measures, including Value at Risk (VaR), the CTE, Conditional Tail Moment (CTM), and Tail Value at Risk (TVaR). Finally, the last section presents distortion risk measures.

- **Chapter 2** : Extreme value analysis.

This Chapter presents important and useful results in the field of the extreme value analysis. We begin by reviewing essential concepts from elementary probability and statistics. Next, we explore the fundamental result of the extreme value distribution. Afterwards, we introduce an important class within EVT known as heavy-tailed distributions. Following this, we interest on the extreme value index and the most commonly used Hill estimator.

- **Chapter 3** : Incomplete data.

The third chapter provides an overview of incomplete data and reviews fundamental concepts in survival data analysis. In Section 1, we introduce the basic concepts of lifetime data. Section 2 discusses the characteristics of truncated and censored data, which can be further classified into three categories: right censoring (right truncation), left censoring (left truncation), and interval censoring (interval truncation). In Section 3, we present nonparametric estimators, specifically the Kaplan-Meier and Nelson-Aalen estimators, as this thesis focuses on data that are randomly right censored.

- **Chapter 4** : Estimation of large risk measures under censorship.

This Chapter offers a review of estimation of risk measures and insurance premiums in the context of random right censored heavy-tailed losses, highlighting methodologies tailored to these complex data structures. The starting point will be a reminder about the tail index estimators in censoring situation. Afterwards, we present estimates of currently available risk measures and insurance premiums, including: VaR, mean, PHP, CTM and our proposed estimator for the CTE. The estimation procedure for the CTE is evaluated through a simulation study and applied to two real datasets of insurance losses and Aids survival time.

- **Chapter 5** : Asymptotic distribution of the CTE estimator.

In the final chapter, we aim to establish the asymptotic distribution of the CTE estimator in the context of randomly right censored heavy-tailed data. We begin by presenting our main result: the asymptotic normality of the proposed estimator, accompanied by a comprehensive proof of this finding. Subsequently, we compile some results that are essential to our analysis in the Appendix.

Lastly, it is worth mentioning that the R statistical analysis software is employed for data manipulation tasks, including graphical representations and numerical calculations.

Chapter 1

Risk measures

Risk measures are statistical and mathematical tools that quantify risk by analyzing the distribution of potential outcomes, represented through a r.v. Their purpose is to evaluate and quantify potential losses associated with decisions, actions, or investments. Following this assessment, organizations can implement control measures to mitigate or eliminate identified risks, facilitating informed decision-making and effective risk management. Risk measures are essential for various fields, including finance, insurance, engineering, environmental sciences and project management. For thorough details on risks and their measures, we refer, for instance, to the books of [Denuit et al.(2005)], [Kaas et al.(2008)] and [Klugman et al.(2019)].

1.1 Basic concepts

Let X be a r.v defined over a probability space $(\Omega, \mathcal{F}, P)$, where:

- Ω represents the set of all possible scenarios,
- $\mathcal{F}$ is a tribe (or σ-algebra),
- P is a probability measure.

Definition 1.1.1 *A risk represents the possibility of the occurrence of an event or outcome.*

In simple terms, the risk X is a r.v that represents the likelihood of an event occurring at some point in the future, typically focusing on negative or undesirable consequences.

Remark 1.1.1 *In this thesis, the risk variable is treated as a positive r.v, which can represent either potential losses (in the context of insurance and finance) or survival time, where it reflects the risk of survival duration in medical or reliability studies.*

To better understand and interpret this risk, one relies on risk measures.

Definition 1.1.2 *A risk measure $\mathcal{R}$ is a function that assigns numerical values to the outcomes of a risk X.*

This description demonstrates that, when it exists, the statistical moments such as expectation, variance, standard deviation, skewness and kurtosis are risk measures. For instance, in the context of insurance, when measuring the level of risk using standard deviation, which is calculated as the square root of the variance, a higher variance shows a higher risk and a higher loss and a low variance shows a lower loss. While these measures are simple and easy to compute, they assume that losses are normally distributed, which may not always be the case in reality, especially in the presence of heavy-tailed or non-normal distributions.

1.1.1 Properties of risk measures

Theoreticians have described a number of properties that a risk measure may possess or lack, as discussed by [Artzner et al.(1997)]. Additionally, [Artzner et al.(1999)] proposed the properties that a reasonable risk measure must satisfy and developed the concept of coherent risk measurements. These desirable characteristics enhance the effectiveness of risk measures in evaluating and managing financial risk, making them essential tools for investors and financial institutions. By comprehensively understanding these properties, practitioners can make informed choices about which risk measures best align with their specific risk management objectives and requirements. Some desirable properties of risk measures include:

Coherent risk measure

A coherent risk measure is a function that satisfies certain properties and axioms related to risk assessment. A risk measure $\mathcal{R}$ is said to be coherent if, for any two positive risks X and Y and for any positive constant c, it satisfies the four fundamental properties:

Positive homogeneity : $\mathcal{R}(cX) = c\mathcal{R}(X)$; this implies that doubling the exposure to a particular risk necessitates a doubling of the risk measure.

Translation invariance : $\mathcal{R}(X+c) = \mathcal{R}(X) + c$; i.e., the risk measure for a given risk remains unchanged when a constant amount is added to the risk. This property ensures that the risk measure is not affected by shifting the origin of the risk.

Monotonicity : if $X \geq Y \Longrightarrow \mathcal{R}(X) \geq \mathcal{R}(Y)$; this property indicates that the risk measure increases as the risks increases. This implies that, if one risk is greater than another, then its risk measure will also be greater.

Sub-additivity : $\mathcal{R}(X+Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$; which means that, the risk of combining two portfolios (such as, company shares, investments and premiums) cannot be greater than the risk of combining the two risks separately.

This property supports the idea that diversification reduces risk, as it allows for the combination of different portfolios to mitigate potential losses. For example, if two portfolios have risks of 11% and 7%, subadditivity guarantees that the risk of holding both portfolios together will not exceed 18%. This encourages investors to diversify their portfolios, thereby potentially reducing their exposure to extreme losses.

Monetary and convex risk measure

Convex risk measures are an important class of risk measures in financial and actuarial contexts. They extend the concept of coherent risk measures by relaxing some of the coherence properties, particularly focusing on convexity. [Föllmer and Schied(2002)] and [Chen and Hu(2018)] presented and studied convex risk measurement concepts.

Definition 1.1.3 (Monetary risk measure) *A risk measure $\mathcal{R}$ is said to be monetary if it is monotonicity and translation invariance.*

Definition 1.1.4 (Convex risk measure) *A convex risk measure is a function $\mathcal{R}$ that, for any X and Y, satisfies the following properties:*

1. *Convexity:* $\mathcal{R}\left(\lambda X+(1-\lambda) Y\right) \leq \lambda \mathcal{R}\left(X\right)+(1-\lambda) \mathcal{R}\left(Y\right), \forall \lambda \in [0,1]$,
2. *Monetary.*

Remark 1.1.2

- *All coherent risk measures are convex, but not all convex measures are coherent.*
- *A convex risk measure is coherent if and only if it is additionally positive homogeneous (see [Chen and Hu(2018)]).*

Additive comonotone and Law-invariant risk measure

Definition 1.1.5 *A risk measure $\mathcal{R}$ is additive comonotonic, if*

$$\mathcal{R}\left(Y_1+Y_2\right)=\mathcal{R}\left(Y_1\right)+\mathcal{R}\left(Y_2\right),$$

for any comonotonic vector $\left(Y_1, Y_2\right)$.

Definition 1.1.6 *A functional $\mathcal{R}$ is called a law-invariant risk measure, if for all risks X and Y, we have*

$$X \stackrel{d}{=} Y \Longrightarrow \mathcal{R}\left(X\right)=\mathcal{R}\left(Y\right).$$

1.2 Premium calculation principles

In insurance, a premium is a positive real number that represents the amount of money that an individual or business pays to an insurance company for insurance policy (coverage against potential risks). The premium is calculated based on various factors such as the type of coverage, the amount of coverage needed, the individual's age, personal information, and location. In actuarial science, the development of premium principles was the first use of risk measures. In order to determine the amount a policyholder should pay for insurance coverage. On the other hand, these principles are functions that assign a real number to the risk, which represents the premium. For more details, we refer, for instance, to [Bühlmann(2007)], [Kaas et al.(2008)], [Montserrat(2014)] and [Dickson(2016)]. In this section, the principles of premium calculation are discussed from a mathematical viewpoint, and various desirable properties for premium principles are considered. There are several premium principles used in insurance pricing, and the specific formula for each one varies. These principles are based on mathematical and statistical concepts and are used to ensure that the premiums charged are fair and adequate to cover the associated risks. Next, we will denote by $\Pi(X)$ the premium that an insurer charges to cover a risk X. When we refer to a risk X, what we mean is that claims from this risk are distributed as the r.v X. Below, we provide the most common premium principles in insurance world.

1.2.1 Some premium principles

Net premium: It is also known as the mean value principle, which is the simplest premium principle and is equal to the expectation of the claim (risk) size variable. In this principle, the premium rate is set equal to the expected value of the risk. The mean value principle can be represented as:

$$\Pi(X) = \mathbf{E}[X],$$

where $\mathbf{E}[X]$ is the expected value of the risk. This principle is based on the idea that the insurer expects to make a profit from the premiums paid by the policyholders, and the premium rate should be set in such a way that it covers the expected costs of the insurer.

Expected value principle: The risk loading can be imposed by the insurer on the net risk premium. This principle with safety is defined by

$$\Pi(X) = (1+\alpha)\,\mathbf{E}[X], \qquad \alpha \geq 0,$$

where α is the loading coefficient (or the surcharge factor). The loading coefficient accounts for various factors such as risk assessment, administrative costs, and profit margin.

Variance principle: This principle takes into account the variance of the claim size variable and is used to calculate the premium based on the variability of the potential losses. Its formula is given by

$$\Pi(X) = \mathbf{E}[X] + \alpha Var(X), \qquad \alpha \geq 0,$$

where $Var(X)$ is the variance of the risk X. This principle incorporates the safety surcharge factor α in order to face random deviations of the r.v losses or loss rates. We may again refer α as the relative security loading which is, in this case, proportional to the variance.

Exponential principle: This premium principle is calculated using exponential utility, it is defined by

$$\Pi(X) = \frac{1}{\alpha}\log\left(\mathbf{E}\left[\exp(\alpha X)\right]\right), \ \alpha > 0,$$

where α is the risk aversion parameter, and $\mathbf{E}[\exp(\alpha X)]$ is the expected value of the claim size variable under the exponential utility function.

Esscher principle: It is calculated using the Esscher transform, this principle is calculated as follows:

$$\Pi(X) = \frac{\mathbf{E}[X \exp(\alpha X)]}{\mathbf{E}[\exp(\alpha X)]}, \ \alpha > 0,$$

where α is the risk aversion parameter. Esscher premium is a principle that used to adjust the premium for the risk aversion of the insured.

Wang distortion functional principle: This principle is based on distortion risk measures and calculates the premium as the expected value of the loss distribution transformed by a distortion function g, which is defined in 1.4.1. This principle can be expressed as follows:

$$\Pi(X) = \int_0^{+\infty} g\left(\overline{F}(x)\right) dx.$$

Proportional hazards premium principle: It is grounded in the proportional hazards model, denoted as PHP, a widely used framework in survival analysis. This model examines how multiple variables influence the time until a specified event occurs. This principle is defined as

$$\Pi(X) = \int_0^{+\infty} \left(\overline{F}(x)\right)^{1/\varrho} dx,$$

where $\varrho \geq 1$ represents the distortion parameter or the risk aversion index. For $\varrho = 1$, $\Pi(X) = \int_0^{\infty} \overline{F}(x)\, dx = \mathbf{E}[X]$. In insurance pricing, the hazard rate is transformed into a premium by the use of the PHP principle (see, [Wang(1995)]).

Each principle has its own properties and applications, and the choice of principle may depend on the specific characteristics of the risks being insured. For more details of the properties of premium principles, see [Reich(1986)].

1.2.2 Properties of premium principles

There are several desirable properties for premium calculation principles. In the following we include most of its basic properties. Since the premium principle is a particular case of a risk measure, it shares many desirable characteristics. Therefore, we briefly mention the basic properties of premium calculation principles include:

• **Coherence:** It is the most important property, and it requires that the premium calculation principle satisfies the following four conditions:

1− **Sub-additivity:** For any two risks X and Y the premium for the sum of X and Y is less than or equal to the sum of the premiums for X and Y separately.

2− **Positive homogeneity:** The premium for a risk is non-negative.

3−**Translation invariance:** The premium for a risk remains unchanged when the location of the risk is shifted.

4− **Monotonicity:** It means that the premium increases as the risk increases.

We refer to [Montserrat(2014)], which analyzed the four properties necessary to meet the coherency criterion of the net premium principle, expected value principle, variance principle, exponential premium principle, Esscher premium principle and Wang distortion functional principle.

• **Convexity:** It is another important property, which ensures that the premium calculation principle is stable.

• **Law invariance:** if premiums for two risks with the same distribution are equal.

• **Additivity:** if the premium assigned to the sum of two independent risks is the sum of the premiums that are assigned to each risk separately. [Gerber(1974)] introduced the concept of an additive premium calculation principle.

• **Comonotonicity:** It requires that the premium for a portfolio of risks is equal to the sum of the premiums for each individual risk.

There are many other desirable properties for premium principles, these properties are some of the most basic and important ones. However, not all principles satisfy all properties, and insurers may choose the most suitable principle based on their specific needs and objectives.

Remark 1.2.1 *Premium calculation principles and risk measures are related but distinct concepts. While premium principles focus on determining the appropriate charges for insurance coverage based on assessed risks, while risk measures quantify the uncertainty as-*

sociated with potential outcomes, such as losses from insurance claims. Both are essential for effective financial management in insurance.

1.3 Usual risk measures

The usual risk measures have been developed to assess risk in various contexts, particularly in actuarial science and finance. These measures are designed to capture different risk characteristics and align with specific applications. However, a key motivation behind their development is the need to address extreme losses that traditional measures fail to capture effectively.

In this section, we provide the most commonly used risk measures, which serve as foundational tools in describing and evaluating risk, particularly in the presence of heavy-tailed distributions and rare but significant events.

1.3.1 Value at Risk

The VaR is a commonly used risk management tool, particularly in the banking and financial sectors. The evolution of VaR as a risk management tool took place over several decades, with its modern usage and recognition emerging in the early 1990s. The VaR is popular in the financial industry due to its simplicity and widespread use.

Definition 1.3.1 *The VaR of a r.v X, for a security level $t \in (0,1)$, is the t-order quantile of the distribution of X. The VaR is defined as*

$$VaR\left(t\right) = Q\left(t\right) = F^{\leftarrow}(t) := \inf\left\{x : F(x) \geq t\right\}.$$

The quantile function corresponding to F is also called the generalized inverse of F.

Remark 1.3.1 *It is important to note that if the df F is strictly increasing (monotonic) and continuous, it becomes a bijective function. In this scenario, we have $F^{-1}\left(t\right) = F^{\leftarrow}\left(t\right)$.*

The VaR represents the maximum potential loss (worst anticipated loss) over a specified time period at a given security level. In other words, the VaR provides a threshold for the worst-case scenario. In the context of risk management, security levels help risk managers determine the reliability of their risk assessments. A high security level indicates that the risk analysis is likely to produce accurate and consistent results (i.e., a greater level of certainty in the risk assessment or prediction), while a low level suggests that the analysis may be less reliable. Commonly, security levels are set at 90%, 95%, or 99%.

Example 1.3.1 *A VaR of one million US$ for one day at the 95% security level implies that there is a 5% chance of experiencing a loss greater than one million US$ over the specified time period.*

Remark 1.3.2 *The VaR is incoherence measure because it does not satisfy the sub-additivity (see [Artzner et al.(1999)]).*

An example showing the incoherence of the VaR can be found in [Klugman et al.(2019)], page 44.
To solve the problem of incoherence VaR, the CTE is introduced as an average of the risk values exceeding the VaR threshold.

1.3.2 Conditional Tail Expectation

The CTE, also known as Expected Shortfall (ES), provides a more comprehensive picture of the risk by considering the magnitude of extreme losses. It has gained popularity among financial institutions and regulators, especially after the 2008 financial crisis highlighted the shortcomings of VaR. Unlike VaR, which focuses on the question "how bad can things get?", CTE measures the average expected outcome (such as a loss or detriment) if things do get bad, offering a more thorough understanding of worst-case scenarios (see [Tasche(2002)]).

Definition 1.3.2 *The CTE represents the mean of the risk values which exceeds a given VaR. In other words, if* $\mathbf{E}|X| < \infty$, *we have*

$$CTE(t) := \mathbf{E}\left[X \mid X > VaR\left(t\right)\right], \qquad 0 < t < 1. \tag{1.1}$$

Remark 1.3.3

1. *The CTE provides information on the tail of distribution of* X *beyond the VaR. It is known as the average of extreme risks.*

2. *By definition, we have, for every security level* $t \in (0,1)$, $CTE(t) \geq VaR(t)$, *for any risk* X.

3. *The CTE is a coherent risk measure only when the underlying distribution of* X *is continuous (see, [Acerbi and Tasche(2002)]).*

In addition to the CTE, we have the CTM, which extends the analysis by incorporating higher-order moments of the upper tail, offering a deeper understanding of extreme data behavior.

1.3.3 Conditional Tail Moment

Introduced by [Methni et al.(2014)], the CTM is a risk measure that provides a tool through which a wide range of risk measures can be written as functions of it.

Definition 1.3.3 *The CTM represents the moment, of order* $\zeta > 0$, *of the loss distribution above the VaR at level* $t \in (0,1)$. *Specifically, it is defined as follows:*

$$CTM_{\zeta}(t) := \mathbf{E}\left[X^{\zeta} \mid X > VaR\left(t\right)\right]. \tag{1.2}$$

Remark 1.3.4 *Note that, for* $\zeta = 1$, *the CTM coincides with the CTE.*

Other existing risk measures include:

• **The Conditional Tail Variance (CTV):** measures the conditional variability of X given $X > VaR(t)$ and indicates how far away the events deviate from $CTE(t)$. The formula of the CTV, for a level $t \in (0,1)$, is given by

$$CTV(t) := \mathbf{E}\left[(X - CTE(t))^2 \mid X > VaR(t)\right].$$

• **The Conditional Tail Skewness (CTS):** is defined as

$$CTS(t) := \frac{\mathbf{E}\left[X^3 \mid X > VaR(t)\right]}{(CTV(t))^{3/2}}.$$

For more details on these measures, we refer to [Valdez(2005)], [Cai and Tan(2007)] and [Hong and Elshahat(2010)].

1.3.4 Tail Value at Risk

The TVaR is a key risk measure used to assess potential extreme adverse outcomes. If the VaR represents the maximum loss when an event of a given probability occur, then the TVaR represents the expected value of the additional potential losses beyond that threshold.

Definition 1.3.4 *The TVaR, for a security level $t \in (0,1)$ associated with risk X, is defined as follows:*

$$TVaR(t) := \frac{1}{1-t}\int_t^1 VaR(s)\,ds.$$

In medical research, the TVaR can be utilized to measure the average expected survival time of patients who survive beyond a specific time frame. For example, if a study indicates that patients receiving a certain treatment achieve the highest survival time, at the level 95%, of 5 years (i.e., $VaR(0.95) = 5$ years), the TVaR can provide insights into the average survival time for those who exceed this 5 year period. This information helps clinicians better understand the long-term outcomes and prognosis for this patient population.

Remark 1.3.5 *The TVaR and the CTE are the same if the distribution of the risk X is continuous.*

1.4 Distortion risk measures

Distortion risk measures are a vital category of risk assessment tools in the insurance and finance sectors, designed to evaluate and manage risk by transforming the probability distribution of potential losses. These measures adjust the underlying probability distribution to emphasize higher-risk outcomes, thus providing a more nuanced view of risk than traditional measures (see, for instance, [Amarante et al.(2023)]).

Definition 1.4.1 (Distortion function) *A function $g : [0,1] \to [0,1]$ is a distortion function if it is non-decreasing with $g(0) = 0$ and $g(1) = 1$.*

Definition 1.4.2 *We define the distortion risk measure as follows:*

$$\mathcal{R}_g(X) = \int_0^1 F^{\leftarrow}(1-s)\, dg(s).$$

Another common representation of distortion risk measures can be given as:

$$\mathcal{R}_g(X) = \int_0^{+\infty} g\left(\overline{F}(x)\right) dx - \int_{-\infty}^{0} \left[1 - g\left(\overline{F}(x)\right)\right] dx.$$

The main distortion risk measures are summarized in Table 1.1, where v is a constant that ranges from 0 to 1.

Risk measure	**Distribution function**
VaR	$g(x) = \mathbb{I}[x \geq v]$
TVaR	$g(x) = \min\{x/v, 1\}$
Proportional Hazard	$g(x) = x^v$
Dual Power	$g(x) = 1 - (1-x)^{1/v}$
Gini	$g(x) = (1+v)\,x - vx^2$
Transformation exponentielle	$g(x) = (1-v)\,/\,(1-v)$

Table 1.1: The main risk measures by distortion.

1.4.1 Wang risk measure

Wang's risk measure is a coherent risk measure that is defined by the Wang transform function, which is a distortion function. It is a specific application of distortion risk measures within the insurance context, using a particular distortion function. The coherence of this risk measure is a consequence of the concavity of the Wang transform function. The concept of a distortion function was proposed by [Wang(1996)] to create a family of risk measures known as distortion risk measures.

The distortion function in Wang's risk measure is mathematically expressed as:

$$g(x) = \Phi\left(\Phi^{-1}(x) + \omega\right),$$

where Φ represents the standard normal df, Φ^{-1} is the inverse standard normal df and $\omega \in \mathbb{R}$ is the risk aversion parameter.

This form of the distortion function reflects the adjustment of probability weights based on the normal distribution, ensuring a consistent and coherent risk evaluation framework.

1.4.2 Spectral measures

[Acerbi(2002)] introduced spectral risk measures, which generalize risk measures such as VaR and CTE. Unlike VaR, which assigns zero weight to losses beyond a certain quantile, and CTE, which applies a constant weight in the tail region, spectral risk measures utilize a risk aversion function to assign varying weights to potential losses. This allows for a more accurate representation of an individual's or institution's risk preferences by giving larger weights to higher losses.

Definition 1.4.3 *A spectral function (or risk aversion function) is defined as a function* $\phi : [0,1] \to \mathbb{R}_+$ *that is non-decreasing and satisfies the normalization condition* $\int_0^1 \phi(x)\, dx = 1$.

Definition 1.4.4 *The spectral risk measure for a r.v* X *is expressed as:*

$$\mathcal{R}_\phi(X) = \int_0^1 F^{\leftarrow}(x)\, \phi(x)\, dx.$$

This formulation allows for a weighted average of potential losses, where worse outcomes are typically assigned greater weights, reflecting the risk aversion of the decision-maker.

Remark 1.4.1

- *The spectral measurements are coherent.*
- *Spectral measures can be characterized as the measurements of the concave distortion function* g, *where* $\phi = g'$.
- *Wang measure is a spectral measure if and only if* $\omega \geq 0$.

Chapter 2

Extreme value analysis

Extreme value analysis, also known as EVT, is a statistical methodology used to estimate the probability of events that are rarest compared to any previously observed. EVT provides a mathematical and probabilistic foundation on which it is possible to build statistical models to predict the size and frequency of these rare phenomena. For full details, we refer to the excellent textbooks of [Embrechts et al.(1997)], [Reiss and Thomas(2007)], [Beirlant et al.(2004)] and [de Haan and Ferreira(2006)].

2.1 Basic concepts

Let X be a continuous r.v defined on a probability space $(\Omega, \mathcal{F}, P)$.

Definition 2.1.1 (Tail quantile functions) *A function denoted by $\mathbb{U}$, known as the tail quantile function, is defined as follows:*

$$\mathbb{U}(z) = Q(1 - 1/z) = \left(1/\overline{F}\right)^{\leftarrow}(z), \qquad 1 < z < \infty.$$

Definition 2.1.2 (Empirical df) *Let $X_1, X_2, ..., X_n$ be a sample of size $n \geq 1$ from a r.v X. The empirical df of the sample $(X_1, X_2, ..., X_n)$ is defined as follows:*

$$F_n(x) := \frac{\textit{number of elements in the sample} \leq x}{n} \qquad (2.1)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{X_j \leq x\right\}, \qquad \forall x \geq 0,$$

where $\mathbf{1}\{B\}$ *is the indicator function of the event* B.

Remark 2.1.1 *The empirical tail function of the sample* $(X_1, X_2, ..., X_n)$ *is defined by*

$$\overline{F}_n(x) = 1 - F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{X_j > x\right\}, \qquad \forall x \geq 0.$$

This estimator possesses several important convergence properties, including: Glivenko-Cantelli for almost sure uniform convergence, asymptotic normality,...

Definition 2.1.3 *The empirical quantile function of a sample* $(X_1, X_2, ..., X_n)$ *is defined as follows:*

$$Q_n(s) = F_n^{\leftarrow}(s) := \inf\left\{x \in \mathbb{R} : \ F_n(x) \geq s\right\}, \ 0 < s < 1.$$

The empirical tail quantile function is defined as:

$$\mathbb{U}_n(z) = Q_n(1 - 1/z), \qquad 1 < z < \infty.$$

Definition 2.1.4 (Arithmetic sum and mean) *Let* $(X_1, X_2, ..., X_n)$ *be a sample of the r.v* X. *For any integer* $n \geq 1$, *the arithmetic sum and the empirical mean are respectively defined as follows:*

$$S_n := \sum_{i=1}^{n} X_i \qquad \textit{and} \qquad \overline{X}_n := S_n/n.$$

Theorem 2.1.1 (Laws of large numbers) *Let* $(X_1, X_2, ..., X_n)$ *be a sample from a r.v* X, *with a finite expected value* $(i.e., \mathbf{E}|X| < \infty)$. *Then we have*

$$\textit{weak law :} \quad \overline{X}_n \xrightarrow{P} \mu, \quad \textit{as} \quad n \to \infty,$$
$$\textit{strong law :} \quad \overline{X}_n \xrightarrow{a.s} \mu, \quad \textit{as} \quad n \to \infty,$$

where $\mu := \mathbf{E}[X]$.

Theorem 2.1.2 (Central limit theorem) *If* $(X_1, X_2, ..., X_n)$ *are independent and identically distributed (iid) random variables (r.v's) with expected value* $\mathbf{E}[X_i] = \mu$ *and finite variance* $Var(X_i) = \sigma^2$, *then*

$$\frac{1}{\sqrt{n}}\left(\frac{S_n - n\mu}{\sigma}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \qquad \text{as } n \to \infty.$$

The proof of this Theorem could be found, for instance, in [Saporta(1990)], page 66.

2.1.1 Order statistics

Since order statistics provide valuable insights into the tail distribution (specifically the right tail), they have gained increasing importance in EVT. For further information, see [David and Nagaraja(2004)].

Definition 2.1.5 *Let* $X_1, X_2, ..., X_n$ *be a sample of size* $n \geq 1$ *from a r.v* X. *The series arranged in increasing orders of the r.v's* $X_1, X_2, ..., X_n$ *is referred to as the order statistics of this sample. They are generally denoted by* $X_{1:n}, X_{2:n}, ..., X_{n:n}$.

For $r = 1, 2, ..., n$, the r.v $X_{r:n}$ is called the rank order statistic of rank r or r-th order statistics. The r.v's $X_{1:n}$ and $X_{n:n}$ represent the smallest and largest observations, respectively, where

$$X_{1:n} := \min(X_1, X_2, ..., X_n) \qquad \text{and} \qquad X_{n:n} := \max(X_1, X_2, ..., X_n),$$

Order statistics distribution

Proposition 2.1.1 (Distribution of $X_{1:n}$ and $X_{n:n}$) *The probability distribution of the r.v's* $X_{1:n}$ *and* $X_{n:n}$ *are given by their respective distribution functions (df's)*

$$F_{X_{1:n}}(x) = 1 - [\overline{F}(x)]^n \qquad \text{and} \qquad F_{X_{n:n}}(x) = [F(x)]^n, \ x \in \mathbb{R}.$$

Proof 2.1.1 *We have that*

$$F_{X_{1:n}}(x) = P(X_{1:n} \leq x) = 1 - P(X_{1:n} > x).$$

It is clear that the event $(X_{1:n} > x)$ *is equivalent to* $(X_1 > x, X_2 > x, ..., X_n > x)$; *thus,*

$$F_{X_{1:n}}(x) = 1 - P(X_1 > x, X_2 > x, ..., X_n > x),$$

which by the independence of the r.v's X_i, *becomes*

$$F_{X_{1:n}}(x) = 1 - P(X_1 > x).P(X_2 > x)...P(X_n > x).$$

Finally, we reach a conclusion based on the equidistribution of observations, leading to the following expression:

$$F_{X_{1:n}}(x) = 1 - [P(X_1 > x)]^n = 1 - [\overline{F}\,(x)]^n.$$

The corresponding result for the maximum can also be derived using the same principle.□

Proposition 2.1.2 (Distribution of $X_{r:n}$) *In general, for* $1 \leq r \leq n$, *the df of the* r*-th order statistic is given by:*

$$F_{X_{r:n}}(x) = \sum_{i=r}^{n} \frac{n!}{i!(n-i)!} F^i(x)[1 - F(x)]^{n-i}, \; x \in \mathbb{R}.$$

Proof 2.1.2 *Let* $x \in \mathbb{R}$ *be fixed. To say that the event* $(X_{r:n} \leq x)$ *occurs is equivalent to stating that among the variables* $X_1, X_2, ..., X_n$, *at least* r *of them are smaller than* x. *In other words, we have*

$$X_{r:n} \leq x \Longleftrightarrow \sum_{j=1}^{n} \mathbf{1}\left\{X_j \leq x\right\} \geq r,$$

which implies

$$P(X_{r:n} \leq x) = P\left(\sum_{j=1}^{n} \mathbf{1}\{X_j \leq x\} \geq r\right).$$

On the other hand, we have that $\sum_{j=1}^{n} \mathbf{1}\{X_j \leq x\}$ *follows a binomial distribution with parameters* n *and* $F(x)$*. Thus we have, for* $i = 1, 2, ..., n$,

$$P\left(\sum_{j=1}^{n} \mathbf{1}\{X_j \leq x\} = i\right) = \frac{n!}{i!(n-i)!} F^i(x)[1 - F(x)]^{n-i}.$$

The result is obtained by taking the sum from $i = r$ *to* n. □

Remark 2.1.2 *The representation of the empirical df* F_n *in terms of order statistics is given by:*

$$F_n(x) = \begin{cases} 0 & \textit{if} \quad x < X_{1:n}, \\ \dfrac{i}{n} & \textit{if} \quad X_{i:n} \leq x < X_{i+1:n}, \quad \textit{for} \;\; i = 1, 2, ..., n-1, \\ 1 & \textit{if} \quad x \geq X_{n:n}. \end{cases} \tag{2.2}$$

Equation (2.2) implies that

$$Q_n(s) = X_{i:n}, \qquad \textit{for} \qquad \frac{i-1}{n} < s \leq \frac{i}{n}, \;\; i = 1, 2, ..., n.$$

Note that, for $0 \leq z \leq 1$, $Q_n(z) = X_{[nz:n]}$, *where* $[nz]$ *denotes the integer part of* nz.

2.2 Extreme value distribution

EVT is used to model the maxima (or minima) of r.v's in a manner analogous to how the central limit theorem is employed to model the sum of r.v's. The fundamental concept of EVT is that when the distribution of the maximum is appropriately normalized, the limiting distribution can only be one of three possible types: the Gumbel distribution, the Fréchet distribution, or the Weibull distribution. These distributions are widely used in various fields, including finance, environmental science, and engineering, to model extreme events and assess tail risks. In this section, we interest on the result of limiting distribution of a suitably normalized maximum.

2.2.1 Domains of attraction

Domains of attraction are groups of probability distributions that share similar tail behaviors. These domains of attraction help in understanding the behavior of extreme events in different types of data. The maximum domains of attraction is associated with the behavior of tail probabilities and the convergence of rescaled maxima to a limiting distribution. Here we introduce the notion of maximum domains of attraction.

Definition 2.2.1 (Maximum domain of attraction) *A df F, is said to be in the maximum domain of attraction of the extreme value distribution $\mathcal{H}_\gamma$, denoted as $F \in D(\mathcal{H}_\gamma)$, if there exist sequences of constants, $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} P\left(\frac{X_{n:n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \mathcal{H}_\gamma(x), \qquad \forall x \in \mathbb{R}, \tag{2.3}$$

where $\mathcal{H}_\gamma$ is a non-degenerate df and $\gamma \in \mathbb{R}$.

Remark 2.2.1 *The parameter γ is called the tail index, the extreme value index or the shape parameter.*

The distribution function $\mathcal{H}_\gamma$ is known as the Generalized Extreme Value (GEV) distribution. The GEV distribution is a family of continuous probability distributions that includes the three fundamental types of extreme value distributions as special cases. It is widely used to model the maxima of long (finite) sequences of r.v's and to assess tail risks in various applications.

A key finding in EVT, the Fisher-Tippett theorem, often called the Gnedenko theorem (see [Fisher and Tippett(1928)] and [Gnedenko(1943)]), characterizes the limiting distribution of the maximum. The theorem is essential in EVT because it provides a way to identify the limiting distribution of the maximum and to estimate the tail probabilities of the distribution. The Fisher-Tippett-Gnedenko theorem is summarized as follows :

Theorem 2.2.1 *Let $(X_1, X_2, ..., X_n)$ be a sequence of iid r.v's with df F. Assume that there exist two normalizing sequences $a_n > 0$ and $b_n \in \mathbb{R}$, and a non-degenerate df $\mathcal{H}$ satisfies (2.3) . Then the limiting distribution $\mathcal{H}$ belongs to one of the following three classe:*

$$\textit{Fréchet } (\gamma > 0) : \Phi_\gamma(x) = \begin{cases} \exp\left(-x^{-\gamma}\right) & x > 0, \\ 0 & x \leq 0. \end{cases}$$

$$\textit{Gumbel } (\gamma = 0) : \Lambda_\gamma(x) = \exp\left(-\exp\left(-x\right)\right), \qquad x \in \mathbb{R}.$$

$$\textit{Weibull } (\gamma < 0) : \Psi_\gamma(x) = \begin{cases} 1 & \text{if} \quad x > 0, \\ \exp\left(-\left(-x^{\gamma}\right)\right) & \text{if} \quad x \leq 0. \end{cases}$$

The sketch of the proof can be found, for instance, on page 122 of [Embrechts et al.(1997)]. Figure 2.1 illustrates the three forms of the limiting df's.

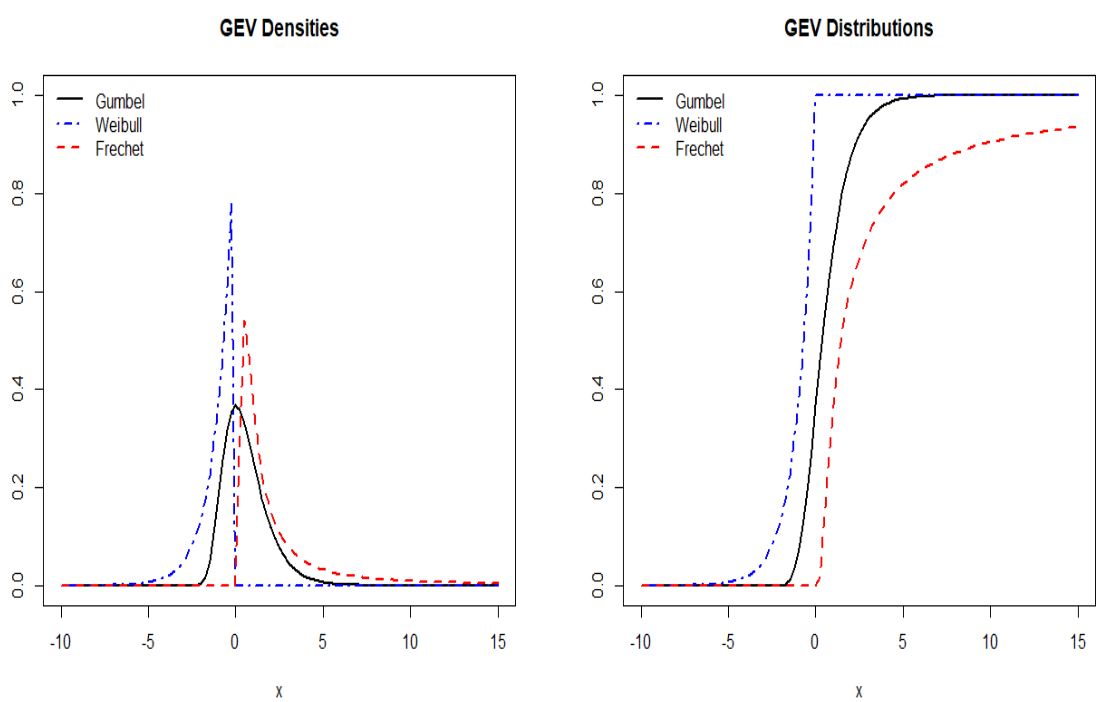

Figure 2.1: Density and distributions of extreme value distributions

To unify the three families into a single framework, [Jenkinson(1955)] and [Von Mises(1936)] developed the family of GEV distributions as follows:

Definition 2.2.2 *The GEV distribution combines the Gumbel, Fréchet and Weibull families, which is defined as*

$$\mathcal{H}_\gamma(x) := \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right) & \textit{if} \quad \gamma \neq 0, \\ \exp(-\exp(-x)) & \textit{if} \quad \gamma = 0, \end{cases}$$

for $\gamma \in \mathbb{R}$ and $1+\gamma x > 0$, where γ is the shape parameter that governs the tail behavior of $\mathcal{H}_\gamma(x)$.

We present the three extreme value distributions in terms of the GEV distribution $\mathcal{H}_\gamma$ as follows :

$$\begin{aligned} \Phi_\gamma(x) &= \mathcal{H}_{1/\gamma}(\gamma(x-1)), & x &> 0, \\ \Lambda_\gamma(x) &= \mathcal{H}_0(x), & x &\in \mathbb{R}, \\ \Psi_\gamma(x) &= \mathcal{H}_{-1/\gamma}(\gamma(x+1)), & x &< 0. \end{aligned}$$

2.2.2 Limit distributions

Let x^* be the right (or upper) endpoint of the df F, where $x^* = \sup\{x \in \mathbb{R} : F(x) < 1\}$.
Since $F_{X_{n:n}}(x) = [F(x)]^n$, it follows that $X_{n:n} \xrightarrow{P} x^*$, as $n \to \infty$.
Depending on the sign of γ, the limiting distribution $\mathcal{H}$ of the normalized sample maximum can be characterized by:

1. Fréchet type: if $\gamma > 0$, the right endpoint of the distribution is infinity $(x^* = +\infty)$. This domain of attraction corresponds to heavy-tailed distributions, which include the Pareto, Burr, log-gamma, inverse gamma, Fréchet, and Cauchy distributions.

2. Gumbel type: if $\gamma = 0$, with the right endpoint x^* can be either finite or infinite. This type includes distributions such as Gaussian, log-normal, exponential, and gamma distributions. In this category, we encounter finitely-tailed distributions.

3. Weibull type: if $\gamma < 0$, the right endpoint of the distribution is finite with $x^* = -1/\gamma$. This domain of attraction corresponds to light-tailed distributions, including the Weibull, Beta (a, b) and uniform distributions.

2.3 Heavy-tailed distributions

In probability theory and statistics, heavy-tailed distributions are those whose tails are not exponentially bounded, meaning they have heavier tails than the exponential distribution. In other words, a heavy-tailed distribution decays more slowly toward zero than the exponential distribution, which results in a greater concentration of extreme values. Any distribution with heavier tails than the normal distribution is occasionally referred to as heavy-tailed (see [Bryson(1974)], [El Adlouni et al.(2007)] and [Cohen et al.(2020)]). In numerous applications, the right tail of a distribution is particularly significant, even when the distribution features a heavy left tail, a heavy right tail, or both. This focus arises from the fact that the right tail is often associated with rare and extreme events that can have significant consequences. In this thesis, we focus specifically on the right tail of the distribution.

Definition 2.3.1 *The distribution of a r.v X with df F is said to have a heavy right tail if the expected value of the exponential function is infinite, specifically:*

$$\mathbf{E}\left[e^{\lambda X}\right] = \int_{-\infty}^{+\infty} e^{\lambda x} dF(x) = \infty, \text{ for all } \lambda > 0.$$

Remark 2.3.1 *If $\mathbf{E}\left[e^{\lambda X}\right] < \infty$, the df F is said to be a light-tailed distribution.*

Such distributions have thinner tails than an exponential distribution and tend to zero more rapidly than the exponential.

Definition 2.3.2 *The kurtosis coefficient $\mathcal{K}$ is a statistical measure used to assess whether a dataset exhibits heavy tails compared to a normal distribution. It is defined as follows:*

$$\mathcal{K} := \mathbf{E}\left[\frac{(X-\mu)^4}{\sigma}\right] = \frac{\mu_4}{\mu_2^2},$$

where $\mu_n = \mathbf{E}\left[(X - \mathbf{E}[X])^n\right]$ represents the nth central moments.

For a distribution to be considered heavy-tailed, its kurtosis must be greater than that of the normal distribution (i.e., $\mathcal{K} > 3$).

Theorem 2.3.1 *For any df F, the following statements are equivalent:*

- *F is a heavy-tailed distribution.*
- *The tail distribution of F is heavy-tailed.*

Proof 2.3.1 *See Theorem* 2.6, *in [Foss et al.(2011)], page* 8.

A prominent subclass of heavy-tailed distributions is the class of regularly varying distributions. Their properties are extensively utilized in analyzing the behavior of estimators within the domain of extreme value analysis. For further details, refer to the works of [Bingham et al.(1987)] and [Mikosch(1999)].

2.3.1 Regularly varying functions

Definition 2.3.3 (Regularly varying function) *A measurable function $V : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be regularly varying (at infinity) with index $\varsigma \in \mathbb{R}$, denote $V \in RV(\varsigma)$, if*

$$\lim_{v\to\infty} \frac{V(vx)}{V(v)} = x^{\varsigma}, \quad \textit{for all } x > 0.$$

Here, ς is called the variation exponent or the index of regular variation.

Remark 2.3.2 (Slowly varying function) *If $\varsigma = 0$, a measurable function L is said to be slowly varying (at infinity) if*

$$\lim_{v\to\infty} \frac{L(vx)}{L(v)} = 1, \quad \textit{for any } x > 0.$$

Example 2.3.1

- *The functions x^{ς}, $x^{\varsigma}\log(1+x)$ and $(x\log(1+x))^{\varsigma}$ are regularly varying functions with index $\varsigma \in \mathbb{R}$.*
- *The functions $\log(x+1)$ and $\log(\log(1+x))$ are slowly varying functions.*
- *The functions $\sin(x+2)$, $\exp(x)$ and $\exp(\log(x+1))$ are not regularly varying functions.*

Next, we will present some important characteristics of regularly varying functions.

Proposition 2.3.1

- *A regularly varying function V with index $\varsigma \in \mathbb{R}$, can be expressed in the form $V(x) = x^{\varsigma}L(x)$.*
- *If V is a regularly varying function with index ς, then V^{-1} is also regular varying, but with index $(1/\varsigma)$.*
- *If V is a regularly varying function with index ς, then $V(1/x)$ is regular variation with index $(-\varsigma)$.*

Proposition 2.3.2 (Potter's inequalities) *Suppose that $V \in RV(\varsigma)$, where $\varsigma \in \mathbb{R}$. Then, for any small $\epsilon > 0$, there exists a value $v_0 = v_0(\epsilon)$ such that for $v > v_0$ and $x \geq 1$, we have*

$$(1-\epsilon)\, x^{\varsigma-\epsilon} \leq \frac{V(vx)}{V(v)} \leq (1+\epsilon)\, x^{\varsigma+\epsilon}. \tag{2.4}$$

Proof 2.3.2 *See [Resnick(2007)], Proposition* 2.6, *page* 32.

Proposition 2.3.3 (First order regular variation condition) *If F is a heavy-tailed df, then the following statements are equivalent:*

1. *$\overline{F}$ regularly varying at infinity with index* $(-1/\gamma)$

$$\lim_{v\to\infty}\frac{\overline{F}(vx)}{\overline{F}(v)} = x^{-1/\gamma}, \quad x > 0.$$

2. *$Q\,(1-s)$ regularly varying at zero with index* $(-\gamma)$

$$\lim_{s\to 0}\frac{Q(1-sx)}{Q(1-s)} = x^{-\gamma}, \quad x > 0.$$

3. *$\mathbb{U}$ regularly varying at infinity with index γ*

$$\lim_{v\to\infty}\frac{\mathbb{U}(vx)}{\mathbb{U}(v)} = x^{\gamma}, \quad x > 0.$$

To investigate the asymptotic normality of tail index estimators, a first order condition for regularly varying functions is often inadequate. Consequently, a second order condition is necessary, as it facilitates weak approximations of statistics derived from EVT.

Definition 2.3.4 (Second order regular variation condition) *If F satisfies one of the following equivalent conditions, we say that it is second order regularly varying at infinity:*

(i) There exists a parameter $\rho \leq 0$ and a function A, which tends to zero and does not change sign near infinity, such that for any $x > 0$, we have

$$\lim_{v\to\infty}\frac{\overline{F}(vx)/\overline{F}(v) - x^{-1/\gamma}}{A\,(v)} = x^{-1/\gamma}\frac{x^{\rho}-1}{\rho}.$$

(ii) *There exists a parameter* $\rho \leq 0$ *and a function* A^*, *which tends to zero and does not change sign near zero, such that for all* $x > 0$, *we have*

$$\lim_{s \to 0} \frac{Q(1-sx)/Q(1-s) - x^{-\gamma}}{A^*(s)} = x^{-\gamma} \frac{x^{\rho} - 1}{\rho}.$$

(iii) *There exist some parameter* $\rho \leq 0$ *and a function* A^{**}, *which tends to zero and does not change sign near zero, such that for any* $x > 0$, *we have*

$$\lim_{v \to \infty} \frac{\mathbb{U}(vx)/\mathbb{U}(v) - x^{\gamma}}{A^{**}(v)} = x^{\gamma_1} \frac{x^{\rho} - 1}{\rho}. \tag{2.5}$$

If $\rho = 0$, *then* $(x^{\rho} - 1)/\rho$ *is interpreted as* $\log x$. *The constant* ρ *serves as a second order parameter that controls the speed of convergence of the first order condition. The functions* A, A^* *and* A^{**} *are regularly varying, with the relationships:*

$$A(v) = A^{**}\left(1/\overline{F}(v)\right) \quad \textit{and} \quad A^*(v) = A^{**}(1/v).$$

They play a role in controlling the rate of convergence in first order regular variation conditions 1, 2 and 3, respectively.

The second order regular variation condition is satisfied by a subclass of heavy-tailed distributions known as the "Hall class", as introduced by [Hall(1982)]. This class is very important in the discussion of the estimators of a positive tail index.

2.3.2 Hall's class

Definition 2.3.5 *We say that* F *belongs to the set of Hall's models if it can be expressed in the following form:*

$$\overline{F}(x) = lx^{-1/\gamma}\left(1 + mx^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right), \ \textit{as } x \to \infty, \tag{2.6}$$

where $m \in \mathbb{R} \backslash \{0\}$ *and* $l > 0$.

This class includes some of the most common distributions, such as Fréchet, Student, Pareto-like distributions, Burr, GEV, generalized Pareto distributions, among others. We can reformulate (2.6) in terms of the functions Q and $\mathbb{U}$ as follows:

$$Q(1-s) = l^{\gamma} s^{-\gamma} \left(1 + \gamma l^{\rho} m s^{-\rho} + o\left(s^{-\rho}\right)\right), \text{ as } s \to 0,$$

and

$$\mathbb{U}(x) = l^{\gamma} x^{\gamma} \left(1 + \gamma m l^{\rho} x^{\rho} + o\left(x^{\rho}\right)\right), \text{ as } x \to \infty.$$

Note that in the Hall model, we have $A(x) = m\gamma\rho x^{\rho/\gamma}$, $A^{**}(x) = m\gamma\rho l^{\rho} x^{\rho}$ as $x \to \infty$, and $A^{*}(s) = m\gamma\rho l^{\rho} s^{-\rho}$, as $s \to 0$.

2.4 Extreme value index

One important parameter in EVT that is essential for describing the right tail behavior of a df and determining its decay rate is the extreme value index. It is employed to evaluate the probability of events that are more extreme than any previously observed. More specifically, a high tail index indicates a significant probability of extreme events. Estimating the tail index is crucial to comprehending and predicting rare and extreme events in various domains, such as finance, engineering, and environmental sciences. Numerous methods are available in the literature for estimating tail indices, among which Hill's estimator ([Hill(1975)]), the moment estimator ([Dekkers et al.(1989)]) and Pickand's estimator ([Pickands(1975)]). Each of these techniques offers unique advantages and can be applied depending on the specific characteristics of the data. Those estimators are based on the largest order statistics $X_{n-k:n} \leq ... \leq X_{n:n}$, where $k = k_n$ is an intermediate integer sequence that depends on n and satisfies:

$$1 < k < n, \; k \to \infty \quad \text{and} \quad k/n \to 0, \; \text{as } n \to \infty. \tag{2.7}$$

The statistic $X_{n-k:n}$ is then said to be intermediate order statistic. The estimator from Hill is the most widely used due to its attractive asymptotic properties and computational simplicity. The calculation of this estimator depends on the selection of the optimal number k of extreme values, which will be denoted by k_{opt}. This choice affects the accuracy of the estimation. Next, we will present the algorithm for determining this optimal number after defining the Hill estimator for the tail index γ.

2.4.1 Hill's estimator

The estimate of γ $(\gamma > 0)$ provided by B. Hill in [Hill(1975)] is the most well-known method for estimating a positive tail index. The Hill estimator of γ, denoted by $\widehat{\gamma}^{(Hill)}$, is determined by:

$$\widehat{\gamma}^{(Hill)} = \widehat{\gamma}^{(Hill)}(k) := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}}.$$

The Hill estimator is consistent for the tail index; however, it is important to note that obtaining accurate estimates requires using a sufficiently large value of k. However, a value of k that is too large can result in a biased estimate. Thus, selecting an optimal value of k is essential to obtain accurate estimates of the tail index. Another limitation of the Hill estimator is that it is restricted to estimating the positive extreme value index, which limits its applicability in certain contexts. Additionally, when γ moves beyond 0.5, Hill estimator may exhibit substancial bias. This influences the related estimations may fail to approach the theoretical value for realistic scenarios. These limitations highlight the need for careful consideration when using the Hill estimator for tail index estimation in heavy-tailed distributions, and the potential for alternative methods or adjustments to address these issues. Among these alternative methods are those proposed by [Kim and Kim(2015)] and [Németh and Zempléni(2020)]. The asymptotic behavior of the estimator was not examined by [Hill(1975)]. [Mason(1982)] established the weak consistency of $\widehat{\gamma}^{(Hill)}$. Additionally, [Deheuvels et al.(1988)] demonstrated the strong consistency by providing an optimal rate of convergence for a suitably selected sequence k_n. Under certain additional conditions on the df F, the asymptotic normality was established in a number of papers, including [Davis and Resnick(1984)] and [Csörgő and Mason(1985)] .

The following theorem summarizes the asymptotic properties of Hill's estimator.

Theorem 2.4.1 *Assume that* $F \in D\left(\Phi_{1/\gamma}\right)$, $\gamma > 0$ *and* k *satisfying* (2.7) . *Then we have*

(i) *Weak Consistency:*

$$\widehat{\gamma}^{(Hill)} \xrightarrow{P} \gamma, \qquad \text{as } n \to \infty.$$

(ii) *Strong consistency: if* $k/\log\log n \to \infty$ *as* $n \to \infty$, *then*

$$\widehat{\gamma}^{(Hill)} \xrightarrow{a.s} \gamma, \qquad \text{as } n \to \infty.$$

(iii) *Asymptotic normality: Assume that* F *satisfies* (2.5) . *If* $\sqrt{k}A^{**}(n/k) \to \lambda \in \mathbb{R}$, *as* $n \to \infty$, *then*

$$\sqrt{k}\left(\widehat{\gamma}^{(Hill)} - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\lambda}{1-\tau}, \gamma^2\right), \qquad \text{as } n \to \infty.$$

It is important to highlight that the most commonly used extreme quantile, $Q(1-s)$, as $s \to 0$, is the one proposed by [Weissman(1978)]. This is outlined in the following definition:

Definition 2.4.1 *The Weissman estimator is defined as follows:*

$$\widehat{Q}(1-s) := X_{n-k+1:n} + \left(\frac{k}{ns}\right)^{\widehat{\gamma}^{(Hill)}}, \qquad s \to 0.$$

Further insights into the properties of this estimator can be found in [Weissman(1978)] and [de Haan and Ferreira(2006)].

Weissman's estimation method extends Hill's technique to provide more accurate estimates of extreme quantiles in heavy-tailed distributions. The method focuses on using available data about the extreme values already present in the distribution to improve the estimation of tail behavior. By combining order statistics with an estimator for the tail index, Weissman enhances the ability to model the behavior of extreme events more accurately, making it a powerful tool for risk assessment and managing extreme events in heavy-tailed distributions.

Now, we address the selection of the number k of upper order statistics.

2.4.2 Optimal sample fraction selection

As previously stated, the estimators of γ are computed using the optimal number k_{opt} of upper order statistics which determines where the distribution tail really starts. In the literature, several algorithms and adaptive procedures are allocated to obtaining reliable estimates, such as [Dekkers and Dehaan (1993)], [Drees and Kaufmann(1998)], [Cheng and Peng(2001)] and [Danielson et al.(2001)]. The selection of the k is not a straightforward operation, because one must deal with one of two scenarios:

- If k is too small the estimation variance becomes too large.
- If too many observations are used in the estimation procedure, i.e., if k is very large, a large bias appears.

Thus, choosing between bias and variace requires compromise. Generally, to get a good estimates for γ (and other estimators that depend on the optimal sample fraction k_{opt}), we have to extract the k highest observations from a r.v, which represents a very small fraction of the entire sample (i.e. satisfying (2.7)).

Numerous studies in the analytical method focus on selecting k to minimize the asymptotic mean squared error of the adopted semiparametric estimator (see, for example, [Hall(1982)], [de Haan and Peng (1998)]). Additionally, there are graphical and numerical techniques for determining the optimal value of k. One notable approach is the algorithm proposed by [Reiss and Thomas(2007)], page 121, which defines this critical number as follows:

$$k_{opt} := \arg\min_{1 \leq k \leq n} \left\{ \frac{1}{k} \sum_{i=1}^{k} i^{\theta} \left| \widehat{\gamma}^{(Hill)}(i) - \text{median}\left(\widehat{\gamma}^{(Hill)}(1), ..., \widehat{\gamma}^{(Hill)}(k) \right) \right| \right\}, \qquad (2.8)$$

where $0 \leq \theta \leq 0.5$. For a discussion on the choice of the constant θ, one should consult [Neves and Fraga Alves(2004)].

Chapter 3

Incomplete data

It is common in the analysis of lifetime data, such as survival analysis, reliability engineering, and insurance, to encounter incomplete data, which can manifest as two distinct events: truncation and censoring. Incomplete data often arise from study design or observational limitations. Key sources include attrition, where participants drop out of longitudinal studies before an event of interest occurs, and delayed entry, where participants enter a study after its commencement. Additionally, selective reporting may occur when researchers choose not to report all outcomes measured in a study.

Ignoring truncation or censoring during data analysis can lead to biased estimates of population parameters and complicate the analytical process. When the sample no longer accurately represents the entire population, the validity of the findings may be compromised, resulting in erroneous conclusions. Proper statistical methods are essential to account for these issues and ensure reliable and unbiased results. For more details, we may cite the books of [Cohen(1991)], [Meeker and Escobar(1998)] and [Klein and Moeshberger(2003)]. Next, we will present an overview of essential definitions related to lifetime data.

3.1 Lifetime data

Definition 3.1.1 (Survival time) *We define survival time as a positive r.v that represents the time elapsed until the occurrence of a specific event of interest.*

Example 3.1.1 *Survival time encompasses various durations, including the time until death, the interval from the initiation of treatment to the observed response, and the time to failure of a system or component. In the context of insurance, it refers to the time until the occurrence of a specific event, such as filing an insurance claim, the occurrence of a covered incident, or exceeding a predefined financial loss threshold.*

3.1.1 Survival time distributions

If the r.v X is non-negative and has a continuous df F, then one of five corresponding functions can be employed to characterize its probability distribution. It is worth noting that, the df F and tail function $\overline{F}$, are introduced here in the context of survival analysis.

Cumulative distribution function

The cumulative distribution function (cdf), denoted as F, represents the probability of an event occurring before a specified time x (the likelihood of dying before time x), i.e.,

$$F(x) = P(X \leq x) = P(X < x), \qquad x \geq 0.$$

Survival function

The survival function, denoted as $\overline{F}$, represents the probability of surviving at least until a specified time x, i.e.,

$$\overline{F}(x) = P(X \geq x) = 1 - F(x).$$

Density function

For a fixed value of x, the density function f represents the probability of dying within a small time interval immediately following x. If the cdf F has a derivative at point x, then f is defined for all $x \geq 0$ as follows:

$$f(x) := \frac{dF(x)}{dx} = \lim_{dy \to 0} \frac{P(x < X < x + dy)}{dy}.$$

Hazard function

The hazard function, denoted by h, for a fixed x, quantifies the instantaneous risk of death in a small time interval following x, given that an individual has survived up to time x. It is defined as:

$$h(x) := \lim_{dy \to 0} \frac{P(x < Y < x + dy / Y > x)}{dy}.$$

Cumulative hazard function

The cumulative hazard function, denoted by Λ, represents the total accumulated risk up to time s and is defined as:

$$\Lambda(s) := \int_0^s h(x)dx = \int_0^s \frac{dF(x)}{\overline{F}(x)}.$$

Remark 3.1.1 *These functions have the following relationships:*

1.

$$\begin{aligned} h(x) &= \lim_{dy \to 0} \left(\frac{\overline{F}(x) - \overline{F}(x + dy)}{dy \overline{F}(x)} \right) \\ &= -\frac{d \log \overline{F}(x)}{dy}. \end{aligned}$$

2.

$$\Lambda(x) = -\log \overline{F}(x). \tag{3.1}$$

3.

$$f(x) = \overline{F}(x)h(x).$$

3.2 Censorship and truncation

In the following we consider the notations X, C and T, which are used to model the censored and truncation data, where

- X_j denotes the variable of the interest for the j-th subject,
- C_j denotes the censoring or truncation threshold for the j-th subject,
- T_j represents the observed variable for the j-th subject.

3.2.1 Censoring

Definition 3.2.1 *Censorship refers to situations where the value of an observation is only partially known.*

In other words, the variable of interest X, is said censored by a variable or a value of censorship, denoted by C, if sometimes observe C instead of X. For instance, in a medical research, individuals may leave the study (called dropout) so we only observe their leaving time instead of the actual death time.
In this subsection, we present the different categories and types of censored data, accompanied by illustrative examples.

Right censoring

An observation is considered right censored if a subject exits the study before the event occurs, or if the study concludes prior to the event's occurrence. Hereafter, we present examples illustrating the contexts in which right censored data can occur and how it is observed.

• In insurance, censorship occurs when a claim amount exceeds the insurer's policy limit. For instance, if policyholders have a limit of C, any loss above this limit is reported as exactly C. Thus, the insurance loss is right censored if it exceeds C, which is the maximum payout. In such cases, especially smaller insurance companies may need to purchase excess-of-loss reinsurance from larger firms.

• In medicine, right censoring occurs when a study concludes before the event of interest, such as death, is recorded, or when a subject exits the study prior to experiencing the event. This means that the full survival times are unknown; only that the survival time exceeds the observation period. For example, in a clinical trial to study the effect of treatments on stroke occurrence, the study may end after 5 years, patients who have not experienced a stroke by that time are considered right censored.

• In reliability studies within electrical engineering (or mechanics), an engine is considered right censored if it is lost to follow-up (for example, if an engine is taken out for maintenance or inspection before it breaks down) or if the study concludes before it experiences a breakdown.

• In sociology, when analyzing the duration of marriages, right censoring occurs when a couple withdraws from the study or when the study concludes before a divorce takes place.

Left censoring

Left censoring occurs when the event of interest has already taken place before data collection begins or the study commences. In such cases, we only know the upper limit of the time at which the event occurred. Below are some examples of left censoring:

• In insurance claims, left censoring occurs when the event of interest, such as the onset of a disease or a significant life event, happens before an individual enrolls in an insurance plan or before a study begins. For instance, if an insurance company is studying the effects of car accidents on long-term health and a participant had an accident in 2017 but did not enroll until 2020, their health data related to that accident would be left censored.

Since the accident occurred prior to enrollment, any claims or health issues arising from it would be excluded from the study, resulting in incomplete data on the impact of car accidents on health.

• In medical study, some patients may have already experienced the event of interest before the study begins, and their event time is unknown. For example, in a study examining the onset of diabetes, researchers track participants over 4 years. If a participant was diagnosed with diabetes before the study began, their exact diagnosis date would be unknown. This represents left censoring, as the event occurred prior to enrollment. Consequently, any health data related to their diabetes before joining the study would be excluded, potentially skewing the results regarding diabetes incidence and effects in the population.

• In reliability engineering, left censored data pertains to items that have already failed prior to the commencement of a study. For instance, in a study assessing the lifespan of light bulbs, some bulbs may have failed before the study began. As a result, the exact failure times for these bulbs are left censored, leading to incomplete data regarding their performance and potentially affecting the overall analysis of bulb longevity.

Mixed Censoring

Mixed censoring occurs when data are subject to both types of censoring, meaning that some values are left censored because they fall below a certain threshold, while others are right censored because they exceed a specific limit.

As an example of this type in the field of insurance, recorded compensation amounts for insured losses may be subject to mixed censoring due to coverage policies. For instance, in vehicle insurance:

compensation is not provided for losses below 500$, as the policyholder must cover them. Consequently, their exact values are not recorded. The insurance company does not pay more than 20000$, even if the actual loss exceeds this amount. Therefore, all values above this threshold are recorded as equal to it. In this case, the recorded data do not reflect the true values of some losses but rather classify them within a specific range, leading to mixed censoring.

Interval censoring

Interval censoring occurs in statistical analysis when the exact timing of an event is not known, but it is known to fall within a specific interval. This situation often arises in survival analysis and clinical studies where observations are conducted at discrete time points. For instance, a subject may have had two tests for Human Immunodeficiency Virus (HIV), where the first test result was negative (not infected), and the second test result was positive (infected). In such cases, determining the exact time of HIV infection is based on periodic blood tests, which cannot be conducted continuously. As a result, the exact time at which the subject became infected with HIV (the event of interest) occurred after the first test and before the second test. This makes the subject interval censored within the time interval between the two tests.

Type I censoring (fixed censorship)

Type I censoring, also known as fixed censoring, occurs in statistical studies when the observation period is predetermined, and any subjects or items that have not experienced the event of interest by the end of this period are right censored. This means that researchers know the event has not occurred up to a certain time, but they do not know when it will happen.

An example of Type I censoring can be found in a clinical trial studying the effectiveness of a new cancer treatment. Suppose researchers follow patients for a maximum of 2 years. If a patient has not experienced disease progression or death by the end of the 2 year period, their data is right censored at that time point. While it is known that these patients survived for at least 2 years, their exact survival times beyond that point remain unknown.

Type II censoring (failure censoring)

In Type II censoring, an experiment continues until a specified number r $(1 \leq r \leq n)$ of failures occurs among the subjects being tested. Once this number of failures is reached, the experiment concludes, and any remaining subjects that have not failed are considered right censored. This means that while the exact failure times for the observed failures are recorded, the exact failure times for the remaining subjects are unknown but are known to exceed the time at which the last failure occurred. In other words, let $X_{j:n}$ and $T_{j:n}$ represent the order statistics of the variables X_j and T_j, respectively. The censored data is denoted as $X_{r:n}$ and we observe the following variables:

$$T_{j:n} = \begin{cases} X_{j:n} & \text{if} \quad j \leq r, \\ X_{r:n} & \text{if} \quad j > r. \end{cases}$$

For instance, if a reliability test involves 80 units and the goal is to observe 30 failures, the test will continue running until 30 units have failed. At that point, the test ends, and the remaining 50 units are right censored.

This type of censoring is particularly useful when the goal is to study the characteristics of failures among a fixed number of subjects rather than observing all subjects for a fixed period.

Random censorship (type III censoring)

In random censoring, the censoring times are independent of the event times. This means that the time at which data is censored does not provide any information about the likelihood of the event occurring at that time.

Each subject in a study has a failure time and a censoring time. The observed outcome is the minimum of these two times. If a subject's failure time occurs after their censoring time, they are considered right censored. In other words, we say that X_j is right censored by C_j if, for $1 \leq j \leq n$, instead of observing the values of X_j directly, we observe a pair of values (T_j, δ_j) defined as follows:

$$T_j := \min(X_j, C_j) \qquad \text{and} \qquad \delta_j := \mathbf{1}\{X_j \leq C_j\}, \tag{3.2}$$

where δ_j serves as the censorship indicator. Specifically:

$$\delta_j = \begin{cases} 0 & \text{if the individual is censored } \ (\text{i.e. } T_j = C_j), \\ 1 & \text{if the event is observed } \ (\text{i.e. } T_j = X_j). \end{cases}$$

In Chapter 4, Subsection 4.6.2, we present two examples of real-world datasets exhibiting random right censoring: one related to insurance losses and the other concerning survival times in AIDS patients.

The figure 3.1 illustrates the different categories of censored data.

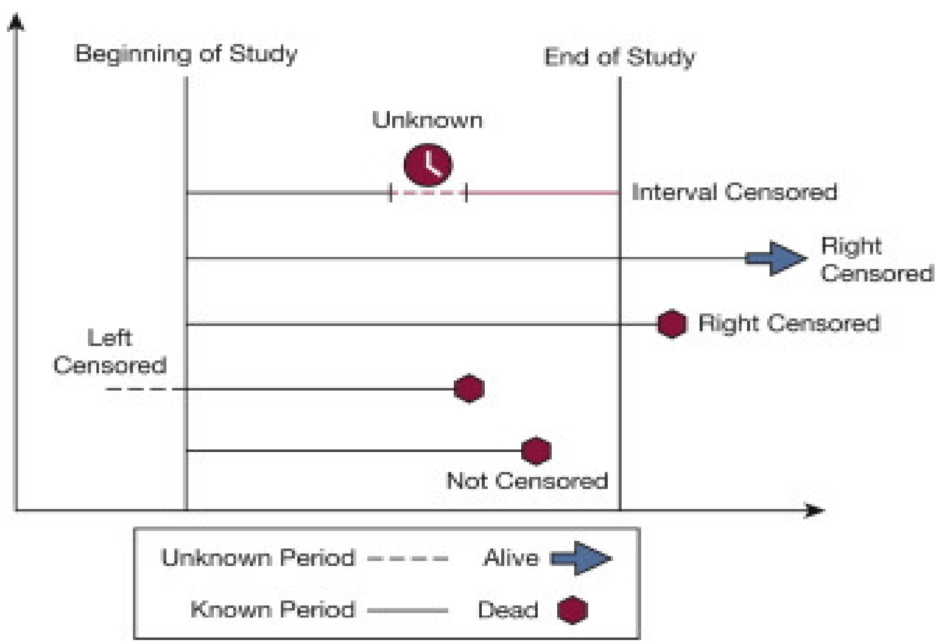

Figure 3.1: A representative example of the censored data cases

3.2.2 Truncation

In survival analysis, truncation occurs when only individuals who have experienced the event of interest are observable, which is often a result of the study's design. This means that only results above (or below) the truncation limit can be observed, leading to data loss and creating a truncated sample.

There are three different types of truncation, as follows:

Right truncation

Right truncation occurs when only individuals whose variable of interest falls below a certain threshold are observable, while those exceeding this threshold remain unobserved. In other words, a random variable is said to be right truncated if its exact value is known for all cases below a specific threshold but unknown for cases above it. Specifically, this means that the value of X is known when $X < C$ but remains unknown when $X \geq C$.

An example of right truncation is the timing of the appearance of primary teeth in infants, specifically the two bottom front teeth, which typically emerge between 6 and 8 months of age. In this case, we have right truncated data because infants whose teeth appear after 8 months are not included in the analysis.

Left truncation

Left truncation occurs when individuals with values below a certain threshold are excluded from the study. In this context, only individuals whose variable of interest exceeds this threshold are observed. This means that the variable of interest is observable only if it exceeds a certain limit (i.e., X is observable only when $X > C$).

Example 3.2.1 *In a clinical trial testing a new medication for hypertension, left truncation occurs when only patients who are diagnosed with high blood pressure after the age of* 40 *are included in the study. If a patient is diagnosed at age* 38 *and starts treatment before turning* 40, *they are excluded from the trial.*

Interval truncation (doubly truncation)

Interval truncation, also referred to as doubly truncation, occurs when observations are only recorded within a specified range, leading to the exclusion of values outside that range.

Example 3.2.2 *When using a mercury thermometer with a measurement range of* $-40°C$ *to* $+80°C$*, any temperature readings outside this range cannot be recorded. Consequently, temperatures below* $-40°C$ *and above* $+80°C$ *are not observed, resulting in interval truncation for values outside this specified range.*

[Lynden-Bell(1971)] proposed a nonparametric estimation of the df of X within the framework of the truncation model,along with its asymptotic properties. The strong law and asymptotic normality were further examined by [Woodroofe (1985)].

Remark 3.2.1 *Truncation is distinct from censoring, as in truncation, observations never result in values outside a given range, and values outside the range are never seen or recorded if they are seen, while in censoring, a note is recorded documenting which bound (upper or lower) had been exceeded and the value of that bound.*

It is necessary to choose appropriate estimation techniques that are unaffected by missing values. Below, we present the nonparametric methods for censored data that are central to the focus of this thesis.

3.3 Nonparametric estimators

In this section, we introduce the main estimators that are crucial for analyzing randomly censored data: the Kaplan-Meier estimator for the df and the Nelson-Aalen estimator for the cumulative hazard function.

3.3.1 Kaplan-Meier estimator

The Kaplan-Meier estimator, also known as the product limit estimators or nonparametric maximum likelihood, was introduced by [Kaplan and Meier(1958)]. It is a nonparametric statistical tool used to estimate the survival function from lifetime data, particularly in the context of censored observations. The Kaplan-Meier estimator is widely utilized across various fields, providing valuable insights into the duration that an event may continue before another event occurs (such as death or failure). In medical research, it is often employed to measure the proportion of patients who survive for a specified period after treatment. For further insights into its real-world applications, we refer to [Goel et al.(2010)] and [Etikan et al.(2017)].

Next, we present the formula for the Kaplan-Meier estimator. Let X and C be two independent r.v's defined on a probability space $(\Omega, \mathcal{F}, P)$, with continuous df's F and G, respectively. The Kaplan-Meier estimators of the df's F and G, denoted as $\widehat{F}_n$ and $\widehat{G}_n$, are defined as follows:

$$\widehat{F}_n(x) := \begin{cases} 1 - \prod_{T_{j:n} \leq x} \left(1 - \dfrac{\delta_{[j:n]}}{n-j+1}\right) & \text{for} \quad x < T_{n:n}, \\ 1 & \text{for} \quad x \geq T_{n:n}. \end{cases} \tag{3.3}$$

and

$$\widehat{G}_n(x) := \begin{cases} 1 - \prod_{T_{j:n} \leq x} \left(1 - \dfrac{1-\delta_{[j:n]}}{n-j+1}\right) & \text{for} \quad x < T_{n:n}, \\ 1 & \text{for} \quad x \geq T_{n:n}. \end{cases} \tag{3.4}$$

where $T_{1:n} \leq T_{2:n} \leq ... \leq T_{n:n}$ represents the order statistics associated with the sample $(T_1, T_2, ..., T_n)$ and $\delta_{[1:n]}, \delta_{[2:n]}, ..., \delta_{[n:n]}$ denotes the corresponding concomitant values, satisfying $\delta_{[j:n]} = \delta_i$ for i such that $T_{j:n} = T_i$. We denote by H the cdf of the observed $T\prime$s, which is defined as follows:

$$H(x) := 1 - (1 - F(x))(1 - G(x)). \tag{3.5}$$

The key requirement for using the Kaplan-Meier estimator is the ordered nature of observed values, allowing for a clear distinction between censored observations and events. Whether these values represent time, financial losses, or any other ordered continuous variable, the estimator remains valid as long as the ordering and censoring information are properly accounted for. However, it is essential to emphasize that the Kaplan-Meier estimator is specifically tailored to address random right censoring scenarios, which aligns with its core theoretical foundation and practical application.

Remark 3.3.1 *The estimators* (3.3) *and* (3.4) *are rewritten as follows:*

$$\widehat{F}_n(x) := \begin{cases} 1 - \prod_{T_{j:n} \leq x} \left(\dfrac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} & \text{for} \quad x < T_{n:n}, \\ 1 & \text{for} \quad x \geq T_{n:n}, \end{cases}$$

$$\widehat{G}_n(x) := \begin{cases} 1 - \prod_{T_{j:n} \leq x} \left(\dfrac{n-j}{n-j+1} \right)^{1-\delta_{[j:n]}} & \text{for} \quad x < T_{n:n}, \\ 1 & \text{for} \quad x \geq T_{n:n}, \end{cases}$$

Remark 3.3.2 *If $X_1, X_2, ..., X_n$ are not censored, the df F can be estimated by F_n, as given in* (2.1) .

The Kaplan-Meier estimator has properties analogous to those of the empirical df. For more information, see the book by [Shorack and Wellner(2009)].

3.3.2 Nelson-Aalen estimator

The Nelson-Aalen estimator, introduced by [Nelson(1972)] and generalized by [Aalen(1978)], is a nonparametric estimator of the cumulative hazard function. The estimator is useful for imputing variables that depend on survival time, and it can be used to estimate the cumulative number of expected events. The method requires that the observations are independent, and the censoring must be independent. According to the independence hypothesis between X and C, we obtain the following equalities:

$$H(x) = H^{(0)}(x) + H^{(1)}(x),$$

where

$$H^{(0)}(x) := P(T \leq x, \delta = 0) = \int_0^x \overline{F}(y)\, dG(y),$$

and

$$H^{(1)}(x) := P(T \leq x, \delta = 1) = \int_0^x \overline{G}(y)\, dF(y).$$

For $x \geq 0$, we can express the cumulative hazard function as follows:

$$\Lambda(x) = \int_0^x \frac{dH^{(1)}(y)}{\overline{H}(y)}.$$

The Nelson-Aalen estimator for Λ is given by

$$\Lambda_n(x) = \int_0^x \frac{dH_n^{(1)}(y)}{\overline{H}_n(y)} := \begin{cases} \sum\limits_{T_{j:n} \leq x}^{n} \dfrac{\delta_{[j:n]}}{n-j+1} & \text{if} \quad x < T_{j:n}, \\ 1 & \text{if} \quad x \geq T_{j:n}, \end{cases}$$

where H_n and $H_n^{(1)}$ represent respectively the empirical df of H and the empirical counterpart of $H^{(1)}$, based on the sample T_j, $1 \leq j \leq n$, which are defined by

$$H_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{T_j \leq x\}$$

and

$$H_n^{(1)}(x) := \frac{1}{n} \sum_{i=1}^{n} \delta_j \mathbf{1}\{T_j \leq x\}.$$

Remark 3.3.3 *By utilizing the relationship defined in (3.1) and the Kaplan-Meier estimator, we can derive another estimator of the df F, known as the Breslow estimator (see*

[Breslow and Crowley(1974)]). This estimator is defined as follows:

$$\widehat{F}_n^{NA}(x) := \begin{cases} 1 - \prod_{T_{j:n} \leq x} \exp\left(-\frac{\delta_{[j:n]}}{n-j+1}\right) & \textit{for} \quad x < T_{n:n}, \\ 1 & \textit{for} \quad x \geq T_{n:n}. \end{cases}$$

Chapter 4

Estimation of large risk measures under censorship

Estimating risk measures is crucial for quantifying and managing risk across various fields, including survival analysis, reliability, and insurance. The selection of an appropriate estimator is vital and depends on the data's characteristics and the underlying distribution. In this chapter, we review estimators available in the literature for some of the risk measures introduced in Chapter 1, as well as for insurance premiums such as the net premium (which corresponds to the mean) and a reinsurance premium calculated using the PHP. The focus is on the case of randomly right censored and heavy-tailed distributions. We will mainly focus on the CTE estimator that we recently introduced in [Guesmia et al.(2024)], which represents the central part of our reseach work. We will reserve the end of the chapter for a detailed discussion on this topic. However, first, we need to discuss the fondamental issue of estimating tail indices within the framework of random right censorship.

4.1 Tail index estimators

In the censorship framework, the tail index is estimated using various methods, which is relevant for understanding the distribution of extreme values in various datasets. The first to mention the subject are [Beirlant et al.(1996)] in Section 2.7 and [Reiss and Thomas(2007)] in Section 6.1, but without asymptotic results. Thereafter, [Beirlant et al.(2007)] and [Einmahl et al.(2008)] were proposed new classes of estimators for the extreme value index when the data are subject to random censorship. These estimators are based on the order statistics of the censored data and are designed to be consistent and asymptotically normal. [Einmahl et al.(2008)] adapted different estimators for γ_1 to the case where the data are censored by a random threshold and proposed a unified method to establish their asymptotic normality. Their estimators are based on a standard estimator of the tail index divided by the estimator of the proportion of uncensored data among the largest k observations of T's:

$$\widehat{\gamma}_1^{(\bullet,c)} := \frac{\widehat{\gamma}^{(\bullet)}}{\widehat{p}},$$

where $\widehat{\gamma}^{(\bullet)}$ can be any estimator that is not adapted to censoring and $\widehat{p} := (1/k)\sum_{i=1}^{k}\delta_{[n-i+1:n]}$ is the estimate of the proportion p of upper non-censored observations, with $p := \gamma_2/\left(\gamma_1+\gamma_2\right)$.

Example 4.1.1 *Hill's estimator of γ_1, adapted for censored data, is then defined as follows:*

$$\widehat{\gamma}_1^{(c)} := \frac{\sum_{i=1}^{k}\left(\log T_{n-i+1:n} - \log T_{n-k:n}\right)}{\sum_{i=1}^{k}\delta_{[n-i+1:n]}}. \tag{4.1}$$

Along with [Beirlant et al.(2007)] and [Einmahl et al.(2008)], a large number of studies, have shown interest in the tail index in the case of random censorship. The particular formulas of these estimators may change depending on the distribution and the method being taken into account. [Worms and Worms(2014)] presented new approaches for estimating the tail index in the context of randomly censored samples, particularly for heavy-tailed distributions. [Beirlant et al.(2016)] and [Beirlant et al.(2018)] proposed a new bias reduced estimator of the tail index for censored Pareto-type data. Additionally, [Beirlant et al.(2019)] introduced a new class of estimators which encompasses earlier proposals given in [Worms and Worms(2014)] and [Beirlant et al.(2018)], which were shown to have good bias properties compared with the maximum likelihood estimator proposed in [Beirlant et al.(2007)] and [Einmahl et al.(2008)]. See also [Brahimi et al.(2015a)], [Brahimi et al.(2015b)], [Brahimi et al.(2018)], [Bladt et al.(2021)] and [Worms and Worms(2021)] for other papers on the subject.

Example 4.1.2 *To see the performance of the adapted Hill estimator, given in* (4.1) , *as a function of the number k of upper order statistics, we carry out a simulation study based on* 100 *samples of size* $n = 2000$ *from Fréchet model with parameter* γ_1, *defined by*

$$F(x) = \exp(-x^{-1/\gamma_1}), \text{ for } x > 0,$$

censored by another variable of Fréchet of parameter $\gamma_2 = p\gamma_1/(1-p)$. *We choose* 0.5 *as a value for the tail index* γ_1. *For the proportion* p, *we take* 0.40, *and* 0.60, *that is, we allow the percentage of censoring in the right tail of* X *to be* 60% *and* 40%.

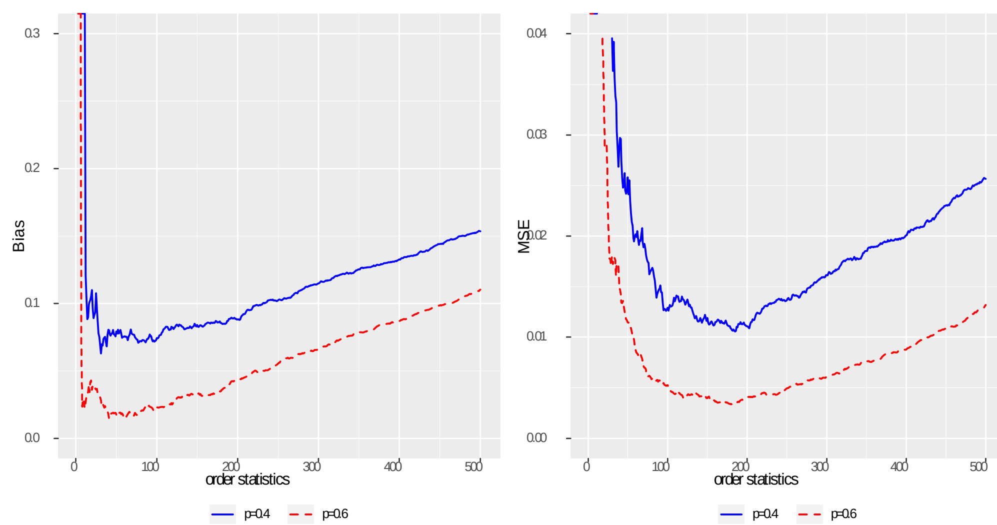

Figure 4.1: Bias (left panel) and MSE (right panel) of $\widehat{\gamma}_1^{(c)}$ based on 100 samples of size 2000 for $\gamma_1 = 0.5$ as a functions of the number k of upper order statistics

Figure 4.1 illustrates the absolute bias (Bias) and the mean squared error (MSE) of $\widehat{\gamma}_1^{(c)}$ as a function of the number k of upper order statistics. We observe that the bias and the

MSE of $\widehat{\gamma}_1^{(c)}$ decrease as the percentage of censoring diminishes. On the other hand, we note that for a large value of k the bias increases significantly.

4.2 Estimating the VaR

We recall that the VaR is the quantile function Q corresponding to df F. Therefore, the nonparametric estimator of the VaR is defined by substituting $Q(s)$ with its empirical counterpart under random censorship, represented as $Q_n(t) := \inf\{x : F_n(x) \geq t\}$ for $0 < t < 1$. From the definition of Kaplan-Meier estimator of F_n, we obtain

$$\widehat{VaR}(t) = Q_n(t) = T_{i:n} \qquad \text{for } w_{i-1} < t \leq w_i, \tag{4.2}$$

where $w_i = 1 - \prod_{j=1}^{i} \left(1 - \delta_{[j:n]} / (n - j + 1)\right)$; $i = 1, ..., n$ (we agree that $w_0 = 0$). The Kaplan-Meier estimator is a widely used method for estimating quantiles from censored data (see [Hong et al.(2013)]), as it is nonparametric and does not require any assumptions about the underlying distribution.

In EVT, estimating the VaR is directly linked to the accurate modeling and estimation of the quantile function $Q(1 - s)$, where $s = 1 - t$. As $s \to 0$, the VaR corresponds to the extreme quantiles of the distribution. Consequently, extreme quantile estimation is the process of estimating the tail probabilities of a distribution, which are associated with rare events.

When data are subject to random censoring, several estimators have been proposed for extreme quantile estimation of heavy-tailed random variables. Noteworthy contributions include the works of [Beirlant et al.(2007)], [Einmahl et al.(2008)], [Bladt et al.(2021)] and [Worms and Worms(2021)]. Hereafter, we present the estimates to the extreme quantiles, of Weissman-type (see, [Weissman(1978)]), for randomly censored data. We suppose that

$$\lim_{v \to \infty} \frac{\overline{F}(vx)}{\overline{F}(v)} = x^{-1/\gamma_1}. \tag{4.3}$$

If we take $v = h = h_n := H^{\leftarrow}(1 - k/n)$ in the relation of (4.3), we deduce that $\overline{F}(x)$ is equivalent to $\overline{F}(h)(x/h)^{-1/\gamma_1}$ as $x \to \infty$. Here $H^{\leftarrow}(s) := \inf\{x : H(x) \geq s\}$, $0 < s < 1$, denotes the quantile function or the generalized inverse of H. Replacing h by its empirical counterpart $T_{n-k:n}$ yields a Weissman-type estimator to the tail $\overline{F}(x)$, for large x, as follows:

$$\widehat{\overline{F}}(x) \sim \left(\frac{x}{T_{n-k:n}}\right)^{-1/\widehat{\gamma}_1^{(c)}} \overline{F}_n(T_{n-k:n}).$$

From (3.3), we have $\overline{F}_n(T_{n-k:n}) = \prod_{i=1}^{n-k}\left(1 - \delta_{[i:n]}/(n-i+1)\right)$. Thus, we get

$$\widehat{\overline{F}}(x) \sim \left(\frac{x}{T_{n-k:n}}\right)^{-1/\widehat{\gamma}_1^{(c)}} \prod_{i=1}^{n-k}\left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right). \tag{4.4}$$

From (4.4), we define a Weissman-type estimator to $Q(1-s)$, denoted by $\widehat{q}_s$, for randomly censored data, as follows:

$$\widehat{q}_s := s^{-\widehat{\gamma}_1^{(c)}} T_{n-k:n} \prod_{i=1}^{n-k}\left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right)^{\widehat{\gamma}_1^{(c)}}, \quad \text{as } s \to 0. \tag{4.5}$$

The convergence in distribution of the extreme quantile estimator $\widehat{q}_s$ is established in Theorem 1 of [Goegebeur et al.(2023)].

4.3 Estimating the mean

The mean (net premium) of the r.v X with df F is defined by

$$\mu = \mathbf{E}[X] := \int_0^\infty x dF(x) = \int_0^\infty \overline{F}(x)\, dx.$$

In the censorship case, the nonparametric estimator of μ, denoted as $\widehat{\mu}_n$, is referred to as the Kaplan-Meier integral, which was introduced by [Stute(1995)]. It is defined as follows:

$$\widehat{\mu}_n = \sum_{i=1}^{n} T_{i:n} W_{i:n},$$

where

$$W_{i:n} := \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}}. \tag{4.6}$$

[Stute(1995)] established the central limit theorem under random censorship, which is satisfied provided that

$$\int_0^\infty x^2 \kappa^2 (x)\, dH^{(1)}(x) < \infty \text{ and } \int_0^\infty x \left(\int_0^x \frac{dG(z)}{\overline{H}(z)\, \overline{G}(z)} \right)^{1/2} dF(x) < \infty, \tag{4.7}$$

where $\kappa(x) := \exp\left\{ \int_0^x dH^{(0)}(z) / \overline{H}(z) \right\}$ with $H^{(j)}$, $j = 0, 1$, being defined in (5.9) . However, this seems constraining for some class of heavy-tailed distributions. Indeed, [Soltane et al.(2015)] showed that, for Pareto models, one of the conditions above (or both) is not met for tail indices satisfying $\gamma_2/(1+2\gamma_2) < \gamma_1 < 1$. [Soltane et al.(2015)] defined an alternative estimator, which is given in Definition 4.3.1.

Definition 4.3.1 *The semiparametric estimator of the mean, denoted as $\widehat{\mu}$, is defined as follows:*

$$\widehat{\mu} := \sum_{i=2}^{n-k} \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} T_{i:n} + \prod_{j=1}^{n-k} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}} \frac{T_{n-k:n}}{1-\widehat{\gamma}_1^{(c)}}.$$

More recently, [Kouider et al.(2024)] introduced an alternative estimator for the mean within the same context and established its asymptotic normality. The proposed estimator relies on a threshold parameter z, which divides the data into two segments: values less than or equal to z and those greater than z. The tail df of F is assumed to start at some threshold z with $z \to +\infty$. The proposed estimator, denoted by $\widehat{M}$, is defined as

$$\widehat{M} := \frac{1}{n} \sum_{i=1}^{n} \frac{T_{i:n} \delta_{[i:n]}}{1 - \widehat{G}_n (T_{i-1:n})} \mathbf{1} \{ T_{i:n} \leq x; x \leq z \} + \frac{z \left(1 - \widehat{F}_n(z) \right)}{1 - \widehat{\gamma}_1^{(KIB)}}$$

where $\widehat{\gamma}_1^{(KIB)}$ is the shape paramete estimator proposed by [Kouider et al.(2023)].

4.4 Estimating the PHP

The PHP estimation in the context of randomly right censored losses is a statistical method used to estimate the excess-of-loss reinsurance premium when the risks are randomly right-censored. Excess-of-loss reinsurance is a specific type of reinsurance where the ceding company is compensated for losses that exceed a specified limit. [Soltane et al.(2016)] focus on this topic, discussing the estimation of the PHP of loss under random censoring and the asymptotic normality of the proposed estimator under mild conditions. This method is particularly relevant in the context of insurance premium principles and risk assessment.

The PHP of risk for the layer from retention level $R \geq 0$ to infinity is defined as

$$\Pi_{\varrho}(R) := \int_{R}^{\infty} \left(\overline{F}(x)\right)^{1/\varrho} dx,$$

where $\varrho \geq 1$. The estimator of $\Pi_{\varrho}(R)$, denoted by $\widehat{\Pi}_{\varrho}(R)$, is defined as follows:

$$\widehat{\Pi}_{\rho}(R) := \frac{\varrho R}{1/\widehat{\gamma}_1^{(c)} - \varrho} \left(\frac{R}{T_{n-k:n}}\right)^{-1/\left(\varrho \widehat{\gamma}_1^{(c)}\right)} \prod_{i=1}^{n-k} \left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right)^{1/\varrho}.$$

4.5 Estimating the CTM

In the context of extreme losses, [Goegebeur et al.(2023)] proposed the estimator

$$\widehat{CTM}_{\zeta}(t) := \frac{1}{\left(1 - \widehat{\gamma}_1^{(c)}\right)\zeta} \left(\frac{T_{n-k:n}}{(1-t)^{\widehat{\gamma}_1^{(c)}}} \prod_{i=1}^{n-k} \left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right)^{\widehat{\gamma}_1^{(c)}}\right)^{\zeta},$$

where $\zeta > 0$, as an estimator for the CTM, which is defined in (1.2). The authors also proved its asymptotic normality in the case of censorship.

Based on Remark 1.3.4, we introduce an estimator of the CTE (i.e., when $\zeta = 1$) for extreme risks. This estimator is denoted as $\widehat{CTM}_1(t)$ and is defined as follows:

$$\widehat{CTM}_1(t) := \frac{T_{n-k:n}}{(1-t)^{\widehat{\gamma}_1^{(c)}}\left(1-\widehat{\gamma}_1^{(c)}\right)} \prod_{i=1}^{n-k} \left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right)^{\widehat{\gamma}_1^{(c)}}.$$

Example 4.5.1 *To assess the performance of the $\widehat{CTM}_1(t)$, we conduct a simulation study based on 1000 samples of size n drawn from a Burr distribution with parameters $\gamma_1 = 0.4$ and $\eta = 0.25$, given by $F(x) = 1 - \left(1 + x^{1/\eta}\right)^{-\eta/\gamma_1}$, censored by a Fréchet variable with parameter γ_2. For the proportion p, we take 0.50, 0.70 and 0.90. We take two levels 0.99 and 0.999. To select the optimal number of top statistics k_{opt}, we apply the adaptive algorithm of [Reiss and Thomas(2007)]. The notations bias and rmse respectively stand for the absolute value of the bias and the root of the mean squared error of $\widehat{CTM}_1(t)$. The simulation results are summarized in Table 4.1.*

$CTM_1(t)$	10.516 ($t = 0.99$)			26.415 ($t = 0.999$)		
			$p = 0.50$			
n	$\widehat{CTM}_1(t)$	bias	rmse	$\widehat{CTM}_1(t)$	bias	rmse
1000	12.474	1.958	10.339	35.009	8.594	27.160
2000	11.905	1.389	4.704	31.609	5.194	18.055
			$p = 0.70$			
1000	11.038	0.522	4.110	28.561	2.146	14.084
2000	10.775	0.259	2.463	28.100	1.685	10.221
			$p = 0.90$			
1000	10.663	0.147	2.838	26.203	0.212	10.950
2000	10.564	0.049	1.985	26.338	0.077	7.463

Table 4.1: $\widehat{CTM}_1$ based on 1000 right-censored samples of size n from Burr model censored by Fréchet model for $\gamma_1 = 0.4$

4.6 Estimating the CTE

In this section, we define our estimator for the CTE under randomly right censored heavy-tailed data. Since the df F is continuous, we readily get

$$CTE\left(t\right)=\frac{1}{1-t}\int_{t}^{1}Q(s)ds, \tag{4.8}$$

which, to be well defined, requires that the index γ_1 be less than 1. This could be easily checked for Pareto's model defined by $F(x)=1-x^{-1/\gamma_1}$, $x\geq 1$.
To construct our estimator, that will be denoted by $\widehat{CTE}\left(t\right)$, we start by writing $CTE\left(t\right)$ as the sum of

$$CTE_{1,n}\left(t\right):=\frac{1}{1-t}\int_{t}^{w_{n-k}}Q(s)ds \quad \text{and} \quad CTE_{2,n}\left(t\right):=\frac{1}{1-t}\int_{w_{n-k}}^{1}Q(s)ds,$$

where $w_{n-k}=F_n\left(T_{n-k:n}\right)=1-\prod_{i=1}^{n-k}\left(1-\delta_{[i:n]}/\left(n-i+1\right)\right)$. From Lemma (5.3.5), we have $w_{n-k}\overset{P}{\rightarrow}1$, as $n\rightarrow\infty$.
We divided the domain of the estimator into two parts to enhance accuracy: the first part handles moderate values using traditional methods, while the second part focuses on extreme values, applying specialized techniques suitable for tail data. In the first integral, we replace $Q(s)$ by its empirical counterpart under random censorship $Q_n(s)$, that is defined in (4.2), to obtain an estimator

$$\widehat{CTE}_{1,n}\left(t\right):=\frac{1}{1-t}\int_{t}^{w_{n-k}}Q_n(s)ds,$$

for $CTE_{1,n}\left(t\right)$.
A change of variables in $CTE_{2,n}\left(t\right)$ yields

$$CTE_{2,n}\left(t\right)=\frac{1}{1-t}\int_{0}^{1-w_{n-k}}Q(1-s)ds. \tag{4.9}$$

By substituting $Q(1-s)$ with $\widehat{q}_s$, as defined in (4.5), in (4.9), we obtain the following after integration:

$$\widehat{CTE}_{2,n}(t) := \frac{T_{n-k:n}}{(1-t)\left(1-\widehat{\gamma}_1^{(c)}\right)} \prod_{i=1}^{n-k}\left(1-\frac{\delta_{[i:n]}}{n-i+1}\right),$$

which serves as an estimator for $CTE_{2,n}(t)$. Finally, by adding $\widehat{CTE}_{2,n}(t)$ to $\widehat{CTE}_{1,n}(t)$ we obtain the estimator of the CTE, for every $t \in (0,1)$, as follows:

$$\widehat{CTE}(t) := \frac{1}{1-t}\int_t^{w_{n-k}} Q_n(s)ds + \frac{T_{n-k:n}}{(1-t)\left(1-\widehat{\gamma}_1^{(c)}\right)} \prod_{i=1}^{n-k}\left(1-\frac{\delta_{[i:n]}}{n-i+1}\right). \tag{4.10}$$

We present our main result in Chapter 5, Section 5.1, which is the asymptotic normality of our estimator and was proven in Section 5.2.

Below, we present the graphs and numerical results of the simulation study which we conducted to check the performance of our estimation procedure of the CTE of randomly censored heavy-tailed data.

4.6.1 Simulation study

We carry out a simulation study to illustrate the performance of our estimator, through two sets of data from Burr and Fréchet models. We consider the following four censoring scenarios with a combination of tail indices and censorship proportions.

- scenario 1: Burr (γ_1, η) censored by Burr (γ_2, η).
- scenario 2: Fréchet (γ_1) censored by Fréchet (γ_2).
- scenario 3: Burr (γ_1, η) censored by Fréchet (γ_2).
- scenario 4: Fréchet (γ_1) censored by Burr (γ_2, η).

We fix $\eta = 1/4$ and we choose 0.3 and 0.5 as values for the tail index γ_1. For the proportion p, we take 0.50, 0.70 and 0.90. Additionally, we consider two levels for t : 0.90 and 0.95. In each case, we generate 1000 independent replicates of samples (of X and C) of sizes $n = 1000$, 1500 and 2000. Our overall results are computed as the empirical means of the results that are obtained through all repetitions.

Defining the nonparametric estimator

The nonparametric estimator $\widetilde{CTE}(t)$ of the CTE at level t is obtained by substituting $Q_n(s)$ for $Q(s)$ in formula (4.8) . In other words, we have

$$\widetilde{CTE}(t) := \frac{1}{1-t}\int_t^1 Q_n(s)\,ds,\ \ 0 < t < 1.$$

Changing variables in equation (4.8) yields the alternative following form to the CTE,

$$CTE(t) = \frac{1}{1-t}\int_{Q(t)}^{\infty} x dF(x) = \int_0^{\infty} \frac{x}{1-t}\mathbf{1}\{Q(t) < x\}\, dF(x).$$

Now, we use Kaplan and Meier df estimator $F_n(x)$ and the corresponding quantile function estimator $Q_n(t)$ to propose, in the spirit of the construction in [Stute(1995)] (page 423). An alternative formulation for the nonparametric estimator of the CTE, is given by:

$$\widetilde{CTE}(t) = \sum_{i=1}^{n} W_{i:n}\varphi(T_{i:n}),$$

where $W_{i:n}$ is defined in (4.6) and

$$\varphi(T_{i:n}) := \frac{T_{i:n}}{1-t}\mathbf{1}\{Q_n(t) < T_{i:n}\}.$$

The simulation results regarding the performance of $\widehat{CTE}(t)$ are summarized in Tables 4.2-4.9. For each censoring scenario, we have two tables corresponding to the different values of the extreme value index. Those pertaining to the comparison (where we only consider the censoring scheme of Senario 1 with $\gamma_1 = 0.5$) between $\widehat{CTE}(t)$ and $\widetilde{CTE}(t)$ are given in Tables 4.11 and 4.12. Graphically, we illustrate the behavior of the estimator $\widehat{CTE}(0.9)$, as a function of the number k of upper order statistics, in Figures 4.2, 4.3, 4.4 and 4.5 corresponding to $\gamma_1 = 0.3$, $\gamma_1 = 0.5$, $\gamma_1 = 0.6$ and $\gamma_1 = 0.8$ respectively.

Conclusion

In light of the simulation results, we may conclude that:

1. regardless of the censoring scenario, the biases and rmse's in Tables 4.2-4.9 indicate that the estimation accuracy increases for:

- larger sample size.
- minor censoring percentage.
- lower security level t.
- smaller tail index value.

2. Figures 4.2, 4.3, 4.4 and 4.5 show the influence of the censoring proportion and the extreme value index on the estimation results.

3. After a great number of trials, we found that, as the value of γ_1 moves beyond $1/2$, our estimation procedure gives unsatisfactory results mainly when the censoring proportion is high as it could be seen in Figures 4.4 and 4.5. This is confirmed by the results of Table 4.10, where we only consider the situation of weak censoring $(p = 0.90)$.

4. The results of Tables 4.11 and 4.12 show that our (semiparametric) estimator $\widehat{CTE}(t)$ performs better then the nonparametric $\widetilde{CTE}(t)$, especially for larger sample sizes.

4.6.2 Case studies

Our estimation procedure is applied to sets of real data, namely losses of an insurance company in the United States and survival time of Aids patients in Australia.

t	0.90			0.95		
$CTE\,(t)$	2.834			3.500		
n	$\widehat{CTE}\,(t)$	bias	rmse	$\widehat{CTE}\,(t)$	bias	rmse
			$p = 0.50$			
1000	2.972	0.139	0.591	3.813	0.313	1.288
1500	2.955	0.121	0.425	3.745	0.245	0.805
2000	2.937	0.103	0.397	3.677	0.176	0.633
			$p = 0.70$			
1000	2.863	0.029	0.310	3.552	0.052	0.521
1500	2.848	0.014	0.231	3.540	0.040	0.406
2000	2.845	0.011	0.200	3.538	0.037	0.352
			$p = 0.90$			
1000	2.847	0.013	0.176	3.513	0.013	0.319
1500	2.840	0.007	0.154	3.509	0.009	0.265
2000	2.836	0.002	0.121	3.505	0.005	0.221

Table 4.2: CTE estimates based on 1000 right censored samples of size n from scenario 1 with $\gamma_1 = 0.3$

Insurance loss

The data, collected by the Insurance Services Office, Inc., are available, under the name "loss", in the package copula of the statistical software R. The datafile, which is made up with two main variables, namely the loss (indemnity payment) and the ALAE (allocated loss adjustment expenses), was processed by several authors as, for instance, [Frees and Valdez(1998)], [Klugman and Parsa(1999)] and [Denuit et al.(2006)], who were interested in modeling the joint distribution of the couple loss-ALAE. In our study, we focus on the first variable consisting of 1500 observations, of which 34 are censored. The censoring variable is defined by the values of the policy limit pertaining to each loss. The censorship occurs when the size of the loss exceeds this limit. In insurance, a policy limit, specific to each contract, is upper-bounded by a maximum claim amount that a company could pay (see [Denuit et al.(2006)]). A summary of the elementary statistics related to the loss can be found in Table 4 of [Frees and Valdez(1998)]. The heavy-tailed nature of the data is checked through two empirical criteria. First, we compute the sample variance

t	0.90			0.95		
$CTE(t)$	2.832			3.498		
n	$\widehat{CTE}(t)$	bias	rmse	$\widehat{CTE}(t)$	bias	rmse
			$p = 0.50$			
1000	2.946	0.113	0.610	3.765	0.267	1.141
1500	2.922	0.090	0.459	3.664	0.166	0.747
2000	2.917	0.085	0.356	3.638	0.140	0.596
			$p = 0.70$			
1000	2.847	0.015	0.291	3.540	0.042	0.523
1500	2.846	0.014	0.245	3.532	0.034	0.408
2000	2.845	0.012	0.204	3.527	0.029	0.346
			$p = 0.90$			
1000	2.840	0.008	0.224	3.513	0.015	0.367
1500	2.838	0.006	0.175	3.512	0.014	0.294
2000	2.833	0.001	0.146	3.504	0.006	0.261

Table 4.3: CTE estimates based on 1000 right censored samples of size n from scenario 2 with $\gamma_1 = 0.3$

of the available (uncensored) loss values and we find an extremely large value (in billions). Second, we plot the (ordered) claim amount and their boxplot in Figure 4.6 which clearly shows the existence of extremes in the dataset.

By applying the adaptive algorithm of [Reiss and Thomas(2007)], we find the optimal sample fraction $k_{opt} = 52$. The corresponding estimates of γ_1 and p are $\widehat{\gamma}_1^{(c)} = 0.63$ and $\widehat{p} = 0.77$. The VaR's of levels 0.90 and 0.95 are respectively equal to 100000 and 200000 US$ and the means of the losses which exceed these amounts are estimated to be 349522 and 562686 US$ respectively.

Aids survival time

The data, known as Australian Aids data, are provided by Dr P.J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research. The datafile is available under the name "Aids" in the package MASS of the statistical software R. They consist of medical observations on 2843 patients (among whom 2754 are male), diagnosed with

t	0.90			0.95		
$CTE(t)$	2.834			3.500		
n	$\widehat{CTE}(t)$	bias	rmse	$\widehat{CTE}(t)$	bias	rmse
			$p = 0.50$			
1000	2.982	0.149	0.591	3.784	0.284	1.053
1500	2.951	0.118	0.469	3.723	0.223	0.865
2000	2.947	0.114	0.409	3.683	0.183	0.638
			$p = 0.70$			
1000	2.862	0.029	0.297	3.559	0.059	0.577
1500	2.854	0.021	0.262	3.556	0.055	0.451
2000	2.853	0.019	0.204	3.536	0.036	0.372
			$p = 0.90$			
1000	2.839	0.006	0.218	3.488	0.012	0.371
1500	2.837	0.003	0.165	3.512	0.012	0.294
2000	2.832	0.002	0.147	3.510	0.010	0.255

Table 4.4: CTE estimates based on 1000 right censored samples of size n from scenario 3 with $\gamma_1 = 0.3$

Aids in Australia before July 1^{st}, 1991. Of these patients, 1761 have died, while the remaining survival times are right-censored. In this study, we only consider male patients due to the small number of women (89 patients). In the literature, these data were analyzed with different prospects by several authors like, for instance, [Ripley and Solomon(1994)] and [Venables and Ripley(2002)] (pages 379-385), [Einmahl et al.(2008)], [Ndao et al.(2014)] and [Stupfler(2016)]. Recently, [Goegebeur et al.(2019)] investigated and described the conditional distribution using survival data of AIDS patients. To check the heavy-tailed nature of the data, we proceed as we did for the insurance loss dataset. On the one hand, the value of the sample variance is around 130000, which is very high and on the other hand, Figure 4.7 clearly indicates that extreme survival times do exist among the data. Analogous steps as those in the insurance example above lead to $k_{opt} = 211$, $\widehat{\gamma}_1^{(c)} = 0.75$ and $\widehat{p} = 0.35$. The respective VaR's (i.e., maximum survival time) of levels 0.90 and 0.95 are estimated to be approximately 5 and 7 years and the estimates of the mean of survival times exceeding these thresholds are about 18 and 31 years repectively.

t	0.90			0.95		
$CTE(t)$	2.832			3.498		
n	$\widehat{CTE}(t)$	bias	rmse	$\widehat{CTE}(t)$	bias	rmse
$p = 0.50$						
1000	2.933	0.101	0.544	3.751	0.253	1.036
1500	2.919	0.086	0.405	3.673	0.174	0.849
2000	2.893	0.061	0.341	3.632	0.133	0.638
$p = 0.70$						
1000	2.849	0.017	0.277	3.526	0.028	0.513
1500	2.847	0.014	0.240	3.523	0.025	0.431
2000	2.841	0.009	0.199	3.515	0.017	0.356
$p = 0.90$						
1000	2.839	0.006	0.174	3.517	0.018	0.313
1500	2.829	0.003	0.143	3.512	0.014	0.269
2000	2.833	0.001	0.126	3.507	0.009	0.233

Table 4.5: CTE estimates based on 1000 right censored samples of size n from scenario 4 with $\gamma_1 = 0.3$

t	0.90			0.95		
$CTE(t)$	6.321			8.943		
n	$\widehat{CTE}(t)$	bias	rmse	$\widehat{CTE}(t)$	bias	rmse
$p = 0.50$						
1000	7.192	0.871	5.118	10.518	1.575	8.956
1500	6.756	0.435	3.190	10.133	1.190	6.924
2000	6.650	0.329	1.837	9.614	0.671	3.854
$p = 0.70$						
1000	6.555	0.233	2.172	9.426	0.483	3.695
1500	6.425	0.104	1.698	9.140	0.197	2.591
2000	6.416	0.094	1.108	9.045	0.101	1.899
$p = 0.90$						
1000	6.395	0.074	1.628	9.058	0.115	2.033
1500	6.360	0.039	1.169	9.024	0.081	1.536
2000	6.338	0.017	0.667	9.005	0.062	1.301

Table 4.6: CTE estimates based on 1000 right censored samples of size n from scenario 1 with $\gamma_1 = 0.5$

t	0.90			0.95		
$CTE\,(t)$	6.271			8.907		
n	$\widehat{CTE}\,(t)$	bias	rmse	$\widehat{CTE}\,(t)$	bias	rmse
			$p = 0.50$			
1000	7.277	1.006	4.067	10.555	1.648	9.807
1500	7.020	0.749	3.324	10.528	1.621	7.041
2000	6.996	0.725	2.346	10.063	1.156	3.888
			$p = 0.70$			
1000	6.590	0.319	2.391	9.608	0.701	5.003
1500	6.513	0.242	1.365	9.238	0.332	2.779
2000	6.488	0.217	1.114	9.306	0.399	2.151
			$p = 0.90$			
1000	6.386	0.115	1.267	9.121	0.214	2.283
1500	6.366	0.095	0.976	9.117	0.210	1.867
2000	6.354	0.083	0.785	9.028	0.121	1.458

Table 4.7: CTE estimates based on 1000 right censored samples of size n from scenario 2 with $\gamma_1 = 0.5$

t	0.90			0.95		
$CTE\,(t)$	6.321			8.943		
n	$\widehat{CTE}\,(t)$	bias	rmse	$\widehat{CTE}\,(t)$	bias	rmse
			$p = 0.50$			
1000	6.985	0.663	5.743	10.162	1.219	8.385
1500	9.149	0.206	4.032	9.743	0.800	5.393
2000	6.518	0.197	2.763	9.546	0.603	3.797
			$p = 0.70$			
1000	6.538	0.217	1.749	9.285	0.341	3.334
1500	6.498	0.176	1.599	9.165	0.222	2.374
2000	6.409	0.088	1.100	9.071	0.128	1.857
			$p = 0.90$			
1000	6.403	0.082	1.113	9.142	0.198	2.132
1500	6.383	0.062	0.888	9.051	0.108	1.856
2000	6.336	0.014	0.759	9.017	0.074	1.391

Table 4.8: CTE estimates based on 1000 right censored samples of size n from scenario 3 with $\gamma_1 = 0.5$

t	0.90			0.95		
$CTE\,(t)$	6.271			8.907		
n	$\widehat{CTE}\,(t)$	bias	rmse	$\widehat{CTE}\,(t)$	bias	rmse
			$p = 0.50$			
1000	7.616	1.345	5.233	10.854	1.948	9.542
1500	6.977	0.706	2.629	10.631	1.724	8.542
2000	6.960	0.689	2.337	10.234	1.328	4.177
			$p = 0.70$			
1000	6.683	0.413	2.028	9.711	0.805	4.529
1500	6.516	0.245	1.369	9.371	0.465	2.705
2000	6.472	0.201	1.079	9.293	0.386	2.106
			$p = 0.90$			
1000	6.362	0.091	1.125	9.093	0.186	1.940
1500	6.337	0.066	0.823	9.085	0.178	1.569
2000	6.303	0.032	0.689	9.057	0.150	1.364

Table 4.9: CTE estimates based on 1000 right censored samples of size n from scenario 4 with $\gamma_1 = 0.5$

γ_1	0.6			0.8		
CTE	9.951			31.548		
n	$\widehat{CTE}$	bias	rmse	$\widehat{CTE}$	bias	rmse
1000	10.502	0.550	5.633	22.791	8.756	244.359
1500	10.289	0.338	2.915	37.652	6.105	142.360
2000	10.230	0.279	2.087	34.807	3.259	136.389

Table 4.10: CTE estimates based on 1000 right censored samples of size n from scenario 3 with p=0.90 and t=0.90

t	0.90					
$CTE\,(t)$	6.321					
n	$\widehat{CTE}\,(t)$	bias	rmse	$\widetilde{CTE}\,(t)$	bias	rmse
$p = 0.50$						
1000	6.878	0.557	4.941	3.879	2.442	3.059
1500	6.806	0.485	3.127	4.121	2.201	2.700
2000	6.791	0.470	2.084	4.237	2.084	2.510
$p = 0.70$						
1000	6.568	0.246	1.695	5.509	0.812	1.840
1500	6.483	0.161	1.302	5.648	0.674	1.831
2000	6.443	0.122	1.029	5.699	0.622	1.182
$p = 0.90$						
1000	6.428	0.106	1.089	6.217	0.105	1.485
1500	6.386	0.064	0.864	6.221	0.101	1.129
2000	6.367	0.045	0.674	6.22	0.101	0.885

Table 4.11: Comparison results between $\widehat{CTE}\,(t)$ and $\widetilde{CTE}\,(t)$ based on 1000 right censored samples of size n from scenario 1 with $\gamma_1 = 0.5$

t	0.95					
$CTE\,(t)$	8.943					
n	$\widehat{CTE}\,(t)$	bias	rmse	$\widetilde{CTE}\,(t)$	bias	rmse
$p = 0.50$						
1000	10.483	1.540	8.512	3.955	4.988	5.753
1500	10.086	1.143	6.206	4.364	4.580	5.346
2000	9.788	0.845	4.472	4.724	4.219	4.957
$p = 0.70$						
1000	9.471	0.528	4.343	7.090	1.853	3.116
1500	9.284	0.341	2.231	7.494	1.449	2.559
2000	9.075	0.132	1.924	7.515	1.428	2.349
$p = 0.90$						
1000	9.169	0.225	2.095	8.627	0.316	2.105
1500	9.056	0.113	1.561	8.641	0.302	1.686
2000	9.023	0.080	1.360	8.674	0.269	1.633

Table 4.12: Comparison results between $\widehat{CTE}\,(t)$ and $\widetilde{CTE}\,(t)$ based on 1000 right censored samples of size n from scenario 1 with $\gamma_1 = 0.5$

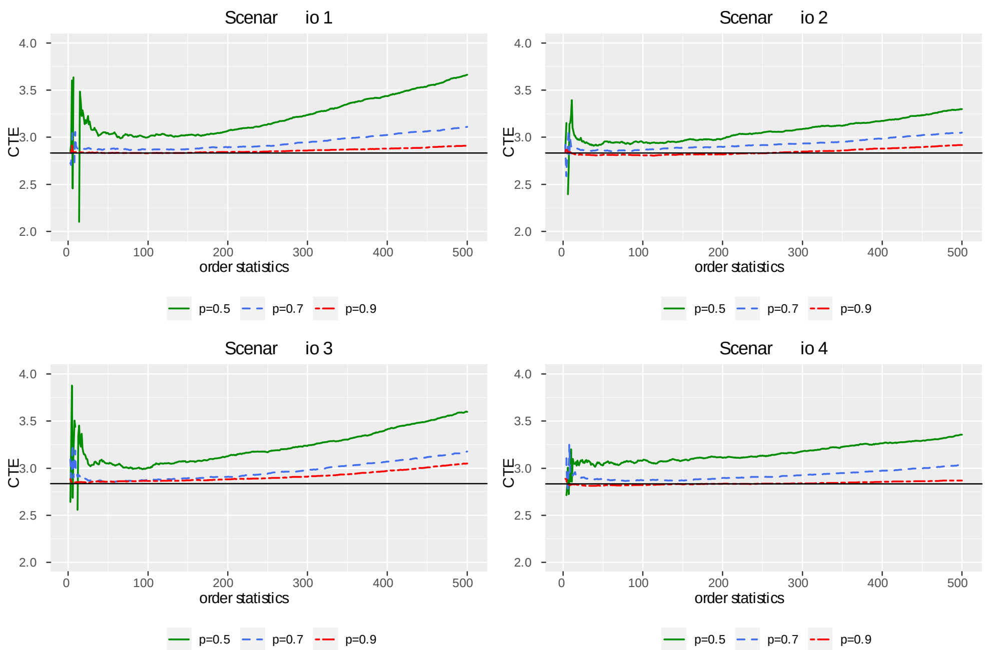

Figure 4.2: Plots of CTE estimator (based on 100 samples of size 2000) for $\gamma_1 = 0.3$ and $t = 0.9$ as functions of the number k of upper order statistics. The horizontal line represents the true value of CTE

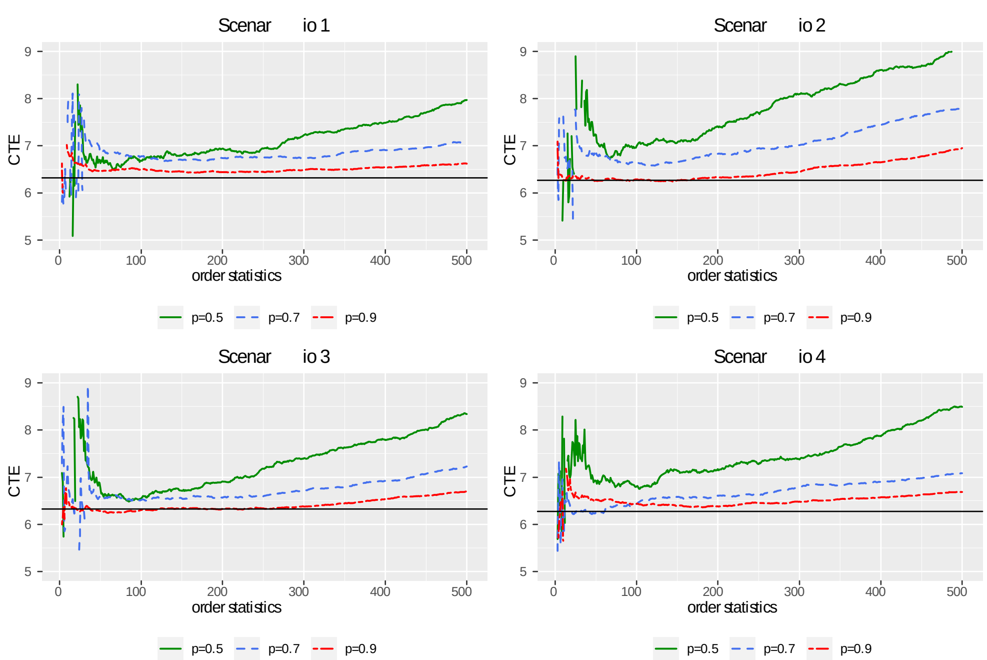

Figure 4.3: Plots of CTE estimator (based on 100 samples of size 2000) for $\gamma_1 = 0.5$ and $t = 0.9$ as functions of the number k of upper order statistics. The horizontal line represents the true value of CTE

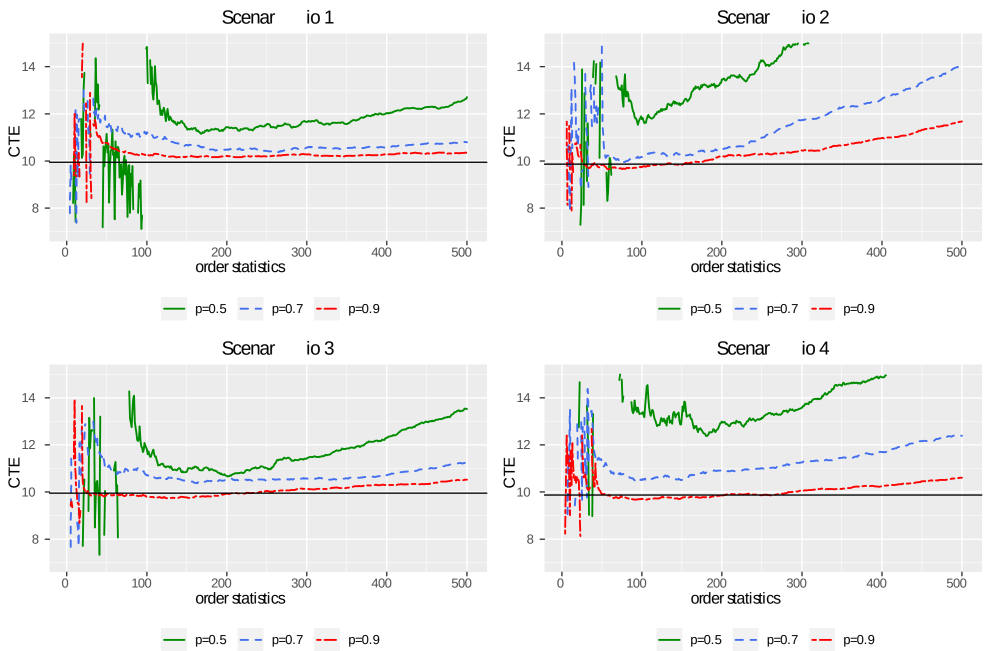

Figure 4.4: Plots of CTE estimator (based on 100 samples of size 2000) for $\gamma_1 = 0.6$ and $t = 0.9$ as functions of the number k of upper order statistics. The horizontal line represents the true value of CTE

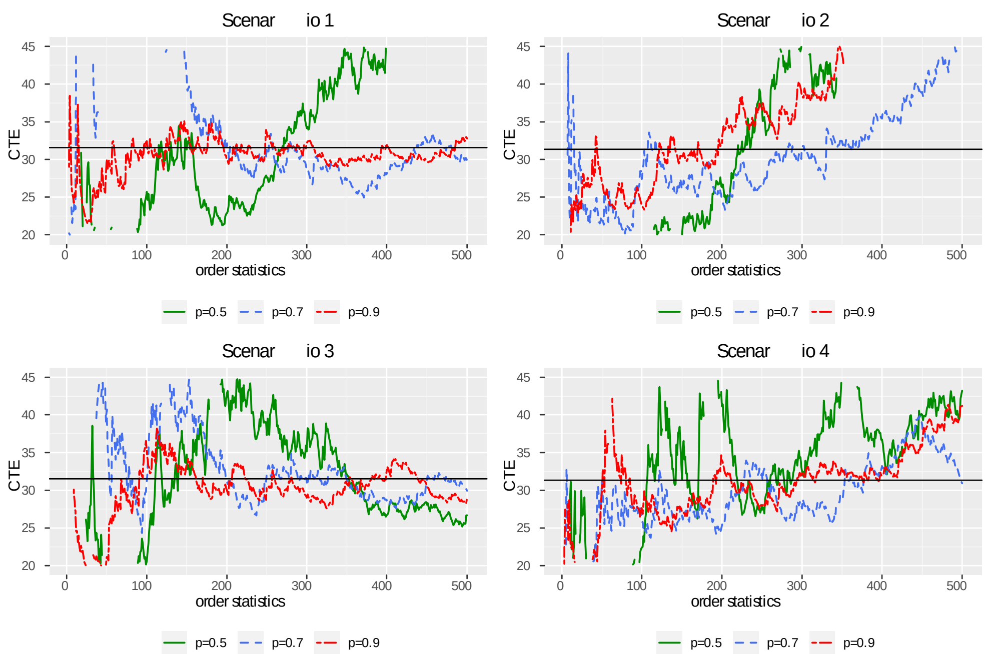

Figure 4.5: Plots of CTE estimator (based on 100 samples of size 2000) for $\gamma_1 = 0.8$ and $t = 0.9$ as functions of the number k of upper order statistics. The horizontal line represents the true value of CTE

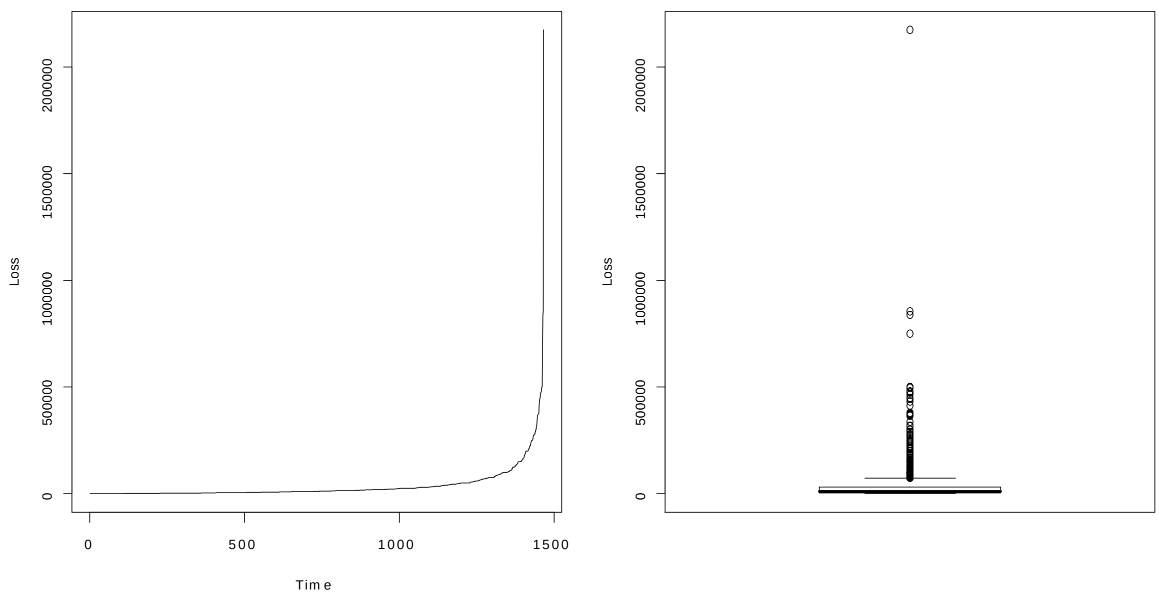

Figure 4.6: Plots of the ordered amounts (left) and boxplot (right) of insurance losses

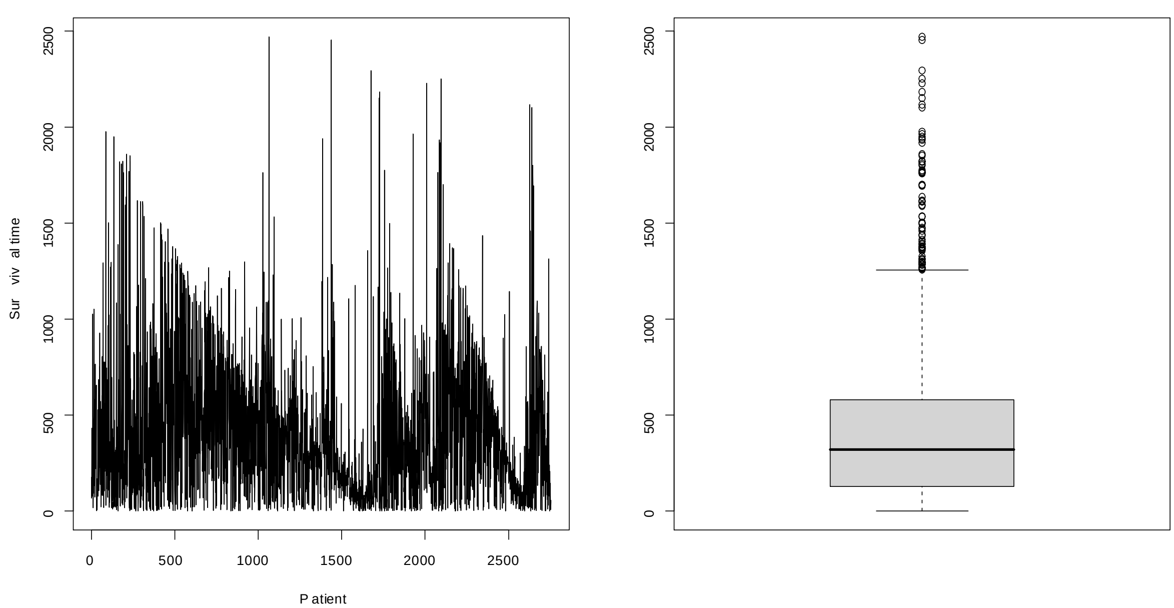

Figure 4.7: Plots of the survival time to Aids (left) and boxplot (right) of Australian male patients

Chapter 5

Asymptotic distribution of the CTE estimator

This chapter examines the asymptotic normality of our proposed CTE estimator, as presented in (4.10), in the case of a randomly censored heavy-tailed distribution, and provides a complete proof of this result.

We suppose both F and G are heavy-tailed or equivalently that $\overline{F}$ and $\overline{G}$ are regularly varying at infinity. That is, there are two tail indices $\gamma_1 > 0$ and $\gamma_2 > 0$, such that for any $x > 0$, we have

$$\lim_{v \to \infty} \frac{\overline{F}(vx)}{\overline{F}(v)} = x^{-1/\gamma_1} \quad \text{and} \quad \lim_{v \to \infty} \frac{\overline{G}(vx)}{\overline{G}(v)} = x^{-1/\gamma_2}. \tag{5.1}$$

Let H denote the df of T, then the independence of X and C yields that the survival function of H is equal to the product of the tails of F and G. That is, we have $\overline{H} = \overline{F} \times \overline{G}$, which yields that H is also heavy-tailed, with extreme value index $\gamma := \gamma_1\gamma_2/(\gamma_1 + \gamma_2)$. We also assume that both F and G belong to the Hall's models (see [Hall(1982)]). These models are defined, as $x \to \infty$, by

$$\begin{aligned} \overline{F}(x) &= l_1 x^{-1/\gamma_1} \left(1 + m_1 x^{\tau_1/\gamma_1} (1 + o(1))\right), \\ \overline{G}(x) &= l_2 x^{-1/\gamma_2} \left(1 + m_2 x^{\tau_2/\gamma_2} (1 + o(1))\right), \end{aligned} \tag{5.2}$$

where $m_i \in \mathbb{R}$, $l_i > 0$ and $\tau_i < 0$, $i = 1, 2$. The latter are called second-order parameters of df's F and G respectively. From (5.2), we infer that H belongs to Hall's class as well. That is

$$\overline{H}(x) = lx^{-1/\gamma}\left(1 + mx^{\tau/\gamma}(1 + o(1))\right), \text{ as } x \to \infty, \tag{5.3}$$

where $l := l_1 l_2$, $\tau := \min(\tau_1, \tau_2)$ and

$$m := m_1 \mathbf{1}\{\tau_1 < \tau_2\} + m_2 \mathbf{1}\{\tau_1 > \tau_2\} + (m_1 + m_2)\mathbf{1}\{\tau_1 = \tau_2\}.$$

In the sequel, we use the representation $\overline{K}(x) := K(\infty) - K(x)$, $x > 0$, for any function K.

Before we present our main result, it is important to note that the asymptotic normality of $\widehat{CTE}(t)$ is established under the same conditions as those outlined by [Soltane et al.(2015)]. Indeed, in our context, since $CTE(t)$ may be rewritten as $\int_{Q(t)}^{\infty} x dF(x)/(1-t)$, the integrals in (4.7) become of the form

$$I_1 := \frac{1}{1-t}\int_{Q(t)}^{\infty} x^2 \kappa^2(x)\, dH^{(1)}(x) \quad \text{and} \quad I_2 := \frac{1}{1-t}\int_{Q(t)}^{\infty} x\left(\int_0^x \frac{dG(z)}{\overline{H}(z)\,\overline{G}(z)}\right)^{1/2} dF(x). \tag{5.4}$$

In Lemma (5.3.1), we will show that, in the case of Pareto-like distributions, the contrary of (4.7), i.e. I_1 or I_2 are infinite, is also satisfied under the condition $\gamma_2/(1+2\gamma_2) < \gamma_1 < 1$.

5.1 Main result

Weak approximations of EVT based statistics are achieved in the second-order framework. Thus, it seems quite natural to suppose that df's F and G satisfy the well-known second-order condition of regular variation, which specify the convergence rates in (5.1). That is, we assume that there exist two constants $\tau_j \leq 0$ and two functions A_j, $j = 1, 2$, tending to zero and not changing sign near infinity, such that for any $x > 0$, we have

$$\lim_{v\to\infty} \frac{\overline{F}(vx)/\overline{F}(v) - x^{-1/\gamma_1}}{A_{\mathbf{1}}(v)} = x^{-1/\gamma_1} \frac{x^{\tau_1/\gamma_1} - 1}{\gamma_1 \tau_1}, \tag{5.5}$$

and

$$\lim_{v\to\infty} \frac{\overline{G}(vx)/\overline{G}(v) - x^{-1/\gamma_2}}{A_{\mathbf{2}}(v)} = x^{-1/\gamma_2} \frac{x^{\tau_2/\gamma_2} - 1}{\gamma_2 \tau_2}, \tag{5.6}$$

where $A_{\mathbf{i}}(v) = m_i \gamma_i \tau_i v^{\tau_i/\gamma_i}$, $i = 1, 2$.

Theorem 5.1.1 *Assume that the second-order conditions of regular variation (5.5) and (5.6) hold, with $0 < \gamma_1 < \gamma_2$ and $\gamma_2/(1+2\gamma_2) < \gamma_1 < 1$. Let $k = k_n$ be an integer sequence satisfying (2.7) and $h = h_n := H^{-1}(1 - k/n)$ be such that $\sqrt{k} A_1(h) \to \lambda \in \mathbb{R}$ and $\sqrt{k} h \overline{F}(h) \to \infty$, as $n \to \infty$. Then, for any fixed $t \in (0,1)$, we have*

$$\frac{\sqrt{k}\left(\widehat{CTE}(t) - CTE(t)\right)}{h\overline{F}(h)} (1-t) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu, \sigma^2\right), \ \text{as } n \to \infty,$$

where

$$\mu = \mu(\lambda, p, \gamma_1, \tau_1) := \frac{\lambda}{(1-\gamma_1)} \left(\frac{1}{(1-\gamma_1)(1-p\tau_1)} + \frac{1}{(\gamma_1 + \tau_1 - 1)} \right), \tag{5.7}$$

and

$$\sigma^2 = \sigma^2(p, \gamma_1) := \frac{2p\gamma_1^2 (1 - 3p + 3p^2 + 2p\gamma_1 - 4p^2\gamma_1 + p^2\gamma_1^2)}{(1-\gamma_1)^2 (1 - p + p\gamma_1)(1 - 2p + 2p\gamma_1)} + \frac{2\gamma_1^2 (1-p)}{(1-\gamma_1)^3} + \frac{\gamma_1^2}{p(1-\gamma_1)^4}. \tag{5.8}$$

Remark 5.1.1 *The assumptions of Theorem 5.1.1 are justified in this remark. The condition $\sqrt{k} A_1(h) \to \lambda$ is a standard requirement of extreme value theory needed in the computation of the asymptotic bias. The hypothesis $\sqrt{k} h \overline{F}(h) \to \infty$ makes the rest term $O_{\mathbb{P}}(1/k)$, in the asymptotic representation (5.15) of Kaplan-Meier estimator, negligible. The assumption $\gamma_1 < \gamma_2$ is required so that the term (5.33), in the first integral in (5.14),* *be independent of the level t. This will be useful in the computation of the asymptotic variance σ^2. In terms of the censorship proportion, this hypothesis is equivalent to $p > 1/2$.*

Remark 5.1.2 *Note that if $\lambda = 0$, then the limiting distribution in Theorem 5.1.1 is centered.*

Remark 5.1.3 *In practice, μ and σ^2 are computed by replacing the parameters by their respective estimates in formulas (5.7) and (5.8).*

5.2 Proof

First, we define, for $j = 0, 1$, the following very crucial subdistribution functions:

$$H^{(j)}(v) := P(T \leq v, \delta = j), \ v \geq 0, \tag{5.9}$$

and their sample counterparts

$$H_n^{(j)}(v) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{T_i \leq v, \delta_i = j\}. \tag{5.10}$$

Then, we have $H(v) = H^{(0)}(v) + H^{(1)}(v)$ and $\overline{H}(v) = \overline{H}^{(0)}(v) + \overline{H}^{(1)}(v)$. The same equalities hold empirically. For $i = 1, ..., n$, let

$$U_i := \delta_i H^{(1)}(T_i) + (1 - \delta_i)\left(\theta + H^{(0)}(T_i)\right), \text{ with } \theta := H^{(1)}(\infty),$$

be iid $(0,1)$-uniform r.v's (see [Einmahl and Koning(1992)]). The uniform empirical df and the uniform empirical process based on $U_1, ..., U_n$ are respectively denoted by

$$\mathbb{L}_n(s) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{U_i \leq s\} \text{ and } \alpha_n(s) := \sqrt{n}\,(\mathbb{L}_n(s) - s), \ 0 \leq s \leq 1.$$

The empirical processes

$$\beta_n^{(j)}(v) := \sqrt{n}\left(\overline{H}_n^{(j)}(v) - \overline{H}^{(j)}(v)\right), j = 0, 1, \tag{5.11}$$

may be represented, almost surely, by a uniform empirical process. Indeed, [Deheuvels and Einmahl(1996)] state that almost surely

$$H_n^{(0)}(v) = \mathbb{L}_n\left(H^{(0)}(v) + \theta\right) - \mathbb{L}_n(\theta), \text{ for } 0 < H^{(0)}(v) < 1 - \theta,$$

and

$$H_n^{(1)}(v) = \mathbb{L}_n\left(H^{(1)}(v)\right), \text{ for } 0 < H^{(1)}(v) < \theta.$$

Then, it readily checked that, almost surely, we have

$$\beta_n^{(1)}(v) = \alpha_n(\theta) - \alpha_n\left(\theta - \overline{H}^{(1)}(v)\right), \text{ for } 0 < \overline{H}^{(1)}(v) < \theta,$$

and

$$\beta_n^{(0)}(v) = -\alpha_n\left(1 - \overline{H}^{(0)}(v)\right), \text{ for } 0 < \overline{H}^{(0)}(v) < 1 - \theta.$$

Our proof highly relies on the famous Gaussian approximation due to [Csörgő et al.(1986)] in Corollary 2.1. There exists a sequence of Brownian bridges $\{B_n(s);\ 0 \le s \le 1\}$ on the probability space $(\Omega, \mathcal{F}, P)$, such that, for every $0 \le \omega < 1/4$,

$$\sup_{1/n \le s \le 1} \frac{n^{\omega}\left|\alpha_n(1-s) - B_n(1-s)\right|}{s^{1/2-\omega}} = O_{\mathbb{P}}(1). \tag{5.12}$$

For the increments $\alpha_n(\theta) - \alpha_n(\theta - s)$, we will need an approximation of the same type as (5.12). Following analogous arguments as those used in the proofs of assertions (2.2) and (2.8) of Theorems 2.1 and 2.2 respectively in [Csörgő et al.(1986)], we may show that, for every $0 < \theta < 1$ and $0 \leq \omega < 1/4$, we have

$$\sup_{1/n \leq s \leq \theta} \frac{n^{\omega} \left| \left[\alpha_n(\theta) - \alpha_n(\theta - s) \right] - \left[B_n(\theta) - B_n(\theta - s) \right] \right|}{s^{1/2-\omega}} = O_{\mathbb{P}}(1). \tag{5.13}$$

The Gaussian processes

$$\mathbf{B}_n(v) := B_n(\theta) - B_n\left(\theta - \overline{H}^{(1)}(v)\right), \text{ for } 0 < \overline{H}^{(1)}(v) < \theta,$$

and

$$\mathbf{B}_n^*(v) := \mathbf{B}_n(v) - B_n\left(1 - \overline{H}^{(0)}(v)\right), \text{ for } 0 < \overline{H}^{(0)}(v) < 1 - \theta,$$

will be critical to our requirements.

The actual starting point of the proof is to rewrite the estimator $\widehat{CTE}(t)$ into another (equivalent) form. A change of variables and an integration by parts in (4.8), yield

$$(1-t)CTE(t) = (1-t)Q(t) + \int_{Q(t)}^{\infty} \overline{F}(x)dx,$$

which we decompose the sum of

$$L_{n,1}(t) := (1-t)Q(t) + \int_{Q(t)}^{h} \overline{F}(x)dx \text{ and } L_{n,2} := \int_{h}^{\infty} \overline{F}(x)dx.$$

By a change of variables in $L_{n,2}$, that does not depend on t, we have

$$L_{n,2} = h\overline{F}(h) \int_{1}^{\infty} \frac{\overline{F}(hx)}{\overline{F}(h)} dx,$$

which, according to the well-known Karamata theorem (see, for instance, page 363 in [de Haan and Ferreira(2006)]), is equivalent, for $0 < \gamma_1 < 1$, to $h\overline{F}(h)\gamma_1/(1-\gamma_1)$, as $n \to \infty$. The difference between the CTE and its estimator is decomposed as

$$(1-t)\left(\widehat{CTE}(t) - CTE(t)\right) = \left(\hat{L}_{n,1}(t) - L_{n,1}(t)\right) + \left(\hat{L}_{n,2} - L_{n,2}\right),$$

where

$$\hat{L}_{n,1}(t) := (1-t)Q_n(t) + \int_{Q_n(t)}^{T_{n-k:n}} \overline{F}_n(x)dx \text{ and } \hat{L}_{n,2} := \frac{\hat{\gamma}_1^{(c)}}{1-\hat{\gamma}_1^{(c)}} T_{n-k:n}\overline{F}_n(T_{n-k:n}).$$

It is easy to check that

$$\begin{aligned}\hat{L}_{n,1}(t) - L_{n,1}(t) &= \int_{Q_n(t)}^{T_{n-k:n}} \left(\overline{F}_n(x) - \overline{F}(x)\right) dx - \int_{T_{n-k:n}}^{h} \overline{F}(x)dx \\ &+ \int_{Q_n(t)}^{Q(t)} \overline{F}(x)dx + (1-t)\left(Q_n(t) - Q(t)\right).\end{aligned} \tag{5.14}$$

By Proposition 5 together with equation (4.9) in [Csörgő(1996)], we have, for any real number $x \leq T_{n-k:n}$,

$$\begin{aligned}\frac{\overline{F}_n(x) - \overline{F}(x)}{\overline{F}(x)} &= \int_0^x \frac{d\left(\overline{H}_n^{(1)}(v) - \overline{H}^{(1)}(v)\right)}{\overline{H}(v)} \\ &- \int_0^x \frac{\overline{H}_n(v) - \overline{H}(v)}{\overline{H}^2(v)} d\overline{H}^{(1)}(v) + O_{\mathbb{P}}\left(\tfrac{1}{k}\right),\end{aligned} \tag{5.15}$$

which, after integrating the first integral by parts, becomes

$$\begin{aligned}\frac{\overline{F}_n(x) - \overline{F}(x)}{\overline{F}(x)} &= \frac{\overline{H}_n^{(1)}(x) - \overline{H}^{(1)}(x)}{\overline{H}(x)} - \left(\overline{H}_n^{(1)}(0) - \overline{H}^{(1)}(0)\right) \\ &+ \int_0^x \frac{\overline{H}_n^{(1)}(v) - \overline{H}^{(1)}(v)}{\overline{H}^2(v)} d\overline{H}(v) - \int_0^x \frac{\overline{H}_n(v) - \overline{H}(v)}{\overline{H}^2(v)} d\overline{H}^{(1)}(v) \\ &+ O_{\mathbb{P}}\left(\frac{1}{k}\right).\end{aligned}$$

By combining (5.9) and (5.10) with the definition (5.11) of the processes $\beta_n^{(j)}(v)$, we get

$$\sqrt{n}\left(\overline{H}_n(v) - \overline{H}(v)\right) = \beta_n^{(1)}(v) + \beta_n^{(0)}(v).$$

On the other hand, by the classical central limit theorem, we have

$$\overline{H}_n^{(1)}(0) - \overline{H}^{(1)}(0) = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

Therefore, we have

$$\begin{aligned}\frac{\overline{F}_n(x) - \overline{F}(x)}{\overline{F}(x)} &= \frac{1}{\sqrt{n}}\frac{\beta_n^{(1)}(x)}{\overline{H}(x)} + \frac{1}{\sqrt{n}}\int_0^x \frac{\beta_n^{(1)}(v)}{\overline{H}^2(v)} d\overline{H}(v) \\ &- \frac{1}{\sqrt{n}}\int_0^x \frac{\beta_n^{(1)}(v) + \beta_n^{(0)}(v)}{\overline{H}^2(v)} d\overline{H}^{(1)}(v) \\ &+ O_{\mathbb{P}}\left(\frac{1}{k}\right) + O_{\mathbb{P}}\left(n^{-1/2}\right).\end{aligned}$$

Now, we consider the following decomposition

$$\frac{\sqrt{k}\left(\hat{L}_{n,1}(t) - L_{n,1}(t)\right)}{h\overline{F}(h)} = \sum_{i=1}^{7} J_{n,i},$$

with

$$J_{n,1} := d_n \int_{Q_n(t)}^{T_{n-k:n}} \frac{\beta_n^{(1)}(x)}{\overline{H}(x)} \overline{F}(x)dx,$$

$$J_{n,2} := d_n \int_{Q_n(t)}^{T_{n-k:n}} \left\{\int_0^x \frac{\beta_n^{(1)}(v)}{\overline{H}^2(v)} d\overline{H}(v)\right\} \overline{F}(x)dx,$$

$$J_{n,3} := -d_n \int_{Q_n(t)}^{T_{n-k:n}} \left\{\int_0^x \frac{\beta_n^{(1)}(v) + \beta_n^{(0)}(v)}{\overline{H}^2(v)} d\overline{H}^{(1)}(v)\right\} \overline{F}(x)dx,$$

$$J_{n,4} := -d_n\sqrt{n} \int_{T_{n-k:n}}^{h} \overline{F}(x)dx,$$

$$J_{n,5} := d_n\sqrt{n} \int_{Q_n(t)}^{Q(t)} \overline{F}(x)dx,$$

$$J_{n,6} := d_n\sqrt{n}\,(1-t)\,(Q_n(t) - Q(t)),$$

and

$$J_{n,7} := \left(O_{\mathbb{P}}\left(\sqrt{n}/k\right) + O_{\mathbb{P}}(1)\right) d_n \int_{Q_n(t)}^{T_{n-k:n}} \overline{F}(x)dx,$$

where $d_n := \sqrt{k/n}/\left(h\overline{F}(h)\right)$. For the last term $J_{n,7}$, we check that it tends to zero in probability. To this end, we write it into the sum of

$$J_{n,7}^{(1)} := O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}h\overline{F}(h)}\right)\int_{Q_n(t)}^{T_{n-k:n}}\overline{F}(x)dx \text{ and } J_{n,7}^{(2)} := O_{\mathbb{P}}(d_n)\int_{Q_n(t)}^{T_{n-k:n}}\overline{F}(x)dx,$$

We have $\int_{Q_n(t)}^{T_{n-k:n}}\overline{F}(x)dx < \mathbf{E}[X]$ and by the assumption $\sqrt{k}h\overline{F}(h) \to \infty$, we obtain $J_{n,7}^{(1)} = o_{\mathbb{P}}(1)$, as $n \to \infty$. On the other hand, from Lemma 3 in [Hua and Joe(2011)], the second-order conditions (5.5)-(5.6) yield that, for some positive constants v_1 and v_2, we have

$$\overline{F}(x) \sim v_1 x^{-1/\gamma_1} \text{ and } \overline{G}(x) \sim v_2 x^{-1/\gamma_2}, \text{ as } x \to \infty. \tag{5.16}$$

Hence $\overline{H}(x) \sim v_1 v_2 x^{-1/\gamma}$, as $x \to \infty$, and therefore $H^{-1}(1-s) \sim (v_1 v_2)^{\gamma} s^{-\gamma}$, as $s \to 0$. It follows that $h \sim (v_1 v_2)^{\gamma}(k/n)^{-\gamma}$, as $n \to \infty$. Thus

$$d_n \sim v_1^{-1}(v_1 v_2)^{\gamma+\gamma/\gamma_1}(k/n)^{1/2+\gamma-\gamma/\gamma_1}, \text{ as } n \to \infty. \tag{5.17}$$

Since $1/2 + \gamma - \gamma/\gamma_1 > 0$ (by assumption we have $\gamma_2/(1+2\gamma_2) < \gamma_1 < 1$), then

$$d_n \to 0, \text{ as } n \to \infty. \tag{5.18}$$

and therefore $J_{n,7}^{(2)} = o_{\mathbb{P}}(1)$, as $n \to \infty$. Consequently, we have

$$J_{n,7} = o_{\mathbb{P}}(1), \text{ as } n \to \infty. \tag{5.19}$$

Now, we examine the first term $J_{n,1}$. We start by applying the Gaussian approximation (5.13), we get, for a fixed real number $0 \le \omega < 1/4$,

$$J_{n,1} = d_n \int_{Q_n(t)}^{T_{n-k:n}} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x)dx \\ + O_{\mathbb{P}}(n^{-\omega}) d_n \int_{Q_n(t)}^{T_{n-k:n}} \frac{\left[\overline{H}^{(1)}(x)\right]^{1/2-\omega}}{\overline{H}(x)} \overline{F}(x)dx, \tag{5.20}$$

provided that $1/n \leq \overline{H}^{(1)}(x) < \theta$. Indeed, we have $T_{n-k:n}/h \xrightarrow{P} 1$, as $n \to \infty$ (see, for instance, Theorem 2.1 in [Brahimi et al.(2015a)]), that is, for a fixed $0 < \epsilon < 1$, the probability of the set $S_n(\epsilon) := \{|T_{n-k:n}/h - 1| \leq \epsilon\}$ is close to 1, for sufficiently large n. Since $\overline{H}^{(1)}$ is a non-increasing function, then, in $S_n(\epsilon)$, we have $\overline{H}^{(1)}((1+\epsilon)h) \leq \overline{H}^{(1)}(T_{n-k:n})$. From Lemma 4.1 in [Brahimi et al.(2015a)], we infer that $\overline{H}^{(1)}(x) \sim p\overline{H}(x)$, as $x \to \infty$, therefore $\overline{H}^{(1)}((1+\epsilon)h) \sim p\overline{H}((1+\epsilon)h)$, as $n \to \infty$. Since $\overline{H}$ is regularly varying at infinity with index $-1/\gamma$ and $\overline{H}(h) = k/n$, thus $\overline{H}^{(1)}((1+\epsilon)h) \sim p(1+\epsilon)^{-1/\gamma} k/n$, as $n \to \infty$. Hence, we have $\overline{H}^{(1)}((1+\epsilon)h) \geq 1/n$, for all large n, i.e., $\overline{H}^{(1)}(T_{n-k:n}) \geq 1/n$. On the other hand, we have $Q_n(t) > 0$, for any fixed $t \in (0,1)$, then we get

$$\overline{H}^{(1)}(Q_n(t)) < \overline{H}^{(1)}(0) = H^{(1)}(\infty) = \theta.$$

Consequently, we have $1/n \leq \overline{H}^{(1)}(x) < \theta$, for any $Q_n(t) \leq x \leq T_{n-k:n}$ and all large n. Next, we show that the second part of (5.20) tends to zero in probability. To this end, we have, for $0 < t < 1$,

$$\int_{Q_n(t)}^{T_{n-k:n}} \frac{\left[\overline{H}^{(1)}(x)\right]^{1/2-\omega}}{\overline{H}(x)} \overline{F}(x)dx \leq \int_0^{T_{n-k:n}} \frac{\left[\overline{H}^{(1)}(x)\right]^{1/2-\omega}}{\overline{H}(x)} \overline{F}(x)dx.$$

Note that $\overline{H}^{(1)} \leq \overline{H}$, then

$$\int_0^{T_{n-k:n}} \frac{\left[\overline{H}^{(1)}(x)\right]^{1/2-\omega}}{\overline{H}(x)} \overline{F}(x)dx \leq \int_0^{T_{n-k:n}} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx,$$

which, in the set $S_n(\epsilon)$, is not greater than $\int_0^{(1+\epsilon)h} \left(\overline{F}(x)/\left[\overline{H}(x)\right]^{1/2+\omega}\right) dx$.

From Lemma 5.3.3, we conclude that

$$n^{-\omega}d_n \int_{Q_n(t)}^{T_{n-k:n}} \frac{\left[\overline{H}^{(1)}(x)\right]^{1/2-\omega}}{\overline{H}(x)} \overline{F}(x)dx \to 0, \text{ as } n \to \infty,$$

for any fixed $0 \leq \omega < 1/4$. Hence, we may write that

$$J_{n,1} = d_n \int_{Q_n(t)}^{T_{n-k:n}} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x)dx + o_{\mathbb{P}}(1), \text{ as } n \to \infty.$$

By Lemma 5.3.2, we have

$$J_{n,1} = d_n \int_{0}^{T_{n-k:n}} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x)dx + o_{\mathbb{P}}(1), \text{ as } n \to \infty,$$

which, according to [Soltane et al.(2015)] (pages 15-16), may be rewritten as

$$J_{n,1} = d_n \int_{0}^{h} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x)dx + o_{\mathbb{P}}(1), \text{ as } n \to \infty. \tag{5.21}$$

The same arguments as the above lead to

$$J_{n,2} = d_n \int_{0}^{h} \left[\int_0^x \frac{\mathbf{B}_n(v)}{\overline{H}^2(v)} d\overline{H}(v)\right] \overline{F}(x)dx + o_{\mathbb{P}}(1), \tag{5.22}$$

and

$$J_{n,3} = -d_n \int_{0}^{h} \left[\int_0^x \frac{\mathbf{B}_n^*(v)}{\overline{H}^2(v)} d\overline{H}^{(1)}(v)\right] \overline{F}(x)dx + o_{\mathbb{P}}(1). \tag{5.23}$$

For the term $J_{n,4}$, equation (4.18) in [Soltane et al.(2015)] says that

$$J_{n,4} = \gamma\sqrt{\frac{n}{k}}\mathbf{B}_n^*(h) + o_{\mathbb{P}}(1). \tag{5.24}$$

Now, we treat the fifth term $J_{n,5}$, which, by a change of variables, becomes

$$J_{n,5} = d_n \sqrt{n} Q(t) \overline{F}(Q(t)) \int_{Q_n(t)/Q(t)}^{1} \frac{\overline{F}(xQ(t))}{\overline{F}(Q(t))} dx.$$

This may be split into the sum of

$$J_{n,5}^{(1)} := d_n \sqrt{n} Q(t) \overline{F}(Q(t)) \int_{Q_n(t)/Q(t)}^{1} \left(\frac{\overline{F}(xQ(t))}{\overline{F}(Q(t))} - x^{-1/\gamma_1} \right) dx,$$

and

$$J_{n,5}^{(2)} := d_n \sqrt{n} Q(t) \overline{F}(Q(t)) \int_{Q_n(t)/Q(t)}^{1} x^{-1/\gamma_1} dx.$$

By using the uniform inequality relative to regularly varying functions to the second-order, given in Proposition 4 in [Hua and Joe(2011)], we have, for all sufficiently large n and $0 < \epsilon < 1$,

$$\left| \frac{\overline{F}(xQ(t))/\overline{F}(Q(t)) - x^{-1/\gamma_1}}{A_{\mathbf{1}}^{*}(Q(t))} - x^{-1/\gamma_1} \frac{x^{\tau_1/\gamma_1} - 1}{\gamma_1 \tau_1} \right| \leq \epsilon x^{-1/\gamma_1 + \epsilon}, \quad x \geq 1, \tag{5.25}$$

where $A_{\mathbf{1}}^{*}(v) \sim A_{\mathbf{1}}(v)$, as $v \to \infty$. We apply the inequality (5.25) to $J_{n,5}^{(1)}$, to get

$$J_{n,5}^{(1)} = d_n \sqrt{n} Q(t) \overline{F}(Q(t)) A_{\mathbf{1}}^{*}(Q(t)) \int_{Q_n(t)/Q(t)}^{1} x^{-1/\gamma_1} \frac{x^{\tau_1/\gamma_1} - 1}{\gamma_1 \tau_1} dx + o_{\mathbb{P}}(1),$$

which, after integration, becomes

$$J_{n,5}^{(1)} = d_n \sqrt{n} Q(t) \overline{F}(Q(t)) A_{\mathbf{1}}^{*}(Q(t)) \{\phi(1) - \phi(Q_n(t)/Q(t))\} + o_{\mathbb{P}}(1),$$

where

$$\phi(s) := \frac{\left(\tau_1 + \gamma_1 + s^{\tau_1/\gamma_1} - \gamma_1 s^{\tau_1/\gamma_1} - 1\right) s^{(\gamma_1 - 1)/\gamma_1}}{\tau_1 (1 - \gamma_1)(\gamma_1 + \tau_1 - 1)},$$

is clearly a continuously differentiable function. Then by using Taylor's expansion, we get

$$J_{n,5}^{(1)} = -d_n \overline{F}\left(Q\left(t\right)\right) A_{\mathbf{1}}^{*}\left(Q\left(t\right)\right) \phi'\left(\zeta_n\left(t\right)\right) \sqrt{n}\left(Q_n\left(t\right) - Q\left(t\right)\right) + o_{\mathbb{P}}\left(1\right),$$

where $\zeta_n\left(t\right)$ lies between $Q_n\left(t\right)/Q\left(t\right)$ and 1. Lemma 3.3 in [Tse(2005)] says that

$$\sup_{0<t_0\leq t\leq t_1<1} \sqrt{n}\left|Q_n\left(t\right) - Q\left(t\right)\right| = O\left(\sqrt{\log\log n}\right), \tag{5.26}$$

with probability one. On the other hand, both $\overline{F}\left(Q\left(t\right)\right)$ and $\phi'\left(\zeta_n\left(t\right)\right)$ are bounded, for any $t\in(0,1)$, and $A_{\mathbf{1}}^{*}\left(Q\left(t\right)\right) \sim A_{\mathbf{1}}\left(Q\left(t\right)\right) = O\left(1\right)$. Then, we can write that

$$\lim_{n\to\infty} \sup_{0<t_0\leq t\leq t_1<1} \left|J_{n,5}^{(1)}\right| = \lim_{n\to\infty} \left|d_n O\left(\sqrt{\log\log n}\right)\right|,$$

By using Lemma 5.3.4, we end up with $J_{n,5}^{(1)} = o_{\mathbb{P}}\left(1\right)$, as $n\to\infty$.
For the term $J_{n,5}^{(2)}$, we apply Taylor's expansion to the function $s \to \int_0^s x^{-1/\gamma_1}dx$ on the interval $\left(Q_n\left(t\right)/Q\left(t\right), 1\right)$. From (5.26), we deduce that $Q_n\left(t\right)/Q\left(t\right) \overset{P}{\to} 1$, as $n\to\infty$, and we get

$$J_{n,5}^{(2)} = -d_n\left(1 + o_{\mathbb{P}}\left(1\right)\right)\overline{F}\left(Q\left(t\right)\right)\sqrt{n}\left(Q_n\left(t\right) - Q\left(t\right)\right).$$

By similar arguments to those used for $J_{n,5}^{(1)}$, we obtain $J_{n,5}^{(2)} = o_{\mathbb{P}}\left(1\right)$, as $n\to\infty$. Therefore, we have

$$J_{n,5} = o_{\mathbb{P}}\left(1\right), \text{ as } n\to\infty. \tag{5.27}$$

For the term $J_{n,6}$, we note that $(1-t) = \overline{F}\left(Q\left(t\right)\right)$, then we have

$$J_{n,6} = d_n \overline{F}\left(Q\left(t\right)\right)\sqrt{n}\left(Q_n\left(t\right) - Q\left(t\right)\right),$$

which resembles to $J_{n,5}^{(2)}$. Hence

$$J_{n,6} = o_{\mathbb{P}}\left(1\right), \text{ as } n\to\infty. \tag{5.28}$$

Finally, we gather the approximations (5.21) , (5.22) , (5.23) and (5.24) with the asymptotic negligibilities (5.19) , (5.27) and (5.28) , to get

$$\begin{aligned}\frac{\sqrt{k}\left(\hat{L}_{n,1}(t)-L_{n,1}(t)\right)}{h\overline{F}\left(h\right)} &= d_n\int_0^h \frac{\mathbf{B}_n\left(x\right)}{\overline{H}\left(x\right)}\overline{F}(x)dx \\ &+ d_n\int_0^h\left[\int_0^x \frac{\mathbf{B}_n\left(v\right)}{\overline{H}^2\left(v\right)}d\overline{H}\left(v\right)\right]\overline{F}(x)dx \\ &- d_n\int_0^h\left[\int_0^x \frac{\mathbf{B}_n^*\left(v\right)}{\overline{H}^2\left(v\right)}d\overline{H}^{(1)}\left(v\right)\right]\overline{F}(x)dx \\ &+ \gamma\sqrt{\frac{n}{k}}\mathbf{B}_n^*\left(h\right)+o_{\mathbb{P}}\left(1\right).\end{aligned} \tag{5.29}$$

Next, recall that

$$\hat{L}_{n,2}-L_{n,2}=\frac{\hat{\gamma}_1^{(c)}}{1-\hat{\gamma}_1^{(c)}}T_{n-k:n}\overline{F}_n(T_{n-k:n})-h\overline{F}(h)\int_1^\infty \frac{\overline{F}(hx)}{\overline{F}(h)}dx.$$

[Soltane et al.(2015)] showed (in pages 18, 19) that

$$\begin{aligned}\frac{\sqrt{k}\left(\hat{L}_{n,2}-L_{n,2}\right)}{h\overline{F}\left(h\right)} &= \frac{\gamma_1}{1-\gamma_1}\sqrt{\frac{k}{n}}\left\{\int_0^h \frac{\mathbf{B}_n\left(x\right)}{\left[\overline{H}\left(x\right)\right]^2}d\overline{H}\left(x\right)-\int_0^h \frac{\mathbf{B}_n^*\left(x\right)}{\left[\overline{H}\left(x\right)\right]^2}d\overline{H}^{(1)}\left(x\right)\right\} \\ &+\sqrt{\frac{n}{k}}\left\{\frac{1}{p\left(1-\gamma_1\right)^2}\int_1^\infty x^{-1}\mathbf{B}_n^*\left(hx\right)dx\right. \\ &\left.-\gamma\mathbf{B}_n^*\left(h\right)-\frac{\gamma_1}{p\left(1-\gamma_1\right)^2}\mathbf{B}_n\left(h\right)\right\} \\ &+\frac{\sqrt{k}A_1\left(h\right)}{\left(1-\gamma_1\right)}\left\{\frac{1}{\left(1-p\tau_1\right)\left(1-\gamma_1\right)}+\frac{1}{\left(\gamma_1+\tau-1\right)}\right\}+o_{\mathbb{P}}\left(1\right).\end{aligned} \tag{5.30}$$

Finally, by adding (5.30) to (5.29) , we get

$$\frac{\sqrt{k}\left(\widehat{CTE}\left(t\right)-CTE\left(t\right)\right)}{h\overline{F}\left(h\right)}\left(1-t\right)=\sum_{i=1}^5 M_{ni}+R_n+o_{\mathbb{P}}\left(1\right),$$

where

$$M_{n1} := d_n \int_0^h \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x) dx, \ M_{n2} := d_n \int_0^h \left[\int_0^x \frac{\mathbf{B}_n(v)}{\overline{H}^2(v)} d\overline{H}(v) \right] \overline{F}(x) dx,$$

$$M_{n3} := -d_n \int_0^h \left[\int_0^x \frac{\mathbf{B}_n^*(v)}{\overline{H}^2(v)} d\overline{H}^{(1)}(v) \right] \overline{F}(x) dx,$$

$$M_{n4} := \frac{\gamma_1}{(1-\gamma_1)} \sqrt{\frac{k}{n}} \left\{ \int_0^h \frac{\mathbf{B}_n(x)}{\left[\overline{H}(x)\right]^2} d\overline{H}(x) - \int_0^h \frac{\mathbf{B}_n^*(x)}{\left[\overline{H}(x)\right]^2} d\overline{H}^{(1)}(x) \right\},$$

$$M_{n5} := \sqrt{\frac{n}{k}} \left\{ \frac{1}{p(1-\gamma_1)^2} \int_1^\infty x^{-1} \mathbf{B}_n^*(hx)\, dx - \frac{\gamma_1}{p(1-\gamma_1)^2} \mathbf{B}_n(h) \right\},$$

and

$$R_n := \frac{\sqrt{k} A_1(h)}{(1-\gamma_1)} \left\{ \frac{1}{(1-p\tau_1)(1-\gamma_1)} + \frac{1}{(\gamma_1+\tau-1)} \right\}.$$

An integration by parts in the term M_{n2} yields

$$M_{n2} = d_n \int_0^h \left[\int_x^\infty \overline{F}(v) dv \right] \frac{\mathbf{B}_n(x)}{\overline{H}^2(x)} d\overline{H}(x)$$
$$- \frac{\sqrt{k/n}}{h\overline{F}(h)} \int_h^\infty \overline{F}(x) dx \left\{ \int_0^h \frac{\mathbf{B}_n(x)}{\left[\overline{H}(x)\right]^2} d\overline{H}(x) \right\}.$$

For the second term in M_{n2}, we use equation $(B.1.9)$ in Theorem $B.1.5$ in [de Haan and Ferreira(2006)], page 363, to obtain $\int_h^\infty \overline{F}(x)dx / \left(h\overline{F}(h)\right) \to \gamma_1 / (1-\gamma_1)$. By applying the same technique to M_{n3}, we get

$$\frac{\sqrt{k}\left(\widehat{CTE}(t) - CTE(t)\right)}{h\overline{F}(h)} (1-t) = \sum_{i=1}^4 N_{ni} + R_n + o_{\mathbb{P}}(1), \tag{5.31}$$

where N_{n1} and N_{n4} are exactly the same as M_{n1} and M_{n5} respectively and

$$N_{n2} := d_n \int_0^h \left[\int_x^\infty \overline{F}(v) dv \right] \frac{\mathbf{B}_n(x)}{\overline{H}^2(x)} d\overline{H}(x),$$

$$N_{n3} := -d_n \int_0^h \left[\int_x^\infty \overline{F}(v)dv \right] \frac{\mathbf{B}_n^*(x)}{\overline{H}^2(x)} d\overline{H}^{(1)}(x).$$

The decomposition (5.31) leads to the asymptotic normality of our estimator $\widehat{CTE}(t)$. The following formulas, obtained from the covariance structure presented in [Csörgő(1996)], page 2768,

$$\begin{cases} \mathbf{E}[\mathbf{B}_n(u)\mathbf{B}_n(v)] = \min\left(\overline{H}^{(1)}(u), \overline{H}^{(1)}(v)\right) - \overline{H}^{(1)}(u)\overline{H}^{(1)}(v), \\ \mathbf{E}[\mathbf{B}_n^*(u)\mathbf{B}_n^*(v)] = \min\left(\overline{H}(u), \overline{H}(v)\right) - \overline{H}(u)\overline{H}(v), \\ \mathbf{E}[\mathbf{B}_n(u)\mathbf{B}_n^*(v)] = \min\left(\overline{H}^{(1)}(u), \overline{H}^{(1)}(v)\right) - \overline{H}^{(1)}(u)\overline{H}(v)., \end{cases} \tag{5.32}$$

will be, in addition to L'Hôpital's rule, very useful in the computation of the variance σ^2 of the limiting distribution. We have $\overline{H}^{(1)}(x) \sim p\overline{H}(x)$ and, in view of (5.16), we can easily show that $k/n = \overline{H}(h) \sim v_1 v_2 h^{-1/\gamma}$. After performing a straightforward calculation, we find, as $n \to \infty$,

$$\mathbf{E}[N_{n1}]^2 \to \frac{2p^3\gamma_1^2}{(1-p+p\gamma_1)(1-2p+2p\gamma_1)},$$

$$\mathbf{E}[N_{n2}]^2 \to \frac{2p\gamma_1^2}{(1-\gamma_1)^2(1-p+p\gamma_1)(1-2p+2p\gamma_1)},$$

$$\mathbf{E}[N_{n3}]^2 \to \frac{2p^2\gamma_1^2}{(1-\gamma_1)^2(1-p+p\gamma_1)(1-2p+2p\gamma_1)},$$

$$\mathbf{E}[N_{n4}]^2 \to \frac{\gamma_1^2}{p(1-\gamma_1)^4}, \quad 2\mathbf{E}[N_{n1}N_{n4}] \to -\frac{2p\gamma_1^2(1-p)}{(1-\gamma_1)^2(1-p+p\gamma_1)},$$

$$2\mathbf{E}[N_{n1}N_{n2}] \to -\frac{4p^2\gamma_1^2}{(1-\gamma_1)(1-p+p\gamma_1)(1-2p+2p\gamma_1)},$$

$$2\mathbf{E}[N_{n1}N_{n3}] \to \frac{4p^3\gamma_1^2}{(1-\gamma_1)(1-p+p\gamma_1)(1-2p+2p\gamma_1)},$$

$$2\mathbf{E}[N_{n2}N_{n3}] \to -\frac{4p^2\gamma_1^2}{(1-\gamma_1)^2(1-p+p\gamma_1)(1-2p+2p\gamma_1)},$$

$$2\mathbf{E}[N_{n2}N_{n4}] \to \frac{2\gamma_1^2(1-p)}{(1-\gamma_1)^3(1-p+p\gamma_1)}, \quad 2\mathbf{E}[N_{n3}N_{n4}] \to 0.$$

Consequently, we infer that $\sqrt{k}\left(\widehat{CTE}\left(t\right)-CTE\left(t\right)\right)\left(1-t\right)/\left(h\overline{F}(x)\right)$ is asymptotically Gaussian with bias $\mu:=\lim_{n\to\infty}R_n$, which can be obtained easily since $\sqrt{k}A_1\left(h\right)\to\lambda$, and variance $\sigma^2:=\lim_{n\to\infty}\mathbf{E}\left[\sum_{i=1}^4 N_{ni}\right]^2$. This completes the proof. □

5.3 Appendix

Lemma 5.3.1 *Assume that F and G satisfy the second order conditions (5.5) and (5.6) with $\gamma_2/(1+2\gamma_2)<\gamma_1<1$. Then one of the integrals I_1 or I_2 defined in (5.4) are infinite.*

Proof 5.3.1 *First, note that the constraint $\gamma_1<1$ is due to the definition of the CTE. For $0<t<1$, we have*

$$I_1\geq\int_c^\infty x^2\kappa^2\left(x\right)dH^{(1)}\left(x\right),$$

where c is a large real number such that $c>Q(t)$. Note that $dH^{(1)}\left(x\right)=\overline{G}\left(x\right)dF\left(x\right)$ and $dH^{(0)}\left(x\right)=\overline{F}\left(x\right)dG\left(x\right)$, therefore

$$\kappa\left(x\right)=\exp\left\{\int_0^x dH^{(0)}\left(z\right)/\overline{H}\left(z\right)\right\}=\frac{1}{\overline{G}\left(x\right)}.$$

It follows that

$$I_1\geq\int_c^\infty x^2\frac{dF\left(x\right)}{\overline{G}\left(x\right)},$$

which, by a change of variables, may be rewritten as

$$I_1\geq c^2\frac{\overline{F}\left(c\right)}{\overline{G}\left(c\right)}\int_1^\infty x^2\frac{\overline{G}\left(c\right)}{\overline{G}\left(cx\right)}\frac{dF\left(cx\right)}{\overline{F}\left(c\right)}.$$

By applying Potter's inequalities (2.4) to $\overline{F}$ and $\overline{G}$, which are regularly varying at infinity with respective indices $-1/\gamma_1$ and $-1/\gamma_2$, we have, for any small $\epsilon>0$ and $x\geq1$

$$\int_1^\infty x^2\frac{\overline{G}\left(c\right)}{\overline{G}\left(cx\right)}\frac{dF\left(cx\right)}{\overline{F}\left(c\right)}\geq\frac{1-\epsilon}{1+\epsilon}\left(\epsilon+1/\gamma_1\right)\int_1^\infty x^{(1/\gamma_2)-(1/\gamma_1)+1}dx.$$

Since $\gamma_2/(1+2\gamma_2) < \gamma_1$, *then* $(1/\gamma_2) - (1/\gamma_1) + 2 > 0$ *and* $\int_1^\infty x^{(1/\gamma_2)-(1/\gamma_1)+1} dx$ *is infinite. By noting that* $c^2 \overline{F}(c) / \overline{G}(c) > 0$, *we infer that, for any* $0 < t < 1$, I_1 *is infinite when the couple* (γ_1, γ_2) *is in the range*

$$R_1 := \left\{ 0 < \gamma_1, \gamma_2 < \infty : \frac{\gamma_2}{1+2\gamma_2} < \gamma_1 \right\}.$$

For the second integral I_2, *the same reason as for* I_1 *and the fact that* $\overline{F}(z) \leq 1$ *yield*

$$I_2 \geq \int_c^\infty x \left(\int_0^x \frac{dG(z)}{\overline{F}(z)\,\overline{G}^2(z)} \right)^{1/2} dF(x) \geq \int_c^\infty x \left(\frac{1}{\overline{G}(x)} - 1 \right)^{1/2} dF(x).$$

For $c < x < \infty$, *we have* $\sqrt{1-\overline{G}(x)} > \sqrt{1-\overline{G}(c)}$. *Hence, we end up with*

$$I_2 \geq \sqrt{1-\overline{G}(c)} \int_c^\infty \frac{x}{\sqrt{\overline{G}(x)}} dF(x),$$

which is equal to

$$c \frac{\overline{F}(c)\sqrt{1-\overline{G}(c)}}{\sqrt{\overline{G}(c)}} \int_1^\infty x \sqrt{\frac{\overline{G}(c)}{\overline{G}(cx)}} \frac{dF(cx)}{\overline{F}(c)}.$$

By similar arguments as those used to show that $I_1 = \infty$, *we find that* $I_2 = \infty$ *when the tail indices belong to*

$$R_2 := \left\{ 0 < \gamma_1, \gamma_2 < \infty : \frac{2\gamma_2}{1+2\gamma_2} < \gamma_1 \right\}.$$

Finally, the union of both ranges is

$$R_1 \cup R_2 = \left\{ 0 < \gamma_1, \gamma_2 < \infty : \frac{\gamma_2}{1+2\gamma_2} < \gamma_1 \right\}.$$

Lemma 5.3.2 *Suppose that $\gamma_1 < \gamma_2$, then we have*

$$J_{n,1} = d_n \int_0^{T_{n-k:n}} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x) dx + r_n,$$

where $d_n = (k/n)^{1/2} / \left(h\overline{F}(h)\right)$ and

$$r_n := -d_n \int_0^{Q_n(t)} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x) dx = o_{\mathbb{P}}(1), \text{ as } n \to \infty. \tag{5.33}$$

Proof 5.3.2 *Observe that*

$$\begin{aligned}
\mathbf{E}\left| \int_0^{Q_n(t)} \frac{\mathbf{B}_n(x)}{\overline{H}(x)} \overline{F}(x) dx \right| &\leq \int_0^{Q_n(t)} \mathbf{E}\left|\mathbf{B}_n(x)\right| \frac{\overline{F}(x)}{\overline{H}(x)} dx \\
&\leq \int_0^{\infty} \mathbf{E}\left|\mathbf{B}_n(x)\right| \frac{\overline{F}(x)}{\overline{H}(x)} dx.
\end{aligned}$$

From the first result of (5.32), we have $\mathbf{E}\left|\mathbf{B}_n(v)\right| \leq \sqrt{\overline{H}^{(1)}(v)}$. Moreover, we have $\overline{H}^{(1)}(v) \leq \overline{H}(v)$, hence

$$\int_0^{\infty} \mathbf{E}\left|\mathbf{B}_n(x)\right| \frac{\overline{F}(x)}{\overline{H}(x)} dx \leq \int_0^{\infty} \frac{\overline{F}(x)}{\sqrt{\overline{H}(x)}} dx.$$

Let $c > 0$ be a large real number, then

$$\int_0^{\infty} \frac{\overline{F}(x)}{\sqrt{\overline{H}(x)}} dx > \int_c^{\infty} \frac{\overline{F}(x)}{\sqrt{\overline{H}(x)}} dx.$$

By using (5.16) we may write

$$\int_c^{\infty} \frac{\overline{F}(x)}{\sqrt{\overline{H}(x)}} dx \sim \left(v_1 v_2^{-1}\right)^{1/2} \int_c^{\infty} x^{\frac{1}{2}\left(-\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right)} dx,$$

where the latter integral is infinite when $\gamma_1 \geq \gamma_2$. Therefore, under the condition $\gamma_1 < \gamma_2$, we find that $\int_0^{Q_n(t)} \left[\mathbf{B}_n(x) / \overline{H}(x)\right] \overline{F}(x) dx < \infty$. From (5.18), we have $d_n \to 0$, then we *get* $r_n = o_{\mathbb{P}}(1)$, *as* $n \to \infty$,

Lemma 5.3.3 *Under the assumptions of Lemma 5.3.1, we have, for any fixed* $0 < \epsilon < 1$ *and* $0 \leq \omega < 1/4$,

$$n^{-\omega} d_n \int_0^{(1+\epsilon)h} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx \to 0, \text{ as } n \to \infty,$$

where $d_n = (k/n)^{1/2} / \left(h\overline{F}(h)\right)$.

Proof 5.3.3 *Since* $\overline{H}(h) = k/n$, *then we have*

$$n^{-\omega} d_n \int_0^{(1+\epsilon)h} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx = k^{-\omega} \frac{\left[\overline{H}(h)\right]^{1/2+\omega}}{h\overline{F}(h)} \int_0^{(1+\epsilon)h} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx.$$

Now $k^{-\omega} \to 0$, *thus it suffices to show that the limit, as* $n \to \infty$, *of*

$$\varphi(h) := \frac{\left[\overline{H}(h)\right]^{1/2+\omega}}{h\overline{F}(h)} \int_0^{(1+\epsilon)h} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx,$$

is finite. By using (5.16), *we show that*

$$\frac{h\overline{F}(h)}{\left[\overline{H}(h)\right]^{1/2+\omega}} \sim v_3 h^{1-1/\gamma_1+(1/2+\omega)/\gamma}, \text{ as } n \to \infty,$$

where $v_3 := v_1 (v_1 v_2)^{-(1/2+\omega)}$, *for some positive constants* v_1 *and* v_2. *It follows that*

$$\varphi(h) \sim \int_0^{(1+\epsilon)h} \frac{\overline{F}(x) / \left[\overline{H}(x)\right]^{1/2+\omega}}{v_3 h^{1-1/\gamma_1+(1/2+\omega)/\gamma}} dx, \text{ as } n \to \infty.$$

For the purpose of applying L'Hôpital's rule, we first have to verify that both

$$\int_0^{(1+\epsilon)h} \overline{F}(x) / \left[\overline{H}(x)\right]^{1/2+\omega} dx \quad \text{and} \quad h^{1-1/\gamma_1+(1/2+\omega)/\gamma},$$

tend to infinity, as $n \to \infty$. *It is obvious that*

$$\int_0^{(1+\epsilon)h} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx \geq \int_h^{(1+\epsilon)h} \frac{\overline{F}(x)}{\left[\overline{H}(x)\right]^{1/2+\omega}} dx,$$

which, by a change of variables, equals $h \int_1^{(1+\epsilon)} \overline{F}(hx) / \left[\overline{H}(hx)\right]^{1/2+\omega} dx$. *Since* $h \to \infty$ *and* $x \to \overline{F}(x) / \left[\overline{H}(x)\right]^{1/2+\omega}$ *is regularly varying at infinity with index* $(1/2+\omega)/\gamma - 1/\gamma_1$, *then, by using Potter's inequalities* (2.4), *we get*

$$\int_1^{(1+\epsilon)} \frac{\overline{F}(hx)}{\left[\overline{H}(hx)\right]^{1/2+\omega}} dx \geq (1-\epsilon) \int_1^{(1+\epsilon)} x^{\frac{1/2+\omega}{\gamma} - \frac{1}{\gamma_1} - \epsilon} dx,$$

which is equal to

$$\frac{(1-\epsilon)(1+\epsilon)^{\frac{1/2+\omega}{\gamma} - \frac{1}{\gamma_1} + 1 - \epsilon}}{(1/2+\omega)/\gamma - 1/\gamma_1 + 1 - \epsilon} =: b(\epsilon) > 0.$$

Thus, we have $\int_0^{(1+\epsilon)h} \overline{F}(x) / \left[\overline{H}(x)\right]^{1/2+\omega} dx \geq b(\epsilon) h$, *which tends to* ∞, *as* $n \to \infty$. *As for the quantity* $h^{1-1/\gamma_1+(1/2+\omega)/\gamma}$, *we note that the assumption* $\gamma_2/(1+2\gamma_2) < \gamma_1$ *implies that* $\gamma - \gamma/\gamma_1 + 1/2 > 0$ *and then* $\gamma - \gamma/\gamma_1 + 1/2 + \omega = (1/\gamma)(1 - 1/\gamma_1 + (1/2+\omega)/\gamma) > 0$. *By taking into account that* $h \to \infty$, *we infer that* $h^{1-1/\gamma_1+(1/2+\omega)/\gamma} \to \infty$, *as* $n \to \infty$. *Now, we are in position to apply L'Hôpital's rule to compute* $\lim_{n\to\infty} \varphi(h)$. *That is*

$$\lim_{n\to\infty} \varphi(h) \sim \frac{1}{(1 - 1/\gamma_1 + (1/2+\omega)/\gamma)} \lim_{n\to\infty} \frac{\overline{F}((1+\epsilon)h) / \left[\overline{H}((1+\epsilon)h)\right]^{1/2+\omega}}{v_3 h^{-1/\gamma_1+(1/2+\omega)/\gamma}}.$$

Once again, by using (5.16), *we write*

$$\frac{\overline{F}((1+\epsilon)h)}{\left[\overline{H}((1+\epsilon)h)\right]^{1/2+\omega}} \sim v_3 \left((1+\epsilon)h\right)^{-1/\gamma_1+(1/2+\omega)/\gamma}, \text{ as } n \to \infty.$$

It follows that

$$\lim_{n\to\infty} \frac{\overline{F}((1+\epsilon)h) / \left[\overline{H}((1+\epsilon)h)\right]^{1/2+\omega}}{v_3 h^{-1/\gamma_1+(1/2+\omega)/\gamma}} = (1+\epsilon)^{-1/\gamma_1+(1/2+\omega)/\gamma},$$

which is indeed finite, as sought.

Lemma 5.3.4 *Assume that $\gamma_2/(1+2\gamma_2)<\gamma_1$, then we have*

$$d_n\sqrt{\log\log n}=o\left(1\right),\ \text{as } n\rightarrow\infty,$$

where $d_n=\left(k/n\right)^{1/2}/\left(h\overline{F}\left(h\right)\right).$

Proof 5.3.4 *For the proof of this lemma, we will need a result that we state in the proposition below. First note that, for any $\epsilon>0$, $n^{-\epsilon}\sqrt{\log\log n}\rightarrow 0$, as $n\rightarrow\infty$. On the other hand, from Proposition 5.3.1, the sample fraction k is such that $k=\left[b_2n^{\alpha}\right]$, where* $0<\alpha:=2\gamma\ell/\left(2\gamma\ell+1\right)<1$. *Then, from (5.17), we may write that, as $n\rightarrow\infty$,*

$$\begin{aligned}n^{\epsilon}d_n &\sim v_1^{-1}\left(v_1v_2\right)^{\gamma+\gamma/\gamma_1}n^{\epsilon}\left(b_2n^{\alpha}/n\right)^{1/2+\gamma-\gamma/\gamma_1}\\ &=\frac{v_1^{-1}\left(v_1v_2\right)^{\gamma+\gamma/\gamma_1}b_2^{(1/2+\gamma-\gamma/\gamma_1)}}{n^{(1-\alpha)(1/2+\gamma-\gamma/\gamma_1)-\epsilon}},\end{aligned}$$

for some positive constants v_1 and v_2. Let us take ϵ such that $0<\epsilon<\left(1-\alpha\right)\left(1/2+\gamma-\gamma/\gamma_1\right)$, thus $n^{(1-\alpha)(1/2+\gamma-\gamma/\gamma_1)-\epsilon}\rightarrow\infty$, as $n\rightarrow\infty$, because $0<\alpha<1$ and $1/2+\gamma-\gamma/\gamma_1>0$ (under the assumption $\gamma_2/(1+2\gamma_2)<\gamma_1$). Consequently, we get $d_n\sqrt{\log\log n}\rightarrow 0$, as $n\rightarrow\infty$.

Proposition 5.3.1 *Suppose that F and G belong Hall's class. Then we have*

$$k=\left[b_2n^{2\gamma\ell/(2\gamma\ell+1)}\right],$$

where $\ell:=\min\left(\ell_1,\ell_2\right)$, $\ell_i:=-\tau_i/\gamma_i>0$, $i=1,2$ and

$$b_2:=\left(\frac{\gamma_1}{2\ell b_1^2p^2}\right)^{1/(2\gamma\ell+1)},$$

for some real constant $b_1=b_1\left(l,m,\gamma,\ell,p\right)$.

Proof 5.3.5 *We will follow the same approach that was applied to determine the optimal number of upper order statistics used in the estimation of the shape parameter of complete heavy-tailed data. The selection of this crucial number was extensively studied in EVT (see, for instance, [Beirlant et al.(2004)], page 123 and page 77 in [de Haan and Ferreira(2006)]). According to [Hall and Welsh(1985)], if X were not censored, then the optimal sample fraction would have been $k = \left[dn^{2\tau_1/(2\tau_1-1)}\right]$, where $d = d(l_1, m_1, \gamma_1, \tau_1) > 0$.*
In the case of random censorship, [Beirlant et al.(2016)] showed the following Gaussian approximation:

$$\sqrt{k}\left(\hat{\gamma}_1^{(c)} - \gamma_1\right) = \frac{\gamma_1}{\sqrt{p}}\mathcal{N}(0,1) + \mathcal{AB}\left(\hat{\gamma}_1^{(c)}\right), \text{ as } n \to \infty,$$

where $\mathcal{AB}\left(\hat{\gamma}_1^{(c)}\right) = b_1 (k/n)^{\gamma\ell}$ is the asymptotic bias. The corresponding asymptotic mean squared error (amse) is equal to

$$amse(k) := \mathbf{E}_\infty \left(\hat{\gamma}_1^{(c)} - \gamma_1\right)^2 = \frac{\gamma_1^2}{pk} + b_1^2 (k/n)^{2\gamma\ell},$$

where $\mathbf{E}_\infty$ stands for the asymptotic expectation. Our goal is to look for the k-value that minimizes the quantity above. After differentiation and calculation, we get the desired result.

Lemma 5.3.5 *Assume that $T_{n-k:n} \xrightarrow{P} \infty$, as $n \to \infty$. Then we have $w_{n-k} \xrightarrow{P} 1$, as $n \to \infty$.*

Proof 5.3.6 *Note that $w_{n-k} = F_n(T_{n-k:n})$. We have, from assertion (1.7) of Theorem 2 in [Csörgő(1996)],*

$$\sup_{x \le T_{n-k:n}} \left| \frac{F_n(x) - F(x)}{1 - F(T_{n-k:n})} \right| = O_p\left(\frac{1}{\sqrt{k}}\right),$$

which implies that

$$\sup_{x \le T_{n-k:n}} \left| \frac{(1 - F_n(x)) - (1 - F(x))}{1 - F(T_{n-k:n})} \right| = o_p(1), \text{ as } k \to \infty.$$

For $x = T_{n-k:n}$ *we have*

$$\left| \frac{\overline{F}_n \left(T_{n-k:n} \right) - \overline{F} \left(T_{n-k:n} \right)}{\overline{F} \left(T_{n-k:n} \right)} \right| = o_p \left(1 \right), \ \ as \ k \rightarrow \infty,$$

thus

$$\left| \frac{\overline{F}_n \left(T_{n-k:n} \right)}{\overline{F} \left(T_{n-k:n} \right)} - 1 \right| = o_p \left(1 \right), \ \ as \ k \rightarrow \infty.$$

This means that

$$\frac{\overline{F}_n \left(T_{n-k:n} \right)}{\overline{F} \left(T_{n-k:n} \right)} \xrightarrow{P} 1.$$

Consequently, we get

$$\overline{F}_n \left(T_{n-k:n} \right) = \left(1 + o_p \left(1 \right) \right) \overline{F} \left(T_{n-k:n} \right).$$

Since $T_{n-k:n} \xrightarrow{P} \infty,$ *thus* $\overline{F} \left(T_{n-k:n} \right) \xrightarrow{P} 0,$ *then we get* $\overline{F}_n \left(T_{n-k:n} \right) \xrightarrow{P} 0.$ *Therefore* $F_n \left(T_{n-k:n} \right) \xrightarrow{P} 1.$

Conclusion

In this thesis, we utilize extreme value theory along with survival analysis to estimate the risk measure Conditional Tail Expectation (CTE), allowing for a comprehensive assessment of extreme risks despite data scarcity and the presence of censoring.

To tackle the challenges presented by random right censored and heavy-tailed datasets in estimating risk measures, we dedicate five chapters to proposing an asymptotic normal estimator of CTE, which are appended to an introduction.

Our findings have significant implications for real-world applications, particularly in insurance and healthcare, offering valuable tools for managing extreme risks, safeguarding against rare but devastating events, and improving decision-making processes. The use of Value at Risk (VaR), CTE, and estimators like Hill and Weissman can help analyze heavy-tailed data and estimate extreme losses across multiple fields. Below, we outline how these results can be applied within the contexts of insurance and healthcare, with the potential for expansion to include finance:

Managing extreme risks in insurance and healthcare: insurance companies and the healthcare sector face significant challenges in estimating extreme losses or outcomes caused by rare but impactful events, whether they are natural disasters or assessing survival probabilities for patients suffering from severe illnesses such as AIDS. By using heavy-tailed distributions and risk measures like VaR and CTE, along with statistical estimators such as Hill and Weissman, it becomes possible to achieve accurate estimates of extreme risks and outcomes. This enables better resource allocation strategies, more precise financial reserves, and the development of effective policies to address exceptional events.

Expanding to other fields like finance: the methodologies and results developed in this study, though applied primarily in insurance and healthcare, can be extended to other fields such as finance. In particular, risk assessment in financial markets, especially in times of economic crises, could benefit from these tools to predict extreme losses and ensure better financial planning and risk management.

To further enhance our findings, future work could focus on studying bias reduction and robustness in our proposed estimator, as well as implementing kernel methods for estimation purposes.

Bibliography

Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. The Annals of Statistics, 6, 534-545.

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. Journal of Banking & Finance, 26, 1505-1518.

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. Journal of Banking & Finance, 26, 1487-1503.

Amarante, M., & Liebrich, F. B. (2024). Distortion risk measures: Prudence, coherence, and the expected shortfall. Mathematical Finance, 34, 1291-1327.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). A characterization of measures of risk. Cornell University Operations Research and Industrial Engineering. Preprint 1996/14.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. Mathematical finance, 9, 203-228.

Beirlant, J., Bardoutsos, A., de Wet, T., & Gijbels, I. (2016). Bias reduced tail estimation for censored Pareto type distributions. Statistics & Probability Letters, 109, 78-88.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2004). Statistics of extremes: theory and applications. John Wiley & Sons.

Beirlant, J., Guillou, A., Dierckx, G., & Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. Extremes, 10, 151-174.

Beirlant, J., Maribe, G., & Verster, A. (2018). Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications. Insurance: Mathematics and Economics, 78, 114-122.

Beirlant, J., Teugels, J. L., & Vynckier, P. (1996). Practical analysis of extreme values. Leuven University Press.

Beirlant, J., Worms, J., & Worms, R. (2019). Estimation of the extreme value index in a censorship framework: Asymptotic and finite sample behavior. Journal of Statistical Planning and Inference, 202, 31-56.

Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). Regular variation. Cambridge University Press.

Bladt, M., Albrecher, H., & Beirlant, J. (2021). Trimmed extreme value estimators for censored heavy-tailed data. Electronic Journal of Statistics, 15, 3112-3136.

Brahimi, B., Meraghni, D., & Necir, A. (2015a). Gaussian approximation to the extreme value index estimator of a heavy-tailed distribution under random censoring. Mathematical Methods of Statistics, 24, 266-279.

Brahimi, B., Meraghni, D., & Necir, A. (2015b). Nelson-Aalen tail product-limit process and extreme value index estimation under random censorship. Preprint arXiv:1502.03955.

Brahimi, B., Meraghni, D., Necir, A., & Soltane, L. (2018). Tail empirical process and weighted extreme value index estimator for randomly right-censored data. Preprint arXiv:1801.00572.

Breslow, N., & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. The Annals of Statistics, 2, 437-453.

Bryson, M. C. (1974). Heavy-tailed distributions: properties and tests. Technometrics, 16, 61-68.

Bühlmann, H. (2007). Mathematical methods in risk theory. Springer Science & Business Media.

Cai, J., & Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. ASTIN Bulletin: The Journal of the IAA, 37, 93-112.

Chen, Z., & Hu, Q. (2018). On coherent risk measures induced by convex risk measures. Methodology and Computing in Applied Probability, 20, 673-698.

Cheng, S. and Peng, L. (2001). Confidence intervals for the tail index. Bernoulli, 7, 751-760.

Cohen, A. C., (1991). Truncated and censored samples: Theory and applications. CRC Press.

Cohen, J. E., Davis, R. A., & Samorodnitsky, G. (2020). Heavy-tailed distributions, correlations, kurtosis and Taylor's law of fluctuation scaling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 476(2244).

Csörgő, S. (1996). Universal Gaussian approximations under random censorship. The Annals of Statistics, 24, 2744-2778.

Csörgő, M., Csörgő, S., Horváth, L., & Mason, D. M. (1986). Weighted empirical and quantile processes. The Annals of Probability, 14, 31-85.

Csörgő, S., & Mason, D. M. (1985). Central limit theorems for sums of extreme values. In Mathematical Proceedings of the Cambridge Philosophical Society, 98, 547-558.

Danielsson, J., de Haan, L., Peng, L. & de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate Analysis, 76, 226-248.

David, H. A., & Nagaraja, H. N. (2004). Order statistics (3rd ed.). Wiley-Interscience.

Davis, R., & Resnick, S. (1984). Tail estimates motivated by extreme value theory. The Annals of Statistics, 12, 1467-1487.

Deheuvels, P., Haeusler, E., & Mason, D. M. (1988). Almost sure convergence of the Hill estimator. In Mathematical Proceedings of the Cambridge Philosophical Society, 104, 371-381.

Deheuvels, P., & Einmahl, J. H. (1996). On the strong limiting behavior of local functionals of empirical processes based upon censored data. The Annals of Probability, 24, 504-525.

Dekkers, A. L., Einmahl, J. H., & De Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. The Annals of Statistics, 17, 1-24.

Dekkers, A. L., & De Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. Journal of Multivariate Analysis, 47, 173-195.

Denuit, M., Dhaene, J., Goovaerts, M. & Kaas, R., (2005). Actuarial theory for dependent risks: Measures, orders and models. Wiley.

Denuit, M., Purcaru, O., & Keilegom, I. V. (2006). Bivariate Archimedean copula models for censored data in non-life insurance. Journal of Actuarial Practice, 13, 5-32.

Dickson, D. C. (2016). Insurance risk and ruin. Cambridge University Press.

Drees, H. & Kaufmann, E. (1998). Selection of the optimal sample fraction in univariate extreme value estimation. Stochastic Processes and their Applications 75, 149-195.

Einmahl, J. H., Fils-Villetard, A., & Guillou, A. (2008). Statistics of extremes under random censoring. Bernoulli, 14, 207-227.

Einmahl, J. H., & Koning, A. J. (1992). Limit theorems for a general weighted process under random censoring. Canadian Journal of Statistics, 20, 77-89.

El Adlouni, S., Bobée, B., & Ouarda, T. B. (2007). Caractérisation des distributions à queue lourde pour l'analyse des crues (Research report No. R-929). INRS-ETE, University of Quebec.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling extremal events for insurance and finance. Springer.

Etikan, I., Abubakar, S., & Alkassim, R. (2017). The Kaplan-Meier estimate in survival analysis. Biometrics & Biostatistics International Journal, 5, 00128.

Fisher, R.A., & Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society, 24, 180-190.

Föllmer, H. & Schied, A. (2002). Robust preferences and convex measures of risk. In Advances in Finance and Stochastics, 1, 39-56. Springer Berlin Heidelberg.

Foss, S., Korshunov, D., & Zachary, S. (2011). An introduction to heavy-tailed and subexponential distributions. Springer.

Frees, E. W., & Valdez, E. A. (1998). Understanding relationships using copulas. North American Actuarial Journal, 2, 1-25.

Gerber, H. U. (1974). On additive premium calculation principles. ASTIN Bulletin: The Journal of the IAA, 7, 215-222.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of Mathematics 44, 423-453.

Goegebeur, Y., Guillou, A., & Qin, J. (2019). Bias-corrected estimation for conditional Pareto-type distributions with random right censoring. Extremes, 22, 459-498.

Goegebeur, Y., Guillou, A., & Qin, J. (2023). Conditional tail moment and reinsurance premium estimation under random right censoring. Test, 33, 230-250.

Goegebeur, Y., Guillou, A., Pedersen, T., & Qin, J. (2022). Extreme-value based estimation of the conditional tail moment with application to reinsurance rating. Insurance: Mathematics and Economics, 107, 102-122.

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. International Journal of Ayurveda Research, 1, 274-278.

Gomes, M. I., & Neves, M. M. (2011). Estimation of the extreme value index for randomly censored data. Biometrical Letters, 48, 1-22.

Guesmia, N, E., Meraghni, D., & Soltane, L. (2024). Estimating the conditional tail expectation of randomly right-censored heavy-tailed data. Journal of Statistical Theory and Practice, 18, 1-36.

De Haan, L., & Ferreira, A. (2006). Extreme value theory: An introduction. Springer.

de Haan, L., & Peng, L. (1998). Comparison of tail index estimators. Statistica Neerlandica, 52, 60-70.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society: Series B (Methodological), 44, 37-42.

Hall, P. & Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. Annals of Statistics 13, 331-341.

Hill, B. M. (1975). Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. The Annals of Statistics, 3, 1163-1174.

Hong, J., & Elshahat, A. (2010). Conditional tail variance and conditional tail skewness. Journal of Financial and Economic Practice, 10, 147-156.

Hong, S., Kim, J., & Kim, C. (2013). Nonparametric estimation of quantile functions for randomly right censored data. Journal of the Korean Statistical Society, 42, 169-176.

Hua, L., & Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks. Insurance: Mathematics and Economics, 49, 537-546.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81, 158-172.

Jones, B. L. & Zitikis, R., (2003). Empirical estimation of risk measures and related quantities. North American Actuarial Journal. 7, 44-54.

Jones, B. L., & Zitikis, R. (2007). Risk measures, distortion parameters, and their empirical estimation. Insurance: Mathematics and Economics, 41, 279-297.

Kaas, R., Goovaerts, M., Dhaene, J., & Denuit, M. (2008). Modern actuarial risk theory: Using R. Springer

Kaiser, T., & Brazauskas, V. (2006). Interval estimation of actuarial risk measures. North American Actuarial Journal, 10, 249-268.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457-481.

Kim, J. H., & Kim, J. (2015). A parametric alternative to the Hill estimator for heavy-tailed distributions. Journal of Banking & Finance, 54, 60-71.

Klein, J. P., & Moeschberger, M. L. (2003). Survival analysis: Techniques for censored and truncated data (2nd ed.). Springer.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2019). Loss models: From data to decisions (5th ed.). Wiley.

Klugman, S. A., & Parsa, R. (1999). Fitting bivariate loss distributions with copulas. Insurance: Mathematics and Economics, 24, 139-148.

Kouider, M. R., Idiou, N., & Benatia, F. (2023). Adaptive estimators of the general Pareto distribution parameters under random censorship and application. Journal of Science and Arts, 23, 395-412.

Kouider, M. R., Kheireddine, S., & Benatia, F. (2024). The adaptive mean estimation of heavy tailed distribution under random censoring. Studies in Engineering and Exact Sciences, 5, 01-31.

Lala Bouali, D., Benatia, F., Brahimi, B., & Chesneau, C. (2021). Robust estimator of conditional tail expectation of Pareto-type distribution. Journal of Statistical Theory and Practice, 15, 1-12.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. Monthly Notices of the Royal Astronomical Society, 155, 95-118.

Mason, D. M. (1982). Laws of large numbers for sums of extreme values. The Annals of Probability, 10, 754-764.

Meeker, W. Q., & Escobar, L. A. (1998). Statistical methods for reliability data. John Wiley & Sons, Inc.

Methni, J. E., Gardes, L., & Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. Scandinavian Journal of Statistics, 41, 988-1012.

Mikosch, T. (1999). Regular variation, subexponentiality and their applications in probability theory. Eurandom, Report 99013, 57.

Montserrat, H. S. (2014). The use of the premium calculation principles in actuarial pricing based scenario in a coherent risk measure. Journal of Applied Quantitative Methods, 9, 34-57.

Ndao, P., Diop, A., & Dupuy, J. F., (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. Computational Statistics & Data Analysis. 79, 63-79.

Necir, A., & Meraghni, D. (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. Insurance: Mathematics and economics, 45, 49-58.

Necir, A., Meraghni, D., & Meddi, F. (2007). Statistical estimate of the proportional hazard premium of loss. Scandinavian Actuarial Journal, 3, 147-161.

Necir, A., Rassoul, A., & Zitikis, R. (2010). Estimating the conditional tail expectation in the case of heavy-tailed losses. Journal of Probability and Statistics, 2010, Article ID 596839.

Nelson, W. (1972). A short life test for comparing a sample with previous accelerated test results. Technometrics, 14, 175-185.

Németh, L., & Zempléni, A. (2020). Regression estimator for the tail index. Journal of Statistical Theory and Practice, 14, 1-15.

Neves, C., & Fraga Alves, M.I. (2004). Reiss and Thomas' automatic selection of the number of extremes. Computational Statistics and Data Analysis 47, 689-704.

Pickands III, J. (1975). Statistical inference using extreme order statistics. the Annals of Statistics, 3, 119-131.

Reich, A. (1986). Properties of premium calculation principles. Insurance: Mathematics and Economics, 5, 97-101.

Reiss, R.D., & Thomas, M., (2007). Statistical analysis of extreme values: With applications to insurance, finance, hydrology and other fields (3rd ed.). Birkhäuser.

Resnick, S. I. (2007). Heavy-tail phenomena: probabilistic and statistical modeling. Springer Science & Business Media.

Reynkens, T., Verbelen, R., Beirlant, J., & Antonio, K. (2017). Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. Insurance: Mathematics and Economics, 77, 65-77.

Ripley, B.D., & Solomon, P.J. (1994). A note on Australian AIDS survival. University of Adelaide, Department of Statistics, Research Report 94/3.

Saporta, G. 1990. Probabilités, analyse des données et statistique. Editions Technip.

Shorack, G. R., & Wellner, J. A. (2009). Empirical processes with applications to statistics. Society for Industrial and Applied Mathematics.

Soltane, L., Meraghni, D., & Necir, A. (2015). Estimating the mean of a heavy-tailed distribution under random censoring. Preprint arXiv:1507.03178.

Soltane, L., Meraghni, D., & Necir, A. (2016). Statistical estimate of the proportional hazard premium of loss under random censoring. Afrika Statistika, 11, 883-899.

Stupfler, G., (2016). Estimating the conditional extreme-value index under random right-censoring. Journal of Multivariate Analysis, 144, 1-24.

Stute, W. (1995). The central limit theorem under random censorship. The Annals of Statistics, 23, 422-439.

Tasche, D. (2002). Expected shortfall and beyond. Journal of Banking & Finance, 26, 1519-1533.

Tse, S. (2005). Quantile process for left truncated and right censored data. Annals of the Institute of Statistical Mathematics, 57, 61-69.

Valdez, E. A. (2005). Tail conditional variance for elliptically contoured distributions. Belgian Actuarial Bulletin, 5, 26-36.

Venables, W.N., & Ripley, B.D. (2002). Modern applied statistics with S (4th ed.). Springer.

Von Mises, R. (1936). La distribution de la plus grande de n valeurs. Revue de Mathématique de l'Union Interbalkanique, 1, 141–160.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. Insurance: Mathematics and Economics, 17, 43-54.

Wang, S. (1996). Premium calculation by transforming the layer premium density. ASTIN Bulletin, 26, 71-92.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association, 73, 812-815.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. The Annals of Statistics, 13, 163-177.

Worms, J., & Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. Extremes, 17, 337-358.

Worms, J., & Worms, R. (2021). Estimation of extremes for heavy-tailed and light-tailed distributions in the presence of random censoring. Statistics, 55, 979-1017.

Appendix: Abbreviations and notations

Abbreviations and notations that is largely confined to sections or chapters is mostly excluded from the list below:

CTE	conditional tail expectation
CTM or CTM_ζ	conditional tail moment
cdf	cumulative distribution function
df or df's	distribution function(s)
$D(\mathcal{H}_\gamma)$	domain of attraction of $\mathcal{H}_\gamma$
EVT	extreme value theory
$\mathbf{E}[X]$	mathematical expectation of X
exp	exponential function
F	distribution function
F_n	empirical distribution function
$\widehat{F}_n$	Kaplan-Meier estimator
$\overline{F}$	survival function, tail of F
F^{-1}	inverse function of F
f	probability density
g	distorsion function
h	hazard function

i.e.	in other words (or which means)
iid	independent and identically distributed
$\inf(A)$	infimum of set A
k_{opt}	optimal number of upper order statistics
L	slowly varying function
log	logarithm function
$\max(A)$ (or $\min(A)$)	maximum of A (or minimum of A)
n	integer number greater than 1
$\mathcal{N}(0,1)$	standard normal law
$\mathcal{N}(\mu,\sigma^2)$	Gaussian distribution of parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$
P	probability measure
PHP	proportional hazards premium
Q or $F^{\leftarrow}$	quantile function, generalized inverse of F
Q_n or $F_n^{\leftarrow}$	empirical quantile function
$\mathcal{R}$	risk measure
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of positive real numbers
$rmse$	root mean squared error
r.v or r.v's	random variable(s)
$RV(\varsigma)$	regularly varying at infinity with the index ς
S_n	arithmetic sum
$\sup(A)$	supremum of set A
t	security level
$TVaR$	tail value at risk
$\mathbb{U}$	tail quantile function
$\mathbb{U}_n$	empirical tail quantile function
VaR	value at risk
$Var(X)$	mathematical variance of X

X	r.v defined on $(\Omega, \mathcal{F}, P)$, population
$\overline{X}_n$	empirical mean
$(X_1, X_2, ..., X_n)$	sample of size n from X
$(X_{1:n}, X_{2:n}, ..., X_{n:n})$	order statistics pertaining to $(X_1, X_2, ..., X_n)$
$X_{i:n}$	i-th order statistic $(i = 1, ..., n)$
$[x]$	integer part of real number x
$\lvert . \rvert$	absolute value
$\mathbf{1}\{B\}$	indicator function of set B
$(\Omega, \mathcal{F}, P)$	probability space
Λ	cumulative hazard function
Λ_n	Nelson-Aalen estimator
γ	extreme value index
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$\overset{d}{=}$	equality in distribution
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{a.s}$	convergence almost sure
$:=$	equality in definition
$\sim$	$f(x) \sim g(x)$ as $x \to x_0 : f(x)/g(x) \to 1$ as $x \to x_0$
$o(.)$	$f(x) = o(g(x))$ as $x \to x_0 : f(x)/g(x) \to 0$ as $x \to x_0$
$O(.)$	$f(x) = O(g(x))$ as $x \to x_0 : \exists M > 0, \lvert f(x)/g(x) \rvert \leq M$ as $x \to x_0$
$o_p(.)$	converges to 0 in probability
$O_p(.)$	be bounded in probability

ملخص

الهدف من هذه الأطروحة هو تطبيق مفهومين أساسيين للإحصاء الرياضي، وهما تحليل البقاء ونظرية القيم المتطرفة، لتقدير مقاييس المخاطر. توفر نظرية القيم المتطرفة أدوات لا غنى عنها لقياس احتمال وقوع حوادث غير عادية، وهو مطلب أساسي لتقدير المخاطر بدقة، حتى في وجود بيانات غير كاملة. لقد اقترحنا مُقدِّرًا لأحد أهم مقاييس المخاطر التي تسمى توقع الذيل الشرطي للبيانات ذات الذيل الثقيل والخاضعة للرقابة عشوائيًا من اليمين وأنشأنا طبيعتها المقاربة. يتم تقييم إجراء التقدير هذا من خلال دراسة محاكاة وتطبيقه على مجموعتين من البيانات الحقيقية لخسائر التأمين ووقت بقاء مرضى الإيدز.

الكلمات المفتاحية: التقارب الطبيعي؛ توقع الذيل الشرطي؛ القيم المتطرفة؛ الذيول الثقيلة؛ مقدر هيل؛ مقدر كابلان ماير؛ الرقابة العشوائية؛ مقاييس المخاطر؛ القيمة المعرضة للخطر.

Abstract

The objective of this thesis is to apply two fundamental concepts of mathematical statistics, namely survival analysis and extreme value theory, to the estimation of risk measures. Extreme values theory provides indispensable tools for measuring the probability of unusual incidents occurring, which is a basic requirement for accurate risk estimation, even in the presence of incomplete data. We proposed an estimator of one of the most important measures of risk called the conditional tail expectation of data that are heavy-tailed and randomly censored to the right and we established its asymptotic normality. This estimation procedure is evaluated through a simulation study and applied to two real datasets of insurance losses and survival time of AIDS patients.

Keywords : Asymptotic normality ; Conditional tail expectation ; Extreme values ; Heavy-tails ; Hill estimator ; Kaplan-Meier estimator ; Random censoring ; Risk measures ; Value at Risk.

Résumé

L'objectif de cette thèse est d'appliquer deux concepts fondamentaux de la statistique mathématique, à savoir l'analyse de survie et la théorie des valeurs extrêmes, à l'estimation des mesures de risque. La théorie des valeurs extrêmes fournit des outils indispensables pour mesurer la probabilité d'occurrence d'incidents inhabituels, ce qui est une exigence de base pour une estimation précise du risque, même en présence de données incomplètes. On a proposé un estimateur d'une des mesures de risque les plus importantes appelée espérance conditionnelle da la queue de données à queue lourde et censurées aléatoirement à droite et on a établi sa normalité asymptotique. Cette procédure d'estimation est évaluée par une étude de simulation et appliquée à deux ensembles de données réels de pertes d'assurance et de temps de survie des patients atteints du SIDA.

Mots Clés : Normalité asymptotique ; Espérance conditionnelle de la queue ; Valeurs extrêmes ; Queues lourdes ; Estimateur de Hill ; Estimateur de Kaplan-Meier ; Censure aléatoire ; Mesures de risque ; Valeur à risque.