DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
58

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
58

ANTS KAASIK

Estimating ruin probabilities in the Cramér-Lundberg model with heavy-tailed claims

TARTU UNIVERSITY
PRESS

Institute of Mathematical Statistics, Faculty of Mathematics and Computer Science, University of Tartu, Tartu

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (Ph. D.) in mathematical statistics on October 16, 2009, by the Council of the Institute of Mathematical Statistics, University of Tartu

Supervisor:

Professor, Cand.Sc.	Kalev Pärna University of Tartu Tartu, Estonia

Opponents:

Professor, Ph.D.	Göran Högnäs Åbo Akademi Turku, Finland
Professor RNDr., Dr.Sc.	Tomáš Cipra Charles University of Prague Prague, Czech Republic

The public defence will take place on November, 26, 2009.

Publication of the dissertation is financed by the Institute of Mathematical Statistics, University of Tartu (ESF grants 7044 and 7313 and target financed research project SF0182724s06).

ISSN 1024-4212
ISBN 978-9949-19-245-8 (trükis)
ISBN 978-9949-19-246-5 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 427

Margitile

Contents

Acknowledgements

I wish to expect my gratitude to my supervisor Professor Kalev Pärna who has helped me make this thesis a reality. Many fruitful (and sometimes very long) discussions and helpful remarks have surely led to a better dissertation. Dr. Jüri Lember has also often showed me the way when I have lost (or have not been able to find the right) direction.

I am also thankful to all my teachers and colleagues for their support and for building the foundation that has enabled my successful studies in the Institute of Mathematical Statistics.

Finally, I wish to thank my family and friends for speeding up the doctoral studies with frequent inquiries about the date of the dissertation defence. Extra motivation was gained this way.

The work is partially supported by Estonian Science Foundation grants 7044 and 7313.

List of original publications

I Kaasik, A. and K. Pärna: 2008, 'Approximating the integrated tail distribution'. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **12**, 79–87.

II Kaasik, A.: 2009a, 'An extreme value approximation to the integrated tail distribution'. In: D. Konstantinides (ed.): *Proceedings of the 5th Conference in Actuarial Science and Finance on Samos.* pp. 1–8, http://www.actuar.aegean.gr/samos2008/ Abstracts/Proceedings08.pdf. 4–7 September 2008, Karlovassi, Greece.

III Kaasik, A.: 2009b, 'Simulating the integrated tail distribution in the heavy-tailed setting'. In: K. Hangos (ed.): *Proceedings of the 28th IASTED International Conference on Modelling, Identification and Control.* pp. 28–31, ACTA Press. 16–18 February 2009, Innsbruck, Austria.

IV Kaasik, A. and K. Pärna: 2009, 'On simulation of ruin using integrated tail distribution'. In: E. K. Z. Leonidas Sakalauskas, Christos Skiadas (ed.): *The XII Conference on Applied Stochastic Models and Data Analysis: Selected papers.* pp. 45–50, Vilnius Gediminas Technical University Press. 30 June – 3 July 2009, Vilnius, Lithuania.

Introduction

Ruin probability is the quantity of prime importance in the risk management of an insurance company and a crucial indicator of an unbalanced cash flow and/or insufficient operating capital. To realistically capture the nature of the cash flow, many non-life insurance risk models have been proposed. The simplest classical model, named Cramér-Lundberg model, considers only the income due to insurance premiums and outcome due to covered claims, but ignores e.g. investment of the surplus capital. Assumption of independent and identically distributed claims is also made while income in a fixed time unit is constant. The history of this model dates back to the thesis [Lundberg, 1903]. In [Back et al., 2004: pp. 128–129] other prominent insurance risk models are discussed. This includes the Sparre-Andresen model and the Markov modulated risk model both possibly enhanced with interest earned on the free operating capital. The former model was introduced in [Sparre Andersen, 1957] and one of the first papers that considers the latter in the insurance context is [Reinhard, 1984], while a similar model was in use in the queueing theory before that as is evidenced by the paper [Ross, 1978].

While a simplified model, there are real-life situations where the Cramér-Lundberg model seems sufficient and is still of practical interest. As was discovered in [Pollaczek, 1930a] and [Pollaczek, 1930b] the ruin probability for the Cramér-Lundberg model can be expressed as the probability that the random sum of independent random variables, each having the integrated tail distribution of the claim distribution, exceeds the operating capital. That is the ruin probability

$$\psi(u) = \mathbb{P}\{X_1 + \ldots + X_N > u\},$$

where u is the operating capital, N is a geometric random variable and the distri-

bution function of each X_i is expressed as

$$F_I(x) = \frac{\int_0^x \overline{F}(y)dy}{\int_0^\infty \overline{F}(y)dy}, \quad x > 0,$$

where F is the distribution function of the claims and $\overline{F} = 1 - F$. As the parameter of the geometric random variable is not hard to estimate, the approximation of the integrated tail distribution is the key component in estimating the ruin probabilities. In the last two decades multitude of important papers like [Heidelberger, 1995], [Asmussen et al., 2000], [Boots and Shahabuddin, 2001], [Asmussen et al., 2005], [Asmussen and Kroese, 2006], [Blanchet and Glynn, 2008] dealing with the simulation of the ruin probabilities which may not be trivial when u is large, have been published. Another direction has been the development of asymptotic approximations, that are expressions $\psi^*(u)$ which satisfy

$$\lim_{u\to\infty} \frac{\psi^*(u)}{\psi(u)} = 1.$$

A crucial paper in this context is [Embrechts and Veraverbeke, 1982] but nowadays a text-book treatment of this direction is available (e.g. [Grandell, 1991] and [Asmussen, 2003]).

In this thesis a different problem – approximation of F_I – is considered. Clearly the approximation of the integrated tail distribution must be based on claim size data. While it may seem that this problem is just a special case of a distribution approximation, results for the integral functionals of the empirical distribution function from [Csörgö et al., 1986] show that the approximating quality is not the same when the ordinary empirical distribution function is considered. In the thesis theoretical and practical aspects of two possible approaches are studied – it is intuitive to consider either the empirical distribution function of the claims or its combination with the generalized Pareto distribution as the candidate for approximating claim size distribution. Generalized Pareto distribution is a basic extreme-value tool that is justified as a model of the tail of of a probability distribution under rather general assumptions.

In the most classical context it is also assumed that the claims come from a probability distribution that has moments of any order which is typically violated in

practice. In the thesis the opposite (i.e. heavy-tailed claim size distribution) is assumed.

The goal of the thesis is to describe and study a complete methodology for estimating the ruin probabilities in the heavy-tailed setting starting from the assumption of the Cramér-Lundberg model and observed claim size data together with time of occurrence. For this the following sub-goals are set.

- to describe the Cramér-Lundberg model and differences in ruin probability estimation for light and heavy-tailed claims;
- to study the properties of the integrated tail distribution of a given original distribution including generation of independent realizations from the former;
- to rigorously define approximations of the integrated tail distribution and to study their approximation qualities theoretically and using simulation;
- to illustrate the methodology proposed in the thesis using real world data as a practical example.

Chapter 1 gives an overview of the used notation, relevant definitions and important known results. This includes the practical differences between the distributions with light and heavy tails and subclasses of the latter of which subexponential distributions are the most prominent. The extreme value theorem essentially equating the generalized extreme value distribution and the generalized Pareto distribution as limit distributions of the sample maximum and conditional tail, respectively, is also presented. Means for comparing estimators in simulation are also given and the role of Brownian bridges in empirical processes theory is highlighted.

Chapter 2 gives the formal definition of the Cramér-Lundberg insurance risk process and ruin probability. Various representations for the latter are presented, most notably the Pollaczeck-Khinchine formula which allows to estimate the ruin probability as a probability of a sum with random number of summands exceeding a fixed threshold. The integrated tail distribution of an original distribution is properly defined. Implications of the tail weight of the claims for ruin probabilities are considered. A key observation is the fact that subexponentiality of the integrated tail distribution allows the approximation of the ruin probability by means

of the tail of that distribution (for large values of initial capital).

Chapter 3 begins with author's results about the relationship between the moments of the original distribution and its integrated tail distribution that are incorporated into Proposition 3.1 and Proposition 3.2 originating from [Kaasik, 2009b]. Three subexponential distributions – Pareto, Weibull and log-normal distribution – used as examples throughout the thesis are formally introduced. A general methodology for simulating from the integrated tail distribution ([Kaasik, 2009b]) is also proposed. This iterative methodology makes use of Newton-Raphson method and has a quadratic rate of convergence as shown by the author with Proposition 3.4. Proposition 3.5 is also a result from [Kaasik, 2009b] about the relation between the integrated tail distribution of Weibull and the transformed gamma distribution. In the remainder a simple conditional Monte-Carlo simulation algorithm from the literature for simulating a sum of independent heavy-tailed random variables exceeding sum fixed value is presented and its application for Pollaczeck-Khinchine formula considered.

Chapter 4 starts with a summary of methods that allow to check whether the integrated tail distribution is subexponential. Empirical approximation of the integrated tail distribution is given and its almost sure uniform convergence is proved in Proposition 4.1 originating from [Kaasik and Pärna, 2008]. Respective empirical process has an approximation process as shown by Proposition 4.2 which is a novel result. Another approximation of the integrated tail distribution is then presented. As shown in Proposition 4.3 from [Kaasik, 2009a], this extreme value approximation is justified when the tail of the original distribution behaves like a generalized Pareto distribution. In addition, the parameters of the approximating distribution can be calculated using the parameters of the distribution that approximates the claim size distribution. The methods of maximum likelihood and probability weighted moments are discussed in the context of the generalized Pareto distribution. Asymptotic normality of the parameter estimates of the integrated tail distribution is shown in Proposition 4.7.

Chapter 5 is dedicated to numerical comparison of the empirical and extreme value approach when the original distribution is subexponential. Results from [Kaasik and Pärna, 2009] show that the approximation making use of the gen-

eralized Pareto distribution outperforms the empirical approximation in the tail region. Combining the two approximations by using the empirical approximation in the main part and the extreme value approximation in the tail part yields a new approximation that might not be continuous when the parameters have been estimated using the method of maximum likelihood as shown in Proposition 5.1. To remedy this yet another approximation is presented. This new approximation coincides with the previous one when the method of probability weighted moments is used for parameter estimation and is continuous when maximum likelihood is used. Two combined approximations are compared numerically at the end of the chapter.

Chapter 6 consists of an example that makes used of the machinery described in the previous chapters for estimating the ruin probability based on a real world insurance claims data. It turns out that the methodology proposed is adequate for modeling the actual insurance risk process and thus it is tempting to believe that the ruin probability estimates obtained by making use of the combined approximations and the simulation algorithm are relatively accurate.

Chapter 7 summarizes the novel results obtained in the thesis and discusses the goals achieved.
In the thesis it is noted on several occasions that the Cramér-Lundberg risk process model is closely related with the workload process of a simple first in first out queueing system with a single server. In modern times when parallel processing is becoming the industry standard for large tasks it might also be of interest to consider the proposed methodology for analyzing and planning the performance of a server network to avoid large delays.

Chapter 1

Preliminaries

1.1. Notation

In the thesis the following notation is used.

- Capital letters typically stand for random variables. Thus X is a random variable. Suppose that X is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The notation X is a shortened notation for $X(\omega)$, where $\omega \in \Omega$. Random variable with a sub-index is just member of a sequence of random variables. Thus Y_i is the i-th random variable in a finite or countable sequence.
- Random processes are also denoted with capital letters but notation also includes an argument as in $N(t)$. Again, the argument ω is omitted.
- Probability distributions (referred to only as distributions) are also denoted with a capital letter, but for that purpose only letters F, G and H (possibly with an index) are used.
- A cumulative distribution function (cdf) can also be denoted with the same capital letters but notation also includes an argument as in $F(x)$. When speaking about the fact that X has distribution G it is also meant that X has a cdf $G(x) = \mathbb{P}\{X \leqslant x\}$ (if X was defined on a probability space with measure $\mathbb{P}$). The converse is also used – if Y has a cdf $F(x)$ then the distribution of Y is F.
- Suppose $X_1, \ldots, X_n$ are independent and identically distributed (iid) random variables with distribution F. Then their sum has distribution F^{*n}, that is F^{*n} denotes the n-fold convolution of F.
- Notation $\overline{F}(x)$ stands for a complementary cdf i.e. $\overline{F}(x) := 1 - F(x)$

- Expectation is denoted with $\mathbb{E}$ and variance with Var. Thus the mean of X is $\mathbb{E}X$ while $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.
- Symbol $\sim$ stands for asymptotic equivalence in the specified process. Thus $a(x) \sim b(x)$, $x \to \infty$ denotes the fact that $\lim_{x\to\infty} a(x)/b(x) = 1$.
- Acronyms GPD and CLM refer respectively to generalized Pareto distribution from Definition 1.9 and Cramér-Lundberg insurance risk model defined in Section 2.1.

Additional notation is explained after its introduction.

1.2. Light and heavy-tailed distributions

The following definitions and results are all well-known and are collected from [Embrechts et al., 1997: pp. 49–57], [Klugman et al., 2004: pp. 39–114] and [Sigman, 1999].

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X with distribution F that has support $(0, \infty)$.

Definition 1.1. *If the moment generating function (mgf) of X defined as $\mathbb{E}\exp\{tX\}$ is finite for some $t > 0$ then we say that X is a light-tailed random variable. Otherwise X is a heavy-tailed random variable. We also say that F is, respectively, a light-tailed or a heavy-tailed distribution.*

Finite mgf for some positive argument guarantees the existence of all moments. However, it is possible that the mgf of a random variable is infinite for all $t > 0$, but the random variable still possesses all moments.

Definition 1.2. *If for all $y \geqslant 0$ it holds that*

$$\overline{F}(x+y) \sim \overline{F}(x), \ x \to \infty, \tag{1.1}$$

then we say that the distribution F belongs to the class of long-tailed distributions and denote this by $F \in \mathcal{L}$. Alternatively, we say that X is a long-tailed random variable.

$F \in \mathcal{L}$ implies that F is a heavy-tailed distribution. Thus long-tailed distributions form a sub-class of heavy-tailed distributions.

Definition 1.3. *If for all $n \geqslant 2$ we have that*

$$\overline{F^{n*}}(x) \sim n\overline{F}(x) \tag{1.2}$$

then we say that F belongs to the class of subexponential distributions and denote it with $F \in \mathcal{S}$. Alternatively, we say that X is a subexponential random variable.

$F \in \mathcal{S}$ implies that $F \in \mathcal{L}$. Thus subexponential distributions form a sub-class of long-tailed distributions. It is also known that $\mathcal{L}$ is a proper sub-class of heavy-tailed distributions and $\mathcal{S}$ is a proper sub-class of $\mathcal{L}$. Still, majority of the well-known heavy-tailed distributions are also subexponential. Class $\mathcal{S}$ includes, for example, Pareto, Weibull, log-normal, Burr, log-gamma and Benktander distributions.

Definition 1.4. *We say that F is a dominatedly varying distribution if*

$$\limsup_{x\to\infty} \frac{\overline{F}(x/2)}{\overline{F}(x)} < \infty. \tag{1.3}$$

and denote it by $F \in \mathcal{D}$.

Class $\mathcal{D}$ is also a sub-class of heavy-tailed distributions.

Definition 1.5. *We say that F is a regularly varying distribution with index $-\alpha$, where $\alpha > 0$, if for every $t > 0$ it holds that*

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha}. \tag{1.4}$$

and denote it by $F \in \mathcal{R}_{-\alpha}$.

Class $\mathcal{R}$ defined as $\cup_{\alpha\geqslant 0}\mathcal{R}_{-\alpha}$ is a proper sub-class of both $\mathcal{S}$ and $\mathcal{D}$.
There are also other means to classify the tails of distributions.

Definition 1.6. *The function*

$$a(x) = \mathbb{E}(X - x|X > x) = \frac{\int_x^\infty \overline{F}(y)dy}{\overline{F}(x)}, \tag{1.5}$$

defined for $x \geqslant 0$ is the mean excess of loss function for X (or mean excess of loss function for F).

Suppose now that X also has a density function $f(x) = F'(x)$.

Definition 1.7. *The function*

$$h(x) = \frac{f(x)}{\overline{F}(x)}, \tag{1.6}$$

is called the hazard rate of X (or hazard rate of distribution F).

Typically a heavy-tailed distribution has an eventually increasing mean excess of loss function (sometimes also referenced as mean residual life function) and an eventually decreasing hazard rate, that is, there exist $x_1 > 0$ and $x_2 > 0$ such that its mean excess of loss function is increasing in (x_1, ∞) and its hazard rate is decreasing in (x_2, ∞). These facts can be easily interpreted if X is considered to be the size of an insurance claim. Consider a threshold $x > 0$ that we can adjust. Increasing $a(x)$ would mean that the average of the claims that exceed a threshold x is growing faster than x itself. Decreasing $h(x)$ would roughly mean that the the claims which exceed a threshold x are less likely to be close to x when we increase the threshold.

1.3. Basics of extreme-value theory

The following definitions and results are all well-known and are collected from [Embrechts et al., 1997: pp. 113–168] and [De Haan and Ferreira, 2006: pp. 6–12].

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X with distribution F that has support $(0, \infty)$.

Definition 1.8. *The conditional distribution function of $X - u$ given that $X > u$ is denoted by $F_u(y)$ and defined as*

$$F_u(y) = \frac{F(u+y) - F(u)}{\overline{F}(u)}, \quad u > 0, y > 0. \tag{1.7}$$

Definition 1.9. *The cdf of the generalized Pareto distribution (GPD) is defined as*

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \sigma > 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0, \sigma > 0, \end{cases} \tag{1.8}$$

where $0 < y < -\sigma/\xi$ for $\xi < 0$ and $0 < y < \infty$ for $\xi \geqslant 0$.

Note that the case $\xi = 0$ can be thought of as a limit because

$$\lim_{\xi \to 0} G_{\xi,\sigma}(y) = G_{0,\sigma}(y), \quad y > 0. \tag{1.9}$$

The mean of the GPD is $\sigma/(1-\xi)$ if $\xi < 1$ and infinite otherwise.

Theorem 1.1. Suppose that there exist constants $a_n > 0$ and b_n such that

$$\lim_{n \to \infty} F^n(a_n x + b_n) = \exp\left(-(1+\xi x)^{-1/\xi}\right) \tag{1.10}$$

for some $\xi \geqslant 0$ and for all x, where the case $\xi = 0$ is considered as $\xi \to 0$. Then and only then exists a positive function $a(x)$ such that

$$\lim_{u \to \infty} \sup_{x>0} |\overline{F_u}(xa(u)) - \overline{G}_{\xi,1}(x)| = 0. \tag{1.11}$$

The distribution function that appears on the right hand side of (1.10) is known as the generalized extreme value distribution and is the only possible non-degenerate distribution that can appear as a limit in (1.10). The parameter ξ can be negative for certain distributions but a generalized extreme value distribution with $\xi < 0$ cannot appear as a limit when F has support $(0, \infty)$ and is thus not relevant here. In the case when $\xi = 0$ the mean excess of loss function can be used in the role of function $a(x)$, while $a(x) = \xi x$ is a possible choice for the case $\xi > 0$.

The existence of constants a_n and b_n so that (1.10) holds is usually referred to as the fact that the extreme value (EV) condition is satisfied (for the distribution F). As a non-degenerate limit is not always attainable, several conditions for the checking the EV condition are available (e.g. in [De Haan and Ferreira, 2006: pp. 15–23]). For example, the distributions satisfying the EV condition with $\xi > 0$ are the ones with regularly varying tail. More precisely $F \in \mathcal{R}_{-\alpha}$ if and only if it satisfies the EV condition with $\xi = 1/\alpha$ and thus when F has finite mean, necessarily $\xi < 1$.

Note that (1.11) can also be presented as

$$\lim_{u \to \infty} \sup_{x>0} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0, \tag{1.12}$$

with $\sigma(u) = a(u)$. This formulation dating back to [Pickands III, 1975] is the basis for approximating the tails of distributions.

In general, the EV condition is satisfied for distributions with continuous cdf and thus the assumption that there exist such ξ and $\sigma(u)$ that (1.12) holds is typically

not restrictive in practice.

1.4. Assessing the quality of an estimator in simulation

Suppose we want to know the probability of some event A_u in a situation which we can simulate and that probability depends on some initial parameter u. Denote this probability as $\gamma(u)$ and further assume that $\gamma(u) \to 0$ as $u \to \infty$. The naive way of estimating the probability $\gamma(u)$ would be to simulate the situation and set $I_1 = 1$ if A_u occurs and $I_1 = 0$ otherwise. This can be repeated n times and we would have an unbiased estimator of $\gamma(u)$ in the form of

$$Q_1(u) := \frac{1}{n} \sum_{i=1}^{n} I_i, \tag{1.13}$$

due to the law of large numbers. Moreover, the central limit theorem allows us to calculate an approximate $(1-\alpha)100\%$ confidence interval (CI) for $\gamma(u)$ as

$$Q_1(u) \pm z_{\alpha/2} \sqrt{\frac{Var(I_i)}{n}}, \tag{1.14}$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.
We now analyze the relative error of the estimator defined as half-width of the CI relative to the size of $\gamma(u)$ to find the required number of replications n. Because $Var(I_i) = \gamma(u)(1-\gamma(u))$, we have that the relative error is expressible as

$$z_{\alpha/2} \sqrt{\frac{1-\gamma(u)}{\gamma(u)n}}. \tag{1.15}$$

The following definitions are from [Juneja and Shahabuddin, 2006].

Definition 1.10. *An unbiased estimator $Q(u)$ of probability $\gamma(u)$ has bounded relative error if the quotient*

$$\frac{Var(Q(u))}{[\gamma(u)]^2} \tag{1.16}$$

is bounded in u.

Definition 1.11. *An unbiased estimator $Q(u)$ of probability $\gamma(u)$ is asymptotically efficient if*

$$\lim_{u \to \infty} \frac{\log \mathbb{E}([Q(u)]^2)}{2 \log \gamma(u)} = 1, \tag{1.17}$$

Bounded relative error means that there is a fixed number of replications that guarantees desired relative accuracy (regardless of the value u). Asymptotic efficiency means that the estimator is asymptotically optimal on a logarithmic scale. Bounded relative error implies asymptotic efficiency. The estimator Q_1 defined in (1.13) is not an estimator with a bounded relative error as is seen from (1.15). Furthermore, we have that Q is not even asymptotically efficient as we have that

$$\lim_{u\to\infty} \frac{\log \mathbb{E}([Q(u)]^2)}{2\log\gamma(u)} = \lim_{u\to\infty} \frac{\log\left(\frac{\gamma(u)(1-\gamma(u))}{n} + [\gamma(u)]^2\right)}{2\log\gamma(u)} = \frac{1}{2}.$$

1.5. Basics of the theory of empirical processes

We now list some of the well-known facts from the theory of empirical processes that are useful in what follows. The key result is from [Komlós et al., 1975].

Definition 1.12. *A Gaussian random process $B(t)$ defined for $t \in [0,1]$ is called a Brownian bridge if $\mathbb{E}B(t) = 0$ for each t and $\mathbb{E}B(s)B(t) = s(1-t)$ whenever $s \leqslant t$ and the process has continuous trajectories.*

If we suppose that $X_1, X_2, \ldots$ are iid with continuous cdf $F(x)$ then we can first define an empirical cdf $F_n(x)$ as

$$F_n(x) = \frac{\{\#X_i : X_i \leqslant x,\ i = 1, \ldots, n\}}{n} \tag{1.18}$$

for each n and then the uniform empirical process

$$\alpha_n(y) = \sqrt{n}[y - F_n(F^{-1}(y))] \tag{1.19}$$

which can be approximated by a standardized sum of independent Brownian bridges as is made precise by the following statement slightly tailored for our context.

Theorem 1.2. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_1, X_2, \ldots$ are iid random variables with a continuous distribution function $F(x)$ and independent Brownian bridges $B_n(t)$ defined on that probability space and

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^2} \sup_{0\leqslant y\leqslant 1} |\alpha_n(y) - B_n(y)| \leqslant C\right) = 1 \tag{1.20}$$

for some positive constant C.

While a much stronger result, one of the key conclusions from Theorem 1.2 is that the absolute difference between $F(x)$ and $F_n(x)$ inflated $\sqrt{n}$ times has roughly the same distribution as a Brownian bridge at $F(x)$ for large values of n.

Chapter 2

Cramér-Lundberg insurance risk model and ruin probabilities

2.1. Definition of the Cramér-Lundberg insurance risk model

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and introduce a stochastic process

$$U_u(t) = u + ct - S(t), \quad S(t) = \sum_{i=1}^{N(t)} X_i, \quad t \geqslant 0, \tag{2.1}$$

where $u \geqslant 0$ and $c > 0$ are constants, X_i are independent and identically distributed (iid) random variables with distribution F that has finite mean and support $(0, \infty)$ while $N(t)$ is a counting process. Process (2.1) is the classical model for insurance risk: u stands for the initial capital of the insurance company, c for the premium rate in a time unit, X_i are the claims from the insured subjects and $N(t)$ equals the number of claims made in time interval $[0, t]$. This model is also known as the Sparre-Andersen model. A simplification of (2.1) is known as the Cramér-Lundberg insurance risk model (CLM) where the process $N(t)$ is assumed to be a Poisson counting process independent of all the X_i with intensity $\lambda > 0$. This means, among other things, that the process has independent increments and the (random) number of claims occurring in a time interval of arbitrary length $t_* \geqslant 0$ has a Poisson distribution with mean λt_* i.e.

$$\forall s \geqslant 0, \quad \mathbb{P}\{N(t_* + s) - N(s) = n\} = \frac{(\lambda t_*)^n}{n!} \exp\{-\lambda t_*\}, \quad n = 0, 1, 2, \ldots.$$

If we denote the arrival time of the i-th claim by T_i (i.e. $T_i = \inf\{t \geqslant 0 | N(t) = i\}$) and take $T_0 = 0$ then the consequence of CLM is that random variables $Y_i := T_i - T_{i-1}$ defined for $i = 1, 2, \ldots$ are iid with exponential distribution and mean

$1/\lambda$ [Klugman et al., 2004: p. 224]. Further let $\Theta := \{T_1, T_2, \ldots\}$. In the following we will assume the CLM.

The object of interest is the probability

$$\psi(u) := \mathbb{P}\{\exists t < \infty : U_u(t) < 0\}, \tag{2.2}$$

known as the (ultimate) ruin probability which can be interpreted as a chance that an insurance company with initial capital u will eventually go bankrupt.

Remark 2.1. It is possible to define "a new time" by changing $t \to t/c$ i.e. we set a new time unit so that the premium rate in one new time unit is equal to one. For this new process we also have the change $\lambda \to \lambda/c$. Probability (2.2), however, remains unchanged, and thus in what follows it is assumed that $c = 1$.

To avoid the situation where $\psi(u) = 1$, it is assumed that $\mathbb{E}[S(t)] < t$ which means that

$$\rho := \frac{1}{\lambda \mathbb{E}X} - 1 > 0. \tag{2.3}$$

Constant ρ is known as the safety loading coefficient and it holds that

$$\psi(0) = \frac{1}{1+\rho} \tag{2.4}$$

(see e.g. [Klugman et al., 2004: pp. 236–237]).

Remark 2.2. One can also consider the finite-time ruin probability i.e.

$$\psi(u, T) := \mathbb{P}\{\exists t < T : U_u(t) < 0\}. \tag{2.5}$$

Clearly $\psi(u, T) \leqslant \psi(u)$.

Infinite time ruin probability can be expressed in a more convenient way than (2.2) and the next section is devoted to these representations.

2.2. Two representations of the ruin probability

Suppose $U_u(t) < 0$ for some $t < \infty$ i.e the company has ruined. Then by (2.1) it is clear that the time instant when the ruin occurs (i.e. $\tau := \inf\{t \geqslant 0 | U_u(t) < 0\}$)

must belong to the set Θ. Denote

$$R_0 = 0, \quad R_n = \sum_{k=1}^{n} (X_k - Y_k), \quad n \geqslant 1, \tag{2.6}$$

which is a random walk with a negative drift due to fact that $\rho > 0$. We can then give the random walk representation of the ruin probability

$$\psi(u) = \mathbb{P}\left\{\sup_{n \geqslant 0} R_n > u\right\} \tag{2.7}$$

to be found e.g. in [Embrechts et al., 1997: p. 26]. Departing from this equation, another useful representation can be deduced by the following argument, presented precisely in [Klugman et al., 2004: pp. 240–241]. Suppose there is a T_i such that $U_u(T_i) < u$ while for $k < i$ we have $U_u(T_k) \geqslant u$. Probability that such a T_i exists is $\psi(0)$ by (2.7). It turns out that the positive random variable $u - U_u(T_i)$, provided it is defined, has a cumulative distribution function (cdf)

$$F_I(x) = \frac{\int_0^x \overline{F}(y)dy}{\mathbb{E}X}, \quad x > 0. \tag{2.8}$$

This distribution is called the integrated tail distribution (itd) of F. Because Y_i are iid exponential random variables and independent of random variables X_i, process $U_u(t)$ can be "restarted" at any instant. Thus the probability that there is a T_j with $j > i$ such that $U_u(T_j) < U(T_i)$ while for $i < k < j$ we have $U_u(T_k) \geqslant U_u(T_i)$ is again $\psi(0)$. Moreover $U_u(T_j) - U_u(T_i)$ has also the itd of F and is independent of $u - U_u(T_i)$. The number of "new records", N, is thus geometrically distributed i.e.

$$\mathbb{P}\{N = n\} = \left(\frac{1}{1+\rho}\right)^n \left(\frac{\rho}{1+\rho}\right), \quad n = 0, 1, 2, \ldots,$$

because of (2.4), and we have

$$\psi(u) = \mathbb{P}\{Z_1 + \ldots + Z_N > u\}, \tag{2.9}$$

where random variables Z_i are iid with distribution F_I. This representation of the ruin probability is known as the Pollaczeck-Khinchine formula and will be one of the starting points of the thesis. We will see in the next section that the ruin probability coincides with another important probability from queueing theory.

2.3. Steady-state waiting time for a $M/G/1$ queue

$M/G/1$ queue process is a probabilistic model for a service system. There is one server and the service times of the customers are iid random variables X_i. The times between arrivals of subsequent customers are iid exponential random variables Y_i independent of service times (Y_i is the time between the arrivals of customers i and $i+1$ and the first customer enters the system instantly, also $\mathbb{E}Y_i > \mathbb{E}X_i$). The serving order principle is first in - first out. Thus if a customer i enters the system and there are still unserved customers there will be a positive waiting time W_i till the service of customer i begins. It can be expected that when i grows large the waiting time W_i will reach an equilibrium distribution i.e. the queue process will enter a steady state (because the workload in the system (the total time required to complete serving all the customers currently in the system) is a Markov process [Asmussen, 2003: p. 96] and the steady state workload and the waiting time have the same distribution [Asmussen, 2003: pp. 108–109]). Let us find that equilibrium distribution using the following reasoning. Denote $\xi_i = X_i - Y_i$ for $i = 1, 2, \ldots$. Now $W_1 = 0$ and $W_{i+1} = (W_i + \xi_i)^+$ for $i = 2, 3, \ldots$, where $W^+ := \max\{W, 0\}$, and thus recursively

$$W_{i+1} = \max\{\xi_1 + \ldots + \xi_i, \xi_2 + \ldots + \xi_i, \ldots, \xi_i, 0\}.$$

Note that ξ_i are iid random variables and thus the distribution of W_{i+1} is that of $\max\{0, \xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \ldots + \xi_i\}$. The steady-state is reached when $i \to \infty$ and this leads us to (2.7) with $R_0 = 0$ and $R_n = \sum_{i=1}^{n} \xi_i$.

We can conclude that whenever we speak about the ruin probability in the CLM context with initial capital u, we might as well think about the probability that the steady state waiting time of a $M/G/1$ queue exceeds u, given that the claim size distribution and service time distribution are the same and the intensities of claim and customer arrival are also equal.

Remark 2.3. It can be shown that the duality discussed previously remains intact when we consider a more general insurance risk model (2.1) where $N(t)$ may not be a Poisson process. The infinite horizon ruin probability then corresponds to the steady-state waiting time of a $G/G/1$ queue – a service system where times between arrivals of subsequent customers are not necessarily exponential [Asmussen, 2003: pp. 399-400].

2.4. Claim size distribution

So far we have mentioned only the requirement $\mathbb{E}X_i < \infty$. In the classical setting it is further assumed that there exists a $\kappa > 0$ such that

$$\lambda \int_0^\infty e^{\kappa x} \overline{F}(x) dx = 1,$$

where κ is known as the adjustment coefficient [Embrechts et al., 1997: p. 32]. With this additional requirement two important results can be proved: firstly the Lundberg inequality stating that

$$\psi(u) \leqslant e^{-\kappa u}$$

for all $u > 0$ and secondly, the Cramér-Lundberg approximation formula

$$\psi(u) \sim C e^{-\kappa u}, \ u \to \infty \tag{2.10}$$

where $0 < C \leqslant 1$ is an explicit constant [Klugman et al., 2004: pp. 230–244]. Existence of the adjustment coefficient implies the existence of the moment generating function for some $t > 0$ [Klugman et al., 2004: p. 226] which in turn implies the existence of all moments. As noted in [Embrechts et al., 1997: pp. 32–34], distributions that satisfy this assumption are usually not good models for real claim data and thus the use of the classical results is limited in practice.

Using (2.9) we may write

$$\psi(u) = \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} \frac{\overline{F_I^{n*}}(u)}{(1+\rho)^n} \tag{2.11}$$

and if we would have $F_I \in \mathcal{S}$ then

$$\psi(u) \sim \frac{\overline{F_I}(u)}{\rho}, \quad u \to \infty, \tag{2.12}$$

because it is known that for $F_I \in \mathcal{S}$ interchanging limit and sum is justified in the previous situation [Embrechts et al., 1997: p. 41].

Remark 2.4. Note that (2.11) holds regardless of the tail of F, but the subexponentiality requirement $F_I \in \mathcal{S}$ is necessary for (2.12). Methodology developed in

the thesis is applicable for both heavy and light tails but in examples only subexponential distributions, as possible candidates for modeling real world data, are used in the role of F.

Remark 2.5. It can be shown that if $F_I \in \mathcal{S}$ and $F \in \mathcal{L}$ then the asymptotic relation (2.12) holds also for the classical insurance risk model (2.1) i.e. even when $N(t)$ is not a Poisson process [Asmussen, 2000: pp. 261–264].

Chapter 3

Ruin probability estimation when the claim size distribution is known

Suppose we have observed a process governed by the CLM for some time during which n claims have arrived. Perhaps we have u units of capital left and we want to determine the premium rate c so that $\psi(u)$ is smaller than some desired level, say α. Another possibility is that we would like to know how much initial capital is required to have the ruin probability smaller than α with c fixed. In either case we expect the relation (2.9) to help us answer these questions as direct simulation using either (2.1) or (2.7) is not possible due to the infinite time horizon and infinite number of summands, respectively. The goal of this chapter is to show how to use the Pollaczeck-Khinchine formula (2.9) for ruin probability estimation via Monte-Carlo method when the distribution of the claims F is known and has a heavy tail. Main ingredients needed are the skill to simulate from the integrated tail distribution of F and a good simulation algorithm to achieve a precise result with minimum effort.

3.1. Some properties of the integrated tail distribution

Suppose we know both F (thus also $\mathbb{E}X$) and λ. Using Cramér-Lundberg approximation (2.10) is one possibility for estimating the ruin probability in the light-tailed context and even better approximations are available for this type of claims [Klugman et al., 2004: pp. 245–250]. If $F_I \in \mathcal{S}$ then we can use (2.12) to provide an estimate for the ruin probability. It can be the case, however, that the approximation (2.12) is rather inaccurate for moderate u as shown in [Mikosch and Nagaev, 1998]. However, in any case (2.9) can be used to estimate $\psi(u)$. As it is typically (but not always, see e.g. [Klugman et al., 2004: p. 154]) impossible

to find an analytical solution, simulation must be used to generate $Z_1, \ldots, Z_N$ with distribution F_I. For a light-tailed distribution efficient simulation is rather straightforward as importance sampling can be used with mgf determining the optimal distribution, see e.g. [Heidelberger, 1995], [Ross, 2002: pp. 166–179]. As the mgf does not exist for heavy-tailed distributions, better understanding of F_I is required for efficient simulation. We will see that the itd F_I typically has an even heavier tail than the original distribution F which should imply that algorithms working well for heavy-tailed F provide good results for F_I.

The fact that F_I typically has a heavier tail can be expected because of the following result presented in [Kaasik, 2009b], which says that the moment of order $n+1$ of the original distribution F always needs to exist for the moment of order n of F_I to exist.

Proposition 3.1. Let X be a non-negative random variable with distribution function F and finite mean. Let X_I be another random variable with the integrated tail distribution of F. Then for any $n \in \{1, 2, \ldots\}$, for which $\mathbb{E}X_I^n < \infty$, it holds that

$$\mathbb{E}X_I^n = \frac{\mathbb{E}X^{n+1}}{(n+1)\mathbb{E}X}. \tag{3.1}$$

For the proof we use the following simple result.

Lemma 3.1. For a non-negative random variable X with distribution function F and finite n-th moment it holds that

$$\mathbb{E}X^n = n \int_0^\infty x^{n-1} \overline{F}(x) dx. \tag{3.2}$$

Proof of Lemma 3.1. Partial integration gives

$$n \int_0^\infty x^{n-1} \overline{F}(x) dx = \lim_{x \to \infty} x^n \overline{F}(x) + \int_0^\infty x^n dF(x),$$

but

$$\lim_{x \to \infty} x^n \overline{F}(x) = \lim_{x \to \infty} x^n \int_x^\infty dF(y) \leqslant \lim_{x \to \infty} \int_x^\infty y^n dF(y) = 0,$$

because $\int_0^\infty y^n dF(y)$ is assumed to be finite. □

Proof of Proposition 3.1. We write

$$\mathbb{E}X_I^n = \frac{\int_0^\infty x^n \overline{F}(x)dx}{\mathbb{E}X} = \frac{\mathbb{E}X^{n+1}}{(n+1)\mathbb{E}X}$$

using Lemma 3.1. □

The result can be extended as follows.

Proposition 3.2. Let X be a non-negative random variable with distribution function F and finite m-th moment. Set $F_{I_0} = F$ and $F_{I_1} = F_I$ and denote by

$$F_{I_m}(x) = \frac{\int_0^x \overline{F}_{I_{m-1}}(y)dy}{\mathbb{E}X_{I_{m-1}}},$$

the m-th order integrated tail distribution of F where $X_{I_{m-1}}$ is a random variable with distribution $F_{I_{m-1}}$. Then for any $n \in \{1, 2, \ldots\}$ for which $\mathbb{E}(X_{I_m})^n < \infty$ it holds that

$$\mathbb{E}X_{I_m}^n = \binom{m+n}{m}^{-1} \frac{\mathbb{E}X^{m+n}}{\mathbb{E}X^m}. \tag{3.3}$$

Proof of Proposition 3.2. The statement can be proved by induction. Proposition 3.1 shows that the statement holds true for $m = 1$. Suppose it holds also for $m = k$. Then we can write

$$\begin{aligned}\mathbb{E}X_{I_{k+1}}^n &= \frac{\mathbb{E}X_{I_k}^{n+1}}{(n+1)\mathbb{E}X_{I_k}} = \frac{k!(n+1)!(k+1)!}{(k+n+1)!k!(n+1)} \frac{\mathbb{E}X^{k+n+1}}{\mathbb{E}X^{k+1}} \\ &= \binom{k+n+1}{k+1}^{-1} \frac{\mathbb{E}X^{k+n+1}}{\mathbb{E}X^{k+1}}\end{aligned}$$

as required. □

Class $\mathcal{L}$ is important because if either $F \in \mathcal{L}$ or $F_I \in \mathcal{L}$ then F_I has a heavier tail than F in a sense that

$$\lim_{x\to\infty} \frac{\overline{F_I}(x)}{\overline{F}(x)} = \infty \tag{3.4}$$

holds [Sigman, 1999]. One of the consequences of (3.4) is the fact that for heavy-tailed claims $\psi(u)$ is not a sharp upper bound for the finite time ruin probability $\psi(u, T)$ when u is large. This is made precise by the following statement which has not been found in the literature.

Corollary 3.1. Suppose CLM holds with $F \in \mathcal{S}$ and $F_I \in \mathcal{S}$. Then $\forall T > 0$ we have

$$\lim_{u\to\infty} \frac{\psi(u,T)}{\psi(u)} = 0. \tag{3.5}$$

Proof. Without a loss of generality we consider the setup of CLM with $c = 1$. A well known fact to be found e.g. in [Sigman, 1999] with notation from (2.1) gives us that

$$\mathbb{P}\{S(T) > u\} \sim \lambda T\overline{F}(u), \quad \forall T > 0,$$

but we have also that

$$\mathbb{P}\{S(T) > u\} \sim \mathbb{P}\{S(T) > u + T\}, \quad \forall T > 0,$$

because $F \in \mathcal{L}$. The probabilities in the previous relation are an upper and a lower bound, respectively, for $\psi(u,T)$. This can be seen by noting that ruin could not have happened in $[0,T]$ if $S(T) \leqslant u$ while if $S(T) > u+T$ then ruin has definitely occurred in $[0,T]$. Hence we have also that

$$\psi(u,T) \sim \lambda T\overline{F}(u), \quad \forall T > 0.$$

This, together with (2.12) and (3.4), proves the statement. □

3.2. Subexponential distributions

If we were to use (2.12) it would be important to know whether F_I is subexponential. A key observation about the class (apart from the fact that it arises naturally in the ruin theory context as was shown in Section 2.4) is the following. Let M_n be the maximum of n iid random variables X_i with a subexponential distribution F. Then we have for all n that

$$\mathbb{P}\{X_1 + \ldots + X_n > x\} \sim \mathbb{P}\{M_n > x\}, \; x \to \infty, \tag{3.6}$$

because $\mathbb{P}\{M_n > x\} \sim n\overline{F}(x)$ in the process $x \to \infty$. This means intuitively that the sum grows very large due to a single large summand which in our context highlights the fact that a single very large claim can be the cause of ruin when we deal with subexponential claims.

3.2.1. Subexponentiality of the integrated tail distribution

Using the definition to check whether a distribution is subexponential is usually not possible. However, there are several sufficient conditions for F to check whether $F_I \in \mathcal{S}$. For example, it suffices to show that $F \in \mathcal{D}$ (which is considerably simpler) to have $F_I \in \mathcal{S}$ [Embrechts et al., 1997: pp. 52–53]. Sufficient conditions based on the hazard rate of F also exist [Klüppelberg, 1989].

Since $H \in \mathcal{S}$ together with $\overline{H}(x) \sim c\overline{G}(x)$ and $c > 0$ implies that $G \in \mathcal{S}$ (see e.g. [Sigman, 1999]), it follows from (2.12) that the random variable $Z_1 + \ldots + Z_N$ in (2.9) is subexponential if $F_I \in \mathcal{S}$.

3.2.2. Examples of subexponential distributions

The three examples used throughout the thesis in the role of F are the Pareto distribution, the (heavy-tailed) Weibull distribution and the log-normal distribution. All three are subexponential distributions, as noted before. Furthermore $F_I \in \mathcal{S}$ in each case [Embrechts et al., 1997: p. 56].

1. Pareto distribution has a tail $\overline{F}(x) = (1+x)^{-\alpha}$ with $\alpha > 0$ and finite mean when $\alpha > 1$.

2. Weibull distribution has a tail $\overline{F}(x) = e^{-x^\beta}$ with $0 < \beta < 1$. It is possible to consider the distribution with $\beta \geqslant 1$ but then the distribution is no longer heavy-tailed.

3. Log-normal distribution has a tail $\overline{F}(x) = \overline{\Phi}((\log x)/\sigma)$ with parameter $\sigma > 0$, where $\Phi(x)$ is the cdf of a standard normal random variable.

Those three distributions illustrate richness of the class $\mathcal{S}$ as the contrast between the Pareto and Weibull distribution is rather strong – for the Pareto distribution with parameter α the moments of order $n \geqslant \alpha$ do not exist [Balakrishnan and Nevzorov, 2003: p. 154] while the Weibull distribution has moments of all orders regardless of the parameter β [Sigman, 1999]. The lognormal distribution is between the two others in terms of heaviness of the tail.

Proposition 3.3. If we have $\overline{F_1}(x) = (1+x)^{-\alpha}$, $\overline{F_2}(x) = \exp\{-x^\beta\}$ and $\overline{F_3}(x) = \overline{\Phi}((\log x)/\sigma)$ for $x > 0$, where $a > 0$, $0 < \beta < 1$ and $\sigma > 0$, then

$$\lim_{x\to\infty} \frac{\overline{F_1}(x)}{\overline{F_3}(x)} = \infty, \quad \lim_{x\to\infty} \frac{\overline{F_2}(x)}{\overline{F_3}(x)} = 0. \tag{3.7}$$

Proof. Using the l'Hôpital's rule repeatedly we obtain that

$$\lim_{x\to\infty}\frac{\overline{F_1}(x)}{\overline{F_3}(x)}=\lim_{x\to\infty}\frac{(1+x)^{-\alpha}}{\overline{\Phi}\left(\frac{\log x}{\sigma}\right)}=\lim_{x\to\infty}\frac{\alpha\sigma\sqrt{2\pi}x\exp\{\frac{(\log x)^2}{2\sigma^2}\}}{(1+x)^{\alpha+1}}=\infty$$

and that

$$\lim_{x\to\infty}\frac{\overline{F_2}(x)}{\overline{F_3}(x)}=\lim_{x\to\infty}\frac{\exp\{-x^\beta\}}{\overline{\Phi}\left(\frac{\log x}{\sigma}\right)}=\lim_{x\to\infty}\beta\sigma\sqrt{2\pi}x^\beta\exp\{-x^\beta+\frac{(\log x)^2}{2\sigma^2}\}=0.$$

□

Remark 3.1. We could consider the mentioned distributions with more parameters by introducing location and scale parameters. However the location parameter would cause the distributions support to change from the desired $(0,\infty)$ and the scale parameter would not change the tail behavior in terms of heaviness which is the main object of interest in the simulations presented in the thesis.

3.3. Simulating from the integrated tail distribution

Relation (2.9) shows that we will have to simulate from the itd which might not be a trivial task. Firstly, the analytical inversion method is not be applicable if the cdf is not analytically invertible. Secondly, the rejection method requires careful selection of the bounding density to guarantee a decent acceptance rate, due to the heavy tail of the distribution (see e.g. [Ross, 2002: pp. 63–72] for the two mentioned random variable generation methods).

Note that because of its definition (2.8) the itd F_I, also known as the equilibrium distribution [Klugman et al., 2004: p. 53], always has a density

$$f_I(x):=\frac{\overline{F}(x)}{\mathbb{E}X},\quad x>0. \tag{3.8}$$

The existence of a monotone density function implies that effective numerical inversion of the cdf is possible as explained in [Kaasik, 2009b]. Namely, the following statement holds.

Proposition 3.4. Suppose the distribution F with support $(0,\infty)$ has a continuous cdf $F(x)$. Let u be a random number from the standard uniform distribution.

Then the iteration process

$$x_{i+1} = x_i + \frac{\mu\left[\overline{F_I}(x_i) - u\right]}{\overline{F}(x_i)}, \quad i = 0, 1, 2, \ldots$$

with initial value $x_0 = \mu(1-u)$ leads to the solution of the equation $\overline{F_I}(x) = u$. The rate of convergence is quadratic.

Proof. The proof is based on the properties of the Newton-Raphson method from the numerical analysis, description of which is available e.g. in [Booth, 1955: pp. 148–151]. Denote $g(x) := \overline{F_I}(x) - u$ and we are interested in finding the x^* for which $g(x^*) = 0$ holds. First note that $g(x)$ is a strictly convex function. This is because

$$\left[\overline{F_I}(x) - u\right]'' = \left[-\overline{F}(x)/\mu\right]' > 0.$$

as $F(x)$ was assumed to be continuous and support of F was assumed to be unbounded. Iteration (3.4), which is just the Newton's method applied for this specific $g(x)$, is now guaranteed to converge for any non-negative starting point in the left neighborhood of x^* because we have further that $g(x) > 0$ and $g'(x) \neq 0$ for $x < x^*$. The starting point proposed in the statement of the theorem is the x_1 of the iteration process when the starting point is zero. As $g(x)$ is continuously differentiable ($F(x)$ is continuous) and the $g'(x^*) \neq 0$ the quadratic rate of convergence follows. □

If u is generated from the standard uniform distribution then the root of the equation $\overline{F_I}(x) - u = 0$ is a random number from the distribution F_I [Ross, 2002: p. 63] and the proposition offers a possible method for simulating form the integrated tail distribution. Because of the quadratic rate of convergence, usually just a couple of steps are required for a good approximation of the root.

Remark 3.2. The described methodology for generating independent random variables from the itd is universal. However, because the method is based on an iterative algorithm, analytical inversion method remains preferable whenever it can be used.

For the Pareto distribution the random variable generation from its itd is straightforward because when $\alpha > 1$ we have from (2.8) by direct calculation that

$$F_I(x) = 1 - \frac{1}{(1+x)^{\alpha-1}}, \tag{3.9}$$

meaning that the itd also has a Pareto distribution but with parameter $\alpha - 1$. Thus we can use the analytic inversion method.

For the simulation of the itd of Weibull distribution one can exploit the fact that its distribution coincides with a transformed gamma random variable as shown in [Kaasik, 2009b]. First recall the gamma function

$$\Gamma(t) = \int_0^\infty t^{k-1} \exp\{-y\} dt$$

and the cdf of a gamma distribution that has the form

$$G(x) = \frac{1}{s^k \Gamma(k)} \int_0^x y^{k-1} \exp\left\{-\frac{y}{s}\right\} dy,$$

for $x > 0$ with $k > 0$, $s > 0$ respectively the shape and the scale parameter.

Proposition 3.5. The integrated tail distribution of the Weibull distribution with parameter $0 < \beta < 1$ is that of $X^{1/\beta}$ if X has the gamma distribution with shape parameter $1/\beta$ and scale parameter 1.

Proof. Let $F_I(x)$ be the cdf of the itd and $G(x)$ the cdf of a gamma random variable with shape parameter $1/\beta$ and scale parameter 1. Then we have by calculation that

$$\begin{aligned} F_I(x) &= \frac{\int_0^x \exp\{-y^\beta\} dy}{\Gamma(1+\frac{1}{\beta})} = \frac{\int_0^{x^\beta} z^{\frac{1-\beta}{\beta}} \exp\{-z\} dz}{\beta \Gamma(1+\frac{1}{\beta})} \\ &= \frac{G(x^\beta)\Gamma(\frac{1}{\beta})}{\beta\Gamma(1+\frac{1}{\beta})} = G(x^\beta). \end{aligned} \tag{3.10}$$

□

Simulation from the gamma distribution is usually incorporated into statistical software, but the algorithm based on rejection sampling is also easy to implement [Ross, 2002: pp. 69–70]

For illustrative purposes, the simulation of the itd of log-normal distribution using the iterative generation algorithm explained in Proposition 3.4 is demonstrated below. First we generate u from the standard uniform distribution. We know that the mean value of the log-normal distribution introduced in the previous section

is $\exp\{\sigma^2/2\}$ (see e.g. [Balakrishnan and Nevzorov, 2003: p. 234]). Thus we can take $x_0 = \exp\{\sigma^2/2\}(1-u)$. Direct calculation yields that

$$F_I(x) = x\exp\{-\sigma^2/2\}\overline{\Phi}[(\log x)/\sigma] + \Phi[(\log x)/\sigma - \sigma], \tag{3.11}$$

as is confirmed by [Asmussen and Binswanger, 1997]. We can then write the iteration step as

$$\begin{aligned} x_{i+1} &= x_i + \frac{\exp\{\sigma^2/2\}\left[1 - x_i\exp\{-\sigma^2/2\}\overline{\Phi}[(\log x_i)/\sigma] - \Phi[(\log x_i)/\sigma - \sigma] - u\right]}{\overline{\Phi}[(\log x_i)/\sigma]} \\ &= \exp\{\sigma^2/2\}\frac{\overline{\Phi}[(\log x_i)/\sigma - \sigma] - u}{\overline{\Phi}((\log x_i)/\sigma)}. \end{aligned} \tag{3.12}$$

Example 3.1. The process of convergence when simulating the itd of the log-normal distribution is now demonstrated for different values of u from the standard uniform distribution. It is clear that when solving $\overline{F_I}(x) - u = 0$ the largest roots will occur when u is close to zero. For simplicity, let $\sigma = 1$. Then $x_0 = \exp\{1/2\}(1-u)$ and the initial value of the iteration process is also decreasing in u i.e. it gets larger when u is closer to zero. For all subexponential distributions, however, the complementary cdf tends to zero slower than any exponential, as noted e.g. in [Sigman, 1999], thus for the inverse of the complementary cdf the decrease is very rapid when the argument is close to zero.

Suppose first that $u = 0.5$. One can check that $\overline{F_I}(1.132244) = 0.5$. Sequential values of the iteration process (starting from x_0) are 0.82436, 1.096885, 1.131761, 1.132244, which means that the iteration process converges in three steps. Based on the previous argumentation, in this example at least half of the random numbers to be generated are obtained in just three steps or less.

Now let $u = 0.0001$. This means that a smaller value of u is to be expected once in every 10 000 simulations. We have $\overline{F_I}(74.027445) = 0.0001$. Now values 1.648556, 3.694176, 6.540136, 10.368623, 15.377210, 21.774585, 29.748997, 39.370493, 50.329222, 61.384745, 69.892441, 73.533508, 74.020051, 74.027444 and 74.027445 form the iteration process and thus the solution is reached in fourteen steps.

□

3.4. Simulating a random sum with heavy-tailed summands

Suppose we know the distribution (or can estimate the distribution) of the number of random summands in the case of estimating $\psi(u)$ with the help of (2.9). Then we can begin each simulation run by generating a number from that distribution. Then we have to simulate

$$\gamma(u) = \mathbb{P}\{X_1 + \ldots + X_k > u\}, \tag{3.13}$$

where the number of terms k is fixed.

This is the reason why in the last ten years a large variety of simulating algorithms to estimate the probability (3.13), where the summands are independent heavy-tailed random variables with distribution G, have been proposed. Because typically u is large, we have that $\gamma(u)$ is a very small probability and its accurate estimation requires a good simulation algorithm in the sense that the simulation error should be as small as possible. Otherwise a huge number of simulation runs would be required. This is why the naive estimator from (1.13) is not suitable.

3.4.1. Algorithm for simulating a sum with heavy-tailed summands

Next we present an easily implementable simulation algorithm that has a bounded relative error whenever the random variables X_i from (3.13) belong to a large subclass of $\mathcal{S}$.

As mentioned in the beginning of the Section 3.1, variance reduction ideas based on the mgf are not feasible in the heavy-tailed case. First asymptotically efficient simulation algorithm which made use of conditional Monte-Carlo was presented in [Asmussen and Binswanger, 1997]. Later arguments arising from either information theory as in [Asmussen et al., 2005] or arguments based on the twisting of the hazard rate (1.6) as in [Juneja and Shahabuddin, 2002] were used. The algorithm presented here was the first to have a bounded relative error for a sub-class of subexponential distributions. It was presented in [Asmussen and Kroese, 2006] and is known as the AK algorithm. The AK algorithm is also based on conditional Monte-Carlo. Even though an alternative algorithm that has bounded relative er-

ror for a larger sub-class of $\mathcal{S}$ has now been proposed in [Blanchet and Glynn, 2008], the AK algorithm still has the benefits of much easier implementation and computational simplicity which we demonstrate in the following.

We note that

$$\mathbb{P}(X_1 + \ldots + X_k > u) = k\mathbb{P}(X_1 + \ldots + X_k > u, X_k = \max\{X_1, \ldots, X_k\}). \quad (3.14)$$

Thus we could also generate k independent random variables $X_1, \ldots, X_k$ from distribution G and set

$$J_1 = \begin{cases} 1, & X_1 + \ldots + X_k > u \quad \text{and} \quad X_k = \max\{X_1, \ldots, X_k\}, \\ 0, & \text{otherwise.} \end{cases} \quad (3.15)$$

Now set $M_{k-1} = \max\{X_1, \ldots, X_{k-1}\}$ and $S_{k-1} = X_1 + \ldots + X_{k-1}$. Then we have that

$$\begin{aligned} \mathbb{E}(kJ_1 | X_1, \ldots, X_{k-1}) &= k\mathbb{P}(X_k > u - S_{k-1}, X_k > M_{k-1}) \\ &= k\overline{G}(\max\{M_{k-1}, u - S_{k-1}\}), \end{aligned}$$

but because

$$\mathbb{E}[\mathbb{E}(kJ_1 | X_1, \ldots, X_{k-1})] = \mathbb{E}(kJ_1) = \gamma(u)$$

we have that

$$Q_2(u) := k\overline{G}(\max\{M_{k-1}, u - S_{k-1}\}) \quad (3.16)$$

is (another) unbiased estimator of $\gamma(u)$.

Let us explain the idea why the estimator Q_2 has very little variability in the presence of subexponential tails and large u. Using the properties of subexponential distributions we can reason as follows. When we generate $X_1, \ldots, X_{k-1}$ there are basically two likely outcomes:

1. Usually all of them are "small" and $S_{k-1} << u$ which means that we have $\max\{M_{k-1}, u - S_{k-1}\} = u - S_{k-1}$ and $Q_2(u) \approx k\overline{G}(u)$.

2. Sometimes one of them is "big" and then typically $S_{k-1} > u$ which means that $\max\{M_{k-1}, u - S_{k-1}\} = M_{k-1}$ while because of (3.6) typically also $M_{k-1} > u$. Then, again $Q_2(u) \approx k\overline{G}(u)$, because $\mathcal{S} \subset \mathcal{L}$ and thus the

probabilities $\mathbb{P}\{X_k > u\}$ and $\mathbb{P}\{X_k > M_{k-1} | M_{k-1} > u\}$ are close for a large u.

When we consider a random sum there are two more things to take into account. To explain them in detail let us now consider our simulation problem (2.9), where Z_i are independent and have distribution F_I, while N has a geometric distribution. Let us develop (2.11) a bit further. We have

$$\begin{aligned} \psi(u) &= \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} \frac{\overline{F_I^{n*}}(u)}{(1+\rho)^n} = \frac{1}{1+\rho} \sum_{n=1}^{\infty} \frac{\rho \overline{F_I^{n*}}(u)}{(1+\rho)^n} \\ &= \frac{1}{1+\rho} \mathbb{P}\{Z_1 + \ldots + Z_{N^*} > u\}, \end{aligned} \tag{3.17}$$

where

$$\mathbb{P}\{N^* = n\} = \frac{\rho}{(1+\rho)^n}, \quad n = 1, 2, 3, \ldots.$$

Implementing the AK algorithm for simulating (2.9) using N^* as a control variate (see e.g. [Ross, 2002: pp. 139–140]) requires the knowledge of $\mathbb{E}N^*$ and (optionally) $Var(N^*)$. Distribution of N^* coincides with $Y + 1$ where Y is a geometric random variable with parameter $(1+\rho)^{-1}$ and we get from [Balakrishnan and Nevzorov, 2003: pp. 64–66] that $\mathbb{E}N^* = (1+\rho)\rho^{-1}$ and $Var(N^*) = (1+\rho)\rho^{-2}$.

Implementation of the AK algorithm can now be summarized as follows.

1. Fix the distribution F_I, the value of the safety loading coefficient ρ and u.

2. Simulate N^* so that $\mathbb{P}(N^* = n) = \rho(1+\rho)^{-n}$ for $n \in \{1, 2, \ldots\}$. If $N^* = 1$ put $Y = (1+\rho)^{-1}\overline{F_I}(u)$, else simulate $N^* - 1$ independent random variables from F_I, calculate their maximum M and sum S and put $Y = N^*(1+\rho)^{-1}\overline{F_I}(\max\{M, u - S\})$.

3. Repeat the previous step $t_1 - 1$ times and estimate the optimality constant c for the control variate defined as $c = -[\mathbb{E}(YN^*) - \mathbb{E}Y\mathbb{E}N^*]/Var(N^*)$ from the sample of size t_1 using the exact values for $\mathbb{E}N^*$ and $Var(N^*)$ and denote it as $\hat{c}$.

4. Simulate N^* as in step 2, but if $N^* = 1$ set $Y = (1+\rho)^{-1}\overline{F_I}(u) - \hat{c}\rho^{-1}$, else simulate $N^* - 1$ independent random variables from F_I, calculate their maximum M and sum S and put $Y = N^*(1+\rho)^{-1}\overline{F_I}(\max\{M, u - S\}) + \hat{c}(N^* - (1+\rho)\rho^{-1})$.

5. Repeat the previous step $t_2 - 1$ times and estimate $\psi(u)$ as an average of Y's over the sample of size t_2.

As noted in [Asmussen and Kroese, 2006], the AK algorithm is specifically constructed with heavy-tails in mind and its authors demonstrate its good performance for the situations when F_I is either Pareto or Weibull distribution. It would make more sense to consider those distributions in the role of F and then their integrated tail distributions in the role of F_I. This is so because as noted in [Asmussen and Kroese, 2006], when F_I is assumed to be a Weibull distribution, there is no corresponding original distribution F in a sense that equation (2.8) would hold. In [Kaasik, 2009b] Pareto, Weibull and log-normal distribution are used in the role of the original distribution F in a simulation study. The study considers the relative error for different combinations of ρ and u and the conclusions are that the algorithm performs very well for the itd of Pareto and log-normal, but no so good for the itd of Weibull when the parameter of the original distribution is close to 1 (which is the borderline with light tails).

Chapter 4

Ruin probability estimation when the claim size distribution is unknown

Suppose we have observed a risk process governed by the CLM for some time. According to (2.9), we would be able to estimate the ruin probability in the CLM if we would know λ – the intensity of the claims and F_I – the itd of the claim size distribution. Because estimation of λ is simply a parameter estimation, e.g. maximum likelihood can be used to obtain the reciprocal of the sample mean as the estimate. At the same time, estimating F_I is not trivial and will be the main challenge tackled in this chapter. As noted in Remark 2.4 we can also use (2.12) for ruin probability estimation if F_I is subexponential. The latter needs to be tested in practice.

4.1. Testing for a heavy-tail of the itd

Suppose we have iid data $X_1, \ldots, X_n$ from the original distribution F and we need to check whether use of the asymptotic approximation (2.12) for estimating the ruin probability is justified. That is, we ask the question: does $F_I \in \mathcal{S}$ hold? There are, however, no known statistical tests to check the subexponentiality of the integrated tail distribution F_I from the sample that is from the distribution F. We could try to check some condition for F, presented in Section 3.2, to have $F_I \in \mathcal{S}$. We could also take heart from the fact that most known heavy-tailed distributions are included in $\mathcal{S}$. Because of Proposition 3.1 we have a justified expectation that a heavy-tailed F would mean a heavy-tailed F_I which in turn is typically a member of $\mathcal{S}$. Testing whether iid data are indeed from a heavy-tailed distribution is typically judgment based.

4.1.1. QQ-plot

A common technique is plotting the sample order statistics against quantiles of the exponential distribution (as exponential distribution is seen as the heaviest of the light-tailed distributions due to its constant hazard rate and mean excess of loss function). That is, increasingly ordered data points are denoted as $X_{(1)}, \ldots, X_{(n)}$ and points

$$\left\{ X_{(i)}, H^{-1}\left(\frac{i}{n+1}\right) \right\}, \quad i = 1, \ldots, n \tag{4.1}$$

are plotted, where H^{-1} is the inverse of the exponential cdf. This way the sample maximum $X_{(n)}$ forms a point with the $n/(n+1)$-th exponential quantile. Heavy-tails are suspected when the points show concave departure from the straight line that joins the first and last point plotted (see e.g. [Embrechts et al., 1997: pp. 290–294] for details). Of course, the inverse cdf need not be that of the exponential distribution – when we know the parametric class of distributions from which the iid data have come from, we can use that specific cdf in the role of H in (4.1). This way it is possible to estimate location and scale parameters from the QQ-plot and for example when regularly varying tails are considered (i.e. data are from a distribution in $\mathcal{R}$) one can estimate the tail index α based on the QQ-plot as demonstrated in [Kratz and Resnick, 1996].

4.1.2. The ratio of maximum and sum

The following reasoning is taken from [Embrechts et al., 1997: pp. 309–310].

It is known that the ratio of the maximum and the sum of the iid sample converges to zero almost surely if and only if the mean of the distribution is finite. We consider only positive random variables and denote

$$S_n(p) = X_1^p + \ldots + X_n^p, \quad M_n(p) = \max\{X_1^p, \ldots, X_n^p\}, \tag{4.2}$$

where $X_1, \ldots, X_n$ is the theoretical sample and $p > 0$. Then the ratio of the maximum and sum $R_n(p) = M_n(p)/S_n(p)$ converges to zero almost surely if and only if the p-th moment of the distribution is finite. This result enables us to plot $R_n(p)$ for different positive values of p and check whether the ratio tends to zero or not as the sample size increases. If $R_n(p)$ deviates from zero significantly for large values of n then this can be seen as a proof that the considered moment p does not exist and thus the distribution has a heavy-tail.

4.1.3. Mean excess of loss plot

As discussed in Section 1.2, heavy-tailed distributions have an (eventually) increasing mean excess of loss function and (eventually) decreasing hazard rate. As mean excess of loss function is simply based on conditional expectation, its empirical counterpart is readily available – we can plot the points

$$\left\{ X_{(i)}, \frac{1}{n-i} \left(X_{(i+1)} + \ldots + X_{(n)} - (n-i)X_{(i)} \right) \right\}, \quad i = 1, \ldots, n \tag{4.3}$$

and perhaps add confidence intervals assuming the (approximate) normality of the conditional means.

The key why mean excess of loss plot is in some sense more informative than the previously considered techniques is the fact that the hazard rate of the itd for each argument x is the reciprocal of the mean excess of loss function of the original distribution (also evaluated at x) as shown e.g. in [Klugman et al., 2004: p. 53]. This means that by checking a property of the original distribution we can be sure that it carries over for the itd of that distribution i.e. if the mean excess of loss function of the original distribution is eventually increasing then the hazard rate of the itd is, in fact, eventually decreasing.

As is usually the case with functions that depend on sample mean, mean excess of loss function is not robust in a sense that it is very sensitive to large values of data. Heavy-tail of the underlying distribution assures that these large values are most likely not outliers but the interpretation of the mean excess of loss plot is more complicated nonetheless. This is why more robust alternatives like the median excess plot described e.g. in [Rootzén and Tajvidi, 1997] can be used.

4.2. Empirical approximation of the integrated tail distribution

Suppose $X_1, \ldots, X_n$ is an iid sample from the original distribution F. An empirical cdf $F_n(x)$ can be constructed as in (1.18) and the Glivenko-Cantelli theorem states that

$$\mathbb{P}\left(\sup_x |F_n(x) - F(x)| \to 0 \right) = 1. \tag{4.4}$$

Thus for a large sample $F_n(x)$ should be good approximation for $F(x)$. However, our aim is to estimate itd defined by (2.8) which we can write as

$$F_I(x) = \frac{\int_0^x \overline{F}(y)dy}{\mathbb{E}X} = \frac{\int_0^x \overline{F}(y)dy}{\int_0^\infty \overline{F}(y)dy}, \quad x > 0. \tag{4.5}$$

In view of (4.4), a natural estimator for $F_I(x)$ would be

$$F_I^{emp_n}(x) = \frac{\int_0^x \overline{F_n}(y)dy}{\int_0^\infty \overline{F_n}(y)dy} = \frac{\int_0^x \overline{F_n}(y)dy}{\mu_n}, \quad x > 0, \tag{4.6}$$

where $\mu_n = (X_1 + \ldots, X_n)/n$ is the theoretical sample mean. This approximation, dubbed the empirical approximation, was analyzed in [Kaasik and Pärna, 2008] and the following result, complementing (4.4), was proved.

Proposition 4.1. Let X_n be a sequence of independent identically distributed positive random variables with a finite mean μ and cumulative distribution function F with F_n its empirical counterpart. Then the following result holds

$$\mathbb{P}\left(\sup_x \left| \frac{\int_0^x \overline{F}_n(y)dy}{\mu_n} - \frac{\int_0^x \overline{F}(y)dy}{\mu} \right| \xrightarrow{n} 0\right) = 1. \tag{4.7}$$

Proof. SLLN and Glivenko-Cantelli theorem hold simultaneously on a set which has probability one. Fix an ω from that set and also fix an $\epsilon > 0$. Because μ is finite, there exists a $K > 0$ such that

$$\int_K^\infty \overline{F}(y)dy < \frac{\epsilon\mu}{6}. \tag{4.8}$$

Due to SLLN we have that there exist n_1, n_2 and n_3 such that when $n > n_1$

$$|\mu_n - \mu| < \min\{\frac{\epsilon\mu}{6}, \mu\frac{-1 + \sqrt{1 + 2\epsilon/3}}{2}\}, \tag{4.9}$$

when $n > n_2$

$$|\overline{F}_n(K) - \overline{F}(K)| < \frac{\epsilon\mu}{12K}, \tag{4.10}$$

and when $n > n_3$

$$\left| \int_K^\infty y d\overline{F}_n(y) - \int_K^\infty y d\overline{F}(y) \right| < \frac{\epsilon\mu}{12}. \tag{4.11}$$

Also, due to Glivenko-Cantelli theorem, there exists n_4 such that when $n > n_4$

$$\sup_y |\overline{F}_n(y) - \overline{F}(y)| < \frac{\epsilon\mu}{6K} \tag{4.12}$$

holds. Thus when $n > \max\{n_1, n_2, n_3, n_4\}$ we have that

$$\begin{aligned}
&\frac{1}{\mu}\sup_x \left|\frac{\mu}{\mu_n}\int_0^x \overline{F}_n(y)dy - \int_0^x \overline{F}(y)dy\right| \\
&< \frac{1}{\mu}\sup_x \left|\frac{\mu}{\mu_n}\int_0^x \overline{F}_n(y)dy - \int_0^x \overline{F}_n(y)dy\right| + \frac{1}{\mu}\sup_x \left|\int_0^x \overline{F}_n(y)dy - \int_0^x \overline{F}(y)dy\right| \\
&\overset{(4.9)}{<} \frac{\epsilon}{6} + \frac{1}{\mu}\sup_x \left|\int_0^x \overline{F}_n(y)dy - \int_0^x \overline{F}(y)dy\right| \\
&\leqslant \frac{\epsilon}{6} + \frac{1}{\mu}\sup_{x\leqslant K}\int_0^x \left|\overline{F}_n(y)dy - \overline{F}(y)\right| dy + \frac{1}{\mu}\sup_{x>K}\left|\int_0^x \overline{F}_n(y)dy - \int_0^x \overline{F}(y)dy\right| \\
&\overset{(4.12)}{<} \frac{\epsilon}{3} + \frac{1}{\mu}\sup_{x>K}\left|\int_0^x \overline{F}_n(y)dy - \int_0^x \overline{F}(y)dy\right| \\
&\leqslant \frac{\epsilon}{3} + \frac{1}{\mu}\sup_{x>K}|\mu_n - \mu| + \frac{1}{\mu}\int_K^\infty \overline{F}_n(y)dy + \frac{1}{\mu}\int_K^\infty \overline{F}(y)dy \\
&\overset{(4.8),(4.9)}{<} \frac{2\epsilon}{3} + \frac{1}{\mu}\int_K^\infty \overline{F}_n(y)dy \\
&= \frac{2\epsilon}{3} - \frac{K}{\mu}\overline{F}_n(K) + \frac{1}{\mu}\int_K^\infty y\, d\overline{F}_n(y) \\
&\overset{(4.10),(4.11)}{<} \frac{5\epsilon}{6} + \frac{1}{\mu}\int_K^\infty \overline{F}(y)dy \\
&\overset{(4.8)}{<} \epsilon,
\end{aligned}$$

as required. □

When we are willing to assume that the distribution of interest has finite variance then a stronger result than Proposition 4.1 can be proved. We first denote the integrated tail process as ψ_n and then make some manipulations

$$\begin{aligned}
\psi_n(x) &:= \sqrt{n}\left(\frac{\int_x^\infty \overline{F}_n(y)dy}{\mu_n} - \frac{\int_x^\infty \overline{F}(y)dy}{\mu}\right) = \frac{\sqrt{n}}{\mu_n}\int_x^\infty \left(\overline{F}_n(y) - \frac{\mu_n}{\mu}\overline{F}(y)\right)dy \\
&= \frac{\sqrt{n}}{\mu_n}\int_x^\infty \left(\overline{F}_n(y) - \overline{F}(y)\right)dy + \frac{\sqrt{n}}{\mu_n}\int_x^\infty \left(\overline{F}(y) - \frac{\mu_n}{\mu}\overline{F}(y)\right)dy \\
&= \frac{1}{\mu_n}\int_x^\infty \sqrt{n}\left(\overline{F}_n(y) - \overline{F}(y)\right)dy + \frac{1}{\mu_n}\overline{F_I}(x)(\mu - \mu_n).
\end{aligned} \tag{4.13}$$

Now we can prove the following statement.

Proposition 4.2. With the setup from Theorem 1.2 and $\mathbb{E}(X^2) < \infty$ we have that

$$\sup_x |\psi_n(x) - \Psi_n(x)| \xrightarrow{P} 0, \tag{4.14}$$

where

$$\Psi_n(x) = \frac{1}{\mu}\int_x^\infty B_n(F(y))dy + \frac{1}{\mu}\overline{F_I}(x)\int_0^\infty B_n(F(y))dy. \tag{4.15}$$

Proof. The key for the proof is Theorem 1.2. Building on that while assuming the finite second moment, [Csörgö et al., 1986: pp. 35–37] proved that

$$\sup_x \left| \sqrt{n}\int_0^x \left(\overline{F}_n(y) - \overline{F}(y)\right) dy - \int_0^x B_n(F(y))dy \right| \xrightarrow{P} 0. \tag{4.16}$$

Now using the representation (4.13) and the definition of $\Psi_n(x)$ we have that

$$\begin{aligned}\sup_x|\psi_n(x) - \Psi_n(x)| \leqslant \sup_x &\left| \frac{1}{\mu_n}\int_x^\infty \sqrt{n}\left(\overline{F}_n(y) - \overline{F}(y)\right) dy - \frac{1}{\mu}\int_x^\infty B_n(F(y))dy \right| \\ &+ \sup_x \left| \frac{1}{\mu_n}\overline{F_I}(x)\sqrt{n}(\mu - \mu_n) - \frac{1}{\mu}\overline{F_I}(x)\int_0^\infty B_n(F(y))dy \right|.\end{aligned} \tag{4.17}$$

Hence it suffices to show that

$$\sup_x \left| \frac{1}{\mu_n}\int_x^\infty \sqrt{n}\left(\overline{F}_n(y) - \overline{F}(y)\right) dy - \frac{1}{\mu}\int_x^\infty B_n(F(y))dy \right| \xrightarrow{P} 0 \tag{4.18}$$

and

$$\sup_x \left| \frac{1}{\mu_n}\overline{F_I}(x)\sqrt{n}(\mu - \mu_n) - \frac{1}{\mu}\overline{F_I}(x)\int_0^\infty B_n(F(y))dy \right| \xrightarrow{P} 0. \tag{4.19}$$

We begin with the former. Adding and subtracting $\mu_n^{-1}\int_x^\infty B_n(F(y))dy$, we obtain that

$$\begin{aligned}&\sup_x \left| \frac{1}{\mu_n}\int_x^\infty \sqrt{n}\left(\overline{F}_n(y) - \overline{F}(y)\right) dy - \frac{1}{\mu}\int_x^\infty B_n(F(y))dy \right| \\ &\leqslant \frac{1}{\mu_n}\sup_x \left| \int_x^\infty \sqrt{n}\left(\overline{F}_n(y) - \overline{F}(y)\right) dy - \int_x^\infty B_n(F(y))dy \right| \\ &+ \left| \frac{1}{\mu_n} - \frac{1}{\mu} \right| \sup_x \left| \int_x^\infty B_n(F(y))dy \right|.\end{aligned} \tag{4.20}$$

As $(\mu_n)^{-1}$ tends to $\mu^{-1} < \infty$ almost surely due to the SLLN, we can make use of

(4.16) twice to see that the first summand converges to zero in probability. For the second summand we note that the random variable

$$\sup_x \left| \int_x^\infty B_n(F(y))dy \right| \tag{4.21}$$

has the same distribution for all n because the integrands are all Brownian bridges. Also, as shown in [Csörgö et al., 1986: p. 35], finite second moment implies that (4.21) is finite almost surely. Thus the second summand from (4.20) converges to zero in probability by Slutsky theorem (and the fact that for a constant limit convergence in probability is equivalent with convergence in distribution) as $\mu_n^{-1} - \mu^{-1}$ tends to zero almost surely (and hence also in distribution). Thus we do indeed have (4.18).

Proving (4.19) is similar. We first note that

$$\begin{aligned} &\sup_x \left| \frac{1}{\mu_n} \overline{F_I}(x) \sqrt{n}(\mu - \mu_n) - \frac{1}{\mu} \overline{F_I}(x) \int_0^\infty B_n(F(y))dy \right| \\ &\leqslant \sup_x \left| \overline{F_I}(x) \right| \left| \frac{1}{\mu_n} \sqrt{n}(\mu - \mu_n) - \frac{1}{\mu} \int_0^\infty B_n(F(y))dy \right| \\ &= \left| \frac{1}{\mu_n} \sqrt{n}(\mu - \mu_n) - \frac{1}{\mu} \int_0^\infty B_n(F(y))dy \right|, \end{aligned} \tag{4.22}$$

but the latter is a special case of (4.18) as the empirical and actual mean can be represented as an improper integral of $\overline{F_n}$ and $\overline{F}$, respectively, and thus (4.19) also holds. □

One of the the practical meanings of Proposition 4.2 is the fact that for large n it allows us to estimate

$$\mathbb{P}\{|F_I^{emp_n}(x) - F_I(x)| > \epsilon\} \tag{4.23}$$

as

$$\mathbb{P}\left\{ n^{-1/2} \left| \frac{1}{\mu} \int_x^\infty B_n(F(y))dy + \frac{1}{\mu} \overline{F_I}(x) \int_0^\infty B_n(F(y))dy \right| > \epsilon \right\}. \tag{4.24}$$

It is known that

$$\mathbb{P}(\sup_x |F(x) - F_n(x)| > \epsilon) \leqslant 2e^{-2n\epsilon^2}, \tag{4.25}$$

as proved in [Dvoretzky et al., 1956] and [Massart, 1990]. This kind of result enables us to construct a "confidence tube" for the whole distribution function $F_n(x)$ such that with high probability the actual cdf $F(x)$ lies inside this tube for all values of x. We can deduce from (4.25) that when the sample size n increases, the radius of the tube (so that it would still include the original cdf with that high (fixed) probability) decreases as $1/\sqrt{n}$. Simulations in [Kaasik and Pärna, 2008] show that the rate of convergence in (4.7) is also of the order $1/\sqrt{n}$ for subexponential distributions as is also suggested by Proposition 4.2. However this rate is not achieved instantly – for very heavy tails the rate is extremely slow at first. This means that a huge sample size may be needed to achieve a decent approximation for the cdf $F(x)$ when using the empirical approximation.

Serious issue with the empirical approximation (4.6) arises in practice, however, when we would like to use it for ruin probability estimation. The estimator Q_2 defined in (3.16) typically uses the complementary cdf of the itd for large argument values that are comparable with the size of initial capital u as explained in Section 3.4. For such arguments the value of the complementary cdf of the itd might perhaps be in the region of $10^{-4} \dots 10^{-6}$ and it would require an outrageously large sample size to provide a confidence tube that would be meaningful for this part of the complementary cdf. For example, in [Kaasik and Pärna, 2008] the 95% confidence tubes produced (using repetitive simulation) had the radius in the region of $10^{-2} \dots 10^{-3}$ when the sample size was 100 000. Thus, at least for the large argument values, a better approximation technique for the cdf of the itd is required.

4.3. GPD approximation of the integrated tail distribution

The idea of using the GPD as an approximation of the conditional distribution of the claim sizes and this way proceeding to the approximation of the distribution function of the itd was introduced in [Kaasik, 2009a]. The theoretical foundations are as follows. Suppose a positive random variable X with distribution F that has finite mean and support $(0, \infty)$ satisfies the EV condition (1.10). Then, according to Theorem 1.1, there exist $\xi \in [0, 1)$ and positive function $\sigma(u)$ such that

$$\lim_{u \to \infty} \sup_{x>0} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0, \tag{4.26}$$

holds with $F_u(x)$ defined as in (1.7). Now if $X_1, \ldots, X_n$ is an independent sample from F we can set the threshold u at a high level, and form a new sample for the estimation of the parameters of the GPD. Namely suppose that $\#\{X_i : X_i > u\} = N$ and let $\Theta^u = \{i : X_i > u\}$ be the set of indexes i for which $X_i > u$. Now set $Y_j = X_{j^*} - u$ for all $j \in \Theta^u$, where X_{j^*} is the member of the original sample whose index is j-th in the set Θ^u, thus forming a new sample of size N which we call the sample of exceedances. If $\hat{\xi}$ and $\hat{\sigma}$ are estimates of the GPD parameters based on the sample $Y_1, \ldots, Y_N$ then we can assume that for $y > 0$ we have

$$\overline{F}(u+y) = \overline{F}(u)\overline{F_u}(y) \approx \frac{N}{n}\left(1 + \frac{\hat{\xi}y}{\hat{\sigma}}\right)^{-1/\hat{\xi}}, \tag{4.27}$$

as N/n should be close to $\overline{F}(u)$. When (4.27) holds as an equality, the following statement from [Kaasik and Pärna, 2009] holds, essentially saying that the tail of the itd has GPD form if this is true for the original distribution.

Proposition 4.3. Let X be a non-negative random variable with distribution F that has support $(0, \infty)$. If there exists $u > 0$ such that the conditional distribution of $X - u$ given that $X > u$ has a general Pareto distribution (that is $F_u(x) = G_{\xi,\sigma}(x)$ holds for all x with $\xi \in [0, 1)$ and $\sigma > 0$), then for every $y > 0$ it holds that

$$\overline{F_I}(u+y) = \frac{\overline{F}(u)\sigma^*\overline{G_{\xi^*,\sigma^*}}(y)}{\mathbb{E}(X|X \leqslant u)F(u) + (u+\sigma^*)\overline{F}(u)}, \tag{4.28}$$

where $\sigma^* = \sigma/(1-\xi)$ and $\xi^* = \xi/(1-\xi)$.

Proof. Let us start analyzing the numerator of (4.28). We can write

$$\overline{F_I}(u+y) = \frac{\int_{u+y}^{\infty}\overline{F}(z)dz}{\mathbb{E}X} = \frac{\int_y^{\infty}\overline{F}(u)\overline{F_u}(z)dz}{\mathbb{E}X} = \frac{\overline{F}(u)\int_y^{\infty}\overline{G_{\xi,\sigma}}(z)dz}{\mathbb{E}X} \tag{4.29}$$

as $F_u(x) = G_{\xi,\sigma}(x)$ for all x. Now suppose first that $\xi \in (0, 1)$ so that $1/\xi \in (1, \infty)$

and we have for all $y > 0$ that

$$\begin{aligned}\int_y^\infty \overline{G_{\xi,\sigma}}(z)dz &= \int_y^\infty \left(1+\frac{\xi z}{\sigma}\right)^{-1/\xi} dz = \frac{\sigma}{\xi}\int_y^\infty \left(1+\frac{\xi z}{\sigma}\right)^{-1/\xi} d\left(1+\frac{\xi z}{\sigma}\right) \\ &= \frac{\sigma}{\xi}\left(-\frac{1}{\xi}+1\right)\left(1+\frac{\xi z}{\sigma}\right)^{-1/\xi+1}\Bigg|_y^\infty = \frac{\sigma}{\xi}\left(\frac{1}{\xi}-1\right)\left(1+\frac{\xi y}{\sigma}\right)^{-1/\xi+1} \\ &= \frac{\sigma}{1-\xi}\left(1+\frac{\xi y}{\sigma}\right)^{-(1-\xi)/\xi} = \sigma^*\left(1+\frac{\xi^* y}{\sigma^*}\right)^{-1/\xi^*} = \sigma^*\overline{G_{\xi^*,\sigma^*}}(y)\end{aligned} \tag{4.30}$$

by making use of the fact that the ratio of ξ and σ is equal to the ratio of ξ^* and σ^* by definition. If instead $\xi = 0$, then

$$\begin{aligned}\int_y^\infty \overline{G_{\xi,\sigma}}(z)dz &= \int_y^\infty \exp\left(-\frac{z}{\sigma}\right) dz = -\sigma\int_y^\infty \exp\left(-\frac{z}{\sigma}\right) d\left(-\frac{z}{\sigma}\right) \\ &= -\sigma\exp\left(-\frac{z}{\sigma}\right)\Bigg|_y^\infty = \sigma\exp\left(-\frac{y}{\sigma}\right) = \sigma^*\exp\left(-\frac{y}{\sigma^*}\right),\end{aligned} \tag{4.31}$$

because $1-\xi = 0$ and thus $\sigma = \sigma^*$ and $\xi = \xi^*$. Now we switch our attention to the denominator of (4.28). First we write the expectation as

$$\mathbb{E}X = \mathbb{E}(X|X \leqslant u)F(u) + \mathbb{E}(X|X > u)\overline{F}(u). \tag{4.32}$$

Then we have $\mathbb{E}(X|X > u) = u + \mathbb{E}(X - u|X > u)$. Noting that the conditional random variable $X - u|X > u$ has a GPD with mean $\sigma/(1-\xi)$ as pointed out in Section 1.3 completes the proof. □

Remark 4.1. The expression in the denominator of (4.28) is in fact basis for an estimator for the mean that was introduced in [Johansson, 2003]. Initially it might seem, as if we could use another type of estimator for $\mathbb{E}X$, e.g. sample mean. However, if we would like to use the GPD as an approximation for the right tail of F in the numerator of (4.29) then it must be also used in the denominator because of (4.5), as otherwise we would not have a guarantee that the approximation would behave like a distribution function in terms of its value range.

The following result describes a natural estimator for the denominator of (4.28).

Proposition 4.4. Denote $M = \mathbb{E}(X|X \leqslant u)F(u) + (u+\sigma^*)\overline{F}(u)$ with threshold $u > 0$ fixed and suppose we have an independent sample $X_1, \ldots, X_n$ from the

distribution F. If $\tilde{\sigma^*}$ is an unbiased estimator of σ^* that is uncorrelated with N, the size of the sample of exceedances $Y_1, \ldots, Y_N$, then

$$\hat{M} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{N} Y_i + \tilde{\sigma^*} N}{n} \tag{4.33}$$

is an unbiased estimator of M.

Proof. Because of the definition of sample of Y-s we can rewrite the proposed estimator as

$$\begin{aligned} \hat{M} &= \frac{\sum_{i:X_i \leqslant u} X_i + \sum_{i:X_i > u} X_i - \sum_{i:X_i > u} X_i + Nu + N\tilde{\sigma^*}}{n} \\ &= \frac{\sum_{i:X_i \leqslant u} X_i}{n} + (u + \tilde{\sigma^*})\frac{N}{n}. \end{aligned} \tag{4.34}$$

It holds that

$$\mathbb{E}X = \mathbb{E}(XI_{\{X \leqslant u\}}) + \mathbb{E}(XI_{\{X>u\}}) = \mathbb{E}(X|X \leqslant u)F(u) + \mathbb{E}(X|X > u)\overline{F}(u) \tag{4.35}$$

and thus $\sum_i X_i I_{\{X_i \leqslant u\}}/n = \sum_{i:X_i \leqslant u} X_i/n$ and $\sum_i X_i I_{\{X_i > u\}}/n = \sum_{i:X_i > u} X_i/n$ are respective unbiased estimators. Also, $\mathbb{E}I_{\{X>u\}} = \overline{F}(u)$ has an unbiased estimator in $\sum_i I_{\{X_i>u\}}/n = N/n$ which is uncorrelated with $\tilde{\sigma^*}$ by the assumption. Hence

$$\hat{M} = \mathbb{E}\left[\frac{\sum_{i:X_i \leqslant u} X_i}{n} + (u + \tilde{\sigma^*})\frac{N}{n}\right] = \mathbb{E}(X|X \leqslant u)F(u) + (u + \sigma^*)\overline{F}(u) = M, \tag{4.36}$$

as stated. □

As typically (4.27) does not hold as an equality for any finite u, and thus it would be important to know whether approximating with the GPD should provide good results. If we suppose first that $\xi \in (0, 1)$ then we can recall from Section 1.3 that necessarily $F \in \mathcal{R}_{-\alpha}$, with $1/\xi = \alpha \in (1, \infty)$. But then according to Karamata's theorem (see e.g. [Embrechts et al., 1997: p. 567]) the integral of the complementary cdf is again regularly varying and $F_I \in \mathcal{R}_{-\alpha+1}$. So the EV condition (1.10) is once again satisfied (this time for F_I) and Theorem 1.1 is applicable.

Previous paragraph also gives us another possible idea how to check the heavy-tailedness (actually, even subexponentiality) of the itd in addition to the approaches discussed in Section 4.1.

Corollary 4.1. Suppose a distribution F satisfies *EV* condition (1.10) with $0 < \xi < 1$. Then $F_I \in \mathcal{S}$

Proof. According to the previous reasoning EV condition with $0 < \xi < 1$ would mean that $F_I \in \mathcal{R}_{-\alpha_1}$ with $\alpha_1 = (1-\xi)/\xi$. But regularly varying distributions are subexponential as noted in Section 1.2. □

Thus we can estimate the GPD parameters based on the sample of exceedances and check whether ξ is positive (e.g. on the basis of a confidence interval) to check whether $F_I \in \mathcal{S}$. Several statistical tests for testing $\xi > 0$ are also available and described in [Neves and Alves, 2008]. However, in such tests the null hypothesis is typically $\xi = 0$ which does not rule out a heavy tail.

Now suppose $\xi = 0$. We can argue that in (4.29) where we need the assumption of the exact GPD tail (i.e. $F_u(x) = G_{\xi,\sigma}(x)$ for all $x > 0$) the question is essentially about the difference between the two integrals $\int_y^\infty \overline{F_u}(z)dz$ and $\int_y^\infty \overline{G_{\xi,\sigma}}(z)dz$. The following result from [Kaasik, 2009a] holds.

Proposition 4.5. Suppose that for a positive random variable with distribution F the EV condition (1.10) holds with $\xi = 0$. Let $a(x)$ be the mean excess of loss function, defined in (1.5), of that random variable. If there exist an u_0 and K such that for $u > u_0$ we have

$$\overline{F_u}(ya(u)) < \overline{F_{u_0}}(ya(u_0)) \tag{4.37}$$

for all $y > K$, then it also holds that $\int_x^\infty \overline{F_u}(ya(u))dy$ converges uniformly to $\int_x^\infty \overline{G_{\xi,1}}(y)dy$ and

$$\sup_{x>0}\left|\int_x^\infty \overline{F_u}(y)dy - \int_x^\infty \overline{G_{\xi,a(u)}}(y)dy\right| = o(a(u)), \tag{4.38}$$

where $a(x)$ is the mean excess of loss function of F.

Proof. By the assumption

$$\lim_{u\to\infty}\sup_{x>0}|\overline{F_u}(xa(u)) - \overline{G_{\xi,1}}(x)| = 0 \tag{4.39}$$

holds with $\xi = 0$. Denote $f_u(x) := \int_x^\infty \overline{F_u}(ya(u))dy$ and $g(x) := \int_x^\infty \overline{G_{\xi,1}}(y)dy$, where $x > 0$. Function $g(x)$ is well-defined (it is a cdf of an exponential distribution

multiplied by a constant) as is $f_u(x)$ for every positive u, because

$$\begin{aligned}\int_x^\infty \overline{F_u}(ya(u))dy &= \int_x^\infty \frac{\overline{F}(u+ya(u))dy}{\overline{F}(u)} = \frac{\int_{u+xa(u)}^\infty \overline{F}(z)dz}{a(u)\overline{F}(u)} \\ &= \frac{\int_{u+xa(u)}^\infty \overline{F}(z)dz}{\int_u^\infty \overline{F}(z)dz} \leqslant 1\end{aligned} \tag{4.40}$$

by using a change of variables and the mentioned representation of the mean excess of loss function. The existence of such u_0 and K that (4.37) holds for all $y > K$ guarantees that we can construct an integrable dominating function and thus allows the application of the dominated convergence theorem. Thus we have

$$\lim_{u\to\infty} f_u(0) = g(0). \tag{4.41}$$

thanks to (4.39). Now Scheffe lemma can be used to deduce that

$$\lim_{u\to\infty} \int_0^\infty \left|\overline{F_u}(ya(u)) - \overline{G}_{\xi,1}(y)\right| dy = 0. \tag{4.42}$$

and because

$$\sup_{x>0}\left|\int_x^\infty \overline{F_u}(ya(u))dy - \int_x^\infty \overline{G}_{\xi,1}(y)dy\right| \leqslant \int_0^\infty \left|\overline{F_u}(ya(u)) - \overline{G}_{\xi,1}(y)\right| dy, \tag{4.43}$$

we have that $f_u(x)$ converges to $g(x)$ uniformly on $(0,\infty)$. A change of variable in (4.42) proves (4.38). □

Remark 4.2. The critical assumption in Proposition 4.5 is the eventual boundedness (4.37). This condition can be translated to the terms of the hazard rate $h(x)$, defined in (1.6). Namely the existence of K and u_0, such that when $u > u_0$ we have

$$\int_u^{u+ya(u)} h(t)dt \geqslant \int_{u_0}^{u_0+ya(u_0)} h(t)dt \tag{4.44}$$

whenever $y > K$. This is based on the representation of the conditional cdf as a function of the hazard rate found e.g. in [Klugman et al., 2004: p. 52]. With $a(x)$ eventually increasing and $h(x)$ eventually decreasing it is somewhat more intuitive to assume that we can choose a large enough K so that the domain of integration in the left-hand side of (4.44) grows fast enough to offset the decrease of the integrand as the argument increases.

Corollary 4.2. The statements of Proposition 4.5 would continue to hold if the assumption of the existence of constants u_0 and K would be replaced by an assumption of the existence of constants u_0^* and K^* such that for every u_1 and u_2 for which $u_1 > u_2 > u_0^*$ holds we have

$$\overline{F_{u_1}}(ya(u_1)) \geqslant \overline{F_{u_2}}(ya(u_2)) \tag{4.45}$$

for $y > K^*$.

Proof. This new assumption means that the convergence of $\overline{F_u}(a(u)y)$ to a complementary cdf of a GPD is monotone (in the tail part) and monotone convergence theorem is applicable, thus (4.41) still holds and the proof of Proposition 4.5 can be repeated. □

4.4. Selections for the GPD approximation

Before comparing the performance of the approximation (4.6) and the one based on (4.28) it is worth noting that while the former is a non-parametric approximation, the latter is a semi-parametric one and decisions about method of inference have to be made. Even more, the GPD approximation will only be used for a tail region starting from some threshold value u. If u is too low then (4.27) is not reliable while higher threshold clearly reduces the size of the sample of exceedances. This aspect is known as the bias versus variance trade-off from [Smith, 1987].

4.4.1. Threshold selection

A variety of judgment based threshold selection techniques are available. Most common of them are presented e.g. in [Coles, 2001: pp. 78–84]. For example, the fact that a GPD has a linear mean excess of loss function [Embrechts et al., 1997: p. 165] is the basis of the technique which selects the appropriate threshold value u as the smallest point after which the empirical mean excess of loss plot is approximately linear. The decision of whether an empirical residual life plot is approximately linear from some point onwards (taking into account the confidence intervals) cannot be easily automated and the concerns about lack of robustness, mentioned in Section 4.1, are also present. An automatic threshold selection algorithm is presented in [Dupuis, 1999] but the author's practical experience has shown that the algorithm lacks in stability.

One straightforward approach for threshold selection can be described as follows. When the conditional tail of F has a GPD for some large u then for any larger threshold the distribution is once again a GPD with the same shape parameter ξ (but a different scale parameter σ). Thus we could decrease the sample of exceedances and every time estimate the GPD parameters. Threshold u can then be selected thresholds after which the estimate of the parameter ξ remains constant. This approach is used e.g. in [Kaasik and Pärna, 2009]. Parameter estimates are complemented with the respective confidence intervals. One of the improvements of this approach is the ease of automation which is important when the performance of the GPD approximation is tested using simulation.

4.4.2. Parameter estimation methodology

The classical tool for inference – maximum likelihood (ML) method – can be used for the estimation of the GPD parameters. That is, the log-likelihood

$$\begin{aligned}\ell(\xi,\sigma) &= \log\prod_{i=1}^{N}\left[\frac{1}{\sigma}\left(1+\frac{\xi\cdot Y_i}{\sigma}\right)^{-(1+\xi)/\xi}\right]\\ &= \log\frac{1}{\sigma^N}+\sum_{i=1}^{N}\log\left(1+\frac{\xi\cdot Y_i}{\sigma}\right)^{-(1+\xi)/\xi}\\ &= -N\cdot\log\sigma-\left(\frac{1+\xi}{\xi}\right)\sum_{i=1}^{N}\log\left(1+\frac{\xi\cdot Y_i}{\sigma}\right),\end{aligned} \tag{4.46}$$

where $Y_1,\ldots,Y_N$ is the sample of exceedances, is numerically maximized for ξ and σ (with possible constraints arising from Definition 1.9) yielding the estimates for the GPD parameters. Once again the case $\xi = 0$ can be considered as a limit. As mentioned in [Smith, 1987], the ML estimators are asymptotically normal for $\xi > -1/2$ which is likely to be satisfied for the typical claim size distributions (in our case the assumption is $\xi \in [0,1)$). While asymptotically efficient the ML estimator might not be the best candidate for small samples. There is a multitude of possible estimators for estimating the shape parameter ξ which are summarized in [Embrechts et al., 1997: pp. 327–340]. Another estimator for both of the GPD parameters is the method of probability-weighted moments (PWM) from [Hosking and Wallis, 1987] for which the asymptotic normality can be established when $\xi \leqslant 1/2$. The method of PWM is based on equating the empirical and theoretical linear combinations of L-moments (of a given distribution) as described

in [Hosking, 1990]. Because L-moments themselves are linear functions of the data, estimators deduced this way are typically more robust than the ones based on the conventional method of moments. Let X have a GPD with parameters σ and ξ. In the GPD context the method of PWM equates

$$\mathbb{E}(X) = \frac{\sigma}{1-\xi} \tag{4.47}$$

and

$$\mathbb{E}\left[X\overline{G_{\sigma,\xi}}(X)\right] = \frac{\sigma}{2(2-\xi)}, \tag{4.48}$$

with respective unbiased empirical estimators that are sample mean and

$$\frac{1}{n}\sum_{j=1}^{n}\frac{n-j}{n-1}X_{(j)}. \tag{4.49}$$

Unbiasedness of the latter was shown in [Greenwood et al., 1979]. In practice, asymptotically equivalent empirical estimators are used as e.g. in [Hosking and Wallis, 1987] it is advised to use

$$\frac{1}{n}\sum_{j=1}^{n}\frac{n-j+0.35}{n}X_{(j)} \tag{4.50}$$

instead of (4.49).

As demonstrated in [Hosking and Wallis, 1987], PWM estimators can perform better than ML estimators when the sample size is not very large. Because typically the threshold value will be set as high as possible, the size of the sample of exceedances is not very large and thus the small sample properties are more crucial than the asymptotic results. Even though the method of PWM can sometimes produce non-sensible parameter estimates (i.e. data lies outside of the support of the distribution defined by the parameter estimates), this problem is unlikely in our case when ξ is non-negative as shown in [Castillo and Hadi, 1997].

Another important concern is that we have assumed that the data are from a distribution which satisfies the EV condition (1.10) with $\xi \in [0, 1)$ and thus not all values obtained in the estimation process are suitable. We can overcome this obstacle as follows. Suppose first that we obtain the point estimate of ξ that is

negative. Then we set $\xi = 0$ and maximizing the log-likelihood $\ell(0, \sigma)$ from (4.46) which is

$$-N \cdot \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^{N} Y_i, \tag{4.51}$$

leads to the estimator

$$\hat{\sigma} = \sum_{i=1}^{N} Y_i / N. \tag{4.52}$$

In the case of $\xi = 0$ the estimator $\hat{M}$ in (4.33) has a simple form.

Corollary 4.3. Suppose $\xi = 0$ and we use estimator $\hat{\sigma}$ from (4.52) to estimate σ^*, that is $\tilde{\sigma^*} = \hat{\sigma}$ in (4.33). Then $\hat{M} = \sum_{i=1}^{n} X_i / n$.

Proof. We write

$$\begin{aligned} \hat{M} &= \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{N} Y_i + \tilde{\sigma^*} N}{n} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{N} Y_i + N \sum_{i=1}^{N} Y_i / N}{n} \\ &= \frac{\sum_{i=1}^{n} X_i}{n}. \end{aligned} \tag{4.53}$$

□

Now if instead the point estimate of ξ is greater or equal than one then we can fix a small $\epsilon > 0$ and set $\xi = 1 - \epsilon$. A new estimate for σ is obtained as

$$\arg\max_{\sigma > 0} \ell(1 - \epsilon, \sigma) \tag{4.54}$$

in the ML context. When using the method of PWM, (4.54) is not used. Instead a single equation is produced that equates the mean of the sample of exceedances with the mean of the GPD yielding an estimator

$$\epsilon \sum_{i=1}^{N} Y_i / N. \tag{4.55}$$

4.5. Asymptotic distribution of the GPD parameter estimates

As stated in the previous section, GPD parameter estimators are asymptotically normal if the actual value ξ is in a suitable region. In our case a function of these parameters is the object of interest as we want to estimate $\sigma^* = \sigma/(1 - \xi)$ and

$\xi^* = \xi/(1-\xi)$. Using the estimators of the original GPD parameters leads to estimators that are also asymptotically normal as is made precise by the following statements.

Proposition 4.6. Consider a vector-valued function $q(x,y) = (x/(1-y), y/(1-y))$ and suppose that a vector estimator $(\dot{\sigma_n}, \dot{\xi_n})$ of the true vector parameter $(\sigma, \xi) \in [0,\infty) \times (0,1)$ is asymptotically normal i.e.

$$\sqrt{n}\left(\dot{\sigma_n} - \sigma, \dot{\xi_n} - \xi\right) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{4.56}$$

where $\mathcal{N}(0,\Sigma)$ denotes a bivariate normal distribution with zero mean and covariance matrix Σ and $\xrightarrow{d}$ denotes convergence in distribution. Then we have

$$\sqrt{n}\left(q(\dot{\sigma_n}, \dot{\xi_n}) - q(\sigma, \xi)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma^q), \tag{4.57}$$

where

$$\Sigma^q = \begin{pmatrix} \frac{1}{1-\xi} & \frac{\sigma}{(1-\xi)^2} \\ 0 & \frac{1}{(1-\xi)^2} \end{pmatrix} \Sigma \begin{pmatrix} \frac{1}{1-\xi} & 0 \\ \frac{\sigma}{(1-\xi)^2} & \frac{1}{(1-\xi)^2} \end{pmatrix}. \tag{4.58}$$

Proof. The statement of the proposition is a direct application of the delta method (see e.g. [Bilodeau and Brenner, 1999: p. 79]) for the specific case when the function applied to the estimator is q. The matrix of partial derivatives of q valued at (σ, ξ) is precisely

$$\begin{pmatrix} \frac{1}{1-\xi} & \frac{\sigma}{(1-\xi)^2} \\ 0 & \frac{1}{(1-\xi)^2} \end{pmatrix},$$

as can be easily verified. As the partial derivatives are all continuous for $(\sigma, \xi) \in [0,\infty) \times (0,1)$, delta method is applicable. □

Proposition 4.7. Let $\hat{\sigma}$ and $\hat{\xi}$ be the ML estimators and $\check{\sigma}$ and $\check{\xi}$ the PWM estimators of the GPD parameters $\sigma \in [0,\infty)$ and $\xi \in (0,1)$. Then

$$\sqrt{n}\begin{pmatrix} \frac{\hat{\sigma}}{1-\hat{\xi}} - \frac{\sigma}{1-\xi} \\ \frac{\hat{\xi}}{1-\hat{\xi}} - \frac{\xi}{1-\xi} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma^{ML}), \tag{4.59}$$

with

$$\Sigma^{ML} = \frac{1+\xi}{(1-\xi)^4}\begin{pmatrix} (1-\xi)^2(1+\xi) - 2(1-\xi)\sigma^2 + 2\sigma^4 & 2\sigma^3 - \sigma(1-\xi) \\ 2\sigma^3 - \sigma(1-\xi) & 2\sigma^2 \end{pmatrix}. \tag{4.60}$$

If $\xi < 1/2$, then

$$\sqrt{n}\begin{pmatrix} \frac{\check{\sigma}}{1-\check{\xi}} - \frac{\sigma}{1-\xi} \\ \frac{\check{\xi}}{1-\check{\xi}} - \frac{\xi}{1-\xi} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma^{PWM}), \tag{4.61}$$

with

$$\Sigma^{PWM} = \frac{1}{C_4(\xi)}\begin{pmatrix} C_1(\xi) + 2\sigma^2 C_2(\xi) + \sigma^4 C_3(\xi) & \sigma C_2(\xi) + \sigma^3 C_3(\xi) \\ \sigma C_2(\xi) + \sigma^3 C_3(\xi) & \sigma^2 C_3(\xi) \end{pmatrix}, \tag{4.62}$$

where

$$\begin{aligned} C_1(\xi) &= (1-\xi)^3(2-\xi)^2(1-\xi+2\xi^2), \\ C_2(\xi) &= (1-\xi)(2-\xi)(2-6\xi+7\xi^2-2\xi^3), \\ C_3(\xi) &= 7-18\xi+11\xi^2-2\xi^3, \\ C_4(\xi) &= (1-\xi)^4(1-2\xi)(3-2\xi). \end{aligned}$$

Proof. The proof is an application of Proposition 4.6 and utilizes the facts that when $\xi > -1/2$ we have that

$$\sqrt{n}\begin{pmatrix} \hat{\sigma} - \sigma \\ \hat{\xi} - \xi \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_1), \tag{4.63}$$

where

$$\Sigma_1 = (1+\xi)\begin{pmatrix} (1+\xi) & -\sigma \\ -\sigma & 2\sigma^2 \end{pmatrix} \tag{4.64}$$

and when $\xi < 1/2$ we have that

$$\sqrt{n}\begin{pmatrix} \check{\sigma} - \sigma \\ \check{\xi} - \xi \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_2), \tag{4.65}$$

where

$$\Sigma_2 = \frac{1}{(1-2\xi)(3-2\xi)}\begin{pmatrix} \frac{C_1(\xi)}{(1-\xi)^2} & \frac{\sigma C_2(\xi)}{1-\xi} \\ \frac{\sigma C_2(\xi)}{1-\xi} & \sigma^2 C_3(\xi) \end{pmatrix}, \tag{4.66}$$

from [Smith, 1987] and [Hosking and Wallis, 1987], respectively. □

Chapter 5

Simulation study

Suppose we have an independent sample $X_1, \ldots, X_n$ from the distribution F and we wish to estimate the cdf of the itd of F (that is $F_I(x)$). We have assumed that the support of the (theoretical) itd is $(0, \infty)$ but it is immediately clear that the approximations proposed in the previous chapter have certain limitations. The empirical approximation $F_I^{emp_n}(x)$ defined in (4.6) is equal to one if $x > X_{(n)}$, where $X_{(n)} = \max\{X_1, \ldots, X_n\}$. This essentially means that the approximation cannot be used outside of the sample range which is a serious drawback. The GPD approximation, based on (4.28) and (4.33), has the form

$$F_I^{gpd_n}(x) = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{N} Y_i + \sigma^* N G_{\xi^*,\sigma^*}(x-u)}{\sum_{i=1}^{n} X_i - \sum_{i=1}^{N} Y_i + \sigma^* N}, \quad x > u, \tag{5.1}$$

where $Y_1, \ldots, Y_N$ is the sample of exceedances (when the threshold value is u), $G(x)$ is the cdf of a GPD distribution and $\sigma^* = \sigma/(1-\xi)$ and $\xi^* = \xi/(1-\xi)$ are given the values according to the point estimates obtained for σ and ξ using the sample of exceedances. Thus the GPD approximation is less than one for $x > X_{(n)}$, but it is not defined for $x < u$. This means that it is only reasonable to compare the two approximations in the region $(u, X_{(n)})$, which will be our aim in this chapter. If the results favor the GPD approximation then it seems natural to make use of them both to produce a combined approximation with support $(0, \infty)$.

5.1. Methodology of comparison

The following methodology and some of the results are presented in [Kaasik and Pärna, 2009]. Fix n and let $x_1, \ldots, x_n$ be the simulated sample from F, which

in sorted form is denoted as $x_{(1)} < \ldots < x_{(n)}$. Our first concern is to specify a threshold u for the GPD approximation. A natural choice for u is the minimum of thresholds after which the shape parameter ξ of the GPD remains constant as explained in Section 4.4. Let us build a sequence of candidate thresholds $\{u_i\}$ as follows. Assume that it is not reasonable to have more than 300 or less than 20 members in the sample of exceedances $y_1, \ldots, y_N$ as parameter estimation should not be attempted using a very small sample while the GPD will surely not fit to the sample of exceedances if that sample makes up a large part of the whole sample. Required size for the sample of exceedances is guaranteed if we require that $x_{(n-300)} < u \leqslant x_{(n-20)}$. Then we split the interval $(x_{(n-300)}, x_{(n-20)})$ of length L into 200 subintervals using cutting points

$$u_i = x_{(n-300)} + \frac{(i-1)L}{200}, \quad i = 1, \ldots, 200. \tag{5.2}$$

Given the sequence of candidate thresholds $\{u_i\}$ we first estimate, using some pre-specified methodology, the GPD parameters based on the sample of exceedances with the largest threshold u_{200} and obtain a point estimate of ξ (which we denote as ξ_{200}) and its standard error (which we denote as $s_{\xi_{200}}$). We fix a $k > 0$ and produce a confidence interval for ξ as $(\xi^l_{200}, \xi^u_{200}) = (\xi_{200} - ks_{\xi_{200}}, \xi_{200} + ks_{\xi_{200}})$. After that we form another interval (ξ^l, ξ^u), where $\xi^l = \max\{\xi^l_{200}, 0\}$ and $\xi^u = \min\{\xi^u_{200}, 1\}$ and start a cycle with $i = 1$ as follows.

1. Set u_{200-i} as the threshold value to find the sample of exceedances and find the confidence interval $(\xi^l_{200-i}, \xi^u_{200-i})$ for ξ analogously as before.
2. Set $\xi^l = \max\{\xi^l, \xi^l_{200-i}\}$ and $\xi^u = \min\{\xi^u, \xi^u_{200-i}\}$.
3. If $\xi^u < \xi^l$ then break the cycle and select $u_{200-i+1}$ as the threshold value that will be used in the simulations, otherwise restart the cycle with $i = i+1$ unless $i = 199$ in which case break the cycle and select u_1 as the threshold value that will be used in the simulations.

Estimation of the GPD parameters and confidence interval estimation is carried out using the tools in [Coles, 2006] when the method of ML is used and those of [McNeil, 2008] when the method of PWM is used. For the latter case the threshold value is still selected according to the standard errors provided by maximum likelihood as the asymptotic normality for parameter estimates has not been established for the method of PWM when $\xi > 0.5$ and the confidence intervals based on bootstrap techniques are reported to provide unreliable confidence intervals for

the GPD parameters as reported by [Tajvidi, 2003]. The confidence intervals produced by the tools used are based on (4.63) and (4.65), respectively.

The error of an approximation is defined as the mean relative error of the approximation in terms of the complementary cdf. That is, if the actual integrated tail distribution has a complementary cdf $\overline{F_I}(x)$ then we are interested in the random variables

$$A_k^{emp_n} = \frac{1}{X_{(n)} - u} \int_u^{X_{(n)}} \frac{|F_I^{emp_n}(x) - F_I(x)|}{\overline{F_I}(x)} dx \tag{5.3}$$

and

$$A_k^{gpd_n} = \frac{1}{X_{(n)} - u} \int_u^{X_{(n)}} \frac{|F_I^{gpd_n}(x) - F_I(x)|}{\overline{F_I}(x)} dx, \tag{5.4}$$

where k is the constant from the previously explained threshold selection algorithm. More specifically we will be estimating $\mathbb{P}(A_k^{gpd_n} < A_k^{emp_n})$ for different values of k and n for our model distributions. Each estimate is based on 1000 replications and is complemented with 95% confidence intervals based on the normal approximation. The model parameters in the tables are of the original distribution F from which the integrated tail distribution is formed.

5.2. Simulation results for the initial approximations

5.2.1. Pareto case

Simulation results for the Pareto case presented in the Tables 5.1 and 5.2 seem to suggest the superiority of the extreme value approximation $F_I^{gpd_n}(x)$ in the upper part of the sample range. Of course, as for the Pareto distribution the integrated tail distribution is also Pareto which in turn is included in the class of generalized Pareto distributions, it comes as a little surprise that the performance of the proposed approximation based on the GPD is solid. It is also intuitive that as we increase the value of k and the size of the sample of exceedances grows, the parameter estimation process gives us more precise results (which for the Pareto distribution are free from the bias that usually arises because the tail is only approximately distributed as the GPD). However even for the smallest used constant value $k = 1/4$, the empirical approximation is typically outperformed. This is even more so when the method of PWM is used for parameter estimation.

The difference is not so strong as reported in the results of [Kaasik, 2009a], however there the supremum error was considered and the methodology was also somewhat different as the automatic threshold selection was not used and in each replication the sample of exceedances consisted of a pre-specified (fixed) number of members. The impression that the GPD approximation, when the basis of inference is the method of PWM, seems to perform better for heavier tails (smaller values of α in the Pareto distribution context), however, was also apparent in the previously mentioned article.

Table 5.1: $\hat{\mathbb{P}}(A_k^{gpd_n} < A_k^{emp_n})$ in the Pareto case with ML

α	n	$k = 1/4$	$k = 1/2$	$k = 1$	$k = 2$
1.5	1000	0.584±0.031	0.563±0.031	0.625±0.030	0.696±0.029
	10000	0.545±0.031	0.548±0.031	0.602±0.030	0.678±0.029
	100000	0.545±0.031	0.578±0.031	0.615±0.030	0.686±0.029
2.5	1000	0.596±0.030	0.574±0.031	0.640±0.030	0.684±0.029
	10000	0.590±0.030	0.558±0.031	0.606±0.030	0.658±0.029
	100000	0.579±0.031	0.574±0.031	0.630±0.030	0.664±0.029

Table 5.2: $\hat{\mathbb{P}}(A_k^{gpd_n} < A_k^{emp_n})$ in the Pareto case with PWM

α	n	$k = 1/4$	$k = 1/2$	$k = 1$	$k = 2$
1.5	1000	0.834±0.023	0.806±0.025	0.783±0.026	0.832±0.023
	10000	0.786±0.026	0.772±0.026	0.802±0.025	0.829±0.023
	100000	0.794±0.025	0.797±0.025	0.799±0.025	0.792±0.025
2.5	1000	0.676±0.029	0.670±0.029	0.701±0.028	0.740±0.027
	10000	0.672±0.029	0.625±0.030	0.702±0.028	0.749±0.027
	100000	0.697±0.028	0.666±0.029	0.685±0.029	0.716±0.028

5.2.2. Weibull case

Things are a little different when the Weibull case is considered. The results presented in Tables 5.3 and 5.4 show that when the threshold value is set too high i.e. for big values of k the GPD no longer has an advantage over the empirical approach. When the overall sample size is in the region of thousand elements then even a small value of k can cause the situation where the bias introduced by approximating the conditional tail distribution with the GPD is big enough to cause the GPD approximation to give poor results. This is especially true for heavier tails (smaller values of β) when the method of ML is used. The method of

PWM seems to give better results, at least when the critical cases (small overall sample size and thus a relatively low threshold) are concerned. The differences from the results in [Kaasik, 2009a] seem to suggest that the automatic threshold selection using the ML method for producing the confidence intervals might not be too useful. However as previously explained, PWM based confidence intervals are not always available and judgement based threshold selection is out of the question when simulation is concerned. The interesting bias-variance trade-off, arguably visible also in [Kaasik, 2009a], for the method of PWM is that the method performs better for heavier tails when the samples of exceedances are small but when the threshold is decreased the situation is reversed as the bias for the heavier tails seems to play a bigger role while in the situation with lighter tails the improvement due to increased information available for the parameter estimation process seems to be more important.

Table 5.3: $\hat{\mathbb{P}}(A_k^{gpd_n} < A_k^{emp_n})$ in the Weibull case with ML

β	n	$k = 1/4$	$k = 1/2$	$k = 1$	$k = 2$
0.25	1000	0.474±0.031	0.467±0.031	0.266±0.027	0.057±0.014
	10000	0.563±0.031	0.525±0.031	0.416±0.031	0.276±0.028
	100000	0.571±0.031	0.564±0.031	0.505±0.031	0.485±0.031
0.50	1000	0.596±0.030	0.544±0.031	0.502±0.031	0.307±0.029
	10000	0.619±0.030	0.603±0.030	0.613±0.031	0.574±0.031
	100000	0.610±0.030	0.634±0.030	0.633±0.030	0.600±0.029
0.75	1000	0.599±0.030	0.598±0.030	0.612±0.030	0.560±0.031
	10000	0.575±0.030	0.608±0.030	0.623±0.030	0.645±0.030
	100000	0.596±0.030	0.596±0.030	0.649±0.030	0.700±0.028

Table 5.4: $\hat{\mathbb{P}}(A_k^{gpd_n} < A_k^{emp_n})$ in the Weibull case with PWM

β	n	$k = 1/4$	$k = 1/2$	$k = 1$	$k = 2$
0.25	1000	0.664±0.029	0.644±0.030	0.559±0.031	0.447±0.031
	10000	0.671±0.029	0.585±0.031	0.566±0.031	0.450±0.031
	100000	0.603±0.030	0.578±0.031	0.573±0.031	0.549±0.031
0.50	1000	0.609±0.030	0.601±0.030	0.514±0.031	0.404±0.030
	10000	0.591±0.030	0.616±0.030	0.555±0.031	0.548±0.031
	100000	0.579±0.030	0.616±0.030	0.603±0.030	0.607±0.030
0.75	1000	0.568±0.031	0.607±0.030	0.595±0.030	0.527±0.031
	10000	0.587±0.031	0.584±0.031	0.644±0.030	0.613±0.030
	100000	0.567±0.031	0.595±0.030	0.647±0.030	0.639±0.030

5.2.3. Log-normal case

The simulation results for the integrated tail of the log-normal distribution are in the Tables 5.5 and 5.6. The conclusions are somewhat similar to the Weibull case. Namely the method of ML based GPD approximation is unreliable for the heavier tails (larger values of σ). This is especially the case when the overall sample is not very large (in thousands) while the sample of exceedances of considerable size (in hundreds). If the sample size is very large then the issue does not seem to exist. Again, the extreme value approach based on the PWM inference seems to outperform the one that uses the method of ML and the dominance over the empirical approximation increases with the tail weight. The probabilities are again somewhat smaller than the ones presented in [Kaasik, 2009a]. The width of the confidence intervals (and the amount of different experiments which roughly speaking allows two confidence intervals in every table that do not include the actual probability) does not allow to draw clear conclusions whether there are situations where lowering the threshold to increase the size of the sample of exceedances pays off. The safe option of keeping the sample of exceedances as small as possible seems to be the rule of thumb.

Table 5.5: $\hat{\mathbb{P}}(A_k^{gpd_n} < A_k^{emp_n})$ in the log-normal case with ML

σ	n	$k = 1/4$	$k = 1/2$	$k = 1$	$k = 2$
1	1000	0.615±0.030	0.566±0.031	0.609±0.030	0.619±0.030
	10000	0.602±0.030	0.606±0.030	0.637±0.030	0.615±0.030
	100000	0.609±0.030	0.631±0.030	0.643±0.030	0.625±0.030
2	1000	0.541±0.031	0.489±0.031	0.398±0.030	0.197±0.024
	10000	0.559±0.031	0.541±0.031	0.506±0.031	0.430±0.031
	100000	0.552±0.031	0.540±0.031	0.527±0.031	0.523±0.031
3	1000	0.465±0.031	0.363±0.030	0.205±0.027	0.088±0.018
	10000	0.569±0.031	0.530±0.031	0.366±0.030	0.170±0.023
	100000	0.649±0.030	0.612±0.030	0.629±0.030	0.605±0.030

Conclusions from all of the simulation results are as follows. Suppose we have an iid sample from distribution F with size n. When we are considering estimating the cdf of the integrated tail distribution of F for the arguments that are below but close to the sample maximum then the extreme value approximation should be preferred to the empirical approximation. However, the minimum sample size for which such an idea can be considered should be in the region of 1000. The value of the threshold, u, can be set so that number of members in the the sample of

Table 5.6: $\hat{\mathbb{P}}(A_k^{gpd_n} < A_k^{emp_n})$ in the log-normal case with PWM

σ	n	$k=1/4$	$k=1/2$	$k=1$	$k=2$
1	1000	0.651±0.030	0.616±0.030	0.607±0.030	0.573±0.031
	10000	0.629±0.030	0.571±0.031	0.619±0.030	0.643±0.030
	100000	0.619±0.030	0.621±0.030	0.646±0.030	0.602±0.030
2	1000	0.694±0.029	0.706±0.028	0.622±0.030	0.569±0.031
	10000	0.709±0.028	0.648±0.030	0.625±0.030	0.613±0.030
	100000	0.671±0.029	0.665±0.029	0.652±0.029	0.654±0.029
3	1000	0.837±0.023	0.849±0.022	0.812±0.024	0.728±0.028
	10000	0.795±0.025	0.780±0.026	0.754±0.027	0.700±0.028
	100000	0.737±0.027	0.753±0.027	0.734±0.028	0.786±0.025

exceedances is in the region of 100, and u should rather be set higher than lower. For such sample size the method of PWM typically gives better results than the method of ML in the estimation process of the parameters of the GPD.

The GPD approximation (5.1) is usable only for arguments that are above the threshold value. It can be used for ruin probability estimation via (2.12) which works with large argument values. However since the AK algorithm requires definition of the cdf on the whole interval $(0, \infty)$ the two approximations need to be combined. In view of the simulation results, we should use the empirical approximation in the interval $(0, u]$ and GPD approximation in the interval (u, ∞).

5.3. Combined approximations of the integrated tail distribution

As the cdf of an itd (2.8) is a continuous function, we expect that its approximation also has this property. We can prove the following.

Proposition 5.1. Suppose that the estimators $F_I^{emp_n}(x)$ and $F_I^{gpd_n}(x)$ have been constructed based on the sample $X_1, \ldots, X_n$ and the sample of exceedances $Y_1, \ldots, Y_N$ with the threshold value equal to u. Then, if ξ is estimated as zero and the estimation of σ is based on maximum likelihood (4.52), the combined approximation EGP1, defined as

$$F_I^{egp1_n}(x) = F_I^{emp_n}(x) I_{\{x \leqslant u\}} + F_I^{gpd_n}(x) I_{\{x > u\}}, \tag{5.5}$$

is continuous.

Proof. The proof begins by noting that the empirical approximation can also be written out with the help of the sample of exceedances. Namely

$$F_I^{emp_n}(u) = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i}{\sum_{i=1}^n X_i}, \tag{5.6}$$

as pointed out in [Asmussen and Klüppelberg, 1996]. From the other side, since $\xi = 0$ and $\sigma = \sigma^* = \sum_{i=1}^N Y_i/N$ we have that

$$\begin{aligned} \lim_{x \to u+} F_I^{gpd_n}(x) &= \lim_{x \to u+} \frac{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i + \sigma^* N G_{\xi^*,\sigma^*}(x-u)}{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i + \sigma^* N} \\ &= \frac{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i}{\sum_{i=1}^n X_i} \end{aligned} \tag{5.7}$$

because the cdf of a GPD, denoted by G, is continuous and equal to zero for non-positive arguments. As the empirical approximation and the extreme value approximation are both continuous, so is the combined approximation. □

In general, however, when the method of ML is used there is no guarantee that the cdf $F_I^{egp1_n}(x)$ that is defined in (5.5) is continuous thus making this combined approximation problematic in practice. A possible solution can be given using the conditional distribution. As we can write

$$\overline{F_I^{gpd_n}}(x) = \frac{\sigma^* N \overline{G_{\xi^*,\sigma^*}}(x-u)}{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i + \sigma^* N}, \quad x > u, \tag{5.8}$$

we might also deduce that the conditional distribution has the form

$$\overline{F_{Iu}^{gpd_n}}(y) = \frac{\overline{F_I^{gpd_n}}(u+y)}{\overline{F_I^{gpd_n}}(u)} = \overline{G_{\xi^*,\sigma^*}}(y) \tag{5.9}$$

when $y > 0$. Now we can define a combined approximation EGP2 with

$$F_I^{egp2_n}(x) = F_I^{emp_n}(x) I_{\{x \leqslant u\}} + \left[1 - \overline{F_I^{emp_n}}(u) \overline{G_{\xi^*,\sigma^*}}(x-u)\right] I_{\{x>u\}} \tag{5.10}$$

which is continuous regardless of the method used for parameter estimation of the GPD because

$$\lim_{x \to u+} \overline{G_{\xi^*,\sigma^*}}(x-u) = 1.$$

When the method of PWM is used for parameter estimation then $F_I^{egp1_n}(x)$ is continuous as well because of the following statement.

Proposition 5.2. Suppose that the estimators $F_I^{emp_n}(x)$ and $F_I^{gpd_n}(x)$ have been constructed based on the sample $X_1, \ldots, X_n$ and the sample of exceedances $Y_1, \ldots, Y_N$ with the threshold value equal to u. Then if the estimation of GPD parameters is based on the method of PWM we have that

$$F_I^{egp1_n}(x) = F_I^{egp2_n}(x). \tag{5.11}$$

Proof. Based on the definitions (5.5) and (5.10) we just have to check whether

$$1 - \overline{F_I^{emp_n}}(u)\overline{G_{\xi^*,\sigma^*}}(x-u) = F_I^{gpd_n}(x) \tag{5.12}$$

holds for all $x > u$ when the parameters are of the GPD are estimated using the method of PWM. Using (5.6) and some simple algebra we see that the left-hand side of (5.12) is equal to

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i \overline{G_{\xi^*,\sigma^*}}(x-u)}{\sum_{i=1}^n X_i}. \tag{5.13}$$

The right-hand side is given by (5.1). If we suppose that $\sigma^* = \sum_{i=1}^N Y_i/N$ then we would have

$$\begin{aligned} F_I^{gpd_n}(x) &= \frac{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i + \sum_{i=1}^N Y_i G_{\xi^*,\sigma^*}(x-u)}{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i + \sum_{i=1}^N Y_i} \\ &= \frac{\sum_{i=1}^n X_i - \sum_{i=1}^N Y_i \overline{G_{\xi^*,\sigma^*}}(x-u)}{\sum_{i=1}^n X_i}, \end{aligned}$$

as required. The fact that we indeed do have $\sigma^* = \sum_{i=1}^N Y_i/N$ for the method of PWM is due to the fact that $\sigma^* = \sigma/(1-\xi)$ is the mean of the GPD while $\sum_{i=1}^N Y_i/N$ is the mean of the sample of exceedances and for PWM the first equation for parameter estimation is obtained by equating the theoretical and sample mean as in the classical method of moments. □

5.4. Numerical comparison of the combined approximations

In the previous section two combined approximations were introduced – EGP1 and EGP2 defined by (5.5) and (5.10) respectively – which give identical results for

arguments that are below the threshold value but typically differ in the tail part when maximum likelihood is used for parameter estimation. As was seen previously, the approximation using the method of PWM seemed to perform better in the tail part when EGP1 was compared to the empirical approximation. However as EGP1 is typically not continuous when the method of ML is used, EGP2 is the approximation that should be used with that method. Thus it is also necessary to compare the EGP2 approximations in the tail region for the two different parameter estimation methods. This is even more so because the comparison with the empirical approximation might not correctly reflect the balance between the two – it might be the case that when the approximation with ML outperforms the empirical approximation, it also typically outperforms the approximation with PWM and the situation typically stays the same when both are outperformed by the empirical approximation.

We do not use the automatic threshold selection as it is unclear whether it helps or hinders the approximation when the parameters are estimated using the method of PWM. Additional simulation results (not presented) indicate that the estimates of the probabilities do not change much when automatic threshold selection is used for both approximations. The comparison is carried out in two intervals as in [Kaasik and Pärna, 2009]. The first one, $(u, X_{(n)})$, still has the same form, but now N is fixed beforehand and u selected accordingly. The second interval has the form $(X_{(n)}, T)$, where T is dependent on the sample size n in such a way that the probability of $X_{(n)}$ exceeding T is fixed at 1/100. This latter region serves as an example when predicting outside of the sample range. We note that it can happen that the second region has, in fact, $X_{(n)} > T$, but this cannot be avoided if we want to keep the length of the region from growing very large. The error of an approximation is defined as before. That is, if the actual integrated tail distribution has a complementary cdf $\overline{F_I}(x)$ then we are interested in the random variables

$$A_N^{egp1_n} = \frac{1}{X_{(n)} - u} \int_u^{X_{(n)}} \frac{|F_I^{egp1_n}(x) - F_I(x)|}{\overline{F_I}(x)} dx \tag{5.14}$$

and

$$A_N^{egp2_n} = \frac{1}{X_{(n)} - u} \int_u^{X_{(n)}} \frac{|F_I^{egp2_n}(x) - F_I(x)|}{\overline{F_I}(x)} dx, \tag{5.15}$$

when the first interval is considered and random variables

$$B_N^{egp1_n} = \frac{1}{T - X_{(n)}} \int_{X_{(n)}}^{T} \frac{|F_I^{egp1_n}(x) - F_I(x)|}{\overline{F_I}(x)} dx \tag{5.16}$$

and

$$B_N^{egp2_n} = \frac{1}{T - X_{(n)}} \int_{X_{(n)}}^{T} \frac{|F_I^{egp2_n}(x) - F_I(x)|}{\overline{F_I}(x)} dx, \tag{5.17}$$

when the second interval is considered, where on both instances $F_I^{egp1_n}$ is found using the method of PWM, $F_I^{egp2_n}$ is found using the method of ML and N is the pre-specified size of the sample of exceedances. We will be estimating $\mathbb{P}(A_N^{egp1_n} < A_N^{egp2_n})$ and $\mathbb{P}(B_N^{egp1_n} < B_N^{egp2_n})$ for different values of N and n for our model distributions. Each estimate is based on 1000 replications and is complemented with 95% confidence intervals based on the normal approximation. The model parameters in the tables are once again of the original distribution F from which the integrated tail distribution is formed.

5.4.1. Pareto case

Results for the Pareto case, presented in the Tables 5.7 and 5.8, show that when the size of the sample exceedances increases the asymptotic properties of the ML method can be seen and EGP2 with method of ML should be preferred. For the Pareto distribution with lighter tail and a sample of exceedances with size not exceeding 50 gives the advantage to EGP1. It seems that the latter typically improves over EGP2 when approximation outside of the sample range is considered.

Table 5.7: $\hat{\mathbb{P}}(A_N^{egp1_n} < A_N^{egp2_n})$ in the Pareto case

α	n	$N = 25$	$N = 50$	$N = 100$	$N = 300$
1.5	1000	0.469±0.031	0.434±0.031	0.403±0.030	0.343±0.029
	10000	0.534±0.031	0.411±0.030	0.381±0.030	0.351±0.030
	100000	0.530±0.031	0.399±0.030	0.365±0.030	0.359±0.030
2.5	1000	0.642±0.030	0.592±0.030	0.519±0.031	0.435±0.031
	10000	0.659±0.029	0.558±0.031	0.486±0.031	0.459±0.031
	100000	0.678±0.029	0.607±0.030	0.520±0.031	0.416±0.031

Table 5.8: $\hat{\mathbb{P}}(B_N^{egp1_n} < B_N^{egp2_n})$ in the Pareto case

α	n	$N = 25$	$N = 50$	$N = 100$	$N = 300$
1.5	1000	0.499±0.031	0.474±0.031	0.424±0.031	0.385±0.030
	10000	0.533±0.031	0.463±0.031	0.412±0.031	0.384±0.030
	100000	0.542±0.031	0.424±0.031	0.393±0.030	0.385±0.030
2.5	1000	0.625±0.030	0.604±0.030	0.551±0.031	0.445±0.031
	10000	0.627±0.030	0.579±0.031	0.504±0.031	0.486±0.031
	100000	0.646±0.030	0.613±0.030	0.529±0.031	0.438±0.031

5.4.2. Weibull case

When the Weibull case is considered the balance is instead shifted in favor of the EGP1 when heavier tails are considered as is demonstrated by Tables 5.9 and 5.10. When $\beta = 0.75$ approximation EGP2 seems to have the upper hand. It is hard to spot any deterministic patterns when comparing the relative performance of the approximations inside and outside of the sample range.

Table 5.9: $\hat{\mathbb{P}}(A_N^{egp1_n} < A_N^{egp2_n})$in the Weibull case

β	n	$N = 25$	$N = 50$	$N = 100$	$N = 300$
0.25	1000	0.652±0.030	0.601±0.030	0.806±0.025	1.000±0.000
	10000	0.697±0.028	0.631±0.030	0.607±0.030	0.813±0.024
	100000	0.688±0.029	0.652±0.030	0.640±0.030	0.644±0.030
0.50	1000	0.589±0.030	0.647±0.030	0.702±0.028	0.868±0.021
	10000	0.558±0.031	0.567±0.031	0.561±0.031	0.565±0.031
	100000	0.549±0.031	0.550±0.031	0.541±0.031	0.508±0.031
0.75	1000	0.503±0.031	0.533±0.031	0.474±0.031	0.458±0.031
	10000	0.459±0.031	0.459±0.031	0.456±0.031	0.454±0.031
	100000	0.500±0.031	0.478±0.031	0.517±0.031	0.474±0.031

5.4.3. Log-normal case

The results for the log-normal case are in Tables 5.11 and 5.12. This time EGP1 seems to perform better on almost every occasion. Contrast is the largest for heavier tails and results seem to suggest that EGP2 with ML is especially poor when the itd is approximated outside of the sample range.

In conclusion, it must be noted that the order of preference between the approximation EGP1 and EGP2 is by no means clear-cut. Instead it seems reasonable to try out both approximations and perhaps use the more conservative end-result (i.e. a larger ruin probability).

Table 5.10: $\hat{\mathbb{P}}(B_N^{egp1_n} < B_N^{egp2_n})$ in the Weibull case

β	n	$N = 25$	$N = 50$	$N = 100$	$N = 300$
0.25	1000	0.745±0.027	0.792±0.025	0.947±0.014	0.988±0.007
	10000	0.715±0.028	0.734±0.029	0.755±0.027	0.942±0.014
	100000	0.681±0.029	0.702±0.029	0.756±0.027	0.829±0.023
0.50	1000	0.529±0.031	0.615±0.030	0.706±0.028	0.953±0.013
	10000	0.533±0.031	0.531±0.031	0.528±0.031	0.520±0.031
	100000	0.524±0.031	0.519±0.031	0.511±0.031	0.481±0.031
0.75	1000	0.441±0.031	0.475±0.031	0.419±0.031	0.385±0.030
	10000	0.422±0.031	0.425±0.031	0.426±0.031	0.427±0.031
	100000	0.446±0.031	0.425±0.031	0.491±0.031	0.455±0.031

Table 5.11: $\hat{\mathbb{P}}(A_N^{egp1_n} < A_N^{egp2_n})$ in the log-normal case

σ	n	$N = 25$	$N = 50$	$N = 100$	$N = 300$
1	1000	0.642±0.030	0.678±0.029	0.680±0.029	0.654±0.029
	10000	0.587±0.031	0.615±0.030	0.668±0.029	0.629±0.030
	100000	0.565±0.031	0.628±0.030	0.622±0.030	0.591±0.030
2	1000	0.608±0.030	0.506±0.031	0.562±0.031	0.917±0.017
	10000	0.636±0.030	0.564±0.031	0.543±0.031	0.632±0.030
	100000	0.676±0.029	0.592±0.030	0.539±0.031	0.530±0.031
3	1000	0.516±0.031	0.560±0.031	0.784±0.026	0.983±0.008
	10000	0.498±0.031	0.520±0.031	0.587±0.031	0.846±0.022
	100000	0.546±0.031	0.477±0.031	0.471±0.031	0.653±0.030

Table 5.12: $\hat{\mathbb{P}}(B_N^{egp1_n} < B_N^{egp2_n})$ in the log-normal case

σ	n	$N = 25$	$N = 50$	$N = 100$	$N = 300$
1	1000	0.587±0.031	0.665±0.029	0.696±0.029	0.744±0.027
	10000	0.595±0.030	0.600±0.030	0.664±0.029	0.656±0.029
	100000	0.599±0.030	0.618±0.030	0.623±0.030	0.595±0.030
2	1000	0.663±0.029	0.631±0.030	0.733±0.027	0.979±0.009
	10000	0.682±0.029	0.650±0.030	0.661±0.029	0.786±0.025
	100000	0.689±0.029	0.643±0.030	0.620±0.030	0.656±0.029
3	1000	0.664±0.029	0.699±0.028	0.937±0.015	0.995±0.004
	10000	0.610±0.030	0.608±0.030	0.722±0.028	0.962±0.012
	100000	0.589±0.030	0.563±0.031	0.596±0.030	0.822±0.024

Chapter 6

Ruin probability estimation using real world data

Finally, we use the developed methodology to approximate ruin probabilities when real world data is considered. The Danish fire insurance data from Copenhagen Re available in [McNeil, 2008] is used for this purpose. These data are well-known in the extreme value analysis and are used as an example in [Embrechts et al., 1997], [McNeil, 1997], [Resnick, 1997], to name a few. The data set consists of 2156 large insurance claims in Danish Krone (currently 1 EUR = 7.460 DKK) that exceed the threshold of one million. The arrival times of the claims are fixed with daily precision. The size of the claims will be considered in millions to avoid large numbers. The claims span the time period from the year 1980 to the year 1990. We will be using only the latter part – namely the last 1323 claims that cover the years from 1985 to 1990. We do this because it has been pointed out e.g. in [Cizek et al., 2005] that the arrival times of the claims cannot be considered as a realization of the homogeneous Poisson process. Also, because of our assumptions about the claim size distribution, we subtract 1 from the data. Our goal is to estimate the ruin probability of the company.

6.1. Overview of the claim data

The time series of the claims is plotted on Figure 6.1. First impression suggests that the data is indeed from a heavy-tailed distribution because of the high peaks that are well above the majority of the data. The assumption of whether the data could be iid is discussed in [Resnick, 1997] and the conclusion is that there is little evidence for the iid assumption to be rejected. As mentioned before, the arrival times of the claims for the whole sample are not in a good agreement with a homo-

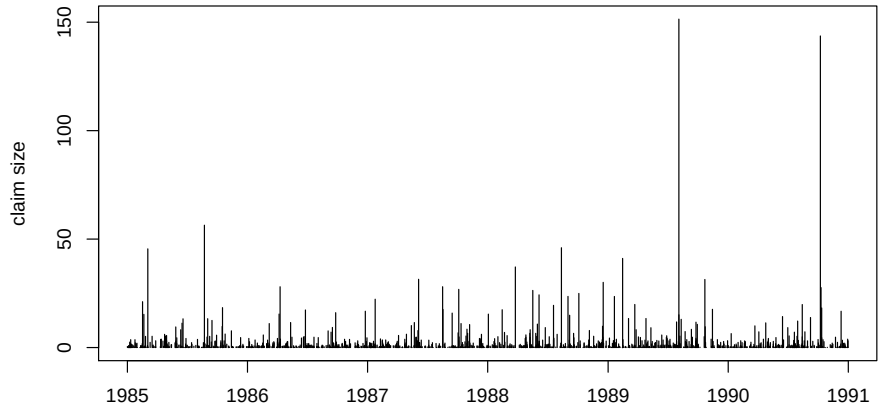

Figure 6.1: Adjusted fire insurance claim size data from the period 1985–1990.

geneous Poisson process. The new time period consists of 2191 days making the expected daily intensity of the Poisson process equal to 0.604. As seen in Figure 6.2 the expected number of claims is well matched by the actual cumulative claim amount making the homogeneous Poisson process plausible for this part of the time series.

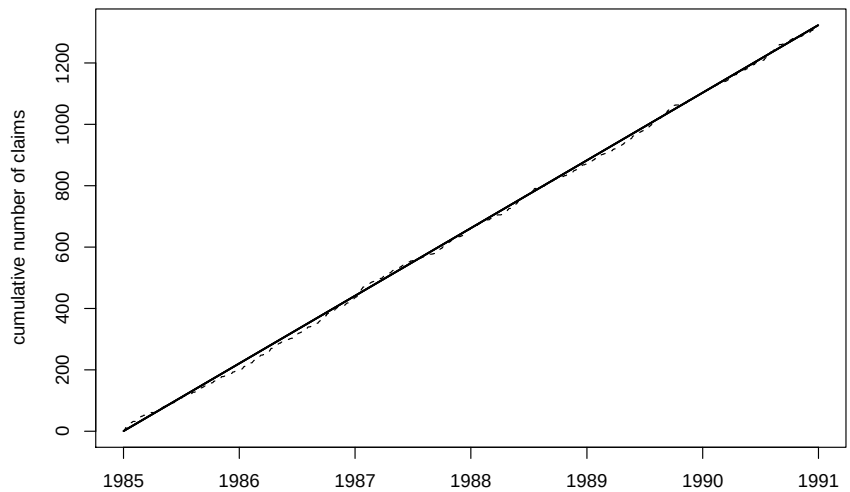

Figure 6.2: Cumulative number of claims from the period 1985–1990. Expected number is represented by the solid and actual number by the dashed line.

The mean of the adjusted claim size is 2.320 and the median is 0.690 with the sample maximum equal to 151.413. These sample characteristics further confirm

that the data is from a heavy-tailed distribution. This conclusion is also suggested by the concave QQ-plot and increasing mean excess of loss function presented in Figure 6.3. When the ratio of sample maximum and sum is plotted heavy-tailedness is once again suggested – according to Figure 6.4 it seems plausible that the distribution of the data does not have finite variance. On the other hand the crucial assumption of finite mean seems plausible because the left-most function seems to be converging to zero.

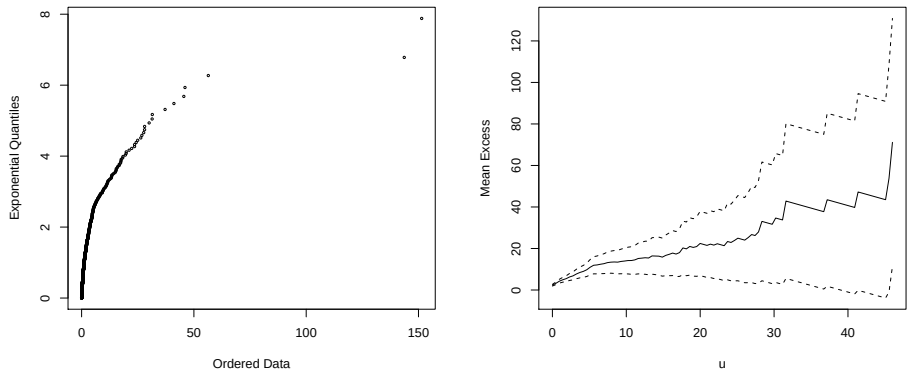

Figure 6.3: QQ-plot (left) and mean excess of loss plot with 95% confidence intervals (right) of the adjusted claim size data from the period 1985–1990. The maximum threshold at which the mean excess of loss function has been calculated is the fourth-largest member of the sample.

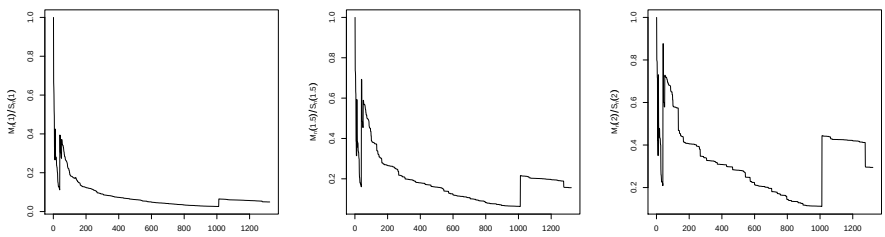

Figure 6.4: The ratio of the maximum and the sum for the sample elements raised to the power 1 (left), 1.5 (center) and 2 (right).

Overall conclusion from the plotted figures is that the claim data are likely to be realizations of a heavy-tailed random variable. That the arrival times of the claims is governed by a homogeneous Poisson process also seems plausible. These facts coupled with the iid assumption of the claims confirm that CLM is valid for this scenario.

The non-restrictive assumption of having the premium rate c equal to one also needs addressing. As the estimated expected daily claim amount is equal to 1.401, having $c = 1$ is not likely in such a situation as that would mean almost sure eventual bankruptcy. To fix this, we double the time between successive claims and thus halving the intensity to 0.302 and then take $c = 1$. This way the estimate of the safety loading coefficient ρ takes the value 0.427. The time period is thus (artificially) doubled by considering different time units, but for convenience we speak of the new time units with old names. That is the length of the time period is 12 years and the daily intensity of the claims is 0.302.

6.2. Threshold selection and parameter estimation

We now proceed to fit the GPD to the right tail of the sample. First possible tool for threshold selection, the mean excess of loss plot, suggests that values 6 and 15 are perhaps likely candidates. The smaller value seems like the first point from which on the function seems roughly linear as there seems to be a decrease in the gradient at this point. The value 15 can arguably also be seen as a point where the gradient of the function changes.

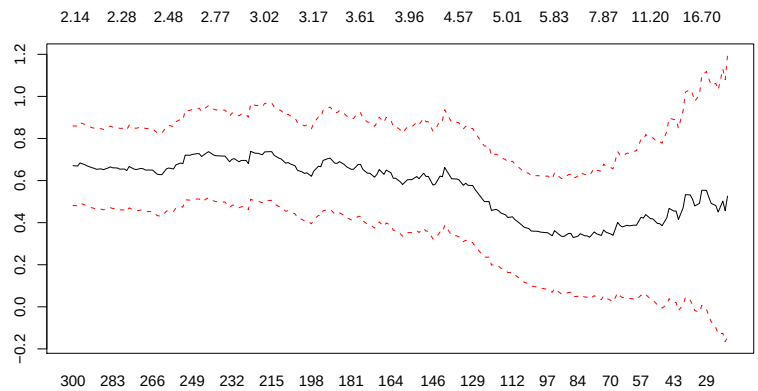

Figure 6.5: Estimates of the GPD shape parameter with 95% confidence intervals by increasing the threshold u (upper horizontal axis) and decreasing number of exceedances (lower horizontal axis). In total 200 models are fitted.

Estimating the shape parameter using the method of ML for different threshold values yields Figure 6.5. Thresholds 6 and 15 are supported by the graph as the estimate of the shape parameter seems rather constant for threshold values larger than 6.

Using the automatic threshold selection methodology described in the previous chapter gives the results in Table 6.1.

Table 6.1: Automatically selected thresholds for different values of constant k

	$k = 1/4$	$k = 1/2$	$k = 1$	$k = 2$
u	20.26	10.14	4.47	2.15

Results obtained when estimating the parameters of the GPD are summarized in the Table 6.2. The contrast between the parameter estimates for the methods of maximum likelihood and probability-weighted moments is rather small. With a smaller threshold value, a less-heavy tailed GPD is deemed better for modeling the tail of the claim data. Achieved fit is analyzed on Figure 6.6. Only the GPD corresponding to the parameter estimates of the ML method are plotted as the PWM variant is hardly separable. It seems that the GPD fits the tail of the data rather well.

Table 6.2: Estimates of the GPD parameters with standard errors (where available)

u	N	ML	PWM
6	95	$\xi = 0.36(0.13),\ \sigma = 7.62(1.24)$	$\xi = 0.38(0.17),\ \sigma = 7.57(1.35)$
15	38	$\xi = 0.56(0.25),\ \sigma = 7.74(2.21)$	$\xi = 0.51,\ \sigma = 7.94$

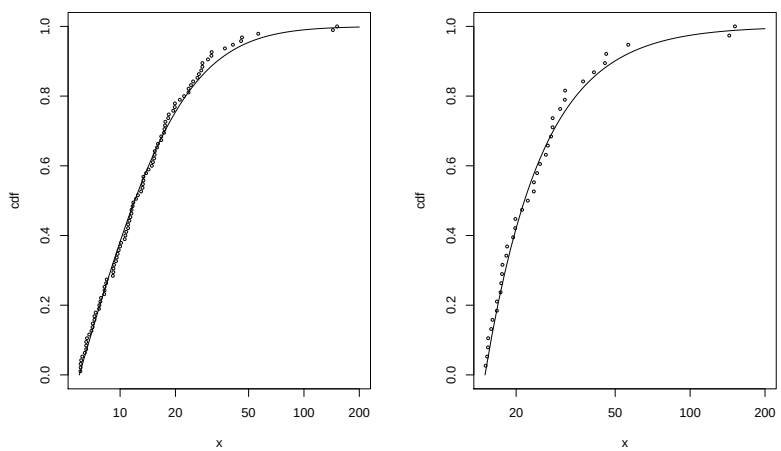

Figure 6.6: In the left-hand plot the GPD is fitted to the 95 exceedances of the threshold 6 and in the right-hand plot the GPD is fitted to the 38 exceedances of the threshold 15. The arguments are plotted on a log-scale. On both occasions the GPD parameters are estimated using the method of ML.

6.3. Constructing the combined approximations

Based on Table 6.2 we can also calculate the estimates for ξ^* and σ^* using the formulae from (4.28). The results are summarized in Table 6.3. We see that for the higher threshold value the mean of the itd does not exist as the estimate of the shape parameter ξ is greater than one. Also, the standard errors accompanied with the parameter estimates are very large caused by estimates of ξ that are close to 1.

Table 6.3: Estimates of the GPD parameters for the integrated tail distribution with standard errors (where available)

u	N	ML	PWM
6	95	$\xi^* = 0.57(3.16),\ \sigma^* = 11.92(23.92)$	$\xi^* = 0.60(3.46),\ \sigma^* = 12.13(26.40)$
15	38	$\xi^* = 1.25(11.18),\ \sigma^* = 17.40(86.25)$	$\xi^* = 1.03,\ \sigma^* = 16.14$

We can now numerically find the four approximations for the itd of the claims (EGP1 and EGP2 for both threshold values). Due to definition, the left part of those distributions is identical. For EGP1 the parameters from the method of PWM are used and for EGP2 the parameters from the method of ML are used. Simulations show that for the threshold $u = 6$ estimates of the mean of the approximations are close to 14.8 while the estimated mean from (3.1) is equal to 12.7. But if we decide that the second moment of the claims does not exist from Figure 6.4 then the mean of the itd would also cease to exist.

Table 6.4: Estimates of the ruin probability using the asymptotic approximation for heavy tails

u_{IC}	u	EGP1	EGP2
50	6	0.129	0.120
	15	0.150	0.171
100	6	0.049	0.044
	15	0.077	0.097
200	6	0.017	0.014
	15	0.040	0.056

To compare the approximations we estimate the ruin probabilities using the asymptotic formula (2.12). The results are given in Table 6.4. Based on these asymptotic results, it seems that the smallest initial capital level considered is not high enough to secure the survival of the company with high probability as even the most opti-

mistic ruin probability estimate exceeds 12%. The highest level on the other hand guarantees a ruin probability below 6% even with the most conservative approximation.

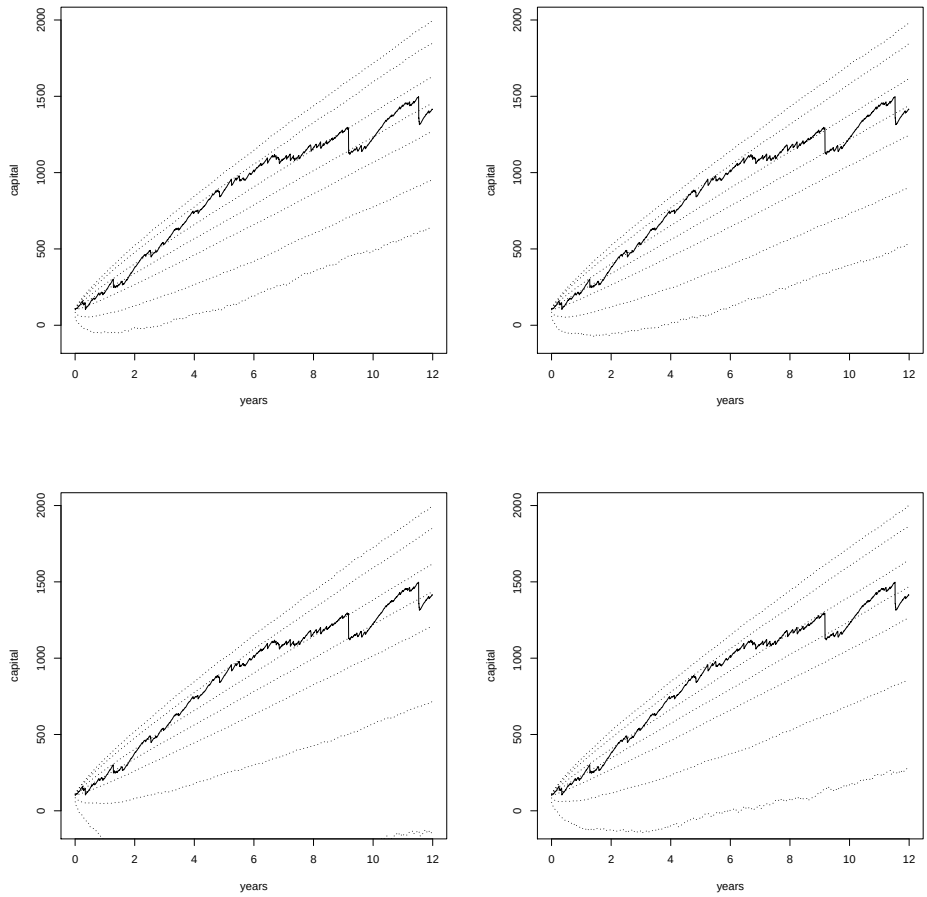

Figure 6.7: Real vs simulated risk processes. Claim size distribution is approximated using thresholds $u = 6$ (up) and $u = 15$ (down) with the method of ML (left) and the method of PWM (right). Thick solid line is the real risk process. The dotted lines are the sample 0.01, 0.05, 0.25, 0.50, 0.75, 0.95, 0.99 quantiles.

Simulated risk process quantiles on Figure 6.7 seem to confirm the higher threshold level as a better choice for approximation. For $u = 6$ the real risk process trajectory is not totally contained between the 0.05 and 0.95 sample quantiles while this is the case for $u = 15$. The difference between EGP1 and EGP2 (respectively based on PWM and ML) is visible (especially for the higher threshold), but the simu-

lation results do not differ too much (as was to be expected based on the rather close GPD parameter estimates as well as the ruin probability approximations). In conclusion, all the four approximations seem consistent with the data (perhaps due to a careful threshold selection process). The heavy-tail characteristic of one large claim causing the ruin can be well illustrated as even though the capital of the insurance company is in the region of 1 300, the probability of a single ruin exceeding that value is still in the region of 0.5% for the approximations using $u = 15$. This is perhaps counterintuitive from the Figure 6.7 according to which it seems (according to the real process trajectory) as if the company is "safe for sure" after 12 years.

Finally, we find the real ruin probabilities corresponding to the approximated claim distributions using the AK algorithm. The results are presented in the Table 6.5 and the first observation is that the asymptotic formula (2.12) does not provide good results in the current case as the results differ significantly from the ones in Table 6.4. This is not surprising in the light of the comments made in the beginning of Section 3.1 about the accuracy of the asymptotic approximation for moderate values of initial capital. Actual ruin probabilities are roughly two times larger than the ones obtained before showing that even with the highest level of initial capital considered ruin can occur with a relatively large probability according to the most conservative estimate.

Table 6.5: Estimates of the ruin probability using simulation

u_{IC}	u	EGP1	EGP2
50	6	0.211	0.235
	15	0.211	0.289
100	6	0.103	0.118
	15	0.121	0.197
200	6	0.037	0.041
	15	0.061	0.121

6.4. Discussion

In conclusion, the real world data example shows the applicability of the proposed methodology in practice. It seems that the GPD fits the tail of the original distribution rather well and the constructed combined estimators are good models for the integrated tail distribution of the claims as suggested by Figure 6.7. Even

though a strong asymptotical result, like Proposition 4.1, might hold, accurately approximating the integrated tail distribution from a limited sample remains a formidable task and the empirical distribution function is not usable in the tail region. The GPD approach also has its limitations – confidence intervals of the modified GPD parameters can be wide and ruin probability estimates can change up to two or three times when a different threshold value is considered. However, in authors opinion, using theoretically justified extreme-value methodology described in the thesis should be preferred tool for estimating the ruin probabilities as long as no additional information is available.

Chapter 7

Conclusions

In the thesis a complete approach for estimating the infinite time ruin probability for a Cramér-Lundberg insurance risk model is proposed with emphasis on heavy-tailed claims. The key part of the approach are the two general Pareto distribution based approximations of the integrated tail distribution, derived for the case when the claim size distribution with finite mean belongs to the domain of attraction of some extreme value distribution and has support that coincides with the positive part of the real line. The following significant contributions by the author should be mentioned:

- ▷ moments of the integrated tail distribution were expressed in terms of the moments of the original distribution;
- ▷ relationship between the integrated tail distribution of a Weibull distribution and a transformed gamma distribution was discovered allowing fast simulation of the former by using known simulation algorithms for the latter;
- ▷ a general iterative simulation algorithm based on the numerical inversion of the cumulative distribution function was proposed for the integrated tail distribution and the rate of convergence was shown to be quadratic with only a few iteration steps typically required;
- ▷ an empirical approximation of the integrated tail distribution was proposed and was shown to be uniformly convergent, moreover the approximation of the integrated tail process was deduced;
- ▷ for the tail part of the integrated tail distribution another approximation based on the generalized Pareto distribution was proposed; asymptotic normality of the parameter estimates of the new approximation was established;

simulation results confirmed that the new approximation is indeed an improvement over the empirical approximation when heavy-tailed distributions are considered;

- ▷ the empirical approximation and the approximation by the generalized Pareto distribution were combined into a joint continuous approximation of the integrated tail distribution;
- ▷ the derived methodology was shown to be applicable for real world insurance data.

The goals set in the introduction where achieved. Finally, it should be mentioned that the extreme value methodology cannot perform miracles – extrapolation of the tail of a distribution based on a small tail sample is always coupled with a moderate amount of uncertainty. This uncertainty remains when the estimates of ruin probabilities are calculated and thus it is always good practice to thoroughly consider all the obtained estimates and perhaps take the conservative route and select the highest of them as the one on which the decisions will be based.

Bibliography

Asmussen, S.: 2000, *Ruin Probabilities*. Singapore: World Scientific Publishing.

Asmussen, S.: 2003, *Applied probability and queues. 2nd revised and extended ed.* New York, NY: Springer.

Asmussen, S. and K. Binswanger: 1997, 'Simulation of ruin probabilities for subexponential claims'. *ASTIN Bulletin* **27**, 297–318.

Asmussen, S., K. Binswanger, and B. Højgaard: 2000, 'Rare events simulation for heavy-tailed distributions.'. *Bernoulli* **6**(2), 303–322.

Asmussen, S. and C. Klüppelberg: 1996, 'Large deviations results for subexponential tails, with applications to insurance risk'. *Stochastic processes and their applications* **64**(1), 103–125.

Asmussen, S., D. Kroese, and R. Rubinstein: 2005, 'Heavy tails, importance sampling and cross-entropy'. *Stochastic Models* **21**(1), 57–76.

Asmussen, S. and D. P. Kroese: 2006, 'Improved algorithms for rare event simulation with heavy tails'. *Advances in Applied Probability* **38**(2), 545–558.

Back, K., T. Bielecki, C. Hipp, S. Peng, and W. Schachermayer: 2004, *Stochastic methods in finance*. Berlin, Germany: Springer-Verlag.

Balakrishnan, N. and V. Nevzorov: 2003, *A primer on statistical distributions*. Hoboken, NJ: Wiley.

Bilodeau, M. and D. Brenner: 1999, *Theory of multivariate statistics*. New York, NY: Springer.

Blanchet, J. and P. Glynn: 2008, 'Efficient rare-event simulation for the maximum of heavy-tailed random walks'. *Annals of Applied Probability* **18**, 1351–1378.

Booth, A. D.: 1955, *Numerical methods.* London, UK: Butterworth's Scientific Publications.

Boots, N. K. and P. Shahabuddin: 2001, 'Simulating ruin probabilities in insurance risk processes with subexponential claims'. In: B. A. Peters, J. S. Smith, D. Medeiros, and M. Rohrer (eds.): *Proceedings of the 2001 Winter Simulation Conference.* pp. 468–476, IEEE Press.

Castillo, E. and A. S. Hadi: 1997, 'Fitting the Generalized Pareto Distribution to Data'. *Journal of the American Statistical Association* **92**(440), 1609–1620.

Cizek, P., W. Härdle, and R. Weron: 2005, *Statistical tools for finance and insurance.* Berlin, Germany: Springer.

Coles, S.: 2001, *An Introduction to Statistical Modeling of Extreme Values.* London, UK: Springer.

Coles, S.: 2006, 'ismev: An Introduction to Statistical Modeling of Extreme Values'. Original S functions by Stuart Coles and R port and R documentation files by Alec Stephenson.

Csörgö, M., S. Csörgö, and L. Horváth: 1986, *An asymptotic theory for empirical reliability and concentration processes.* New York, NY: Springer-Verlag.

De Haan, L. and A. Ferreira: 2006, *Extreme value theory : an introduction.* New York, NY: Springer Science+Business Media.

Dupuis, D.: 1999, 'Exceedances over high thresholds: A guide to threshold selection'. *Extremes* **1**(3), 251–261.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz: 1956, 'Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator'. *The Annals of Mathematical Statistics* **27**(3), 642–669.

Embrechts, P., T. Mikosch, and C. Klüppelberg: 1997, *Modelling extremal events: for insurance and finance.* London, UK: Springer-Verlag.

Embrechts, P. and N. Veraverbeke: 1982, 'Estimates for the probability of ruin with special emphasis on the possibility of large claims'. *Insurance Math. Econom* **1**(1), 55–72.

Finkenstadt, B. and H. Rootzen (eds.): 2004, *Extreme Values in Finance, Telecommunications and the Environment*. Boca Raton, FL: Chapman and Hall/CRC Press.

Grandell, J.: 1991, *Aspects of risk theory*. New York, NY: Springer.

Greenwood, J., J. Landwehr, N. Matalas, and J. Wallis: 1979, 'Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form'. *Water Resources Research* **15**(5), 1049–1054.

Heidelberger, P.: 1995, 'Fast simulation of rare events in queueing and reliability models'. *ACM Transactions on Modeling and Computer Simulation* **5**(1), 43–85.

Hosking, J.: 1990, 'L-moments: analysis and estimation of distributions using linear combinations of order statistics'. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 105–124.

Hosking, J. and J. Wallis: 1987, 'Parameter and quantile estimation for the generalized pareto distribution'. *Technometrics* **29**(3), 339–349.

Johansson, J.: 2003, 'Estimating the Mean of Heavy-Tailed Distributions'. *Extremes* **6**, 91–109.

Juneja, S. and P. Shahabuddin: 2002, 'Simulating heavy tailed processes using delayed hazard rate twisting'. *ACM Transactions on Modeling and Computer Simulation* **12**(2), 94–118.

Juneja, S. and P. Shahabuddin: 2006, 'Rare-event simulation techniques: An introduction and recent advances'. *Simulation* pp. 291–350.

Kaasik, A.: 2009a, 'An extreme value approximation to the integrated tail distribution'. In: D. Konstantinides (ed.): *Proceedings of the 5th Conference in Actuarial Science and Finance on Samos*. pp. 1–8, http://www.actuar.aegean.gr/samos2008/Abstracts/Proceedings08.pdf. 4–7 September 2008, Karlovassi, Greece.

Kaasik, A.: 2009b, 'Simulating the integrated tail distribution in the heavy-tailed setting'. In: K. Hangos (ed.): *Proceedings of the 28th IASTED International Conference on Modelling, Identification and Control*. pp. 28–31, ACTA Press. 16–18 February 2009, Innsbruck, Austria.

Kaasik, A. and K. Pärna: 2008, 'Approximating the integrated tail distribution'. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **12**, 79–87.

Kaasik, A. and K. Pärna: 2009, 'On simulation of ruin using integrated tail distribution'. In: E. K. Z. Leonidas Sakalauskas, Christos Skiadas (ed.): *The XII Conference on Applied Stochastic Models and Data Analysis: Selected papers.* pp. 45–50, Vilnius Gediminas Technical University Press. 30 June – 3 July 2009, Vilnius, Lithuania.

Klugman, S. A., H. H. Panjer, and G. E. Willmot: 2004, *Loss Models: From Data to Decisions, 2nd Ed.* New York, NY: Wiley.

Klüppelberg, C.: 1989, 'Estimation of ruin probabilities by means of hazard rates'. *Insurance: Mathematics and Economics* **8**(4), 279–285.

Komlós, J., P. Major, and G. Tusnády: 1975, 'An approximation of partial sums of independent RV'-s, and the sample DF. I'. *Probability Theory and Related Fields* **32**(1), 111–131.

Kratz, M. and S. Resnick: 1996, 'The qq-estimator and heavy tails'. *Stochastic Models* **12**(4), 699–724.

Lundberg, F.: 1903, *Approximerad framställing av sannolikhetsfunktionen. Återförsäkring av kollektivrisker.* Uppsala: Almqvist & Wiksell. PhD thesis.

Massart, P.: 1990, 'The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality'. *The Annals of Probability* **18**(3), 1269–1283.

McNeil, A.: 1997, 'Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory'. *Astin Bulletin* **27**(1), 117–138.

McNeil, A.: 2008, 'evir: Extreme Values in R'. S original (EVIS) by Alexander McNeil and R port by Alec Stephenson.

Mikosch, T. and A. Nagaev: 1998, 'Large Deviations of Heavy-Tailed Sums with Applications in Insurance'. *Extremes* **1**, 81–110.

Neves, C. and M. Alves: 2008, 'Testing extreme value conditions – an overview and recent approaches'. *REVSTAT–Statistical Journal* **6**(1), 83–100.

Pickands III, J.: 1975, 'Statistical inference using extreme order statistics'. *The Annals of Statistics* **3**, 119–131.

Pollaczek, F.: 1930a, 'Über eine Aufgabe der Wahrscheinlichkeitstheorie. I'. *Mathematische Zeitschrift* **32**, 64–100.

Pollaczek, F.: 1930b, 'Über eine Aufgabe der Wahrscheinlichkeitstheorie. II'. *Mathematische Zeitschrift* **32**, 729–750.

Reinhard, J.: 1984, 'On a class of semi-Markov risk models obtained as classical risk models in a Markovian environment'. *Astin Bulletin* **14**(1), 23–43.

Resnick, S. I.: 1997, 'Discussion of the Danish Data on Large Fire Insurance Losses'. *Astin Bulletin* **27**(1), 139–151.

Rootzén, H. and N. Tajvidi: 1997, 'Extreme value statistics and wind storm losses: a case study'. *Scandinavian Actuarial Journal* **1997**, 70–94.

Ross, S.: 1978, 'Average delay in queues with non-stationary Poisson arrivals'. *Journal of Applied Probability* pp. 602–609.

Ross, S. M.: 2002, *Simulation, 3rd Ed.* Orlando, FL: Academic Press, Inc.

Sigman, K.: 1999, 'Appendix: A primer on heavy-tailed distributions'. *Queueing Systems* **33**, 261–275.

Smith, R. L.: 1987, 'Estimating Tails of Probability Distributions'. *The Annals of Statistics* **15**(3), 1174–1207.

Sparre Andersen, E.: 1957, 'On the collective theory of risk in case of contagion between claims'. In: *Transactions XV International Congress of Actuaries II.* p. 219Ũ229.

Tajvidi, N.: 2003, 'Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions'. *Extremes* **6**(2), 111–123.

Kokkuvõte

Laostumistõenäosuste hindamine raske sabaga kahjunõuetega Cramér-Lundbergi riskiprotsessi korral

Riskiteoorias on levinuimaks kindlustuskompanii varanduslikku seisu kirjeldavaks mudeliks Cramér-Lundbergi riskiprotsess. Olulisim huvipakkuv suurus on seejuures kompanii laostumistõenäosus, mille leidmisel on võtmekomponendiks kahjunõuete jaotuse integreeritud saba jaotus. Mitmed allikad nagu [Asmussen, 2003], [Asmussen and Kroese, 2006], [Blanchet and Glynn, 2008] on muuhulgas pühendatud laostumistõenäosuse leidmisele kui kahjunõuete jaotuse integreeritud saba jaotus või selle jaotuse saba-osa asümptootika on küllalt täpselt teada.

Käesoleva dissertatsiooni põhieesmärgiks on kahjunõuete integreeritud saba jaotuse modelleerimine empiiriliste andmete põhjal, kusjuures sobiv mudeljaotus peab olema teoreetiliselt õigustatud (ning seejuures ka pidev ja kandjaga, mis langeb kokku reaaltelje positiivse osaga). Seejuures on kahjunõuete jaotusele seatud eeldused võimalikult üldised – eeldatakse, et kahjud on lõpliku keskväärtusega pideva mittenegatiivse juhusliku suuruse realisatsioonid. Ehkki kogu metoodika jääb kasutatavaks ka kergete sabadega kahjuandmete korral, on teravdatud tähelepanu all (ja näiterakendustes kasutusel) raskete sabadega jaotused kui realistlikud kahjujaotuste kandidaadid.

Eesmärgi saavutamiseks kasutatakse artiklis [Pickands III, 1975] välja pakutud ekstremaalväärtuste teooria üht põhiteoreemi, mis muuhulgas põhjendab üldistatud Pareto jaotuse sobivust pideva juhusliku suuruse sabaosa tingliku jaotuse mudelina. Üldistatud Pareto jaotus on kaheparameetriline tõenäosusjaotus, mida rakendatakse väga erinevates eluvaldkondades esile kerkivate juhuslike suuruste jaotuse sabaosade modelleerimiseks. Rakendustest saab põgusa ülevaate näiteks

raamatust [Finkenstadt and Rootzen, 2004].

Esimene peatükk annab lühiülevaate töös kasutatavatest tähistustest ja vajalikest mõistetest ja tulemustest. Kirjeldatakse kergete ja raskete sabadega tõenäosusjaotuste erinevust ning raskete sabadega jaotuste alamklasse, muuhukgas regulaarselt varieeruva sabaga jaotuseid ja subeksponentsiaalseid jaotuseid. Ära on toodud üks ekstremaalväärtuste teooria põhiteoreeme, mis sätestab samaväärsena üldistatud ekstremaalväärtuste jaotuse valimimaksimumi piirjaotusena ja üldistatud Pareto jaotuse saba tingliku jaotuse piirjaotusena.

Teises peatükis antakse Cramér-Lundbergi riskiprotsessi ja laostumistõenäosuse definitsioon. Vaadeldakse laostumistõenäosuse erinevaid võimalikke esitusi, nende hulgas ka Pollaczeck-Khinchine'i valemit. Esitatakse laostumistõenäosuse duaalne interpretatsioon järjekorrateoorias. Analüüsitakse kahju jaotuse raske saba mõju laostumistõenäosusele ja selle hindamisele.

Kolmandas peatükis tuletatakse ja sõnastatakse mitmed integreeritud saba jaotuse teoreetilised omadused. Autori panuseks on Teoreem 3.1 ja selle üldistus Teoreem 3.2 artiklist [Kaasik, 2009b], mis näitavad kuidas avalduvad integreeritud saba jaotuse momendid originaaljaotuse momentide kaudu ja kinnitavad fakti, et integreeritud saba jaotus on tüüpiliselt orginaaljaotusest raskema sabaga. Tutvustatakse põgusalt kolme subeksponentsiaalsete jaotuste klassi kuuluvat tõenäosusjaotust: Pareto, Weibulli ja lognormaalset jaotust. Artikli [Kaasik, 2009b] põhjal pakutakse välja iteratiivne eeskiri integreeritud saba jaotusest genereerimiseks ning tõestatakse iteratsiooniprotsessi ruutkoonduvus. Ka seos Weibulli jaotuse ja gamma jaotuse teisenduse vahel on uus tulemus, mis pärineb eelnimetatud artiklist. Peatüki lõpus käsitletakse artiklis [Asmussen and Kroese, 2006] välja pakutud Pollaczeck-Khinchine'i valemist lähtuvat, niinimetatud AK algoritmi, mis võimaldab laostumistõenäosust raske sabaga integreeritud saba jaotuse korral efektiivselt leida simuleerimise teel.

Neljandas peatükis uuritakse esmalt, kuidas kontrollida, kas integreeritud saba jaotus on raske sabaga. Seejärel vaadeldakse integreeritud saba jaotuse lähendi leidmist empiirilise jaotusfunktsiooni abil. Autor tõestab antud lähendi ühtlase koon-

duvuse Teoreemis 4.1, mis pärineb artiklist [Kaasik and Pärna, 2008] ja vaatleb vastava empiirilise protsessi lähendamist Teoreemis 4.2. Järgnevalt tõestab autor Teoreemis 4.3, et integreeritud saba jaotuse sabaosa modelleerimine üldistatud Pareto jaotusega on õigustatud, kui originaaljaotuse sabaosa käitub nagu üldistatud Pareto jaotus, seejuures avalduvad integreeritud saba jaotuse lähendjaotuse parameetrid originaaljaotuse lähendjaotuse parameetrite kaudu. Viimane tulemus on esmakordselt publitseeritud artiklis [Kaasik, 2009a]. Peatüki lõpus vaadeldakse suurima tõepära meetodi ja tõnäosusega kaalutud momentide meetodi kasutamist üldistatud Pareto jaotuse parameetrite hindamisel ning autor tuletab Teoreemis 4.7 integreeritud saba jaotuse lähendjaotuse parameetrite hinnagute asümptootilise normaalsuse.

Viies peatükk on pühendatud empiirilise jaotusfunktsiooni abil ja üldistatud Pareto jaotuse abil saadud lähendite võrdlemisele subeksponentsiaalse originaaljaotuse korral. Võrdlemine viiakse läbi simulatsioonieksperimendina, mille tulemused näitavad üldistatud Pareto jaotusel põhineva lähendi paremust integreeritud saba jaotuse sabaosas. Eksperimendi tulemused on avaldatud artiklis [Kaasik and Pärna, 2009]. Nende kahe lähendi baasil on konstrueeritud uus lähend, mis paraku ei pruugi olla pidev, kui üldistatud Pareto jaotuse parameetrite hinnangud on leitud suurima tõepära meetodi abil. Sellest tulenevalt pakutakse välja ka teine kombineeritud lähend, mis tõenäosusega kaalutud momentide meetodi kasutamise korral langeb kokku eelnevaga, suurima tõepära meetodi korral aga on alati pidev. Peatüki lõpus võrreldakse kahte kombineeritud lähendit simuleerimise teel.

Kuuendas peatükis kasutatakse eespool tuletatud metodoloogiat reaalsetel kahjukindlustuse andmetel kindlustuskompanii laostumistõenäosuse leidmiseks. Selgub, et töös välja pakutud metoodika võimaldab hästi modelleerida tegelikku riskiprotsessi ja seeläbi on alust arvata, et ka kombineeritud lähendite ja AK algoritmi koostöös leitud laostumistõenäosuste hinnangud on küllalt täpsed.

Viimases peatükis võetakse kokku töö käigus saavutatu ning tõdetakse, et igasugune jaotuse sabaosa lähendamine väljaspool tegelikku valimit (ja seeläbi ka näiteks laostumistõenäosuse hindamine kui kahjude jaotus ei ole teada) on oma olemuselt riskantne tegevus, ent töös kasutatud ekstremaalväärtuste teooria vahendid pakuvad selleks parimad võimalikud tööriistad.

Publications

Curriculum Vitae

Name: Ants Kaasik
Citizenship: Estonian Republic
Date of Birth: January 13, 1982
Place of Birth: Tallinn, Estonia
Marital Status: Married
Contact Address: J. Liivi 2-517, Institute of Mathematical Statistics, University of Tartu, Estonia
E-mail: ants.kaasik@ut.ee

Education:
2000–2005 University of Tartu, Bachelor's studies in mathematical statistics
2005–2009 University of Tartu, PhD student in mathematical statistics

Professional Employment:
Since 2007 specialist at the Institute of Mathematical Statistics
Since 2009 data analysis specialist at the excellence centre Frontiers in Biodiversity Research

Curriculum Vitae

Nimi: Ants Kaasik
Kodakondsus: Eesti Vabariik
Sünniaeg: 13. jaanuar 1982
Sünnikoht: Tallinn, Eesti
Perekonnaseis: abielus
Kontaktaadress: J. Liivi 2-517, TÜ matemaatilise statistika instituut
E-posti aadress: ants.kaasik@ut.ee

Hariduskäik:
2000–2005 TÜ matemaatika-informaatikateaduskond, bakalaureuseõpe matemaatilise statistika erialal
2005–2009 TÜ matemaatika-informaatikateaduskond, doktorant matemaatilises statistika erialal

Erialane teenistuskäik:
Alates 2007. aastast spetsialist TÜ matemaatilise statistika instituudis
Alates 2009. aastast andmetöötlusspetsialist bioloogilise mitmekesisuse tippkeskuses FIBIR

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS

1. **Mati Heinloo.** The design of nonhomogeneous spherical vessels, cylindrical tubes and circular discs. Tartu, 1991, 23 p.
2. **Boris Komrakov.** Primitive actions and the Sophus Lie problem. Tartu, 1991, 14 p.
3. **Jaak Heinloo.** Phenomenological (continuum) theory of turbulence. Tartu, 1992, 47 p.
4. **Ants Tauts.** Infinite formulae in intuitionistic logic of higher order. Tartu, 1992, 15 p.
5. **Tarmo Soomere.** Kinetic theory of Rossby waves. Tartu, 1992, 32 p.
6. **Jüri Majak.** Optimization of plastic axisymmetric plates and shells in the case of Von Mises yield condition. Tartu, 1992, 32 p.
7. **Ants Aasma.** Matrix transformations of summability and absolute summability fields of matrix methods. Tartu, 1993, 32 p.
8. **Helle Hein.** Optimization of plastic axisymmetric plates and shells with piece-wise constant thickness. Tartu, 1993, 28 p.
9. **Toomas Kiho.** Study of optimality of iterated Lavrentiev method and its generalizations. Tartu, 1994, 23 p.
10. **Arne Kokk.** Joint spectral theory and extension of non-trivial multiplicative linear functionals. Tartu, 1995, 165 p.
11. **Toomas Lepikult.** Automated calculation of dynamically loaded rigidplastic structures. Tartu, 1995, 93 p. (in Russian).
12. **Sander Hannus.** Parametrical optimization of the plastic cylindrical shells by taking into account geometrical and physical nonlinearities. Tartu, 1995, 74 p.
13. **Sergei Tupailo.** Hilbert's epsilon-symbol in predicative subsystems of analysis. Tartu, 1996, 134 p.
14. **Enno Saks.** Analysis and optimization of elastic-plastic shafts in torsion. Tartu, 1996, 96 p.
15. **Valdis Laan.** Pullbacks and flatness properties of acts. Tartu, 1999, 90 p.
16. **Märt Põldvere.** Subspaces of Banach spaces having Phelps' uniqueness property. Tartu, 1999, 74 p.
17. **Jelena Ausekle.** Compactness of operators in Lorentz and Orlicz sequence spaces. Tartu, 1999, 72 p.
18. **Krista Fischer.** Structural mean models for analyzing the effect of compliance in clinical trials. Tartu, 1999, 124 p.

19. **Helger Lipmaa.** Secure and efficient time-stamping systems. Tartu, 1999, 56 p.
20. **Jüri Lember.** Consistency of empirical k-centres. Tartu, 1999, 148 p.
21. **Ella Puman.** Optimization of plastic conical shells. Tartu, 2000, 102 p.
22. **Kaili Müürisep.** Eesti keele arvutigrammatika: süntaks. Tartu, 2000, 107 lk.
23. **Varmo Vene.** Categorical programming with inductive and coinductive types. Tartu, 2000, 116 p.
24. **Olga Sokratova.** Ω-rings, their flat and projective acts with some applications. Tartu, 2000, 120 p.
25. **Maria Zeltser.** Investigation of double sequence spaces by soft and hard analitical methods. Tartu, 2001, 154 p.
26. **Ernst Tungel.** Optimization of plastic spherical shells. Tartu, 2001, 90 p.
27. **Tiina Puolakainen.** Eesti keele arvutigrammatika: morfoloogiline ühestamine. Tartu, 2001, 138 p.
28. **Rainis Haller.** *M(r,s)*-inequalities. Tartu, 2002, 78 p.
29. **Jan Villemson.** Size-efficient interval time stamps. Tartu, 2002, 82 p.
30. **Eno Tõnisson.** Solving of expession manipulation exercises in computer algebra systems. Tartu, 2002, 92 p.
31. **Mart Abel.** Structure of Gelfand-Mazur algebras. Tartu, 2003. 94 p.
32. **Vladimir Kuchmei.** Affine completeness of some ockham algebras. Tartu, 2003. 100 p.
33. **Olga Dunajeva.** Asymptotic matrix methods in statistical inference problems. Tartu 2003. 78 p.
34. **Mare Tarang.** Stability of the spline collocation method for volterra integro-differential equations. Tartu 2004. 90 p.
35. **Tatjana Nahtman.** Permutation invariance and reparameterizations in linear models. Tartu 2004. 91 p.
36. **Märt Möls.** Linear mixed models with equivalent predictors. Tartu 2004. 70 p.
37. **Kristiina Hakk.** Approximation methods for weakly singular integral equations with discontinuous coefficients. Tartu 2004, 137 p.
38. **Meelis Käärik.** Fitting sets to probability distributions. Tartu 2005, 90 p.
39. **Inga Parts.** Piecewise polynomial collocation methods for solving weakly singular integro-differential equations. Tartu 2005, 140 p.
40. **Natalia Saealle.** Convergence and summability with speed of functional series. Tartu 2005, 91 p.
41. **Tanel Kaart.** The reliability of linear mixed models in genetic studies. Tartu 2006, 124 p.

42. **Kadre Torn.** Shear and bending response of inelastic structures to dynamic load. Tartu 2006, 142 p.
43. **Kristel Mikkor.** Uniform factorisation for compact subsets of Banach spaces of operators. Tartu 2006, 72 p.
44. **Darja Saveljeva.** Quadratic and cubic spline collocation for Volterra integral equations. Tartu 2006, 117 p.
45. **Kristo Heero.** Path planning and learning strategies for mobile robots in dynamic partially unknown environments. Tartu 2006, 123 p.
46. **Annely Mürk.** Optimization of inelastic plates with cracks. Tartu 2006. 137 p.
47. **Annemai Raidjõe.** Sequence spaces defined by modulus functions and superposition operators. Tartu 2006, 97 p.
48. **Olga Panova.** Real Gelfand-Mazur algebras. Tartu 2006, 82 p.
49. **Härmel Nestra.** Iteratively defined transfinite trace semantics and program slicing with respect to them. Tartu 2006, 116 p.
50. **Margus Pihlak.** Approximation of multivariate distribution functions. Tartu 2007, 82 p.
51. **Ene Käärik.** Handling dropouts in repeated measurements using copulas. Tartu 2007, 99 p.
52. **Artur Sepp.** Affine models in mathematical finance: an analytical approach. Tartu 2007, 147 p.
53. **Marina Issakova.** Solving of linear equations, linear inequalities and systems of linear equations in interactive learning environment. Tartu 2007, 170 p.
54. **Kaja Sõstra.** Restriction estimator for domains. Tartu 2007, 104 p.
55. **Kaarel Kaljurand.** Attempto controlled English as a Semantic Web language. Tartu 2007, 162 p.
56. **Mart Anton.** Mechanical modeling of IPMC actuators at large deformations. Tartu 2008, 123 p.
57. **Evely Leetma.** Solution of smoothing problems with obstacles. Tartu 2009, 81 p.