# Comparison of ruin probability approximations in case of real data

Sergidou, E.K.

*Award date:*
2015

# Comparison of ruin probability approximations in case of real data

**Eleni Konstantina Sergidou**

A thesis presented for the degree of
Master of Science


Mathematics and Computer Science Department
Eindhoven University of Technology
The Netherlands
March 30, 2015

I dedicate my master's thesis work
to my parents, Anastasia Kampnou
and Andreas Sergidis. It is only a meager
effort to reciprocate the endeavor
they have made to raise me.

I would like to express my sincere gratitude
to my supervisor Dr. Maria Vlasiou,
to my boyfriend and to all of those that
support me throughout this process.

**Abstract**

The evaluation of the ruin probability in the classical Cramér-Lundberg risk model is one of the most interesting and challenging problems in the area of risk theory, especially in the presence of heavy-tailed claims. In this research, we study and evaluate the performance of various ruin probability approximations using real data. More specifically, we consider the heavy-tailed asymptotic, phase-type and corrected phase-type approximations. Moreover, we apply the corresponding approximations to the distribution of the loss on the portfolio and Value at Risk. To be able to use these approximations, we need to have the appropriate claim size distribution. Therefore, we model the claim sizes with a heavy-tailed, a phase-type and a mixture of these two distributions, while we assume that the inter-arrival time of the claims follows an exponential distribution. Since we cannot calculate the exact ruin probability based on our data, we assume that the true distribution of the claim sizes is a mixture of a phase-type and a heavy-tailed distribution. Finally, we perform a comparison of the outcomes of each approximation with the empirical distribution of the loss and the Value at Risk derived by the real data and the ruin probability obtained by the simulations.

# Contents

**1 Introduction** **1**

**2 Model description** **3**

**I Fitting a distribution** **7**

**3 Heavy-tailed distribution** **7**
3.1 Pareto distribution . . . . . 8
3.1.1 Estimating the parameters of the Pareto distribution . . . . . 8
3.2 Log normal distribution . . . . . 12
3.2.1 Estimating the parameters of the Log normal distribution . . . . . 12
3.3 Weibull distribution . . . . . 13
3.3.1 Estimating the parameters of the Weibull distribution . . . . . 13
3.4 Choosing a heavy-tailed distribution . . . . . 14
3.4.1 Quadratic EDF statistics . . . . . 14
3.4.2 QQ-plot . . . . . 15

**4 Phase-type distributions** **15**
4.1 Special cases of phase-type distributions . . . . . 16
4.1.1 Hyperexponential distribution . . . . . 16
4.1.2 Coxian distribution . . . . . 17
4.2 Fitting phase-type distributions . . . . . 18
4.2.1 EM algorithm . . . . . 18
4.2.2 An EM-based technique . . . . . 19
4.2.3 FW algorithm . . . . . 19
4.2.4 An alternative EM-based technique . . . . . 19
4.3 Choosing the number of phases . . . . . 20

**5 Mixture model: Light-tailed and heavy-tailed distribution** **20**
5.1 Estimating the lower bound and shape parameter of the Pareto distribution 21
5.2 Goodness-of-fit test . . . . . 22

**II Literature overview on the approximations** **23**

**6 Heavy-tailed asymptotic** **23**

**7 Phase-type approximations** **24**

**8 Corrected phase-type approximations** **25**

**III Danish data** **27**

**9 Arrival rate for the Danish data** **27**

**10 Fitting a heavy-tailed distribution** **28**
10.1 Fitting a Pareto distribution . . . . . 28
10.2 Fitting a Log normal distribution . . . . . 30
10.3 Fitting a Weibull distribution . . . . . 31
10.4 QQ-plots . . . . . 31

**11 Phase-type distributions** **33**
11.1 Using all the claims in the Danish data . . . . . 33
11.1.1 Fitting a Hyperexponential distribution . . . . . 33
11.1.2 Fitting a Coxian distribution . . . . . 34
11.2 Using the light-tailed part of the Danish data . . . . . 35
11.2.1 Fitting a Hyperexponential distribution . . . . . 35
11.2.2 Fitting a Coxian distribution . . . . . 36
11.3 Using the heavy-tailed part of the Danish data . . . . . 38
11.3.1 Fitting a Hyperexponential distribution . . . . . 38
11.3.2 Fitting a Coxian distribution . . . . . 39
11.4 Combination of the two parts . . . . . 40

**12 Mixture model: Phase-type and Pareto distribution** **42**

**IV Numerical results** **45**

**13 Total loss and Value at Risk** **45**
13.1 Evaluation of the phase-type approximations . . . . . 46
13.2 Evaluation of the corrected discard approximations . . . . . 49
13.3 Comparison of the different approximations . . . . . 50
13.3.1 In case of the Danish data . . . . . 51
13.3.2 In case of the Erl+Par model . . . . . 52

**14 Ruin probability** **54**

**15 Conclusions** **58**

**A Appendix** **60**

# 1 Introduction

The numerical evaluation of the ruin probability and the steady-state waiting-time distribution of a queue are two important problems in the areas of risk and queueing theory, respectively, and therefore they have been widely studied in the literature ([5], [6] and references therein). The ruin probability is defined as the probability that the reserve of an insurance company will ever drop below zero, while, the time an arriving customer in steady-state has to wait in a queue until his service starts is called waiting time. It can be shown that the ruin probability with initial reserve $u$ is equivalent to the stationary probability that the waiting time of a customer exceeds $u$ of an $G/G/1$ queue, where the service times correspond to the random claim sizes ([5], [6]). This implies that the ruin probability and the waiting time distribution can be calculated interchangeably, and henceforth we restrict ourselves to the notations of the risk theory.

In this research, we study the numerical evaluations of the ruin probability in the classical Cramér-Lundberg risk model. In this model, let $u$ be the initial capital of an insurance company where claims for money arrive according to a Poisson process and the total premium rate is $p$. Under the assumption of exponentially decaying claim size distributions, the ruin probability is available via the classical Pollaczek-Khinchine formula ([6]), since the Laplace transform of the claim size distributions in these cases is available in closed form. Hence, numerical transform inversion can be used on the Pollaczek-Khinchine formula.

However, in insurance models this assumption often is not applicable and heavy-tailed distributions are more appropriate for modeling claim sizes ([16]). In such distributions, an extremely large claim size can have a nontrivial probability since heavy-tailed distributions decay more slowly than any exponential function. In the presence of heavy-tailed claim sizes, numerical methods become challenging or even problematic. This happens because they use the Laplace transform and oftentimes the heavy-tailed distributions do not have a closed-form, analytic Laplace transforms. Thus, since the ruin probability cannot be computed explicitly, it needs to be approximated.

When the average amount of claim per time unit is smaller, but almost equal, to the premium rate and the first two moments of the claims size distribution are finite, heavy-traffic approximations can be used ([26]). The disadvantage of this approximation is that for many heavy-tailed distributions the condition on the first two moments is not satisfied. On the other hand, when on average, the expected premiums are much larger than the expected claims, the use of the light-traffic approximations is more appropriate ([3], [9], [13], [14], [34]). However, in risk theory, the light-traffic approximations are of limited interest, since the heavy-traffic scenario, where the premium rate is slightly bigger than the average claim per unit of time, is most often argued to be the typical case.

The approximation proposed by Beekman in [8] is based on the idea of matching moments of the ruin probability, however several heavy-tailed distributions may have infinite higher-order moments. The ruin probability approximation in [38] is a refinement of the central limit theorem.

A stream of papers ([20], [33]) proposed to approximate directly the Laplace transform of the claim sizes rather than the claim size distribution, using the Transform Approximation

Method. Even though these approximations are quite effective in handling distributions with infinite moments, their accuracy cannot be predetermined.

As mentioned earlier, heavy-tailed distributions may lack a Laplace transform in closed form. One way to get around this difficulty is to approximate the heavy-tailed distribution with a phase-type distribution whose Laplace transform is available in closed form ([7], [19], [22], [25], [32], [36]). Even though these methods can approximate the claim size distribution with high accuracy, the number of phases needed in order to achieve a desired accuracy has to be found by trial and error.

When the claim size distribution belongs to the subexponential class (a subclass of the heavy-tailed distributions), the so-called heavy-tailed approximations can be used to approximate the ruin probability ([18], [29]). The main drawback of these approximations is that they are asymptotic approximations. This means that they provide a good fit only at the tail of the ruin probability, especially when the average claim per unit of time is close to the premium rate. Moreover, if the claim size distribution is a regularly varying distribution, an approximation which can transit from the heavy-traffic regime to the heavy-tailed asymptotic is proposed in [28].

A stream of papers suggests that only a fraction of a sample might follow a heavy-tailed distribution ([11], [21]), while the rest could follow a light-tailed distribution. This implies that the claim size distribution can be modeled as a mixture of a phase-type distribution and a heavy-tailed distribution. Based on this assumption for the claim size distribution, Vatamidou et al. propose in [37] the use of the corrected phase-type approximations, aiming to combine the best elements of phase-type and heavy-tailed asymptotic approximations.

In this research, our purpose is to evaluate the performance of various ruin probability approximations using real data. For this reason, we use the Danish data which consist of 2167 claims over one million Danish Krone (DKM) and their arrival date from the years 1980 to 1990 inclusive. In order to be able to apply any of the existing ruin probability approximations, it is necessary to know the inter-arrival time and claim size distributions. However, since we have real data, these distributions are not a-priori known and thus, they have to be evaluated from our data.

By assumption, the claims arrive to the insurance company according to a Poisson process, whose parameter is estimated by the data we have. The distributions used to model the claim size distribution can be split into three main categories: heavy-tailed, phase-type and a mixture of a phase-type and a heavy-tailed. For each category we follow different procedures to fit a distribution on our data and to choose the most appropriate distribution inside a category.

After the fitting of various distributions on the Danish data and based on the fitted distribution, we calculate the ruin probability using the appropriate approximation. When the claim size distribution is approximated with a phase-type one, we can use directly the Pollaczek-Khinchine formula to evaluate the ruin probability ([6]). In the case of a heavy-tailed claim size distribution, we use the heavy-tailed asymptotic approximation for the ruin probability ([18]). Finally, when the Danish data impose a mixture of a phase-type and a heavy-tailed distribution as the claim size distribution, we can apply the corrected phase-

type approximations ([37]).

Besides the ruin probability, a very popular tool in real-world applications to measure the operational risk is the Value at Risk (VaR). For a given portfolio, a VaR with a probability level $q$ and fixed time horizon is defined as the threshold value such that the loss on the portfolio over the given time horizon exceeds this value with probability $1-q$. In other words, VaR is the $(1-q)$-quantile of the loss on the portfolio in a fixed time interval and can be derived from our data. Moreover, for each approximation we use for the ruin probability, we can apply the corresponding approximation for the distribution of the loss on the portfolio in a fixed time horizon and derive the associated VaR. We compare the outcomes of each approximation with the empirical distribution of the loss on the portfolio and the VaR obtained by the Danish Data.

Finally, since we cannot extract the exact ruin probability using the Danish data, we assume that the true claim size distribution is one of the distributions which we fit on our data. Based on this assumption, we derive the exact ruin probability using simulations and we compare the performance of the ruin probability approximations we consider in this research.

The rest of the thesis is organized as follows. In section 2, a detailed description of the model we use is given. In sections 3, 4 and 5, we study various fitting techniques for each category of distributions we use to model our claim sizes and procedures to choose the most appropriate distribution in each category. In sections 6, 7 and 8, we describe the ruin probability and loss on the portfolio approximations whose performance is evaluated in this research. In sections 10, 11 and 12, we apply to the Danish data the fitting procedures and goodness-of-fit techniques described in previous sections and we give the related results. In sections 13 and 14, the outcomes of applying different approximations for the distribution of the loss on the portfolio, VaR and ruin probability, to our data, are given. Additionally, we compare these outcomes with the ones we derived from the Danish data or simulations. Finally, in section 15, we summarize and discuss our final conclusions for the approximations we considered in this research.

# 2 Model description

The model we use in this research is based on the classical Cramér-Lundberg risk model, also called compound Poisson model ([6]). This model assumes that the claims arrive to the insurance company according to a Poisson process $\{N_t\}_{t\geq 0}$ with rate $\lambda$. Let $U_i$ denote the size of the $i^{th}$ claim, then the $U_i$'s are strictly positive i.i.d. with common distribution $G$. In addition, premiums flow in at rate $p > 0$ per unit of time. The risk reserve process, $\{R_t\}_{t\geq 0}$, has the following form,

$$R_t = u + pt - \sum_{i=1}^{N_t} U_i,$$

where $N_t$ is the number of claims that arrived in $[0,t]$, and $u = R_0$ is the initial reserve of the risk process. The probability of ultimate ruin of the insurance company is defined as the probability of that its reserve ever drops below zero

$$\psi(u) = \mathbb{P}(\inf_{t\geq 0} R_t < 0) = \mathbb{P}(\inf_{t\geq 0} R_t < 0 | R_0 = u).$$

The only difference between the model described here and the classical Cramér-Lundberg risk model is that, in the latter model, the premiums flow in at rate $p = 1$ per unit of time. However, in [6], it is proven that for any $p \neq 1$ one can define $\tilde{R}_t = R_{\frac{t}{p}}$ which yields $\tilde{\psi}(u) = \psi(u)$. This implies that the two models actually are equivalent.

Usually it is more convenient to work with the claim surplus process $\{S_t\}_{t\geq 0}$, defined as $S_t = u - R_t$. Let $M = \sup_{t\geq 0} S_t$, then the probability that the supremum $M$ exceeds $u$ is equivalent to the ruin probability $\psi(u)$, i.e.

$$\psi(u) = \mathbb{P}(M > u).$$

Mathematically, the ruin probability of the classical Cramér-Lundberg risk model is related to the workload (virtual waiting time) process $\{W_t\}_{t>0}$ of an initially empty M/G/1 queue by means of

$$\psi(u) = \mathbb{P}(W_t > u),$$

since the claim sizes can be considered as the service times of customers who arrive to the system according to the same Poisson process as the claims.

In risk theory, there is a constant named $\rho$, such that

$$\rho = \lambda \mathbb{E}[U],$$

which is interpreted as the average amount of claim per unit of time. It should not be confused with the corresponding quantity $\rho_p$ in queuing theory, which represents the traffic intensity of the system, even though both are calculated in the same way.

A further basic quantity in risk theory is the safety loading

$$\eta = \frac{p - \rho}{\rho},$$

i.e. $\eta$ is defined as the relative amount by which the premium rate $p$ exceeds $\rho$. Obviously, an insurance company should try to ensure that $\eta > 0$ ([6]).

When $p = 1$, $\rho$ is equivalent to $\rho_p$ and the assumption $\eta > 0$ yields $\rho < 1$. Otherwise, when $p \neq 1$, we have that $\rho_p = \frac{\rho}{p}$ and $\eta > 0$ implies $\rho < p$.

As we mentioned in the introduction, apart from the ruin probability, VaR is also widely used in real-world applications. For a fixed period of time $[0,t]$, let

$$Loss_t(x) := \mathbb{P}(\sum_{i=1}^{N_t} U_i > x),$$

where $\sum_{i=1}^{N_t} U_i$ is called aggregate loss or total loss and basically represents the loss on the portfolio in the time interval $[0, t]$. Then, the VaR, which is defined as the threshold value such that the probability of the aggregate loss to exceed this value over a given time horizon is less than a given level $q$, is actually equal to the $(1-q)$-quantile of $Loss_t(x)$.

We continue by introducing some common notation we use throughout this report. For a distribution $F$, we denote by

(i) $f(x)$, its probability density function (PDF),

(ii) $F(x)$, its cumulative distribution function (CDF),

(iii) $\bar{F}(x) = 1 - F(x)$, its complementary cumulative distribution function (CCDF),

(iv) $F^{*l}$, its $l$-th convolution,

(v) $f_{LT}(s) = \int_0^\infty e^{-sx} d(F(x))$, its Laplace transform and

(vi) $m_F$, its mean.

In addition, if the distribution has finite mean, the stationary excess distribution is defined as

$$F^e(t) = \frac{1}{m_F} \int_0^t \bar{F}(x) dx.$$

As we commented in the introduction, the claim size distribution $G$ is not known in advance. Therefore, in the following part, before we examine the ruin probability and total loss approximations, we give the definitions of various distributions and study fitting procedures and goodness-of-fit tests that we can apply to the Danish data.

# Part I
# Fitting a distribution

In the following sections, we give the definitions of the distributions we use in this study and we consider fitting techniques for each distribution. More specifically, in section 3, we define the heavy-tailed distributions and two main subclasses of them. Then, the definitions of the Pareto, Log normal and Weibull distributions, and techniques to estimate their parameters based on a data set are given. In section 3.4, we study different goodness-of-fit tests which help us determine which of the heavy-tailed distributions is the most plausible one for our data. Similarly, in section 4, the definition of the phase-type distributions is provided, and several algorithms to approximate a non-negative distribution with a phase-type one are described. Moreover, in section 4.3, we discuss how we can choose how many phases our phase-type distribution needs in order to obtain an accurate approximation of the real distribution. Finally, in section 5, we sketch the procedure which fits a Pareto distribution on the tail of a data set. Moreover, the corresponding goodness-of-fit test is given which could be used to answer the question of the plausibility for the tail of this data set to follow a Pareto distribution.

## 3 Heavy-tailed distribution

In simple words, a distribution is heavy-tailed if its tail is heavier than the tail of an Exponential distribution. Mathematically, a distribution $P$ is said to be heavy-tailed if $\bar{P}(x) > 0$ for all $x \geq 0$, and

$$\lim_{x\to\infty} \frac{\bar{P}(x+y)}{\bar{P}(x)} = 1, \quad \forall y \geq 0.$$

Since the class of the heavy-tailed distributions is quite large, usually one works with a subclass of it. A well known subclass is called subexponential. A distribution $P$ belongs to this subclass if $\bar{P}(x) > 0$ for all $x \geq 0$, and it holds that

$$\lim_{x\to\infty} \frac{\bar{P}^{*m}(x)}{\bar{P}(x)} = m, \quad \forall m \geq 2.$$

Additionally, a subclass of the subexponential class, and hence of the heavy-tailed distributions, is the class of the regularly varying distributions. A distribution $P$ is regularly varying if

$$\bar{P}(x) = x^{-\alpha} L(x),$$

where $L(x)$ is slowly varying, i.e. we have

$$\lim_{x\to\infty} \frac{L(xy)}{L(x)} = 1, \quad \forall y > 0.$$

For more information regarding the heavy-tailed distributions the reader is referred to [35] and the references therein.

In the literature, the heavy-tailed distributions used most often to model the claim sizes are the Pareto, Lognormal and Weibull distributions ([6], [18]). Thus, in the rest of this section, the definitions of these three distributions are given, along with different techniques to estimate their parameters given a data set. Let $Data = \{d_1, d_2, ..., d_n\}$ be the observed data set and $n$ its size.

A common fitting method is the Maximum Likelihood Estimator (MLE) since it can provide unbiased, consistent and efficient estimators of the true parameters of the distribution, when the sample size is bigger than 30. Given a set of observed data and an underlying statistical model, the MLE method selects the set of values for the parameters of the statistical model that maximizes the likelihood function.

## 3.1 Pareto distribution

The Pareto distribution, also known as the *power law* distribution, is a regularly varying distribution, and its PDF is given by

$$p(x) = \frac{\alpha - 1}{\beta} \left( \frac{x}{\beta} \right)^{-\alpha}, \quad \text{for } x \geq \beta,$$

where $\alpha$ is called the shape parameter, and $\beta > 0$ is some lower bound to the power law behaviour. Usually it holds that $2 < \alpha < 3$ ([11]), and here we assume that $\alpha > 2$. The CCDF of the Pareto distribution is then

$$\bar{P}(x) = \left( \frac{x}{\beta} \right)^{-\alpha+1}, \quad \text{for } x \geq \beta,$$

and its mean is equal to

$$m_P = \frac{(\alpha - 1)\beta}{\alpha - 2}.$$

### 3.1.1 Estimating the parameters of the Pareto distribution

When the purpose is to fit a Pareto distribution to the observed data set $Data$, first, we set the lower bound of the distribution: $\beta = \min_{i=1,..,n} d_i$. Hence, we only need to estimate the shape parameter $\alpha$. In the literature, there are several methods to estimate $\alpha$, so in the rest of this section we give a brief survey of the available methods, and we propose some variations as well.

1. Using MLE ([24]). When the underlying statistical model is the Pareto distribution, the MLE for the shape parameter, $\alpha$, is:

$$\alpha_{MLE} = \frac{n}{\sum_{i=1}^{n} \log\left(d_i/\beta\right)} + 1.$$

However, the MLE method tries to adjust $\alpha$ in order to achieve a good fit in the center of the distribution, without caring a lot about the tail where there are fewer observations. In contrast, the Hill estimator, which we explain below and can be used when the distribution is regularly varying, is designed especially to achieve a better fit in the tail.

2. Using Hill Plot. Initially the observations are ordered such that $d_{(1)} \geq d_{(2)} \geq ... \geq d_{(n)}$, where $d_{(i)}$ represents the $i$-th largest value in the data set. Then

$$\alpha_{Hill}(k) = \frac{k-1}{\sum_{i=1}^{k-1} \log\left(d_{(i)}/d_{(k)}\right)} + 1, \quad k = 2, ..., n,$$

is called the Hill estimator based on the $k$ upper order statistic ([21]). Notice that for $k = n$, the Hill estimator is equivalent to the MLE. Obviously, the Hill estimator depends on the choice of $k$, however there is little guidance available for choosing $k$ in the case of real data (see [15]).

Therefore, in practice, a plot of the points $(k, \alpha_{Hill}(k))$, with $2 \leq k \leq n$, is constructed. This plot is called Hill Plot ([15]). The value of the estimate $\alpha_{Hill}$ is inferred from the stable region in that graph. The Hill plot is quite effective when the underlying distribution is Pareto or quite close to Pareto. When another distribution is considered, the Hill plot might not be so informative and an alternative plot can be constructed, like the one proposed in [31]. The issue of this method is that a point is picked from the stable region of the plot manually and this could affect the precision of the estimation.

3. Using rank-frequency plots. If the observed data follows a Pareto distribution, plotting the points $(d_k, \bar{P}(d_k))$ on a doubly logarithmic scale, results in a graph which shows a straight line since

$$\begin{aligned} \bar{P}(x) &= \left(\frac{x}{\beta}\right)^{-\alpha+1} \\ \log \bar{P}(x) &= \log\left(\frac{x}{\beta}\right)^{-\alpha+1} \\ \log \bar{P}(x) &= (1-\alpha)\log x + (\alpha - 1)\log \beta. \end{aligned}$$

Notice that the straight line on a doubly logarithmic scale is a necessary condition but not a sufficient one, since data coming from a different distribution could mimic this behaviour (see [11] and [30]).

Based on the rank-frequency plots there are two possible ways to estimate the shape parameter:

(a) By performing least-squares linear regression on the logarithms of the points, i.e. on the points $(\log d_i, \log \bar{P}(d_i))$, where the points $\log \bar{P}(d_i)$ are the dependent

variables. Using this method, the slope of the straight line can be found ([24]). Let $sl_{ccdf}$ denote the estimated slope, then for our estimate $\alpha_{ccdf,l}$ we have that

$$1 - \alpha_{ccdf,l} = sl_{ccdf} \Rightarrow \alpha_{ccdf,l} = 1 - sl_{ccdf}.$$

Notice that since by assumption $\alpha > 2 \Rightarrow 1 - \alpha < 0$, $sl_{ccdf}$ is expected to be a negative number.

(b) By using a variation of the previous one. Instead of performing a linear regression on the logarithmic points, we can perform a non-linear regression on the actual points. The function $g_{ccdf}(x) = \delta_{ccdf} x^{\gamma_{ccdf}}$ is used as the model function, since the CCDF of Pareto has basically this form. Then, for the estimated shape parameter $\alpha_{ccdf,nl}$ we have

$$1 - \alpha_{ccdf,nl} = \gamma_{ccdf} \Rightarrow \alpha_{ccdf,nl} = 1 - \gamma_{ccdf}.$$

Similar to the previous method, we expect that $\gamma_{ccdf} < 0$.

These methods, as well as other variations on the same theme like the ones we describe later, can generate significant systematic errors. Moreover, since the assumptions of the regression do not hold, it is hard to estimate these errors. For a more detailed explanation on this problem and other problems related to these methods, the reader is referred to [11] and the references therein.

4. Using histograms. Construct a histogram using bins of constant linear or logarithmic width which represents the density distribution of the quantity of interest. Let $bins$ denote the number of bins used in the histogram and $x_i, i = 0, ..., bins$, where $x_0 = 0$, be the bin limits. Then, for a histogram with constant linear width holds that $x_{i+1} - x_i = w$ for all $i = 0, ..., bins - 1$, while, when constant logarithmic width is used, it holds that $\log x_{i+1} - \log x_i = lw$ for all $i = 0, ..., bins - 1$.

In case of heavy-tailed distributions, an advantage of having constant logarithmic width, is the reduction of the number of zero and low-count bins at larger values of $x$. This occurs because the linear width, $w_i$, of bin $i$ increases linearly with $x$ since it holds

$$\log x_{i+1} = \log x_i + lw \Rightarrow x_{i+1} = x_i e^{lw},$$

which yields

$$w_i = x_{i+1} - x_i \Rightarrow w_i = x_i(e^{lw} - 1).$$

This linear dependency of the linear width with $x$ yields that the number of observations lying in each bin is not determined only by $x$ but also by the linear width. Therefore, when the histogram using constant logarithmic width represents the number of counts lying in each bin or the frequency distribution instead of the density, an estimate of the quantity $1 - \alpha$ is provided (for the proof the reader is referred to [10] and [39]). This, however, does not occur when the bins have constant linear width. In other words, an estimate of $\alpha$ is given regardless of what the histogram represents.

If the data follows a Pareto distribution, then by plotting the points $(mean_i, Den_i)$ for $i = 1, ..., bins$ where $mean_i = \frac{x_{i-1}+x_i}{2}$ and $Den_i$ is the density value corresponding to bin $i$, on a doubly logarithmic scale, as in (3), this graph shows a straight line regardless of the type of histogram and the number of bins used. Thus, the same methods used in (3), can be applied to the points $(mean_i, Den_i), i = 1, ..., bins$ excluding the zero counts bins. So, we have:

(a) The shape parameter is equal to the absolute value of the slope, which can be found by performing least-squares linear regression on the logarithms of the points $(mean_i, Den_i)$, i.e.

$$\alpha_{hist,l} = |sl_{hist}|,$$

where $sl_{hist}$ denotes the estimated slope, and it holds that $sl_{hist} < 0$.

(b) Perform non-linear regression on the points $(mean_i, Den_i)$, using the function $g_{hist}(x) = \delta_{hist} x^{\gamma_{hist}}$ as model function. Then

$$\alpha_{hist,nl} = |\gamma_{hist}|.$$

Once again, we expect $\gamma_{hist} < 0$.

Aside from the problems related to the regression, which we mentioned before, an additional issue of the last procedures is the choice of the number of bins, since it can influence the estimate of the shape parameter (see [11]). However, there is no optimal method for choosing the number of bins. Therefore the following formulas can be applied:

- Square-root choice: $bins = \lceil \sqrt{n} \rceil$,
- Sturges' formula: $bins = \lceil \log_2 n + 1 \rceil$ and
- Rice Rule: $bins = \lceil \frac{1}{2} n^{\frac{1}{3}} \rceil$, which is basically a simple alternative to Sturges' formula,

where $\lceil x \rceil$ returns the smallest integer greater than or equal to $x$.

5. Using a plot of the points $(bins, \alpha_{bins})$. Since there is no right choice for how many bins to use in a histogram and inspired by the idea of Hill Plot, in this method we construct a plot of the points $(bins, \alpha_{bins})$, for $bins = 2, ..., B$, where $\alpha_{bins}$ is the estimated $\alpha$ using $bins$ number of bins (the choice of $B$ is based on the size of the sample and the range of the values). Then, we check if there is a point $(bins, \alpha_{bins})$ in the graph for which the value of $\alpha_{bins}$ becomes stable. If $\alpha^*_{bins}$ is such a value, we set

$$\alpha_H = \alpha^*_{bins}.$$

As mentioned earlier, there are two possible ways to construct a histogram (with constant linear or logarithmic width), and two methods to estimate the shape parameter (by linear or non-linear regression), thus the method described above is applied for each kind of histogram using both of the estimation methods for the shape parameter.

## 3.2 Log normal distribution

The Log normal distribution has PDF

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \quad \text{for } x \geq 0,$$

where $\mu$ and $\sigma$ are, respectively, the mean and standard deviation of the variable's natural logarithm. In other words, if $X \sim$ Log normal $(\mu, \sigma^2)$, then $Y = \log X \sim N(\mu, \sigma^2)$. The CDF of the Log normal distribution is

$$P(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad \text{for } x \geq 0,$$

where $\Phi(x)$ is the CDF of the standard Normal distribution. Finally its mean is equal to

$$m_P = e^{(\mu + \frac{1}{2}\sigma^2)}.$$

### 3.2.1 Estimating the parameters of the Log normal distribution

Given the observed data set *Data* and Log normal distribution as the underlying statistical model, the following fitting procedures can be used to estimate the true parameters of the model.

1. Using MLE ([24]). In order to estimate the parameters $\mu$ and $\sigma^2$ of the Log normal distribution we can use the values $\log d_i$ and thus the problem is reduced to that of estimation of the parameters of the Normal distribution. Hence for the parameters of the Log normal distribution, estimated according to the MLE method, we have that:

   $$\mu_{MLE} = \frac{\sum_{i=1}^{n} \log d_i}{n},$$

   and

   $$\sigma^2_{MLE} = \frac{\sum_{i=1}^{n} (\log d_i - \mu_{MLE})^2}{n}.$$

2. Using graphical estimation ([24]). This method is based on the observation that the tangents at two points of inflexion of the PDF remain close to the curve for a considerable length. So one has to calculate the ratio of the observed slopes of the inflection tangents and then use the for this purpose available tables to estimate $\sigma$. Finally, $\mu$ can be estimated using the formula

   $$\mu_{graph} = \frac{3}{2}\sigma^2_{graph} + \frac{1}{2}\log x_1 x_2,$$

   where $x_1$ and $x_2$ are the points of the $x$-axis where the tangents cut the horizontal axis and $\sigma_{graph}$ is the estimated $\sigma$.

## 3.3 Weibull distribution

The PDF of the Weibull distribution is

$$p(x) = \frac{\alpha_w}{\beta_w}\left(\frac{x}{\beta_w}\right)^{\alpha_w - 1} e^{-\left(\frac{x}{\beta_w}\right)^{\alpha_w}}, \quad \text{for } x \geq 0,$$

where $\alpha_w$ and $\beta_w$ are called the shape and scale parameters of the Weibull distribution respectively. The CCDF of the distribution is given by

$$\bar{P}(x) = e^{-\left(\frac{x}{\beta_w}\right)^{\alpha_w}}, \quad \text{for } x \geq 0,$$

which implies that if the random variable $X$ follows a Weibull distribution, then $Y = \left(\frac{X}{\beta_w}\right)^{\alpha_w}$ has the Exponential distribution with PDF $p_Y(y) = e^{-y}$ for $y > 0$. The Weibull distribution is a heavy-tailed distribution if the shape parameter is less than one, i.e. if $\alpha_w < 1$. Finally the mean of the Weibull distribution is

$$m_P = \beta_w \Gamma\left(1 + \frac{1}{\alpha_w}\right),$$

where $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx$, for $s > 0$ is the Gamma function.

### 3.3.1 Estimating the parameters of the Weibull distribution

Using one of the following methods on the data set *Data*, the true parameters of the Weibull distribution can be estimated.

1. Using MLE ([24]). The MLE of the shape parameter of the Weibull distribution is:

$$\alpha_{w,MLE} = \left[\left(\sum_{i=1}^{n} d_i^{\alpha_{w,MLE}} \log d_i\right)\left(\sum_{i=1}^{n} d_i^{\alpha_{w,MLE}}\right)^{-1} - n^{-1}\sum_{i=1}^{n}\log d_i\right]^{-1},$$

   which is an implicit equation that in general must be solved numerically. Then for the scale parameter the MLE is:

$$\beta_{w,MLE} = \left[n^{-1}\sum_{i=1}^{n} d_i^{\alpha_{w,MLE}}\right]^{\frac{1}{\alpha_{w,MLE}}}.$$

2. Using the Method of Moments ([24]). Because of the fact that the coefficient of variation of the Weibull distribution depends only on the shape parameter $\alpha_w$, we can determine $\alpha_w$ numerically. The coefficient of variation is given by the following formula

$$CV_w(\alpha_w) = \frac{\sqrt{\Gamma\left(1 + \frac{2}{\alpha_w}\right) + \left[\Gamma\left(1 + \frac{1}{\alpha_w}\right)\right]^2}}{\Gamma\left(1 + \frac{1}{\alpha_w}\right)},$$

and so a table with $CV_w(\alpha_w)$ values can be formed for various values of the shape parameter. Then, we need to calculate the coefficient of variation of the data on hand and compare this value with the values in the table. The corresponding $\alpha_w$ is the estimated one. Let $\alpha_{w,mom}$ denote the estimated shape parameter of the Weibull distribution using the Method of Moments. The scale parameter can be determined by the following formula

$$\beta_{w,mom} = \frac{\sum_{i=1}^{n} d_i}{n\Gamma\left(1 + \frac{1}{\alpha_{w,mom}}\right)}.$$

## 3.4 Choosing a heavy-tailed distribution

As we saw in the previous sections, we can fit any distribution on our data, regardless of the true distribution. However, the fitting procedures tell us nothing about how good the fittings are. Hence, after fitting a distribution, a goodness-of-fit test is needed in order to check if it is plausible that our data is drawn from this probability distribution. In other words, we test the null hypothesis

$$H_0 : \text{the true distribution of the data set is distribution } P,$$

against the alternative

$$H_1 : \text{the true distribution of the data set is not distribution } P.$$

There are two types of tools that we use to assess whether our sample could come from a specified distribution: the quadratic EDF statistics (i.e. tests which are based on the empirical distribution function) and graphical techniques. In the rest of this section we describe goodness-of-fit tests of both types which we use in this research.

### 3.4.1 Quadratic EDF statistics

In this study, we use the Anderson-Darling and Cramér-von Mises statistics, which are both quadratic EDF statistics. These tests measure the distance between the empirical distribution $F$ of a sample of size $n$ and the hypothesized distribution $P$ by

$$n \int_{-\infty}^{+\infty} (F(x) - P(x))^2 w(x) dP(x),$$

where $w(x)$ is a suitable function which gives weights to the squared difference $(F(x)-P(x))^2$. When the weighting function is equal to 1, the statistic

$$W^2 := n \int_{-\infty}^{+\infty} (F(x) - P(x))^2 dP(x),$$

is called Cramér-von Mises statistic. The Anderson-Darling statistic is defined as

$$A^2 := n \int_{-\infty}^{+\infty} \frac{(F(x) - P(x))^2}{P(x)(1 - P(x))} dP(x),$$

and it is obtained when the weighting function is $w(x) = [P(x)(1-P(x))]^{-1}$. Thus, compared to the Cramér-von Mises distance, the Anderson-Darling distance places more weight on observations in the tail of the distribution.

Once the EDF test statistics are computed, the associated probability values ($p-$values) can be found. The hypothesis that the data are drawn from $P$ is rejected with a given level of significance. An advantage of these tests is that they are consistent under any alternative. This means that if the true distribution is not $P$, we will eventually reject the null hypothesis as $n \to \infty$, no matter what the true distribution is. For more information we refer the reader to [12] and the references therein.

### 3.4.2 QQ-plot

The QQ-plot is a graphical technique for determining the validity of a distributional assumption for a data set. It plots the quantiles of the data set against the quantiles of the distribution in question. It can also be used to check if two data sets come from populations with a common distribution.

In each case, a reference line for which holds $y = x$ is also plotted. If the null hypothesis is correct, in other words if the data set follows that specific distribution, then the points of the QQ-plot should fall approximately along this reference line. In case the alternative hypothesis is valid, there is a departure of the points from this reference line.

In the following section, we continue with the definition and some characteristics of the phase-type distributions, which belong to the second category of distributions we use to model the claim size distribution.

# 4 Phase-type distributions

Let $\{J(t)\}_{t\geq 0}$ be a Markov process on the finite state-space $\{0, 1, ..., ph\}$, where states $1, ..., ph$ are transient and state 0 is absorbing. Then, $\{J_t\}_{t\geq 0}$ has an intensity matrix of the form

$$Q = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\mathbf{T}$ is a $ph \times ph$ dimensional matrix called the phase-type generator, $\mathbf{t}$ is a $ph$ dimensional column vector called exit vector and $\mathbf{0}$ is the $ph$ dimensional row vector of zeros. Since each row of $Q$ sums to zero, it holds that $\mathbf{t} = -\mathbf{Te}$, where $\mathbf{e}$ is a $ph$ dimensional column vector of ones. Let $\boldsymbol{\pi} = (\pi_1, ..., \pi_{ph})$ denote the initial distribution, where $\pi_i = \mathbb{P}[J_0 = i]$ for all $i = 1, ..., ph$. By assumption the probability of starting in the absorbing state is equal to zero, i.e. $\pi_0 = \mathbb{P}[J_0 = 0] = 0$.

Then, the time until absorption $X = \inf\{t \geq 0 | J_t = 0\}$ has a phase-type distribution. The dimension of $\boldsymbol{\pi}$ is the dimension of the phase-type distribution and the pair $(\boldsymbol{\pi}, \mathbf{T})$ is a representation of the phase-type distribution.

Some basic characteristics of a phase-type distribution are:

(i) the PDF $p(x) = \boldsymbol{\pi} \exp\{\mathbf{T}x\}\mathbf{t}$,

(ii) the CDF $P(x) = 1 - \boldsymbol{\pi} \exp\{\mathbf{T}x\}\mathbf{e}$ and

(iii) the Laplace transform $p_{lt}(s) = \int_0^\infty e^{-sx} d(P(x)) = \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{t}$.

For a more detailed description of the phase-type distributions and their characteristics, the reader is referred to [7], [32] and the references therein. In this rest of this section, the definitions of two special cases of phase-type distributions are given. Additionally, we provide a brief survey of different algorithms to approximate a data set or a distribution with a phase-type one. Finally, in section 4.3, a procedure to choose how many phases are needed is described.

## 4.1 Special cases of phase-type distributions

Two of the best known phase-type distributions are the Hyperexponential and the Coxian distributions. Asmussen et al. in [7] establish that all the phase-type distributions corresponding to acyclic distributions coincide with the Coxian distributions. This implies that a Coxian distribution provides almost always a fit as good as a general phase-type distribution. Moreover, Feldmann et al. proved in [19] that any distribution with completely monotone density can be approximated by a Hyperexponential distribution. In this section the structure of these two types of phase-type distributions is given.

### 4.1.1 Hyperexponential distribution

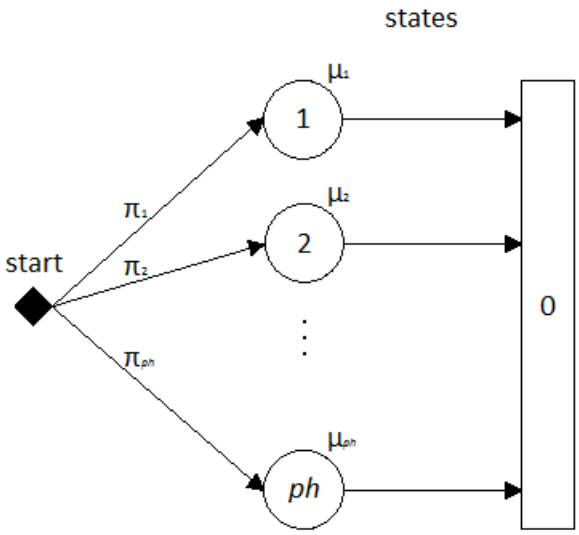
A Hyperexponential distribution, i.e. a mixture of exponentials, represents a Markov process which may start in any state and gets absorbed without visiting any other state. This means that $\boldsymbol{\pi} = (\pi_1, ..., \pi_{ph})$ with $\sum_{i=1}^{ph} \pi_i = 1$ and $\mathbf{T}$ is written as

$$\mathbf{T} = \begin{pmatrix} -\mu_1 & 0 & 0 & \dots & 0 \\ 0 & -\mu_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\mu_{ph} \end{pmatrix},$$

where $\mu_i > 0$ is the rate of state $i, i = 1, ..., ph$, i.e. the Markov process spends on average $\frac{1}{\mu_i}$ time in state $i$. An illustration of a Hyperexponential distribution with $ph$ phases can be found in figure 1. Since, the Hyperexponential distribution can be defined by $\boldsymbol{\pi}$ and $\mu_i, i = 1, ..., ph$, we denote this distribution as

$$\text{Hyperexponential}[(\pi_1, ..., \pi_{ph}), (\mu_1, ..., \mu_{ph})].$$

Figure 1: Representation of the Hyperexponential distribution.




### 4.1.2 Coxian distribution

A Coxian distribution describes a Markov process which starts in the first state, i.e. we have $\boldsymbol{\pi} = (1, 0, ..., 0)$. Moreover, from a state $i$, it is allowed to jump to state $i+1$ with probability $tr_i$ or to the absorption state with probability $1 - tr_i$ for each $i = 1, ..., ph - 1$ while from state $ph$, it can only be absorbed. Let $\mu_i$ for $i = 1, ..., ph$ be the rates of each state $i$, then

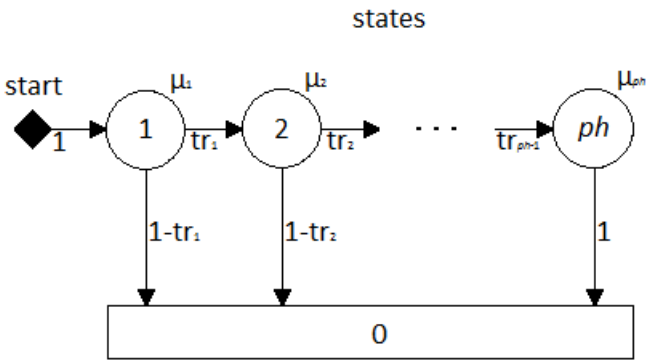
$$\mathbf{T} = \begin{pmatrix} -\mu_1 & tr_1\mu_1 & 0 & \dots & 0 & 0 \\ 0 & -\mu_2 & tr_2\mu_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\mu_{ph-1} & tr_{ph-1}\mu_{ph-1} \\ 0 & 0 & 0 & \dots & 0 & -\mu_{ph} \end{pmatrix}.$$

When $tr_i = 1$ for all $i = 1, ..., ph$ the distribution is called Hypoexponential distribution (or generalized Erlang). If, in addition, $\mu_i = \mu$ for all $i = 1, ..., ph$ then the distribution is called Erlang. A diagram of a Coxian distribution with $ph$ phases can be found in figure 2. We use the notation

$$\text{Coxian}[(tr_1, ..., tr_{ph-1}), (\mu_1, ..., \mu_{ph})],$$

to describe a Coxian distribution, since it provides us with the necessary information in order to construct its phase-type generator matrix $\mathbf{T}$.

Figure 2: Representation of the Coxian distribution.




## 4.2 Fitting phase-type distributions

There are various fitting techniques for phase-type distributions available based on maximum likelihood estimators. Some of these methods can be applied when an approximation of a distribution with a phase-type one is required ([19], [22] and [36]). On the other hand, the techniques in [25] and [32] can be applied for fitting a phase-type distributions to a data set. Finally, the proposed method in [7] can be used for both data sets and distributions. A brief summary of three articles, one from each category, is given hereafter.

### 4.2.1 EM algorithm

The EM algorithm, proposed in [7], is an iterative maximum likelihood method whose goal is to estimate the parameters of the phase-type distributions, $\boldsymbol{\pi}$ and $\mathbf{T}$. The EM algorithm starts with an initial guess for these parameters and in each iteration, a new estimate for the pair $(\boldsymbol{\pi}, \mathbf{T})$ is produced such that the value of the Likelihood function using these new parameters is greater than the previous ones. Ultimately, after some iterations, the sequence of the estimates $(\boldsymbol{\pi}, \mathbf{T})$ should converge to a stationary point which maximizes the Likelihood function. Since there is no guarantee that the derived solution is optimal, the authors suggest to try different initial values in order to avoid local maxima and saddle points. It is worth mentioning that the EM algorithm allows to fit any phase-type distribution to any non-negative distribution or data set. Additionally, the number of phases, $ph$, is considered as an input parameter of the EM algorithm. Finally, the mean of the fitted phase-type distribution is equal to the mean of the distribution or the data set. An implementation of the EM algorithm, called the EMpht-program, is given by M. Olsson and can be found in [27].

### 4.2.2 An EM-based technique

Horváth et al. in [22] notice that in case of long-tailed data sets[1], the EM algorithm may not be able to capture the tail correctly since it searches for a global maximum. Based on this observation, Riska et al. in [32] propose a divide-and-conquer technique to fit a data set into a mixture of Erlang and Hyperexponential distributions if the data set has a non-monotone density and into a Hyperexponential distribution if the data set has a completely monotone density. More specifically, the data set is divided into smaller sets based on its histogram such that each partition has coefficient of variation greater than a certain value. Then, using the EMpht-programme, an Erlang or a Hyperexponential distribution is fitted to each partition. The final result is generated by properly combining the fitted distributions. Note that, even though the number of phases used for each partition is a known parameter of the algorithm, the number of phases, $ph$, of the final result is not known a priori.

### 4.2.3 FW algorithm

The algorithm proposed by Feldmann and Whitt in [19] is a recursive heuristic approach used for approximating a completely monotone non-negative distribution with a Hyperexponential distribution. The algorithm fits the most right portion of the tail of the distribution into a single Exponential distribution which basically represents a phase of the Hyperexponential distribution. In addition, it calculates the probability of entering this particular phase. Then, the weight of the fitted portion is subtracted from the distribution and the process is repeated. The algorithm terminates when the parameters of all the phases of the Hyperexponential distribution are known. Similar to the EM algorithm, the number of phases, $ph$, is an input parameter of the algorithm.

### 4.2.4 An alternative EM-based technique

In case the data set cannot be partitioned based on the criteria proposed by Riska et al. in [32] an alternative technique is required. For this purpose, the data set is partitioned into two parts based on [11] (a brief description is given in section 5). One part includes the so-called heavy-tailed claims, while the other part contains the rest of the observations. For each part, the EMpht-programm is used for fitting a Coxian or a Hyperexponential distribution. The final result is obtained by merging the fitted distributions properly using the weight of each partition. Note that in contrast to the algorithm in [32], the number of phases used for each partition does not have to be the same and their sum gives the total number of phases, $ph$. Moreover, it is not necessary to fit the same type of phase-type distribution to each part. For example, a Coxian distribution with 3 phases can be used for one part and a Hyperexponential distribution with 2 phases for the other part.

Thus, in short, the procedure consists of the following basic steps:

1. Apply the method proposed in [11].

[1] A data set is called long-tailed when it is plausible to fit a heavy-tailed distribution on it.

2. Fit a Coxian or a Hyperexponential on each part.

3. Combine the two fitted distributions to obtain the final result.

### 4.3 Choosing the number of phases

All the algorithms described in section 4.2 require as an input parameter the number of phases, $ph$. Therefore, the fitting is done in a trial-and-error fashion starting with one phase. The process continues by adding one phase each time until a certain degree of fitness is obtained. The purpose is to derive the most parsimonious model for the data. In other words, we search for a model which can approximate the original distribution of the data and, at the same time, should not be too complex. Asmussen et al. in [7] proposed the use of Akaike Information Criterion ($AIC$) for this purpose. The reason is that $AIC$ deals with the trade-off between the goodness-of-fit of the model and its complexity. However, in this research, the main criterion chosen for model selection is the corrected Akaike Information Criterion or $AICc$. Basically, $AICc$ is equal to $AIC$ with a greater penalty to the models with more parameters (for more information regarding $AIC$ and $AICc$ the reader is referred to [2] and [23] and the references therein).

The formulas of $AIC$ and $AICc$ are:

$$AIC = 2k - 2logLik,$$

and

$$AICc = AIC + \frac{2k(k+1)}{n-k-1},$$

where $n$ is the size of the sample, $k$ is the number of parameters of the fitted model and $logLik$ the value of the log-Likelihood function. When a Hyperexponential distribution is used, $k = 2ph$ while using a Coxian distribution implies $k = ph(ph-1)$.

Besides $AICc$, the log-Likelihood value ($logLik$) is also used as an auxiliary goodness-of-fit score. Hence, the preferable model is the one that maximizes the $logLik$ while it minimizes $AICc$. Finally, the plot of the CDF of the empirical distribution of the data set and the CDF of the fitted distribution is used as a visual criterion of choosing how many phases are required.

In the following section, we describe the distributions in the last category of distributions we use to model the claim size distribution. In addition, we briefly explain the corresponding fitting procedure and goodness-of-fit test.

# 5 Mixture model: Light-tailed and heavy-tailed distribution

Let $Data = \{d_1, d_2, ..., d_n\}$ be the observed data set of which we would like to estimate its distribution. The methods described in section 3.1 fit a Pareto distribution to all observed

values. They set the lower bound $\beta$ of the Pareto distribution equal to the minimum observed value in *Data*. However, it is possible that only the tail of the true distribution follows a Pareto distribution. This implies that the lower bound of the Pareto distribution can be greater than $\min_{i=1,\dots,n} d_i$ and that the observed values smaller than $\beta$ come from a population with a different distribution. Hence, the true distribution of the data set has the form

$$F(x) = (1-\epsilon)F_{lt} + \epsilon F_{ht},$$

where $F_{lt}$ is the CDF of a distribution, $F_{ht}$ is the CDF of the Pareto distribution and $\epsilon$ is the probability to obtain a heavy-tailed value, i.e. an observed value from the Pareto distribution.

In the following section we describe the method proposed by [11] for finding where the tail of the distribution starts, i.e. the lower bound of the Pareto, as well as, the shape parameter of it. Notice that, applying the following method results in the separation of the data into two subsets. One subset contains all the observed values smaller than the lower bound of the Pareto distribution and we refer to this set as the light-tailed part of the data set. The rest of the elements of *Data* belong to the second subset, namely the heavy-tailed part of the data set.

Moreover, even though we fit a Pareto distribution on the heavy-tailed part, the distribution $P_{lt}$ remains unknown. However, in case of non-negative values in the light-tailed part we can approximate its distribution with a phase-type one using the EM algorithm described in section 4.2.1.

Finally, the probability of having a heavy-tailed value $\epsilon$ is calculated by dividing the number of the observed values above the lower bound over the size of the data set.

## 5.1 Estimating the lower bound and shape parameter of the Pareto distribution

Initially, a set which contains all the unique values of *Data* is formed. Then, we try as the lower bound of the Pareto distribution, $\beta$, one by one all these values. The process continues by calculating the MLE of the shape parameter corresponding to each value

$$\alpha = 1 + n_{ab}\left[\sum_{i=1}^{n_{ab}} \log\left(\frac{d_j}{\beta}\right)\right]^{-1}, \quad d_i, \text{ such that } d_i \geq \beta,$$

where $n_{ab}$ is the number of observations that are greater than the lower bound we use. Afterwards, using the Kolmogorov-Smirnov (KS) statistic we quantify the distance between the empirical distribution of the observations with value at least $\beta$, $F_{ht}$, and the fitted Pareto distribution, denoted by $P_{ht}$,

$$D = \max_{x \geq \beta} |F_{ht}(x) - P_{ht}(x)|.$$

Having calculated the KS statistic for each element in the set that we form in the beginning, the lower bound of the Pareto distribution is that value with the minimum KS statistic value and the shape parameter is the corresponding MLE.

The next step is to apply an appropriate goodness-of-fit test to verify if it is plausible for the tail of the true distribution to follow this Pareto distribution. Therefore, in the next section, we describe a goodness-of-fit test constructed for this purpose.

## 5.2 Goodness-of-fit test

In order to test if it is plausible to model the distribution of a data set with a mixture model where one of the involved distributions is the Pareto distribution, Clauset et al. in [11] provide the corresponding goodness-of-fit test.

In this test, after finding the lower bound and the shape parameter of the Pareto distribution, we generate a large number of synthetic data sets that follow the same Pareto distribution above the lower bound and have the same distribution as the observed data below the lower bound. This can be done by simply selecting one element uniformly at random from the light-tailed part of the data set with probability $1 - \epsilon$ or by generating a random number drawn from the Pareto distribution with probability $\epsilon$. Then, for each synthetic data set, we apply the method described in the previous section and we calculate its KS statistic. The $p-$value of this goodness-of-fit test is the fraction of time the resulting statistic is larger than the value for the empirical data. Hence, the hypothesis that the tail of the true distribution is Pareto is ruled out when the $p-$value is smaller than 0.1.

Based on an analysis of the expected worst-case performance of the test, Clauset et al. in [11] propose that if we wish the $p-$value to be accurate to about 2 decimal places, then we should generate about 2500 synthetic sets.

# Part II
# Literature overview on the approximations

In this part, a survey is given of the approximations we compare in this project. As we described in the previous part, we can model the claim size distribution using a heavy-tailed, a phase-type or a mixture of a phase-type and a Pareto distribution. Based on the claim size distribution used, we have to apply the appropriate ruin probability approximation. Therefore, we describe the heavy-tailed asymptotic, the phase-type and the corrected phase-type approximations for the ruin probability. Moreover, for each approximation we give the corresponding approximation for the distribution of the total loss and thus the corresponding approximation of the VaR.

## 6 Heavy-tailed asymptotic

Oftentimes, in financial applications, it is suitable to model the claim size distribution using a heavy-tailed distribution (see [6], [16] and the references therein). However, in many cases, the Laplace transform of a heavy-tailed distribution is not available in closed form. This implies, that techniques which use the Laplace transform like the one in [1] are difficult or impossible to use in practice.

However, when the claim size distribution belongs to the subexponential class, asymptotic approximations for the ruin probability are available ([18], [28], [29]). More specifically, for the Cramér-Lundberg risk model, Embrechts and Veraverbeke in [18] prove that for a claim size distribution $G$ with finite first moment $m_G$ such that $\rho < p$, if the integrated tail distribution $G^e(t) := \frac{1}{m_G}\int_0^t \bar{G}(x)dx$ belongs to the subexponential class then

$$\psi(u) \sim \hat{\psi}_{hta}(u) := \frac{\lambda}{p-\rho}\int_u^\infty \bar{G}(x)dx, \quad \text{as } u \to \infty. \tag{6.1}$$

Notice that, due to the asymptotic character of this approximation, it is possible for $\hat{\psi}_{hta}(u)$ to give values bigger than 1 for small $u$'s. For example, let the claim size distribution be a Pareto distribution as we defined in section 3.1 with shape parameter $\alpha > 2$ then the ruin probability can be approximated by

$$\begin{aligned}
\hat{\psi}_{hta}(u) &= \frac{\lambda}{p-\rho}\int_u^\infty \left(\frac{x}{\beta}\right)^{-\alpha+1} dx \\
&= \frac{\lambda\beta^{\alpha-1}}{p-\rho}\int_u^\infty x^{-\alpha+1}dx \\
&= \frac{\lambda\beta^{\alpha-1}u^{2-\alpha}}{(p-\rho)(\alpha-2)}.
\end{aligned}$$

Solving $\hat{\psi}_{hta}(u) \leq 1$ with respect to $u$ gives

$$\begin{aligned} \hat{\psi}_{hta}(u) &\leq 1 \\ u^{2-\alpha} &\leq \frac{(\alpha-2)(p-\rho)}{\lambda\beta^{\alpha-1}} \\ u &\geq \left(\frac{\lambda\beta^{\alpha-1}}{(\alpha-2)(p-\rho)}\right)^{\frac{1}{\alpha-2}} > 0, \end{aligned} \tag{6.2}$$

since all parameters are positive by assumption and $\lambda\beta \neq 0$. Let $u_{hta} = \left(\frac{\lambda\beta^{\alpha-1}}{(\alpha-2)(p-\rho)}\right)^{\frac{1}{\alpha-2}}$, then, in words, the inequality in (6.2) means that the heavy-tailed asymptotic approximation can only provide an estimate for $u$'s greater than $u_{hta}$.

Additionally, using the same method for proving the asymptotic relationship in (6.1), Embrechts et al. (see [16]) show that if the claim size distribution $G$ belongs to the subexponential class then the aggregate claim distribution can be estimated by

$$Loss_t(x) = \mathbb{P}(\sum_{i=1}^{N_t} U_i > x) \sim \lambda t\bar{G}(x), \quad \text{as } x \to \infty. \tag{6.3}$$

Similar to the ruin probability approximation, it is expected that the approximation of $Loss_t(x), \lambda t\bar{G}(x)$, gives values bigger than 1 for small values of $x$. Finally, as we mentioned in section 2, VaR is the $(1-q)-$quatile of $Loss_t(x)$. Thus by solving the equation

$$\lambda t\bar{G}(x) = 1 - q,$$

with respect to $x$, we can obtain the heavy-tailed asymptotic approximation of VaR for a given level $q$.

# 7 Phase-type approximations

Several papers ([7], [19], [25], [36]) propose the use of a phase-type distribution for fitting the claim size distribution since otherwise the ruin probability may not be available in closed form. In other words, it tends to be easier to analyze the behaviour of the ruin probability if the claim size distribution is a phase-type one. This is the case because they have simple Laplace transforms and so they are algorithmically tractable. Moreover, any distribution on $[0, \infty)$ can be approximated arbitrarily close by a phase-type distribution since for any finite number of phases, the class of phase-type distributions is dense ([4]). This implies that one does not expect an essential difference between the actual distribution and the approximated one.

When the claim size distribution is a phase-type disrtibution and under the assumption that $\rho < p$, the approximation of the ruin probability can be obtained by the known Pollaczek-Khinchine formula (see [6]):

$$1-\psi(u)=(1-\rho_p)\sum_{n=0}^{\infty}\rho_p^n(G^e)^{*n}(u),$$

where $\rho_p=\frac{\rho}{p}$ and $(G^e)^{*n}$ represents the $n$-th convolution of the stationary excess distribution of the claim sizes. By using Laplace transforms, one gets

$$m(s):=\mathbb{E}e^{-sM}=\frac{1-\rho_p}{1-\rho_p g_{LT}^e(s)},$$

where $g_{LT}^e(s)$ is the Laplace transform of the stationary excess distribution of the claim sizes. Now, by inverting the Laplace transform $\frac{1-m(s)}{s}$, the phase-type approximation of the ruin probability, $\hat{\psi}_{ph}(u)$, can be derived.

The basic idea of this kind of approximations is to approximate the claim size distribution instead of approximating the ruin probability directly. After approximating the claim size distribution with a phase-type one, we can obtain the distribution of the aggregate claim by simulations. Afterwards, by taking the $(1-q)$-quantile of the CCDF of the aggregate claim, we can derive the corresponding phase-type approximation of VaR at level $q$.

# 8 Corrected phase-type approximations

Vatamidou et al. in [37] use a mixture model for the claim size distribution based on heavy-tailed statistical analysis which suggests that only a small fraction of the upper-order statistics of the sample is relevant for estimating tail probabilities ([21]). This implies that the distribution of the claim sizes can be written as

$$G(x)=(1-\epsilon)G_{lt}(x)+\epsilon G_{ht}(x),\quad x\geq 0,$$

where $G_{lt}$ and $G_{ht}$ are the CDF of a phase-type and a heavy-tailed distribution with means $m_{lt}$ and $m_{ht}$, respectively. Hence, a random claim is heavy-tailed with probability $\epsilon$ and light-tailed with probability $1-\epsilon$. Moreover, we assume that $m_{ht}<\infty$ and set $\delta=\lambda m_{lt}$ and $\theta=\lambda m_{ht}$ . Based on this model, two ruin probability approximations are proposed in [37].

The first approximation, called the corrected discard approximation, is

$$\hat{\psi}_{cd}(u):=\left(1-\frac{\epsilon\theta}{p-\delta+\epsilon\delta}\right)\psi^d(u)+\frac{\epsilon\theta}{p-\delta+\epsilon\delta}\mathbb{P}(M^d+M^d+H^e>u),$$

where $\psi^d(u)$ is the approximation of the exact ruin probability when we discard the heavy-tailed claim sizes (i.e. $G(x)=(1-\epsilon)G_{lt}(x)+\epsilon$), $M^d$ is the supremum of its corresponding claim surplus process and $H^e$ follows the excess distribution of $G_{ht}$. A necessary and sufficient condition in order to get a valid approximation for all values of $u$ is: $|\epsilon\theta|<|p-\delta+\epsilon\delta|$ (for more details the reader is referred to [37]).

The second approximation is the corrected replace approximation and it is defined as

$$\hat{\psi}_{cr}(u) := \left(1 - \frac{\epsilon\delta - \epsilon\theta}{p - \delta}\right)\psi^r(u) + \frac{\epsilon\theta}{p-\delta}\mathbb{P}(M^r + M^r + H^e > u) - \frac{\epsilon\delta}{p-\delta}\mathbb{P}(M^r + M^r + L^e > u),$$

where $\psi^r(u)$ is the approximation of the exact ruin probability when we replace the heavy-tailed claim sizes with phase-type ones (i.e. $\epsilon = 0$) and $M^r$ is the supremum of its corresponding claim surplus process. Finally $H^e$ and $L^e$ follow the excess distributions of $F_{ht}$ and $F_{lt}$ respectively. Similarly, a sufficient condition for this approximation to be valid for all $u$'s is: $\epsilon < \frac{|p-\delta|}{\max\{\delta,\theta\}}$.

Based on the results in [37], the corrected replace approximation tends to give better numerical estimations. However, the corrected discard approximation is simpler and always underestimates the exact ruin probability with error $O(\epsilon^2)$.

Besides the approximations for the ruin probability, the corresponding corrected discard approximation for the total loss is given. Based on this approximation, the distribution of the aggregated claim in a fixed time interval $[0, t]$ is defined as

$$Loss_t^{cd}(x) := e^{-\lambda\epsilon t}\mathbb{P}\left(\sum_{i=1}^{N_P} L_i > x\right) + (1 - e^{-\lambda\epsilon t})\mathbb{P}\left(\sum_{i=1}^{N_P} L_i + H > x\right), \tag{8.1}$$

where $L_i$ and $H$ follow, respectively, $G_{lt}$ and $G_{ht}$ and $N_P$ is a Poisson process with rate $\lambda(1-\epsilon)t$ which represents the number of phase-type claims in the time period $[0, t]$. Finally, the corrected discard approximation of VaR at level $q$ is obtained by taking the $(1 - q)$-quantile of $Loss_t(x)$ as given in (8.1).

# Part III
# Danish data

The Danish data consist of 2167 claims over one million Danish Krone (DKM) from the years 1980 to 1990 inclusive. Table 1 shows the first 10 entries of the Danish data. The average claim size per customer is 3.38509 millions, while the average claim size per day is 1.82702 millions.

In this part, we apply the techniques, described in sections 3, 4 and 5 to estimate the claim size distribution. More specifically, we fit on the claim sizes of the Danish data heavy-tailed, phase-type and mixtures of these two types of distributions. Afterwards, we apply the appropriate goodness-of-fit tests for each fitted distribution.

However, as mentioned before, we assume that the claims arrive to the insurance company according to a Poisson process. Therefore, in section 9, we first estimate the rate of that Poisson process using the arrival dates of the Danish data.

Having estimated the arrival rate and the claim size distribution, we can apply the appropriate approximation for the ruin probability, the distribution of the aggregate claim and VaR in the next part.

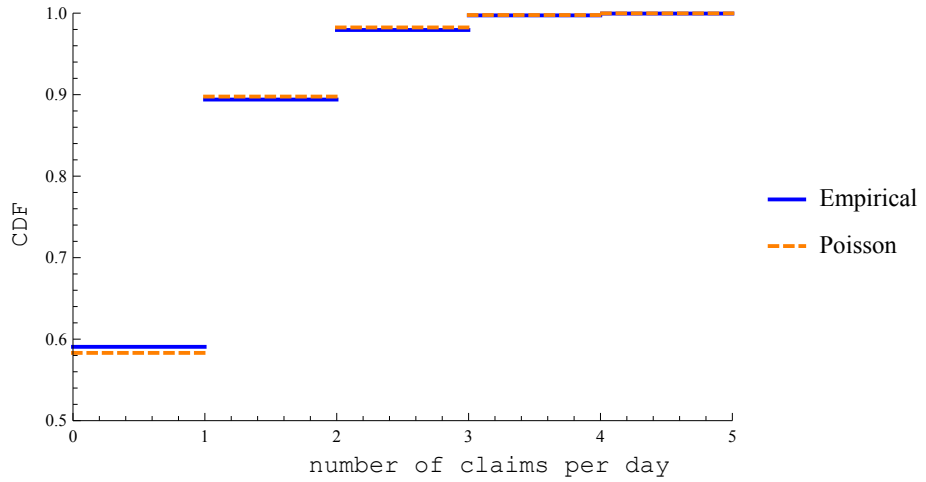
| | Arrival date | Claim size |
|---|---|---|
| 1 | 03/01/1980 | 1.683748 |
| 2 | 04/01/1980 | 2.093704 |
| 3 | 05/01/1980 | 1.732581 |
| 4 | 07/01/1980 | 1.779754 |
| 5 | 07/01/1980 | 4.612006 |
| 6 | 10/01/1980 | 8.725274 |
| 7 | 10/01/1980 | 7.898975 |
| 8 | 16/01/1980 | 2.208045 |
| 9 | 16/01/1980 | 1.486091 |
| 10 | 19/01/1980 | 2.796171 |
| ⋮ | | |

Table 1: The first 10 entries of the Danish data.

## 9 Arrival rate for the Danish data

As mentioned earlier, we assume that the claims arrive according to a Poisson process. In this section we estimate the rate of this Poisson process. Notice that in our data we only have the day of the arrival of the claims, but not the exact time of the arrival. Thus, first we have to calculate how many customers arrived during each day in the years 1980 to 1990 inclusive. Then, we compute the parameter $\lambda$ of the Poisson process using the MLE. The

Figure 3: CDF of the empirical distribution of the number of claims per day and of the Poisson process with $\lambda = 0.539323$.




result we obtain is:

$$\lambda = 0.539323.$$

The graph, in figure 3, shows the empirical distribution of the number of claims per day and the Poisson process estimating this distribution.

In order to verify the assumption that the claims arrive according to a Poisson process with rate $\lambda = 0.539323$, we use the Pearson's chi-squared test. This test is a statistical test which can be applied on categorical data. Using this test, we find that our assumption is not rejected at the 5% level since the test returns a $p-$value of 0.24134.

| | Statistic | $p-$value |
|---|---|---|
| Pearson $\chi^2$ | 4.19326 | 0.24134 |

Table 2: Results of the Pearson's chi-squared test.

# 10 Fitting a heavy-tailed distribution

In the following sections we apply the fitting procedures described in section 3 and for each fitted distribution we test if it is a plausible model for our data. In section 10.4, we choose the heavy-tailed distribution that fits the best according to the results we derive and the QQ-plots.

## 10.1 Fitting a Pareto distribution

In order to select which method can estimate the shape parameter efficiently for our data set, we perform a small experiment. We draw 25 random samples of size 2167 from a Pareto

distribution where its lower bound is $\beta = 1$. We use the values $\alpha = 2.1, 2.3, 2.5, 2.7, 2.9$ as the shape parameter of the Pareto distribution. For each of these samples, we estimate the shape parameter using all the described methods in section 3.1.1. The results of the methods and our conclusions can be found in the Appendix.

Based on the experiment, the methods we select to estimate the shape parameter of the Pareto distribution are the MLE, the non-linear regression using the CCDF points and the Hill Plot. Using these methods, we calculate the shape parameter:

$$\alpha_{MLE} = 2.27073,$$

$$\alpha_{ccdf,nl} = 2.28267,$$

and

$$\alpha_{hill} = 2.413.$$

Now, we take the average value of these three estimates as the shape parameter of the Pareto distribution to estimate the distribution of the Danish Data, hence we have

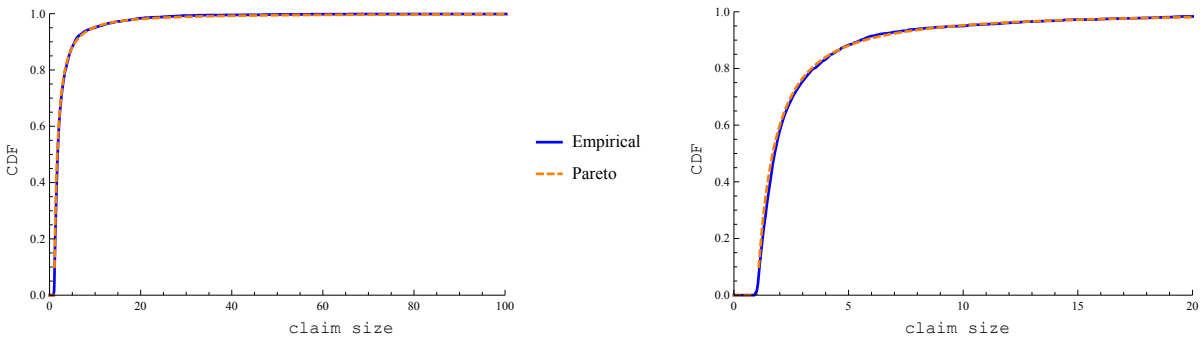
$$\alpha = 2.32213.$$

For the lower bound $\beta$ of the Pareto distribution, we take the smallest value of the Danish data, hence $\beta = 1$. When the claim size distribution is a Pareto one with $\alpha = 2.32213$ and $\beta = 1$, the mean average claim size per customer is: 4.10434.

In figure 4, one can see the empirical distribution of the claim sizes against the fitted Pareto distribution. The range of the $x$-axis in the left figure is from 0 to 100 while in the right figure the range is from 0 to 20 for better visualization of the results.

The results of the quadratic EDF goodness-of-fit tests we describe in section 3.4 are given in table 3. Based on these results the null hypothesis that the data are distributed according

Figure 4: CDF of the empirical distribution of the Danish data and the fitted Pareto distribution.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

| | Statistic | $p-$value |
|---|---|---|
| Anderson-Darling | 134.741 | 0. |
| Cramér-von Mises | 2.744 | $2.793 \times 10^{-7}$ |

Table 3: Results of the quadratic EDF tests for the Pareto distribution.

to this Pareto distribution is rejected, since the $p-$values of both tests are smaller than the significance level 0.05. This implies that the Pareto distribution is not a plausible model for the Danish data.

## 10.2 Fitting a Log normal distribution

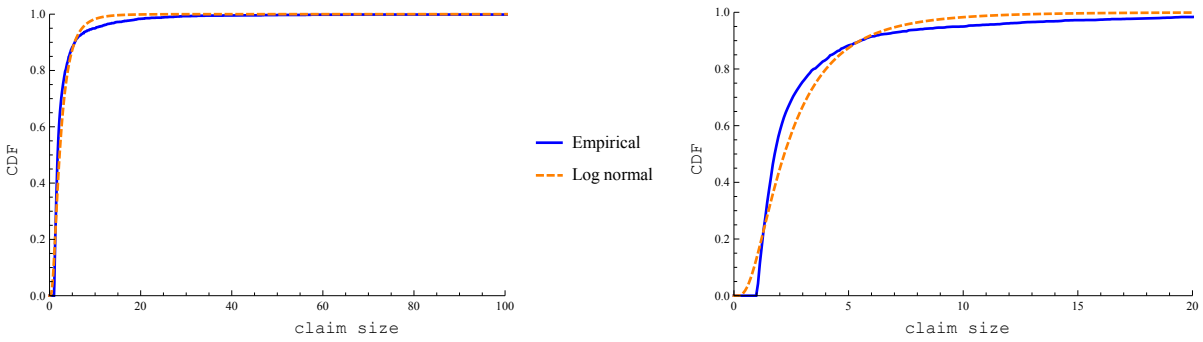
Using the MLE for the Log normal distribution we derive that

$$\mu = 0.78695 \quad \text{and} \quad \sigma^2 = 0.51369.$$

Hence, in this case, the average claim size per customer is: 2.83997. In figure 5, we can see the CDF's of the claim sizes and the Log normal distribution with parameters $\mu = 0.78695$ and $\sigma^2 = 0.513687$.

Based on the results of the quadratic EDF tests, which can be found in table 4, this Log normal distribution is a poor model for our data set, since both tests return a $p-$value significantly smaller than 0.05.

Figure 5: CDF of the empirical distribution of the Danish data and the fitted Log normal distribution.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

| | Statistic | $p-$value |
|---|---|---|
| Anderson-Darling | 87.193 | 0. |
| Cramér-von Mises | 14.791 | $1.110 \times 10^{-16}$ |

Table 4: Results of the quadratic EDF tests for the Log normal distribution.

## 10.3 Fitting a Weibull distribution

Using the MLE formulas we gave in section 3.3.1, we derive that the shape parameter of the Weibull distribution is

$$a_w = 0.95852,$$

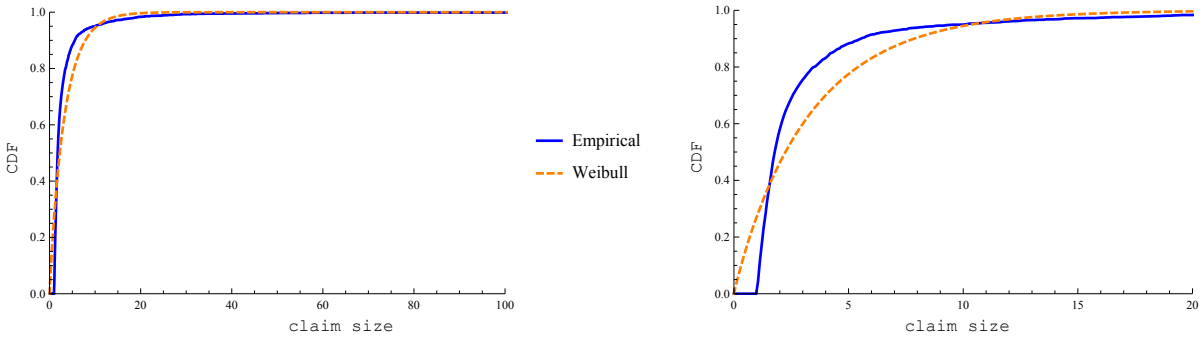
and the scale parameter is

$$b_w = 3.29075.$$

The average claim size per customer, when we use a Weibull distribution to model the claim size distribution, is: 3.35352. A plot of the CDF of the fitted Weibull distribution and the empirical distribution of the claim sizes can be found in figure 6.

However, this Weibull distribution is also rejected at 5% level of significance based on the Anderson-Darling and Cramér-von Mises tests, because in both tests the resulting $p-$values are smaller than 0.05 as shown in table 5.

Figure 6: CDF of the empirical distribution of the Danish data and the fitted Weibull distribution.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

| | Statistic | $p-$value |
|---|---|---|
| Anderson-Darling | 210.919 | 0. |
| Cramér-von Mises | 36.254 | $1.776 \times 10^{-15}$ |

Table 5: Results of the quadratic EDF tests for Weibull distribution

## 10.4 QQ-plots

As we can notice from the results in the previous sections, all the heavy-tailed distributions we fitted on the Danish data are rejected as poor models. However, to get an insight into the performance of the heavy-tailed asymptotic approximation we choose the best heavy-tailed distribution among the fitted ones.

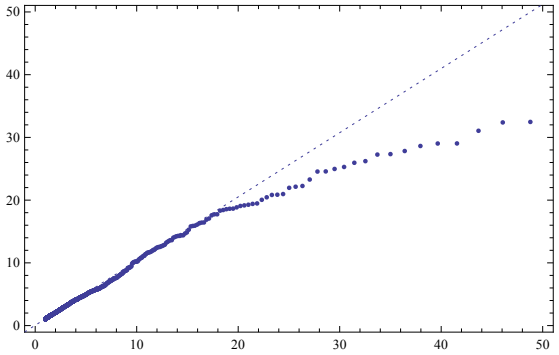
The plots 7a, 7b and 7c show the QQ-plots of the Pareto, Log normal and Weibull distribution respectively. Notice that all the QQ-plots have the same scale in both axes such that we are able to compare them. We notice that in none of these plots the points fall even approximately along the reference line.

More specifically, in case of the Pareto distribution we see that until value 18 the points form a straight line that approximately coincides with the reference line. However, for values larger than 18 the distance of the points from the reference line is increasing, implying that the tail of the Pareto distribution is heavier than the tail of the empirical distribution of the Danish data.

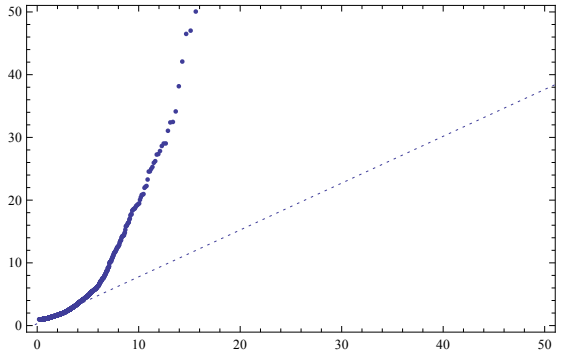
From the QQ-plot for the Log normal distribution we notice that only the points with value less than 4 fall near the reference line. Similarly, for the Weibull distribution we have an approximately straight line near the reference line for values smaller than 6. In both cases, it appears that the tail of the distribution of the claim sizes is heavier than the tails of the Log normal and Weibull distributions.

Therefore, we conclude that the best fitting distribution is the Pareto distribution with shape parameter $\alpha = 2.32213$ and lower bound $\beta = 1$.

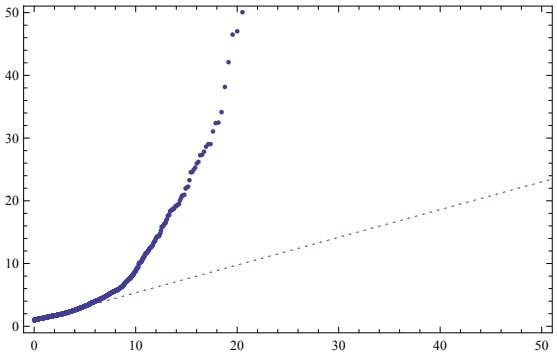
Figure 7: QQ-plots




(a) Danish data against Pareto distribution.




(b) Danish data against Log normal distribution.




(c) Danish data against Weibull distribution.

# 11 Phase-type distributions

In this section, we provide the results of applying the EM algorithm and the alternative EM-based technique we described in sections 4.2.1 and 4.2.4. More specifically, in section 11.1, the results of the EM algorithm using all the claim sizes of the Danish data can be found.

In section 12, the method in [11] is applied and thus the Danish data is split into two parts. The first part, called the light-tailed part of Danish data, includes the claims smaller than 1.37622. All the other claims in the Danish data belong to the second part, namely the heavy-tailed part of the Danish data. Following the alternative EM-based procedure, a Hyperexponential and a Coxian distribution are fitted to each part and the results of these fittings are given in sections 11.2 and 11.3. Finally, the results of these methods are given in section 11.4.

The EM-based technique proposed in [32] cannot be applied because regardless of the histogram used (constant linear or logarithmic width and number of bins) the Danish data cannot be partitioned with respect to the condition required in this paper. Moreover, the FW algorithm is not used since it is not designed to treat data. In theory, we could first fit a suitable heavy-tailed distribution to our data and then apply the FW algorithm. However, this implies that the approximation depends on the quality of the statistical model.

Note that, in this section, there is no comparison between two different types of phase-type distributions. This means, for example, that a fitted Hyperexponential distribution is not compared with a fitted Coxian distribution.

## 11.1 Using all the claims in the Danish data

In this section, the results of fitting Hyperexponential and Coxian distributions to all claim sizes of the Danish data can be found. For this fittings, we used the EMpht-program ([27]), which is an available implementation of the EM algorithm.

### 11.1.1 Fitting a Hyperexponential distribution

| phases | $logLik$ | $AICc$ |
|---|---|---|
| 1 | -4809.38 | 9622.77 |
| 2 | -4556.63 | 9121.27 |
| 3 | -4556.63 | 9125.29 |

Table 6: Results for the Hyperexponential model used to fit all Danish data.

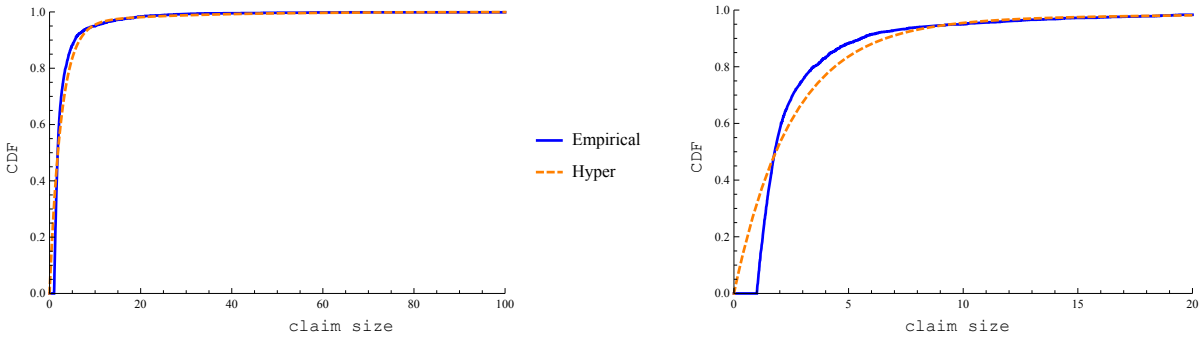
Table 6 gives the values of the $logLik$ and the $AICc$ for fitting a Hyperexponential distribution with 1, 2 and 3 phases to the Danish data. Based on these results, we conclude that a Hyperexponential distribution with 2 phases is the most appropriate choice for our data. This is the case because the Hyperexponential distribution with 1 phase (i.e. an Exponential) has greater $AICc$ and at the same time significant smaller $logLik$ value than the Hyperexponential with 2 phases. Additionally, the Hyperexponential distribution with 3

phases has bigger $AICc$ than the 2-phases Hyperexponential while there is no improvement for the $logLik$ value. The Hyperexponential distribution with 2 phases that the EMpht-program gave is:

$$\text{Hyperexponential}[(0.956892, 0.043108), (0.401224, 0.043102)].$$

The plot in figure 8 represents the CDF of the empirical distribution of the claim sizes and the CDF of the fitted Hyperexponential distribution.

Figure 8: CDF of the empirical and the chosen Hyperexponential distribution for all the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

### 11.1.2 Fitting a Coxian distribution

Table 7 provides the $logLik$ and $AICc$ values for fitting a Coxian distribution with $ph$ phases to the Danish data, where $ph = 1, ..., 11$. Using the same reasoning as in the previous section, the 10-phase Coxian distribution defined by

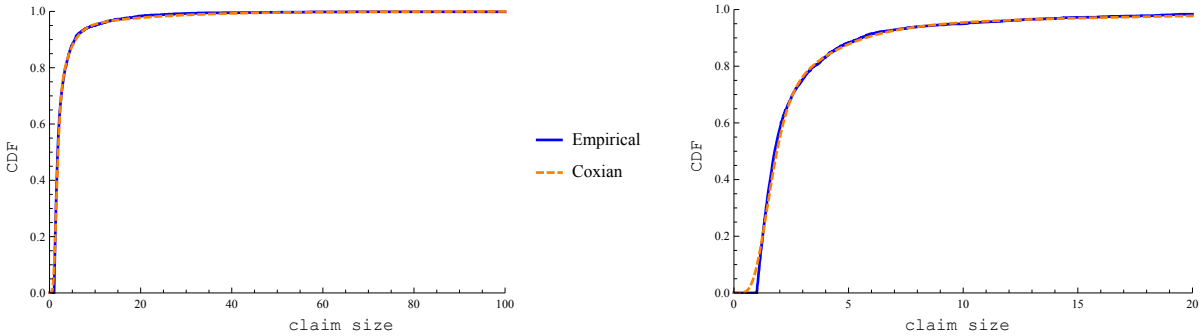
$$\text{Coxian}[(4.199325, 4.199325, 4.199325, 4.199325, 4.199325, 4.199325, 4.199325, 0.501553, 1.164039, 0.057310), (1, 1, 1, 1, 1, 1, 0.272706, 1, 0.197573)],$$

is the chosen Coxian distribution. Figure 9 shows the CDF of the empirical distribution of the Danish data and the CDF of the chosen Coxian distribution.

| phases | *logLik* | *AICc* |
|---|---|---|
| 1 | -4809.38 | 9620.77 |
| 2 | -4576.32 | 9158.64 |
| 3 | -4107.37 | 8224.78 |
| 4 | -3933.44 | 7880.93 |
| 5 | -3848.75 | 7715.58 |
| 6 | -3793.61 | 7609.35 |
| 7 | -3753.45 | 7533.08 |
| 8 | -3722.61 | 7475.44 |
| 9 | -3698.01 | 7430.31 |
| 10 | -3654.59 | 7347.53 |
| 11 | -3752.52 | 7547.48 |

Table 7: Results for the Coxian model used to fit all Danish data.

Figure 9: CDF of the empirical and the chosen Coxian distribution for all the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

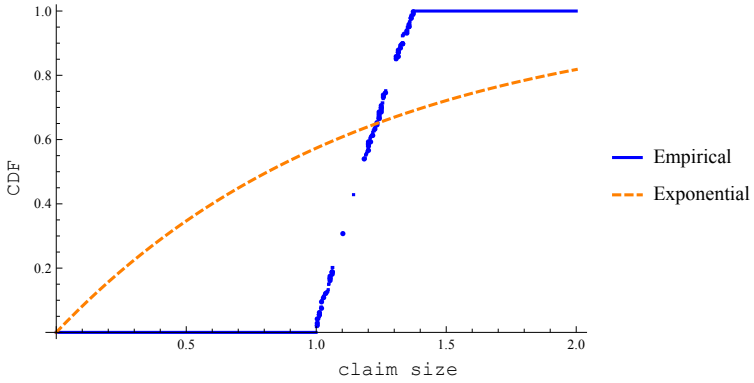
## 11.2 Using the light-tailed part of the Danish data

In this section, the EMpht program is applied to the light-tailed part of the Danish data only. In other words, we are at the second step of the alternative EM-based procedure (see 4.2.4).

### 11.2.1 Fitting a Hyperexponential distribution

Based on table 8 we conclude for the light-tailed part of the Danish data that the preferable model is the Hyperexponential distribution with 1 phase, i.e. the Exponential distribution with rate $\mu_1 = 0.851929$. Figure 10 shows the plot of the empirical distribution of the light-tailed part of the data against the Exponential distribution with rate 0.851929.

Figure 10: CDF of the empirical and the chosen Hyperexponential distribution for the light-tailed part of the Danish data.




| phases | $logLik$ | $AICc$ |
|---|---|---|
| 1 | -706.60 | 1417.21 |
| 2 | -706.60 | 1421.26 |

Table 8: Results for the Hyperexponential model used to fit the light-tailed part of the Danish data.

### 11.2.2 Fitting a Coxian distribution

In case of fitting a Coxian distribution to the light-tailed part of the Danish data the result of the EM algorithm is always an Erlang distribution. As in the previous fittings, starting for one phase we calculate for each Coxian distribution its parameters and its $logLik$ and $AICc$ values. However, from the plot of the CDF of the fitted Erlang distribution for $ph = 1, 2, 3$ and 4 against the empirical distribution of the light-tailed part, we see that these two distributions are not close. Therefore we start increasing the number of phases instead of by 1, by 5 until we reach 20 phases and after that by 10 until we have 60 phases.

As we can see in table 9, the $AICc$ decreased and the $logLik$ value increased as we added phases. Hence, we cannot apply the same reasoning as in the previous sections. We decided to stop at 60 phases since it became very time consuming and there is no indication of how many phases are needed in order to reach a desired result.

Hence, based on figure 11 and the fact that it has the highest $logLik$ value and the smallest $AICc$ value, the chosen Coxian distribution for the light-tailed part is an Erlang distribution with 60 phases and parameter 51.115591. Figure 12 shows the CDF of the chosen Erlang distribution against the emprirical distribution of the light-tailed part of the Danish data.

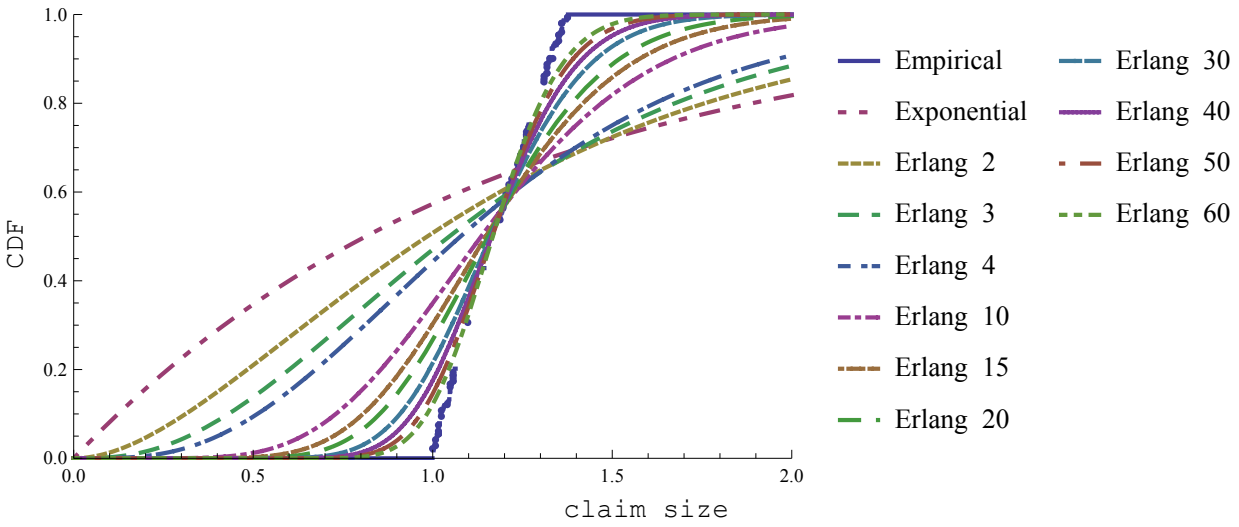



Figure 11: CDF of the empirical and the fitted Erlang distributions for the light-tailed part of the Danish data.

| phases | $logLik$ | $AICc$ |
|---|---|---|
| 1 | -706.60 | 708.60 |
| 2 | -473.96 | 475.96 |
| 3 | -344.79 | 346.80 |
| 4 | -255.62 | 257.62 |
| 5 | -187.76 | 189.77 |
| 10 | 15.27 | -13.27 |
| 15 | 127.34 | -125.34 |
| 20 | 202.69 | -200.69 |
| 30 | 300.82 | -298.82 |
| 40 | 362.66 | -360.66 |
| 50 | 404.69 | -402.68 |
| 60 | 434.19 | -432.19 |

Table 9: Results for the Coxian model used to fit the light-tailed part of the Danish data.

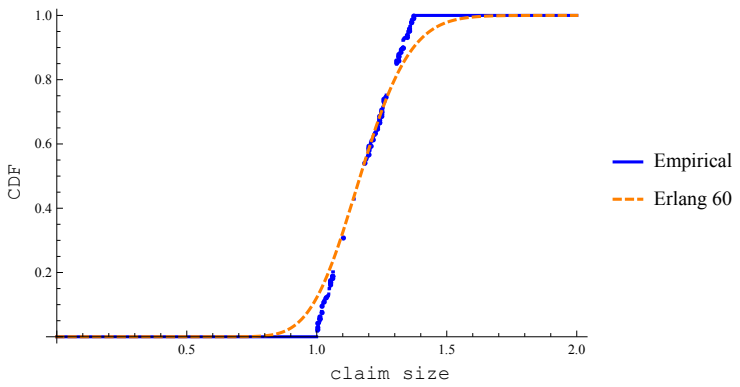



Figure 12: CDF of the empirical and the chosen Erlang distribution for the light-tailed part of the Danish data.

## 11.3 Using the heavy-tailed part of the Danish data

In this section, we approximate the empirical distribution of the heavy-tailed part of the Danish data with a phase-type distribution using the EMpht program. The purpose is to combine these results with the ones we derived in the section 11.2 in order to proceed in section 11.4 with the third and last step of the alternative EM-based technique.

### 11.3.1 Fitting a Hyperexponential distribution

Table 10 gives the values of $logLik$ and $AICc$ when we fit to the heavy-tailed part of the Danish data a Hyperexponential distribution with 1, 2 and 3 phases. The best fit is obtained by a Hyperexponential distribution with 2 phases and parameters:

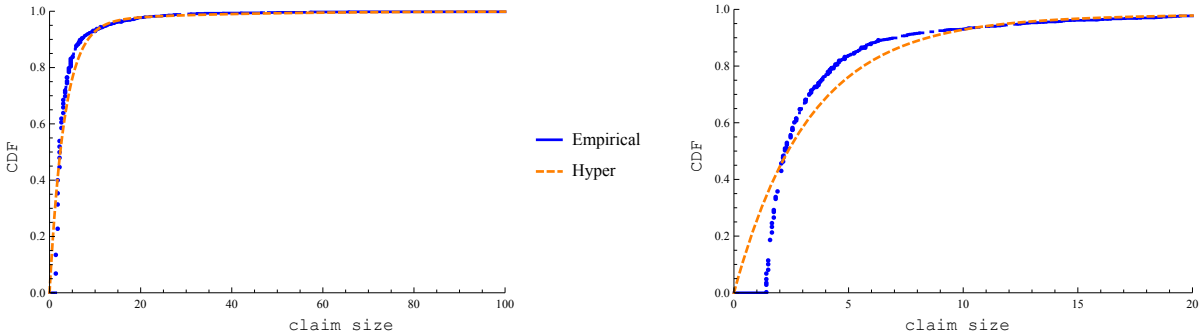
$$\text{Hyperexponential}[(0.039418, 0.960582), (0.034461, 0.309305)]$$

because compared to the other two Hyperexponential distributions this one has the lowest $AICc$ and at the same time the highest $logLik$ value. Figure 13 shows the empirical distribution of the heavy-tailed part of the data and the fitted Hyperexponential distribution.

| phases | $logLik$ | $AICc$ |
|---|---|---|
| 1 | -3812.10 | 7628.20 |
| 2 | -3651.20 | 7310.43 |
| 3 | -3651.20 | 7314.46 |

Table 10: Results for the Hyperexponential model used to fit the heavy-tailed part of the Danish data.

Figure 13: CDF of the empirical and the chosen Hyperexponential distribution for the heavy-tailed part of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

### 11.3.2 Fitting a Coxian distribution

Based on the results in table 11 the preferable Coxian model for the heavy-tailed part of the Danish data is a Coxian distribution with 7 phases since its $AICc$ is the smallest among the fitted Coxian distributions while its $logLik$ value is the highest. Thus, the chosen Coxian distribution is:

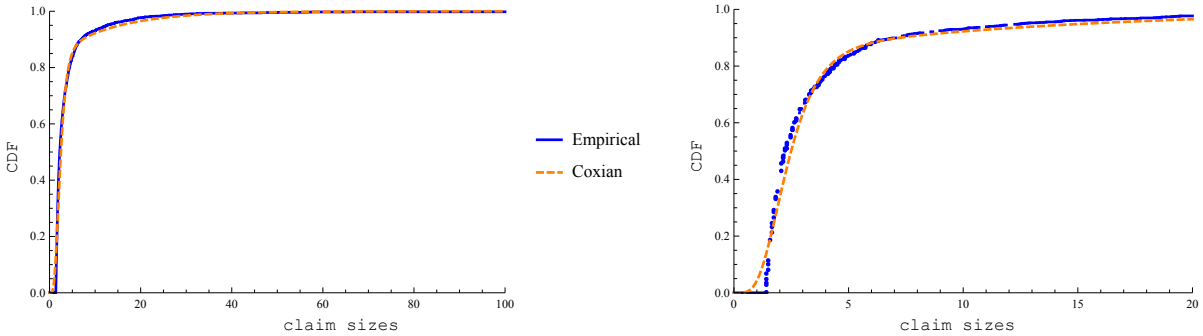
$$\text{Coxian}[(3.278436, 3.278436, 3.278436, 1.025819, 3.278436, 3.278436, 0.080553), \\ (1, 1, 1, 1, 1, 0.140928)].$$

The graph in figure 14 shows the CDF of the empirical distribution of the heavy-tailed part of the Danish data and the CDF of the Coxian distribution described before.

| phases | $logLik$ | $AICc$ |
|---|---|---|
| 1 | -3812.10 | 7626.19 |
| 2 | -3635.48 | 7276.98 |
| 3 | -3322.46 | 6654.95 |
| 4 | -3182.73 | 6379.50 |
| 5 | -3111.11 | 6240.30 |
| 6 | -3070.12 | 6162.36 |
| 7 | -3039.58 | 6105.34 |
| 8 | -3110.85 | 6251.92 |

Table 11: Results for the Coxian model used to fit the heavy-tailed part of the Danish data.

Figure 14: CDF of the empirical and the chosen Coxian distribution for the heavy-tailed part of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

## 11.4 Combination of the two parts

In this section, we basically perform the final step of the alternative EM-based technique. In other words, we try to combine the distributions we used to approximate the empirical distribution of the light- and heavy-tailed parts of the Danish data in a proper way.

Before we can do that we need to know the probability of having a light- or a heavy-tailed claim. In section 12, the probability $\epsilon$ of having a heavy-tailed claim is calculated and it is equal to $\epsilon = 0.718966$, thus the probability for the claim size to be light-tailed is $1 - \epsilon = 0.281034$.

Now, based on the fact that for each part of the Danish data we use two different phase-type distributions to model their empirical distribution, there are four final estimates for the distribution of the claim sizes of the Danish data.

Note that the Exponential distribution with rate 0.851929, which is the chosen Hyperexponential model for the light-tailed part of the Danish data in section 11.2.1, is clearly less accurate than the Erlang distribution with 60 phases and rate 51.115591 found in section 11.2.2. However, both distribution are used in the mixture models. The reason for this is that we would like to determine the influence of the distribution of the light-tailed part in the estimation of the VaR and the ruin probability. We expect this influence to be small since almost 70% of the claims are heavy-tailed. However, we want to verify this since in section 11.2.2 the number of phases used to model the light-tailed part stops at 60, without finding the best phase-type distribution.

Let $Exp$ and $Erl$ denote the chosen Hyperexponential and Coxian distributions for the light-tailed part of the Danish data, respectively. Moreover, let $Hyp_2$ and $Cox_2$ be the Hyperexponential and Coxian distributions we used to model the distribution of the heavy-tailed part of the Danish data. Now, using the alternative EM-based technique, the phase-type distributions that approximate the distribution of the Danish data can be written as:

1. $P = (1 - \epsilon)Exp + \epsilon Hyp_2$,

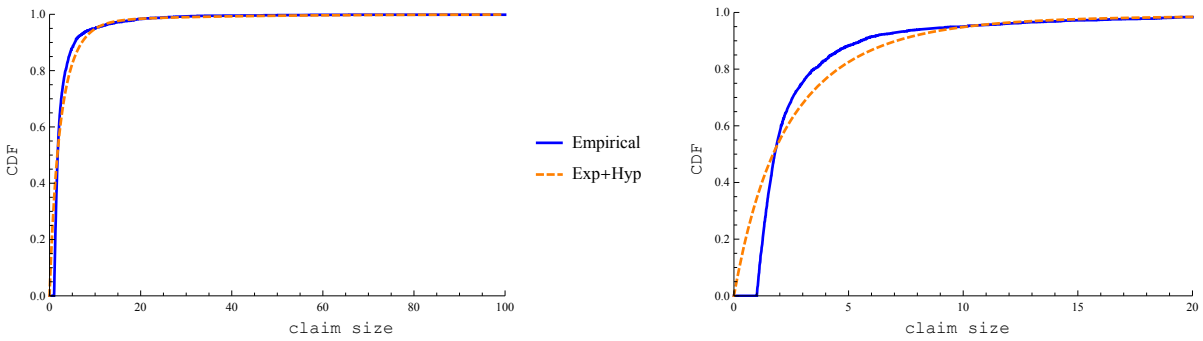
2. $P = (1-\epsilon)Exp + \epsilon Cox_2$,

3. $P = (1-\epsilon)Erl + \epsilon Hyp_2$ and

4. $P = (1-\epsilon)Erl + \epsilon Cox_2$.

The plots in figures 15−18 show the empirical distribution of the Danish data against the results of the alternative EM-based technique we just described.

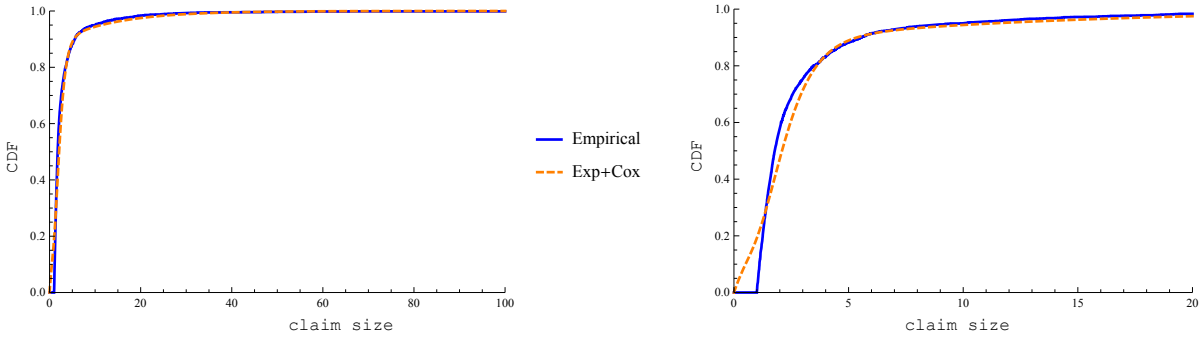
Figure 15: CDF of the mixture $(1-\epsilon)Exp + \epsilon Hyp_2$ against the CDF of the empirical distribution of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

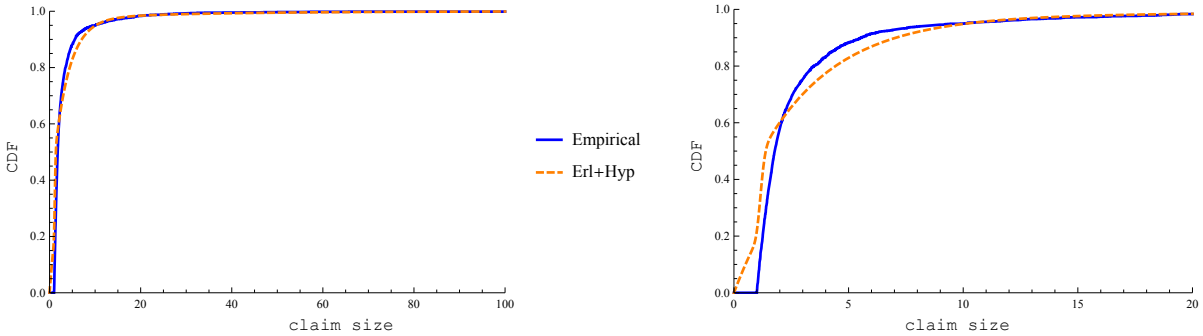
Figure 16: CDF of the mixture $(1-\epsilon)Exp + \epsilon Cox_2$ against the CDF of the empirical distribution of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

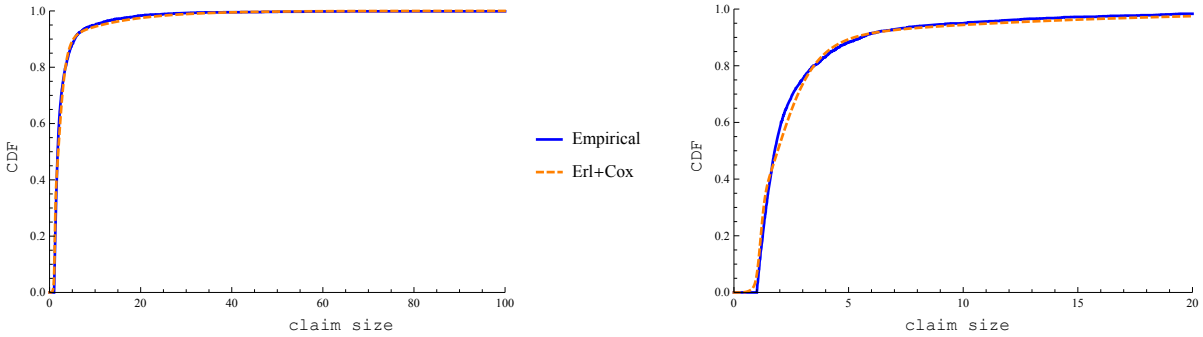
Figure 17: CDF of the mixture $(1-\epsilon)Erl + \epsilon Hyp_2$ against the CDF of the empirical distribution of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

Figure 18: CDF of the mixture $(1-\epsilon)Erl + \epsilon Cox_2$ against the CDF of the empirical distribution of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

# 12 Mixture model: Phase-type and Pareto distribution

In this section we fit a mixture model on the distribution of the Danish data. The mixture model consists of a phase-type distribution and a Pareto distribution. More precisely, we fit a Pareto distribution on the tail of our data. Having estimated this Pareto distribution, a proper goodness-of-fit test is used to answer if it plausible for the tail of this data set to follow this distribution. The distribution of the rest of the claims can be approximated with a phase-type distribution since we only have positive claim sizes.

Following the procedure proposed in [11] and briefly described in section 5 we derive that the lower bound and the shape parameter of the Pareto distribution are:

$$\beta = 1.37622 \quad \text{and} \quad \alpha = 2.40006.$$

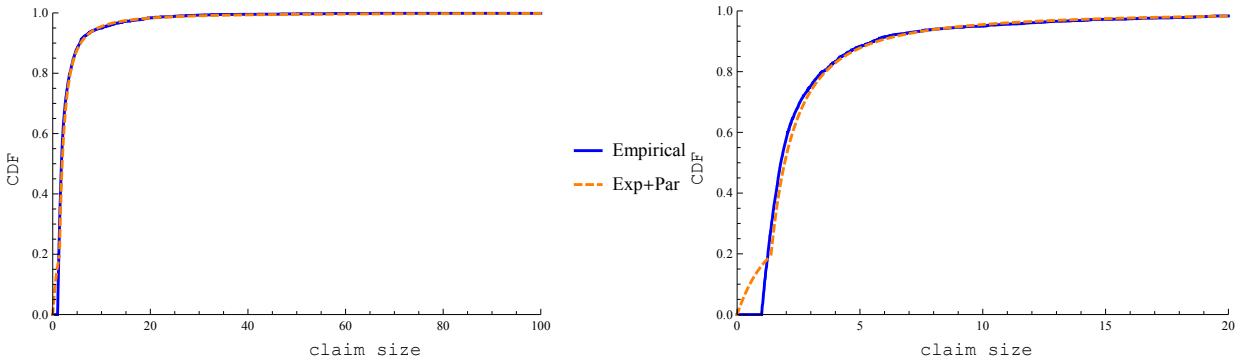
The corresponding KS statistic value is: 0.014717. Additionally, by applying the corresponding goodness-of-fit test we derive that the $p-$value is 0.4092. This implies that it is plausible that the tail of the Danish data follows this Pareto distribution and thus it makes sense to consider a mixture model for our data.

This yields that our data set can be split into two subsets: the light-tailed part and the heavy-tailed part which follows a Pareto distribution with shape parameter 2.40006 and lower bound 1.37622. The distribution of the light-tailed part is approximated by a Hyperexponential and by a Coxian distribution using the EMpht program of which the results can be found in section 11.2. Hence, we have two mixture models: Exponential with Pareto and Erlang with Pareto.

Finally, in order to estimate the probability of having a heavy-tailed claim we compute the fraction of the claims in the Danish data with value at least 1.37622. The result is that with probability $\epsilon = 0.718966$ we obtain a heavy-tailed claim and so with probability $1 - \epsilon = 0.281034$ we get a light-tailed claim.

In figures 19 and 20 one can see the plots of the CDF of the empirical distribution of the Danish data and the CDF of the mixtures models.

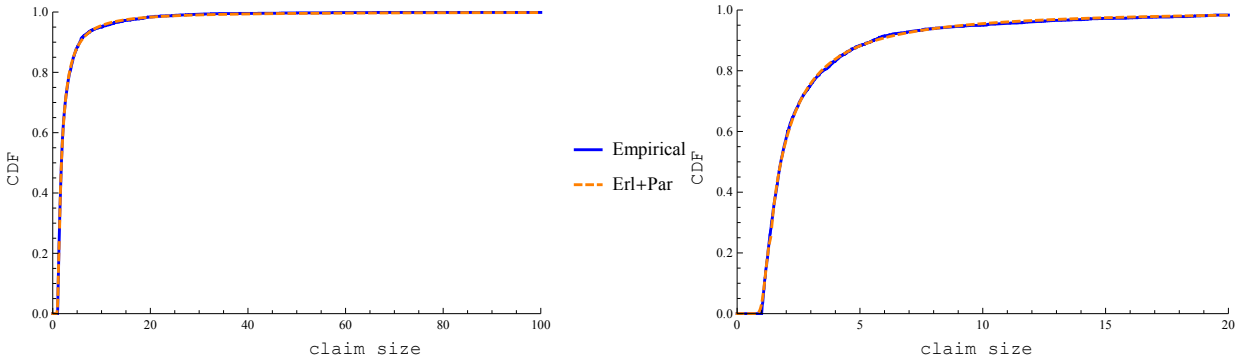
Figure 19: CDF of the mixture Exponential and Pareto against the CDF of the empirical distribution of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

Figure 20: CDF of the mixture Erlang and Pareto against the CDF of the empirical distribution of the Danish data.




(a) Range of the $x$-axis: from 0 to 100.

(b) Range of the $x$-axis: from 0 to 20.

# Part IV
# Numerical results

Having estimated the claim size distribution in the previous part, we can now apply for each fitted distribution the appropriate approximation for the total loss, the VaR and finally for the ruin probability. More specifically, in section 13 we compare the empirical distribution of the aggregate claims with the distributions which we obtained by using the heavy-tailed asymptotic, phase-type and corrected discard approximations. Additionally, we derive the associated VaR's to each approximation and we compare them to the VaR's we obtained from the Danish data. We continue in section 14 with the ruin probability approximations. Since we cannot extract the exact ruin probability by using our data, we assume that the true distribution of the claim sizes is the mixture of the Erlang and the Pareto distribution we found in section 12. Using this distribution, we run simulations and we evaluate the performance of the various ruin probability approximations. Finally, we summarize our conclusions for the approximations in section 15.

## 13 Total loss and Value at Risk

As we mentioned in section 2, for a fixed period of time $[0, t]$ we have that the CCDF of the total loss is:

$$Loss_t(x) = \mathbb{P}(\sum_{i=1}^{N_t} U_i > x),$$

where $\sum_{i=1}^{N_t} U_i$ is the aggregate claim in $[0, t]$. In this research, we consider the fixed period time equal to 1 day, i.e. $t = 1$. Using the arrival dates of the claims in the Danish data we can derive the aggregate claim for each date between the years 1980 to 1990 inclusive. Hence, we can calculate the CCDF of the empirical distribution of the total loss in one day, $Loss_1(x)$.

For each approximation we can apply to the total loss we obtain its estimated values for $x_v = 0.01v, v = 0, ..., 5800$ and we calculate the corresponding relative and absolute errors. In other words, let $\widehat{Loss}_1(x_v)$ denote the approximated value of the $Loss_1(x_v)$ at $x_v$ for one approximation, we compute the corresponding relative error by

$$re_v := \frac{|\widehat{Loss}_1(x_v) - Loss_1(x_v)|}{Loss_1(x_v)},$$

and the absolute error using

$$ae_v := |\widehat{Loss}_1(x_v) - Loss_1(x_v)|.$$

Based on the average relative and maximum absolute errors the approximations return for the total loss, we evaluate their performance.

Furthermore, we can compute the exact VaR for one day for any given probability level $q$. Embrechts et al. mention in [17] that usually, in operational risk, the VaR is calculated for a probability level $q$ for which it holds that $0.001 \leq 1 - q \leq 0.0025$. In our research, however, we obtain the VaR for the probability levels $q_j$ such that

$$1 - q_j = 0.001j, \text{ for } j = 1, ..., 10.$$

For all the probability levels $q_j$ we calculate the relative error between the exact VaR and the approximated VaR. In this case though we do not require the absolute value of the difference, since we do not take the average of the relative errors. Thus, for any given level $q_j$ we can also see if the approximation under- or over-estimates the VaR.

In the following sections, we apply the heavy-tailed asymptotic, phase-type and corrected discard approximations in order to estimate the distribution of the total loss in one day. Moreover, we derive the associated VaR's for the probability levels we mentioned earlier. More specifically, in section 13.1, the performance of the phase-type approximations is evaluated while in section 13.2 we analyze the results of the corrected discard approximations. Finally, in section 13.3, an overall comparison of the heavy-tailed asymptotic, phase-type and corrected discard approximations is given. Two different approaches are used for this comparison. We compare the results of the approximations to the results obtained by the Danish data in section 13.3.1 and to the simulations results of a certain mixture model in section 13.3.2. In section 14, we assume that the true distribution of the claim sizes is a mixture of an Erlang and a Pareto distribution and use simulations to find the ruin probability. Therefore, we also derive the distribution of the total loss and VaR by simulations when the claim size distribution is this particular mixture.

## 13.1 Evaluation of the phase-type approximations

In section 11, we approximate the claim size distribution by using six different phase-type distributions. More precisely, using a Hyperexponential (Hyper), a Coxian, a mixture of an Exponential and a Hyperexponential (Exp+Hyp), a mixture of an Exponential and a Coxian (Exp+Cox), a mixture of an Erlang and a Hyperexponential (Erl+Hyp) and finally using a mixture of an Erlang and a Coxian (Erl+Cox). For each of these phase-type distributions we calculate the distribution of the total loss for one day by simulations and we derive the VaR for the probability levels $q_j$.

Table 12 shows the average relative and maximum absolute errors obtained by the phase-type approximations for $Loss_1(x)$. In figure 21, the points $(x_v, re_v), v = 0, ..., 5800$ are plotted for all the approximations we consider in this section. Furthermore, table 13 gives the VaR based on the Danish data for the different probability levels and the relative errors when the claim size distribution is approximated with a phase-type distribution.

As shown in table 12, the relative errors of all the phase-type approximations are quite small. This implies that the shape of the approximated total loss distribution can be close to the shape of the true distribution. Moreover, the average relative errors of the phase-type distributions Coxian, Exp+Cox and Erl+Cox are smaller than the other three where

| | Average relative error | Maximum absolute error |
|---|---|---|
| Hyper | $0.5403 \times 10^{-2}$ | $9.6588 \times 10^{-2}$ |
| Coxian | $0.2076 \times 10^{-2}$ | $2.2608 \times 10^{-2}$ |
| Exp+Hyp | $0.5661 \times 10^{-2}$ | $10.6081 \times 10^{-2}$ |
| Exp+Cox | $0.3549 \times 10^{-2}$ | $5.4774 \times 10^{-2}$ |
| Erl+Hyp | $0.4885 \times 10^{-2}$ | $6.9401 \times 10^{-2}$ |
| Erl+Cox | $0.2477 \times 10^{-2}$ | $2.1239 \times 10^{-2}$ |

Table 12: Average relative and maximum absolute errors of the phase-type approximations for the distribution of the total loss in one day.

| $1-q$ | VaR | Hyper | Coxian | Exp+Hyp | Exp+Cox | Erl+Hyp | Erl+Cox |
|---|---|---|---|---|---|---|---|
| 0.001 | 57.4106 | 0.2933 | 0.1340 | 0.4135 | -0.0596 | 0.4015 | -0.0483 |
| 0.002 | 46.5000 | 0.2665 | 0.1391 | 0.2604 | 0.0111 | 0.2869 | -0.0049 |
| 0.003 | 36.5849 | 0.3515 | 0.2436 | 0.3243 | 0.1240 | 0.3075 | 0.1350 |
| 0.004 | 32.2525 | 0.3556 | 0.2814 | 0.1948 | 0.1413 | 0.2833 | 0.1365 |
| 0.005 | 29.0371 | 0.2774 | 0.2868 | 0.1734 | 0.1882 | 0.1764 | 0.1555 |
| 0.006 | 27.0186 | 0.1895 | 0.2538 | 0.0905 | 0.1678 | 0.0898 | 0.1815 |
| 0.007 | 24.9703 | 0.2009 | 0.2397 | 0.0248 | 0.2016 | 0.0226 | 0.2072 |
| 0.008 | 24.5555 | 0.0753 | 0.1660 | -0.0537 | 0.1567 | -0.0504 | 0.1229 |
| 0.009 | 21.4864 | 0.1416 | 0.2331 | -0.0051 | 0.2469 | 0.0018 | 0.2510 |
| 0.010 | 19.6336 | 0.0956 | 0.2353 | -0.0215 | 0.2664 | -0.0014 | 0.2751 |

Table 13: The VaR based on the data and the relative errors of the different phase-type approximations.

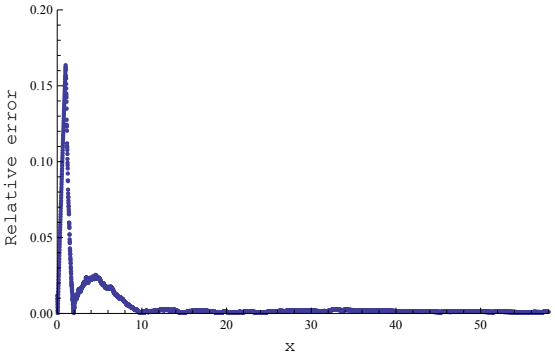
a Hyperexponential distribution is used instead of a Coxian. Similar remarks apply on the maximum absolute errors. It is worth mentioning that the large absolute errors for the Hyper, Exp+Hyp and Exp+Cox occur for small values of $x_v$, as we can notice from the plots in figure 21.

From the results in table 13, we can indicate that in most cases we have an overestimate of VaR. Moreover, it is interesting to compare Exp+Hyp to Erl+Hyp and Exp+Cox to Erl+Cox. In both cases using an Erlang distribution instead of an Exponential to approximate the light-tailed part does not really influence the estimation of the VaR. This occurs probably because only 30% of our data can be characterized as ligth-tailed claims. On the other hand, we reach the opposite conclusion when we compare Exp+Hyp to Exp+Cox and Erl+Hyp to Erl+Cox. In other words, the phase-type distribution used to approximate the heavy-tailed claims has a strong influence in the estimation of VaR. Finally, we indicate that by using a mixture model, we can almost always improve the accuracy of the estimates.

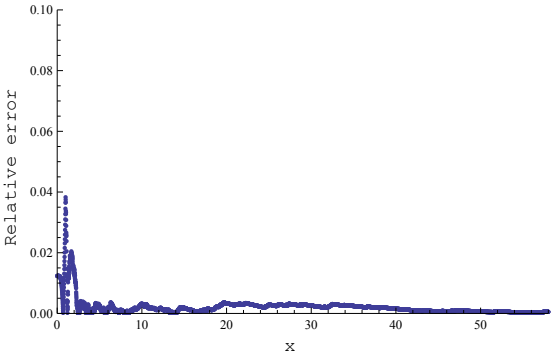
Additionally, for smaller values of $1-q$, Erl+Cox provided the best estimated VaR, while, for bigger values, we obtain better estimates from Erl+Hyp. If we take into consideration that, in risk theory, it is common to have $1 - q \in [0.001, 0.0025]$, we can conclude that

from all the phase-type approximations Erl+Cox is the best approximation for the Danish data. However, in the following sections we compare both Erl+Hyp and Erl+Cox to other approximations.

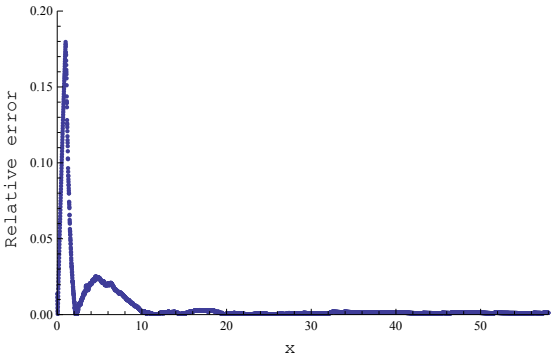
Figure 21: Plots of the relative errors for the distribution of the total loss.



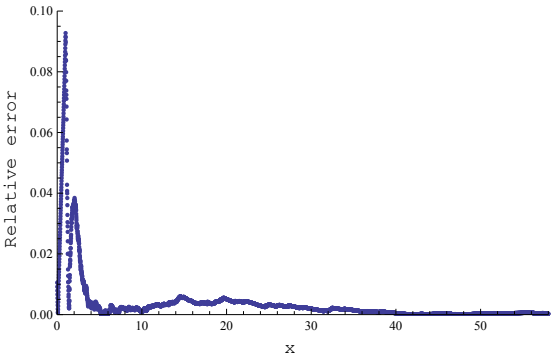

(a) Using the Hyper model.



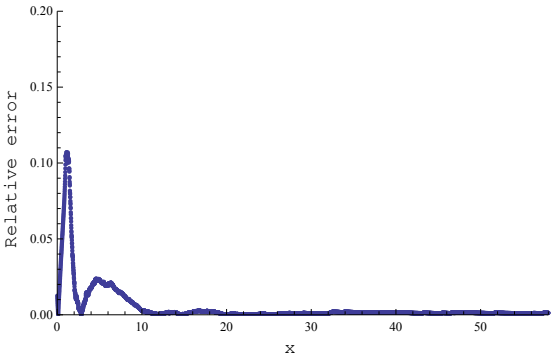

(b) Using the Coxian model.



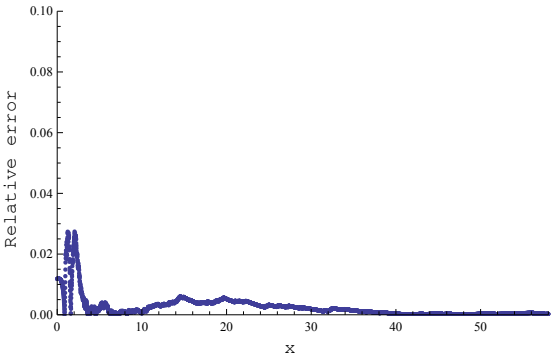

(c) Using the Exp+Hyp model.




(d) Using the Exp+Cox model.




(e) Using the Erl+Hyp model.




(f) Using the Erl+Cox model.

## 13.2 Evaluation of the corrected discard approximations

In section 11.2.2, we approximated the light-tailed part of the Danish data once with an Exponential distribution and once with an Erlang distribution. Therefore, there are two mixture models for the claim size distribution where the tail follows a Pareto distribution (see also section 12). We refer to them as the Exp+Par and the Erl+Par model.

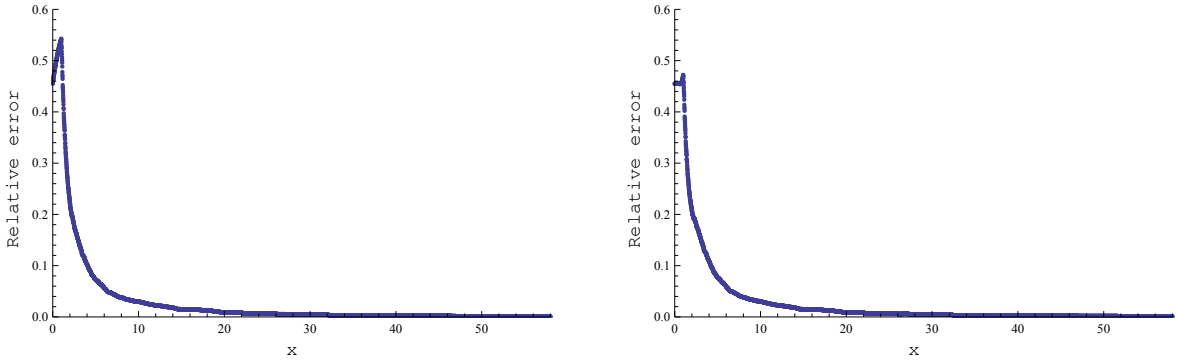
In this section, we apply the corrected discard approximation in order to estimate the distribution of the total loss and VaR. Furthermore, we discuss the related results.

Table 14 presents the average relative and maximum absolute errors. We obtain these values by using the Exp+Par or Erl+Par model as the claim size distribution in the corrected discard approximation for the total loss and comparing these with the empirical total loss distribution. Similar to the previous section, in figure 22, the plots of the points $(x_v, re_v), v = 0, ..., 5800$ for the two different models are given. Last, in table 15, we give the exact VaR and the relative errors of the two different applications of the corrected discard approximations.

| | Average relative error | Maximum absolute error |
|---|---|---|
| Exp+Par | $3.0314 \times 10^{-2}$ | $32.0530 \times 10^{-2}$ |
| Erl+Par | $2.9335 \times 10^{-2}$ | $27.8850 \times 10^{-2}$ |

Table 14: Average relative and maximum absolute errors of the corrected discard approximations for the distribution of the total loss in one day.

Figure 22: Plots of the relative errors for the distribution of the total loss.




(a) Using the Exp+Par model.

(b) Using the Erl+Par model.

As shown in table 14, both approximations provide large average relative and maximum absolute errors. According to [37], the approximation error of the distribution of the total loss is of order $O(\epsilon^2)$. In other words, the performance of the approximation depends on the probability of having heavy-tailed claims, $\epsilon$. This implies that for large $\epsilon$ the approximation is expected to perform poorly. In our case $\epsilon$ is relative large ($\epsilon = 0.718966$), which can explain the poor results we obtain from this approximation.

| $1-q$ | VaR | Exp+Par | Erl+Par |
|---|---|---|---|
| 0.001 | 57.4106 | -0.6114 | -0.6131 |
| 0.002 | 46.5000 | -0.6936 | -0.6981 |
| 0.003 | 36.5849 | -0.6941 | -0.7039 |
| 0.004 | 32.2525 | -0.7036 | -0.7188 |
| 0.005 | 29.0371 | -0.7063 | -0.7270 |
| 0.006 | 27.0186 | -0.7113 | -0.7364 |
| 0.007 | 24.9703 | -0.7097 | -0.7391 |
| 0.008 | 24.5555 | -0.7227 | -0.7535 |
| 0.009 | 21.4864 | -0.7000 | -0.7361 |
| 0.010 | 19.6336 | -0.6874 | -0.7270 |

Table 15: The VaR based on the data and the relative errors of the two different corrected phase-type approximations.

Once again we notice that using an Exponential distribution for the light-tailed part of the Danish data only increases the errors. However, there is no significant difference between the results of the two models.

Furthermore, as we can see in table 15, both approximations provide really poor estimates for the VaR for all values of $1-q$. In addition, there is no essential difference in the estimate of the two approximations. This occurs because the Pareto distribution is much heavier than the Exponential as well as the Erlang distributions. This implies that Pareto distribution influence more the estimation of the VaR. Another factor is that the probability of having a light-tailed claim is 0.281034, while the probability of having a heavy-tailed claim is 0.718966.

## 13.3 Comparison of the different approximations

In this section, we compare the various approximations for the distribution of the aggregate claims and VaR. However, from the phase-type approximations, we consider only the Erl+Hyp and the Erl+Cox models which perform the best as we saw in section 13.1. Additionally, for the corrected discard approximation we view only the results of the Erl+Par model. The reason is that even though both models (Exp+Par and Erl+Par) provide quite similar results, it appears from section 11.2.2 that an Erlang distribution is a more appropriate choice for the light-tailed part of the Danish data.

The rest of this section is organized as follows. In section 13.3.1, the results of the approximations are compared to the results we derive from the Danish data. In section 13.3.2, the performance of the approximations is evaluated with respect to the assumption that the true distribution of the claim sizes is the Erl+Par model. This means, that we obtain the distribution of the total loss and VaR by simulations for the case that the Erl+Par model is the claim size distribution and the arrival rate remains the same. The performance of the approximations is evaluated based on these results. We perform this comparison because in section 14 we use this model as the true claim size distribution to derive the exact ruin

probability and to evaluate the approximations. Moreover, we check if we obtain similar conclusions to the ones in section 13.3.1. Additionally, we would like to examine if we can determine the performance of a ruin probability approximation by the performance of the corresponding approximation for the distribution of the total loss and VaR.

### 13.3.1 In case of the Danish data

In this section, we evaluate the performance of the heavy-tailed asymptotic approximation (denoted as Pareto) and we compare it with the phase-type and corrected discard approximations. Due to the asymptotic relation in (6.3) the heavy-tailed asymptotic approximation of the total loss return values bigger than 1 when $x < 0.627$ since it holds that

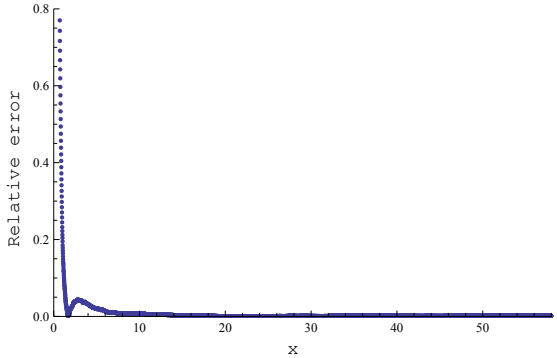
$$\lambda \bar{G}(x) = \lambda * x^{-1.32213} \leq 1 \Rightarrow x \geq 0.627.$$

Therefore we calculate the relative and absolute errors only for $x_v = 0.01v, v = 70, ..., 5800$. In table 16, the average relative and maximum absolute errors of the approximations are given. Note that also for the results of the other approximations, we only take into account the errors calculated for $x_v \geq 0.7$. In figure 23, the plot of the relative errors is provided for the heavy-tailed asymptotic approximation of $Loss_1(x)$. Last, table 17 gives the relative errors of the various approximations for the VaR.

| | Average relative error | Maximum absolute error |
|---|---|---|
| Pareto | $0.7253 \times 10^{-2}$ | $45.4862 \times 10^{-2}$ |
| Erl+Hyp | $0.4611 \times 10^{-2}$ | $6.9401 \times 10^{-2}$ |
| Erl+Cox | $0.2370 \times 10^{-2}$ | $2.1239 \times 10^{-2}$ |
| Erl+Par | $2.4136 \times 10^{-2}$ | $27.8850 \times 10^{-2}$ |

Table 16: Average relative and maximum absolute errors of the different approximations for the distribution of the total loss in one day.

Figure 23: Plot of the relative errors for the distribution of the total loss, using the Pareto model.



A first conclusion is that the average relative errors of the phase-type and corrected discard approximations decrease when we exclude the points $x_v, v = 0, ..., 70$. This happens, because as we can see in figures 21e, 21f and 22b, we obtain large errors for smaller values of the $x_v$. On the other hand, the excluding the points $x_v, v = 0, ..., 70$ does not affect the maximum absolute errors of these approximations.

Furthermore, we notice that the heavy-tailed asymptotic approximation of the $Loss_1(x)$ provides large relative error, especially, in comparison to the phase-type approximation. This occurs probably because the Pareto distribution is not a plausible model for our data set. Moreover, we derive quite large errors for small values of $x$, as we can see in figure 23. However, it performs on average better than the correct discard approximation.

Similar remarks apply to the results for the VaR. The heavy-tailed asymptotic approximation provides bad estimates of the VaR for smaller values of $1 - q$. However, we indicate that it performs considerably better for larger values of $1-q$. Actually, sometimes it performs even better than the phase-type approximations.

In general, the phase-type approximations give the most accurate results for $Loss_1(x)$ and VaR while the corrected discard perform the worst among the approximations. Therefore, one could expect the same conclusions for the corresponding ruin probability approximations. In other words, it is expected that the phase-type approximations for the ruin probability achieve the most accurate results among the different approximations. In section 14, we examine whether or not this is true.

| $1 - q$ | VaR | Pareto | Erl+Hyp | Erl+Cox | Erl+Par |
|---|---|---|---|---|---|
| 0.001 | 57.4106 | 1.0289 | 0.4015 | -0.0483 | -0.6131 |
| 0.002 | 46.5000 | 0.4829 | 0.2869 | -0.0049 | -0.6981 |
| 0.003 | 36.5849 | 0.3870 | 0.3075 | 0.1350 | -0.7039 |
| 0.004 | 32.2525 | 0.2657 | 0.2833 | 0.1365 | -0.7188 |
| 0.005 | 29.0371 | 0.1875 | 0.1764 | 0.1555 | -0.7270 |
| 0.006 | 27.0186 | 0.1118 | 0.0898 | 0.1815 | -0.7364 |
| 0.007 | 24.9703 | 0.0706 | 0.0226 | 0.2072 | -0.7391 |
| 0.008 | 24.5555 | -0.0159 | -0.0504 | 0.1229 | -0.7535 |
| 0.009 | 21.4864 | 0.0288 | 0.0018 | 0.2510 | -0.7361 |
| 0.010 | 19.6336 | 0.0397 | -0.0014 | 0.2751 | -0.7270 |

Table 17: The VaR based on the data and the relative errors of the different approximations.

### 13.3.2 In case of the Erl+Par model

In this section, we assume that the true claim size distribution is the Erl+Par model, as we do in the next section for the ruin probability. Based on this assumption and using simulations, we derive the distribution of the aggregate claim and VaR. Now, the results we derived from the approximations are compared to the results of the simulations. Using this comparison we will later be able to determine the validity of the proposition that the best type of approximation for $Loss_1(x)$ and VaR is also the best type for the ruin probability.

There are two reasons why we choose this particular model as the true distribution of the claim sizes. First, in section 12, the results of the goodness-fit-test tell us that it is plausible for the tail of the Danish data to follow a Pareto distribution. Second, in section 11.2.2, it appears that it is more appropriate to approximate the light-tailed part of the Danish data with an Erlang distribution.

In figure 24, the CCDF of the empirical distribution of the total loss is plotted against the CCDF of the aggregate claim when we use the Erl+Par model as the claim size distribution. Graphically, at least, we could say that these two lines are approximately the same. In figure 25, the CCDF of the distribution of the total loss based on the simulations and the different approximations are plotted.

It seems that the phase-type approximations capture better the curve of the $Loss_1(x)$, especially the Erl+Cox model. As we explained in the previous section, the heavy-tailed asymptotic approximation does not provide an estimate for $x < 0.627$. However, it appears that it performs better for $x$'s bigger than 20. Finally, we notice that the corrected discard approximation is the most inaccurate approximation for the $Loss_1(x)$.

Table 18 presents the VaR we derive using simulations as well as the relative errors of the different approximations for it. Once again we notice that we obtain the worst results by the corrected discard approximation. Furthermore, it is worth mentioning that even though the phase-type approximation using the Erl+Cox model is one of the best with respect to the Danish data, the same thing does not hold in this case. We obtain almost always better estimates from the heavy-tailed asymptotic approximation. It appears that we achieve more accurate results by the phase-type approximation when the Erl+Hyp model is used.

In the following section, we apply the various ruin probability approximations and we discuss the related results. Moreover, we examine if indeed we achieve a better approximation for the ruin probability by using the phase-type or the heavy-tailed asymptotic approximation.

Figure 24: Plot of the CCDF of empirical distribution of the aggregate claim and of the simulations.

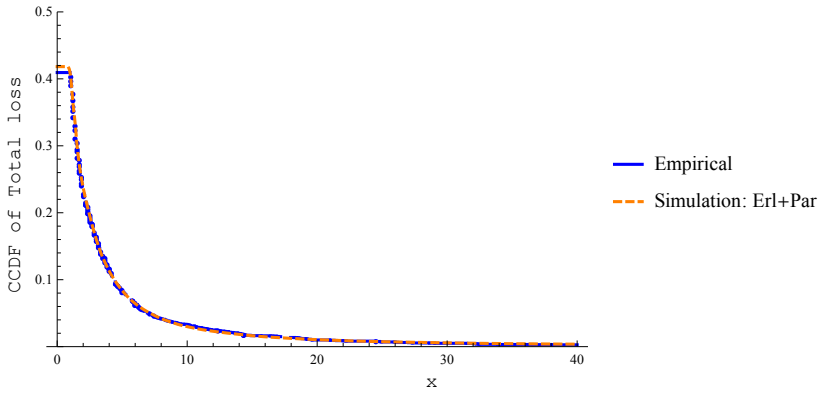
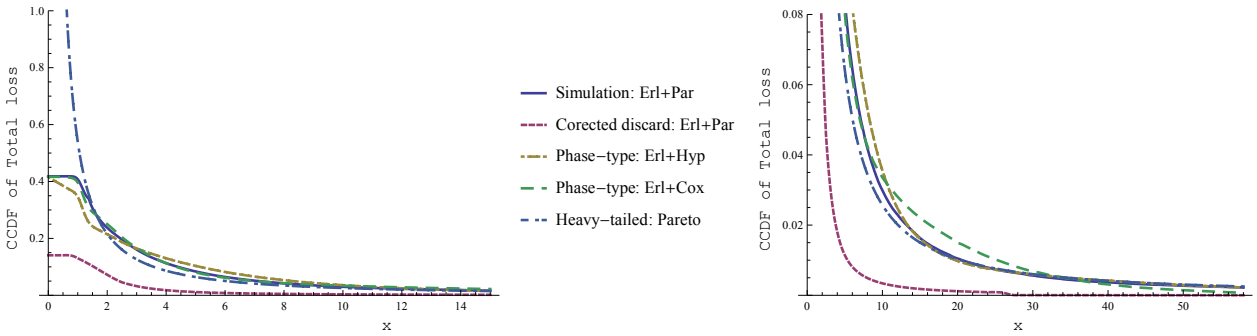


Figure 25: Plots of CCDF of distribution of the aggregate claim using simulations and the approximations.




(a) Range of the $y$-axis: from 0 to 1. Range of the $x$-axis: from 0 to 15.

(b) Range of the $y$-axis: from 0 to 0.08. Range of the $x$-axis: from 0 to 58.

| $1-q$ | simulation | Pareto | Erl+Hyp | Erl+Cox | Erl+Par |
|---|---|---|---|---|---|
| 0.001 | 101.4090 | 0.1486 | -0.2066 | -0.4612 | -0.7810 |
| 0.002 | 62.2783 | 0.1072 | -0.0391 | -0.2570 | -0.7746 |
| 0.003 | 46.6180 | 0.0885 | 0.0261 | -0.1093 | -0.7676 |
| 0.004 | 37.5235 | 0.0879 | 0.1030 | -0.0231 | -0.7583 |
| 0.005 | 32.3518 | 0.0658 | 0.0559 | 0.0371 | -0.7550 |
| 0.006 | 28.9310 | 0.0383 | 0.0177 | 0.1034 | -0.7538 |
| 0.007 | 26.0085 | 0.0279 | -0.0182 | 0.1590 | -0.7495 |
| 0.008 | 23.7626 | 0.0169 | -0.0187 | 0.1604 | -0.7453 |
| 0.009 | 22.0525 | 0.0024 | -0.0239 | 0.2188 | -0.7429 |
| 0.010 | 20.5919 | -0.0087 | -0.0479 | 0.2158 | -0.7397 |

Table 18: The VaR based on simulations and the relative errors of the different approximations.

# 14 Ruin probability

In this section, we evaluate the ruin probability approximations we consider in this research. We assume that the true distribution of the claim sizes is a mixture of an Erlang distribution with 60 phases and rate equal to 51.115591 and a Pareto distribution with shape parameter $\alpha = 2.40006$ and scale parameter $\beta = 1.37622$. Note that this distribution is one of the distributions we derived in section 12, while, in section 13.3.2 we explained why we choose this particular distribution. The average claim size using this distribution is 3.7926 which is higher than 3.38509, the mean of the claim sizes of the Danish data.

Given the Danish data, we require that the premium rate is higher than $\rho = \lambda\mathbb{E}(U) = 1.82702$ since otherwise the safety loading $\eta$ is negative. Using the mixture of the Erlang and Pareto distribution as the claim size distribution, it must hold for the premium rate that

$$p > 2.04544 = p_1.$$

A premium rate within the interval $(1.82702, 2.044544]$ implies that either the mixture model is not the true distribution of the claim sizes or that the system is not stable. Moreover, by construction, all the phase-type distributions have the same mean as the Danish data. This implies that for $1.82702 < p \leq 2.04544$ only the phase-type approximations are available. However, since by assumption the true claim size distribution is the mixture of an Erlang and a Pareto distribution and the system is stable, we do not consider a premium rate smaller than 2.04544.

When we model the claim size distribution with a Pareto distribution the average claim size is 4.10434 and thus the premium rate must be higher than $p_2 = 2.21356$. In section 8, we saw that additional conditions are required for the corrected phase-type approximations. Namely, for the corrected discard we have that

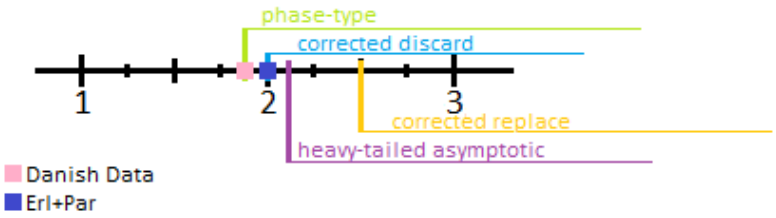
$$|\epsilon\theta| < |p - \delta + \epsilon\delta| \Rightarrow p > 2.04544 = p_3,$$

and for the corrected replace, solving $\epsilon < \frac{|p-\delta|}{\max\{\delta,\theta\}}$ with respect to $p$ and for given $\epsilon = 0.718966$ we obtain that

$$p > 2.50104 = p_4.$$

Figure 26 illustrates for which premium rates we can apply the approximations to our data. In this research we use as premium rate the values $p_j, j = 1, ..., 4$ rounded up to the first decimal, i.e. $p = 2.1, 2.3$ or $2.6$. Hence, for $p = 2.1$ we can apply only the phase-type and corrected discard approximations. For premium rate equal to 2.3 the heavy-tailed asymptotic approximation can be applied too, while for $p = 2.6$ all the ruin probability approximations can be used.

Figure 26: Premium rate condition for the approximations.




For each premium rate we demonstrate the corresponding results in figures 27, 28 and 29. Keep in mind that the corrected phase-type approximations do not include any statistical error since they use as the claim size distribution the true distribution.

**For p = 2.1:** First, we notice that for this premium rate, we actually have a heavy-traffic regime. Therefore, as it can be shown from the simulations the ruin probability of an insurance company is high in this scenario. The phase-type approximations provide more accurate results when $u$ takes values in the interval $(3, 40)$. However, soon they provide values near zero while the real ruin probability is larger than 0.7. Moreover, even though the

corrected discard approximation provides a better estimate of the ruin probability, there is a significant difference between the approximation and the real ruin probability. This might occur because by construction this approximation takes into consideration the appearance of at most one heavy-tailed claim.

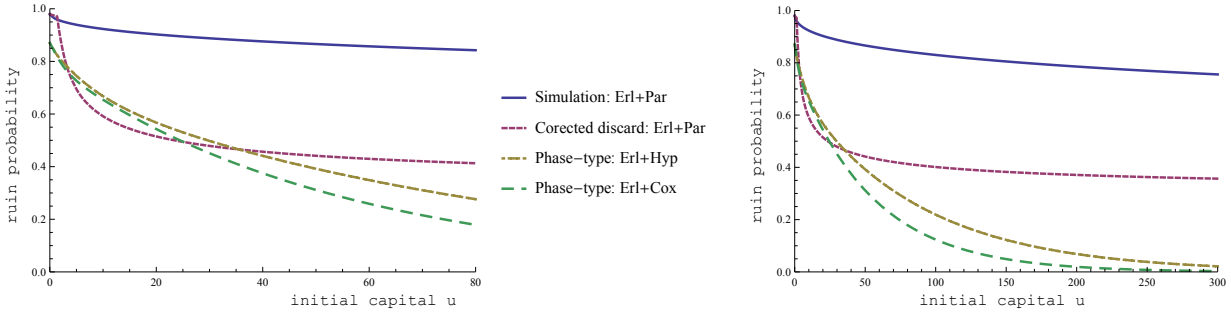
**For p = 2.3:** The first observation is that the heavy-tailed asymptotic provides an estimate for the ruin probability only when $u \geq 9900.27$, according to the relation in (6.2). Once again, we observe that the corrected discard approximation is closer to the exact ruin probability. Finally, the phase-type approximations estimate that the ruin probability is equal to zero for $u$'s bigger than 150.

**For p = 2.6:** For this premium rate we derive conclusions similar to the previous ones. For the heavy-tailed distribution, using (6.2) we derive that $u \geq 94.77$ in order to obtain valid estimates for the ruin probability. Moreover, we indicate that the corrected replace approximation is actually increasing until the lower bound of the Pareto, i.e. for $u \in (0, 1.37622)$. Besides that, however, it estimates the ruin probability more accurately than the other approximations.

In section 13.3, the corrected discard approximation was characterized as the worst approximation for the distribution of the total loss and VaR for both Danish data and simulations. However, in this section, it provided more accurate results for the ruin probability than the phase-type and heavy-tailed asymptotic approximations. This implies that we cannot really determine the performance of a ruin probability approximation by the performance of the corresponding approximation for $Loss_t(x)$ and VaR.

Even though the ruin probability approximations are compared given only one data set, we are still able to have some insights regarding their performance. In the next section, we summarize our research and we discuss the performance of the approximations we consider in this study.

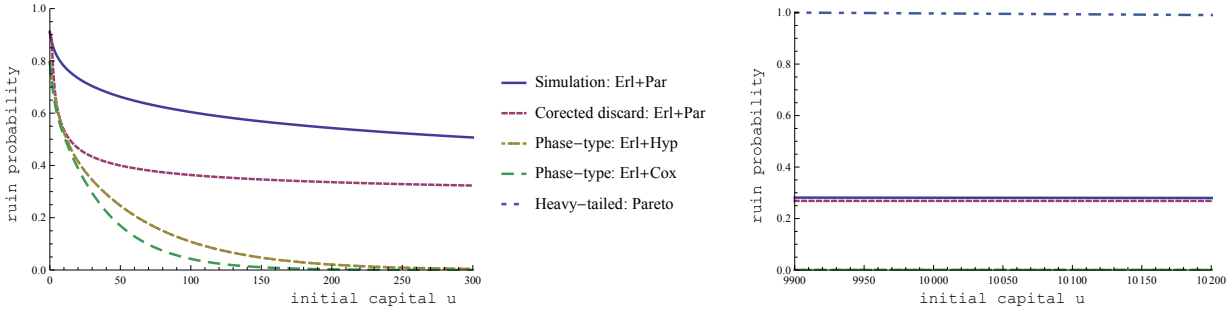
Figure 27: Plots of ruin probability using simulations and the approximations for premium rate p=2.1.




(a) Range of the $x$-axis: from 0 to 80.

(b) Range of the $x$-axis: from 0 to 300.

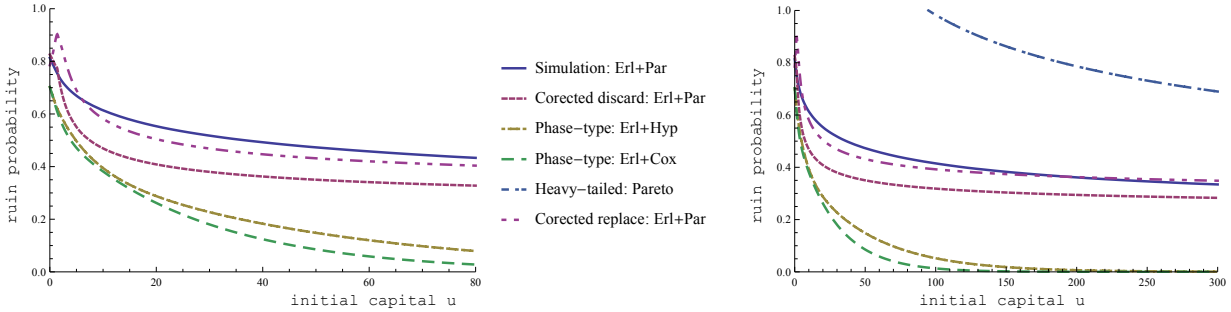
Figure 28: Plots of ruin probability using simulations and the approximations for premium rate p=2.3.




(a) Range of the $x$-axis: from 0 to 300.

(b) Range of the $x$-axis: from 9900 to 10200.

Figure 29: Plots of ruin probability using simulations and the approximations for premium rate p=2.6.




(a) Range of the $x$-axis: from 0 to 80.

(b) Range of the $x$-axis: from 0 to 300.

# 15 Conclusions

The purpose of this research is to evaluate the performance of various ruin probability approximations using real data. More specifically, we considered the heavy-tailed asymptotic, phase-type and corrected phase-type approximations for the ruin probability.

All these approximations assume that the claims arrive to an insurance company according to a Poisson process while the premiums flow with a constant rate. However, for the claim size distribution each approximation considers a different type of distribution. More precisely, the heavy-tailed asymptotic approximation considers a heavy-tailed distribution, the phase-type approximation a phase-type distribution and finally the corrected phase-type approximation considers a mixture of a phase-type and a heavy-tailed distribution. Therefore, to be able to apply these approximations, we estimated the parameter of the Poisson process using the arrival dates of our data. Additionally, we fitted appropriate distributions on the claim sizes of our data set.

Unfortunately, we cannot calculate the exact ruin probability using our data. Hence, we assumed that the true claim size distribution is one of the distributions we used to model our data. Basically, we used the distribution that is most likely to be the true distribution based on our results. Using this distribution, the performance of the approximations is evaluated with respect to the ruin probability we obtained by simulations.

Furthermore, we considered the corresponding heavy-tailed asymptotic, phase-type and corrected discard approximations for the distribution of the total loss and VaR. The reason is that we can derive the empirical distribution of the total loss and VaR from our data and hence we can compare them with the results we obtained by these approximations.

Additionally, we examined the possibility to determine the performance of a ruin probability approximation by the performance of the corresponding approximation for the total loss and VaR. However, according to our results this is not really possible. In other words, a type of approximation which performs poorly for the estimates of the total loss and VaR might provide good estimates for the ruin probability.

More specifically for the performance of the approximations we could say that approximating the claim size distribution with a phase-type one can yield a good approximation for the distribution of the total loss. Therefore, it is possible to obtain good estimates of the VaR by the phase-type approximation even for probability levels really close to 1. It is advisable to divide the data into subsets for better approximations. However, it is possible to overfit the data, especially, in the tail. Moreover, as it was expected, the approximation of the smaller observed values does not influence the estimation of the VaR. Finally, even though they can provide quite accurate results for the distribution of the total loss and VaR, it seems that the same thing does not hold for the ruin probability. As we noticed, they decay a lot faster than the actual ruin probability.

Furthermore, it seems more natural to model the claim size distribution with a mixture of light- and heavy-tailed distributions rather than with a heavy-tailed only. However, when the probability of having a heavy-tailed claim is relatively large, the heavy-tailed asymptotic approximation can provide a good estimate for the distribution of the total loss and VaR. However, due to its asymptotic character, the heavy-tailed asymptotic approximation cannot

always provide an approximation for the distribution of the aggregate claims or the ruin probability. In addition, it is not possible to determine where the asymptotic behavior starts. In other words, even though we can calculate from which point on the approximation produces a valid estimate, we can not detect in which point exactly the tail of the aggregate claims or of the ruin probability starts.

For the corrected phase-type approximations we have that, by construction, they take into account the appearance of at most one heavy-tailed claim. This seems to be enough to approximate the tail of the total loss. However, when the probability of having a heavy-tailed claim is high, their approximation error is not negligible. We may have to consider more heavy-tailed claims in order to improve the performance of the approximations. Similar remarks apply to the performance of the approximations for the ruin probability, especially, in a heavy-traffic regime. However, it is worth mentioning that the corrected discard approximation, regardless of its bad performance for the total loss and VaR, provided the most accurate results for the ruin probability after the corrected replace approximation.

Finally, we indicate that when real data are used then inevitably the ruin probability approximations include statistical errors. Therefore, it is important for a proper statistical analysis to precede the application of any ruin probability approximation. Furthermore, one should re-evaluate the findings when more data become available.

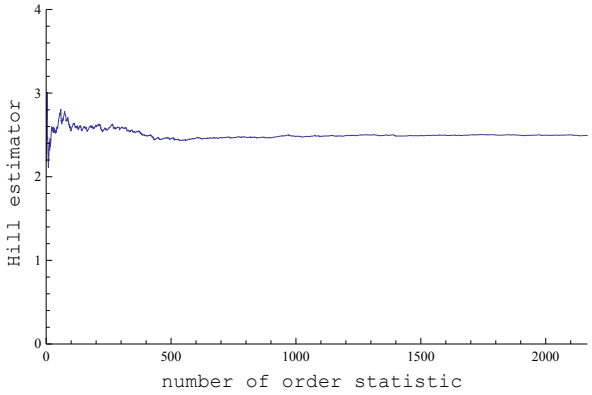
# A Appendix

In section 3.1.1, we described several methods to estimate the shape parameter $\alpha$ of a Pareto distribution given that the lower bound $\beta$ is known. In this section, we apply all of these methods in 25 different data sets of size 2167. Their true distribution is a Pareto with lower bound equal to 1 and shape parameter $\alpha = 2.1, 2.3, 2.5, 2.7$ or 2.9. The results of each approach for each data set can be found in tables 19−23. The estimates with absolute error less than 0.01 are given with bold characters.

For better understanding of our conclusions, before we analyze the results, we demonstrate for one specific sample ($\alpha = 2.5$), the results we obtain by each method:

- Using the MLE we derive that $\alpha_{MLE} = 2.493$.
- The Hill plot of this data set can be found in figure 30. Based on this plot we have that $\alpha_{Hill} = 2.485$.

Figure 30: Hill plot.



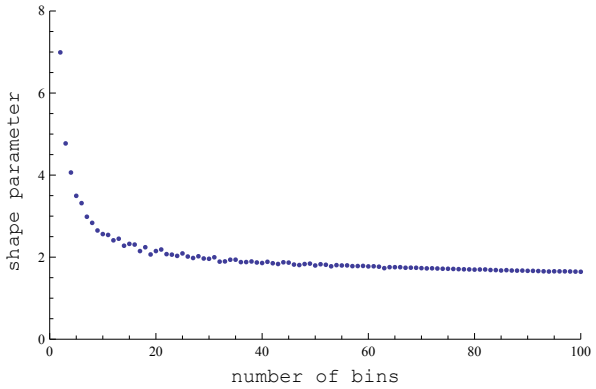

- By applying non-linear regression on the points $(d_i, \bar{P}(d_i))$, we derive that $\alpha_{ccdf,nl} = 2.504$. Applying linear regression on the logarithms of those points yields $\alpha_{ccdf,l} = 2.639$.
- The estimated shape parameter of the Pareto distribution using histogram with constant linear width is denoted as: $\alpha_{bins,method}$, where *bins* is the number of bins used and *method* stands for linear (l) or non-linear (nl) regression. The estimates for our test case are:
  - for $bins = 13, \alpha_{13,nl} = 4.570$ and $\alpha_{13,l} = 2.449$,
  - for $bins = 26, \alpha_{26,nl} = 3.313$ and $\alpha_{26,l} = 2.016$ and
  - for $bins = 49, \alpha_{49,nl} = 2.532$ and $\alpha_{49,l} = 1.808$.

  Similarly, we denote by $\alpha_{log,bins,method}$ the estimates we obtain by using linear or non-linear regression on the points produced from a histogram with constant logarithmic width and number of bins equal to *bins*. The results of these procedures are:

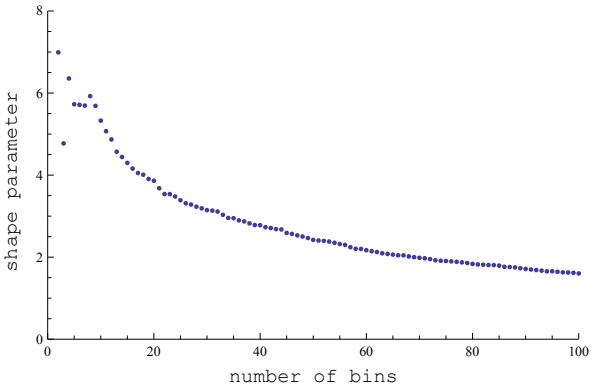
- for $bins = 13, \alpha_{log,13,nl} = 2.492$ and $\alpha_{log,13,l} = 2.399$,
- for $bins = 26, \alpha_{log,26,nl} = 2.483$ and $\alpha_{log,26,l} = 2.401$ and
- for $bins = 49, \alpha_{log,49,nl} = 2.462$ and $\alpha_{log,49,l} = 2.339$.

- Last, in figures 31 and 32, one can find the plots we obtain by estimating, respectively, $\alpha_{bins,method}$ and $\alpha_{log,bins,method}$ for $bins = 2, ..., 100$. The resulting shape parameters are denoted as $\alpha_{loop,method}$ and $\alpha_{log,loop,method}$. The estimates of the shape parameter based on these methods are:
  - using constant linear width: $\alpha_{loop,l} = 1.983$ while we can not determine the shape parameter using the plot in figure 31b, since there is no stable region, and
  - using constant logarithmic width: $\alpha_{log,loop,l} = 2.313$ and $\alpha_{log,loop,nl} = 2.439$.

Figure 31: Using histograms with constant linear width.

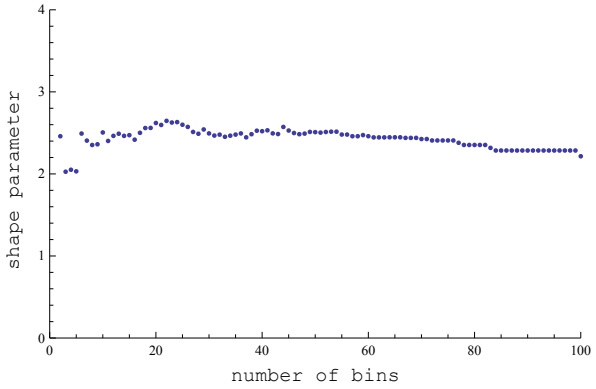



(a) Performing linear regression.

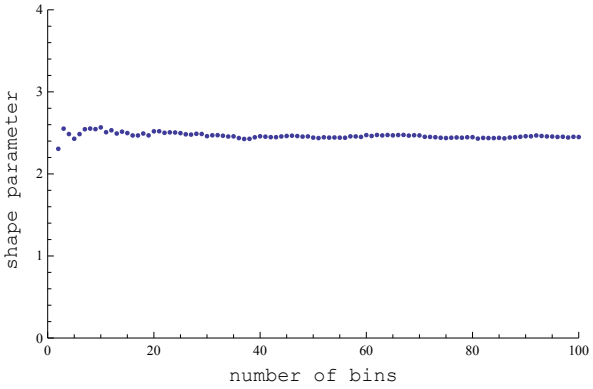



(b) Performing non-linear regression.

Figure 32: Using histograms with constant logarithmic width.




(a) Performing linear regression.




(b) Performing non-linear regression.

It is quite obvious that the techniques based on histograms with constant linear width perform poorly, regardless of whether we perform linear or non-linear regression and their estimates are highly variable. Additionally, the number of bins used has a strong influence on the estimated shape parameter. It appears that having less bins can yield a better estimate for the shape parameter

The methods using the histogram of the data with constant logarithmic width perform better in comparison to constant linear width. However their estimates are also dependent on the width of the bins. Applying non-linear regression to the histogram with constant logarithmic width improve its performance. We observe the exact opposite behaviour when we use constant linear width.

The graphs, based on the histogram techniques, do not provide reliable estimates of the shape parameter. However, they provide insights regarding the sensitivity of the histogram methods to a variety of decisions (like number of bins used, etc). This implies that histogram based methods are not able to produce consistent estimates. For other issues with regression techniques the reader is referred to [11].

The methods using the CCDF points can provide us with good estimates of the shape parameter of the Pareto distribution. Moreover, it appears that their performance is consistent. It is worth mentioning that by performing non-linear regression instead of linear the accuracy of the estimates increases.

Even though the estimates of the Hill plot are consistent, they can be described as biased since it is required to infer the estimate for the shape parameter from the graph. Finally, the MLE is proven mathematically to be one of the best approaches for estimating the shape parameter since it is the minimum variance unbiased estimator of the shape parameter $\alpha$ ([24]).

| Method | $\alpha = 2.1$ | | | | |
|---|---|---|---|---|---|
| MLE | **2.133** | **2.097** | **2.087** | **2.097** | **2.093** |
| CCDF, non linear | **2.112** | **2.089** | **2.089** | **2.099** | **2.094** |
| CCDF, linear | 2.277 | **2.192** | 2.206 | **2.191** | 2.212 |
| 13, non linear | 4.176 | 5.727 | 6.986 | 3.411 | 4.780 |
| 13, linear | 2.463 | 2.266 | **2.168** | 2.410 | 2.296 |
| 26, non linear | 4.053 | 5.448 | 5.971 | 2.860 | 4.306 |
| 26, linear | 1.991 | 1.682 | 1.668 | 1.948 | 1.761 |
| 47, non linear | 3.266 | 4.861 | 5.675 | 2.310 | 3.212 |
| 47, linear | 1.745 | 1.388 | 1.405 | 1.728 | 1.501 |
| log, 13, non linear | 1.972 | **2.080** | **2.111** | **2.150** | **2.057** |
| log, 13, linear | **2.074** | 1.950 | 1.911 | **2.198** | **2.117** |
| log, 26, non linear | 1.992 | **2.168** | **2.196** | **2.146** | 1.986 |
| log, 26, linear | **2.049** | 1.922 | 1.886 | **2.114** | **2.031** |
| log, 47, non linear | **2.069** | **2.170** | **2.166** | **2.081** | **2.031** |
| log, 47, linear | **2.050** | 1.926 | 1.841 | **2.054** | 1.986 |
| Hill plot | 2.202 | **2.081** | **2.081** | **2.111** | **2.111** |
| Loop, non linear | - | - | - | - | - |
| Loop, linear | 1.904 | 1.429 | 1.22 | 1.746 | 1.666 |
| log, Loop, non linear | **2.065** | 2.218 | **2.092** | **2.061** | 1.998 |
| log, Loop, linear | **2.042** | 1.945 | 1.84 | 1.998 | 1.966 |

Table 19: Estimation of the shape parameter of Pareto distribution based on different methods. The true parameter, in these cases, is equal to 2.1.

| Method | $\alpha = 2.3$ | | | | |
|---|---|---|---|---|---|
| MLE | **2.263** | **2.322** | **2.298** | **2.293** | **2.284** |
| CCDF, non linear | **2.258** | **2.336** | **2.282** | **2.289** | **2.283** |
| CCDF, linear | **2.395** | 2.460 | **2.391** | 2.442 | **2.380** |
| 13, non linear | 6.991 | 6.991 | 3.248 | 3.929 | 4.969 |
| 13, linear | 2.064 | 2.064 | **2.353** | 2.46513 | **2.362** |
| 26, non linear | 5.972 | 6.335 | **2.366** | 2.969 | 4.653 |
| 26, linear | 1.623 | 1.595 | 1.960 | 2.106 | 1.938 |
| 47, non linear | 5.350 | 6.330 | 1.638 | **2.340** | 3.480 |
| 47, linear | 1.470 | 1.331 | 1.745 | 1.836 | 1.688 |
| log, 13, non linear | **2.281** | **2.357** | 2.404 | **2.351** | **2.353** |
| log, 13, linear | 1.970 | 1.960 | **2.329** | **2.295** | 2.123 |
| log, 26, non linear | **2.319** | **2.352** | 2.481 | 2.166 | 2.446 |
| log, 26, linear | 2.014 | 1.985 | **2.270** | **2.266** | 2.146 |
| log, 47, non linear | **2.278** | 2.426 | 2.473 | 2.174 | 2.425 |
| log, 47, linear | 2.000 | 1.947 | 2.197 | 2.198 | 2.112 |
| Hill plot | **2.232** | **2.296** | **2.258** | **2.353** | **2.258** |
| Loop, non linear | - | - | - | - | - |
| Loop, linear | 1.283 | 1.429 | 1.825 | 1.851 | 1.746 |
| log, Loop, non linear | **2.313** | **2.345** | 2.502 | 2.155 | 2.471 |
| log, Loop, linear | 1.966 | 1.935 | **2.218** | 2.124 | 2.061 |

Table 20: Estimation of the shape parameter of Pareto distribution based on different methods. The true parameter, in these cases, is equal to 2.3.

| Method | $\alpha = 2.5$ | | | | |
|---|---|---|---|---|---|
| MLE | **2.518** | **2.493** | **2.469** | **2.540** | **2.530** |
| CCDF, non linear | **2.495** | **2.504** | **2.472** | **2.506** | **2.517** |
| CCDF, linear | 2.656 | 2.639 | **2.589** | 2.729 | 2.735 |
| 13, non linear | 4.378 | 4.570 | 3.433 | 3.675 | 6.335 |
| 13, linear | 2.232 | **2.449** | **2.504** | **2.442** | 2.166 |
| 26, non linear | 3.318 | 3.313 | **2.500** | **2.542** | 4.358 |
| 26, linear | 1.876 | 2.016 | 2.073 | 1.992 | 1.887 |
| 47, non linear | 2.323 | **2.532** | 1.752 | 1.547 | 3.511 |
| 47, linear | 1.758 | 1.808 | 1.843 | 1.845 | 1.712 |
| log, 13, non linear | **2.577** | **2.492** | **2.517** | **2.492** | **2.452** |
| log, 13, linear | 2.368 | 2.399 | **2.414** | **2.533** | 2.330 |
| log, 26, non linear | 2.681 | **2.483** | **2.561** | **2.499** | **2.521** |
| log, 26, linear | 2.338 | **2.401** | 2.368 | **2.495** | 2.360 |
| log, 47, non linear | 2.670 | **2.462** | **2.548** | **2.490** | **2.498** |
| log, 47, linear | 2.275 | 2.339 | 2.341 | **2.415** | 2.297 |
| Hill plot | **2.561** | **2.485** | **2.447** | **2.564** | **2.561** |
| Loop, non linear | - | - | - | - | - |
| Loop, linear | 1.825 | 1.983 | 1.904 | 1.851 | 1.746 |
| log, Loop, non linear | 2.660 | **2.439** | **2.534** | **2.502** | **2.502** |
| log, Loop, linear | 2.353 | 2.313 | 2.376 | **2.408** | 2.313 |

Table 21: Estimation of the shape parameter of Pareto distribution based on different methods. The true parameter, in these cases, is equal to 2.5.

| Method | $\alpha = 2.7$ | | | | |
|---|---|---|---|---|---|
| MLE | **2.707** | **2.702** | **2.678** | **2.718** | **2.755** |
| CCDF, non linear | **2.714** | **2.710** | **2.680** | **2.735** | **2.742** |
| CCDF, linear | 2.842 | 2.840 | 2.864 | 2.887 | 2.914 |
| 13, non linear | 3.350 | 4.452 | 2.945 | 3.934 | 4.074 |
| 13, linear | 2.458 | 2.411 | 2.476 | 2.439 | 2.531 |
| 26, non linear | 2.255 | 3.379 | 1.761 | 2.837 | **2.702** |
| 26, linear | 2.104 | 2.068 | 2.098 | 2.066 | 2.126 |
| 47, non linear | 1.393 | 2.450 | 0.9341 | 1.855 | 1.760 |
| 47, linear | 1.956 | 1.861 | 1.902 | 1.910 | 1.966 |
| log, 13, non linear | 2.813 | 2.829 | 2.568 | 2.848 | **2.766** |
| log, 13, linear | 2.593 | 2.511 | **2.689** | 2.542 | 2.549 |
| log, 26, non linear | **2.737** | 2.921 | 2.515 | **2.661** | 2.805 |
| log, 26, linear | 2.556 | 2.496 | 2.584 | 2.486 | 2.566 |
| log, 47, non linear | **2.763** | **2.746** | 2.486 | **2.605** | 2.803 |
| log, 47, linear | 2.534 | 2.405 | 2.569 | 2.486 | 2.542 |
| Hill plot | **2.685** | **2.655** | **2.776** | **2.715** | **2.746** |
| Loop, non linear | - | - | - | - | - |
| Loop, linear | 1.914 | 1.914 | 1.977 | 1.851 | 2.04 |
| log, Loop, non linear | **2.723** | **2.754** | 2.471 | **2.628** | **2.786** |
| log, Loop, linear | 2.502 | 2.376 | 2.534 | 2.439 | 2.502 |

Table 22: Estimation of the shape parameter of Pareto distribution based on different methods. The true parameter, in these cases, is equal to 2.7.

| Method | $\alpha = 2.9$ | | | | |
|---|---|---|---|---|---|
| MLE | **2.878** | **2.921** | **2.938** | **2.931** | **2.876** |
| CCDF, non linear | **2.884** | **2.915** | **2.925** | **2.895** | **2.885** |
| CCDF, linear | 3.0389 | 3.133 | 3.211 | 3.125 | **2.994** |
| 13, non linear | **2.885** | 4.164 | 4.235 | 3.719 | 2.796 |
| 13, linear | 2.480 | 2.443 | 2.411 | 2.411 | 2.353 |
| 26, non linear | 1.690 | 2.786 | 2.562 | 2.413 | 1.539 |
| 26, linear | 2.099 | 2.132 | 2.034 | 2.118 | 2.034 |
| 47, non linear | 0.832 | 1.680 | 1.460 | 1.390 | 0.718 |
| 47, linear | 1.910 | 1.859 | 1.945 | 1.932 | 1.798 |
| log, 13, non linear | **2.862** | **2.844** | **2.906** | **2.886** | **2.827** |
| log, 13, linear | 2.759 | 2.672 | 2.692 | 2.733 | 2.720 |
| log, 26, non linear | **2.833** | **2.892** | 2.792 | **2.946** | **2.826** |
| log, 26, linear | 2.699 | 2.644 | 2.738 | 2.606 | 2.654 |
| log, 47, non linear | 2.797 | **2.838** | 2.753 | 3.029 | **2.832** |
| log, 47, linear | 2.636 | 2.538 | 2.603 | 2.596 | 2.602 |
| Hill plot | **2.867** | **2.957** | 3.018 | **2.927** | **2.867** |
| Loop, non linear | - | - | - | - | - |
| Loop, linear | 2.103 | 2.04 | 1.977 | 2.103 | 2.04 |
| log, Loop, non linear | 2.786 | **2.818** | 2.723 | 3.007 | **2.881** |
| log, Loop, linear | 2.565 | 2.534 | 2.628 | 2.565 | 2.534 |

Table 23: Estimation of the shape parameter of Pareto distribution based on different methods. The true parameter, in these cases, is equal to 2.9.

# References


[1] J. Abate and W. Whitt. Computing Laplace transforms for numerical inversion via continued fractions. *INFORMS Journal on Computing*, 11(4):394–405, 1999.

[2] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, Dec. 1974.

[3] S. Asmussen. Light traffic equivalence in single-server queues. *The Annals of Applied Probability*, 2(3):555–574, 1992.

[4] S. Asmussen. Matrix-analytic models and their analysis. *Scandinavian Journal of Statistics*, 27(2):193–226, 2000.

[5] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.

[6] S. Asmussen and H. Albrecher. *Ruin Probabilities*. Advanced Series on Statistical Science & Applied Probability, 14. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2010.

[7] S. Asmussen, O. Nerman, and M. Olson. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

[8] J. A. Beekman. A ruin function approximation. *Transactions of the Society of Actuaries*, 21:41–48, 1969.

[9] P. Bloomfield and D. R. Cox. A low traffic approximation for queues. *Journal of Applied Probability*, 9:832–840, 1972.

[10] E. Bonnet, O. Bour, N. E. Odling, P. Davy, I. Main, P. Cowie, and B. Berkowitz. Scaling of fracture systems in geological media. *Reviews of Geophysics*, 39(3):347–383, 2001.

[11] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[12] R. B. D'Agostino and M. A. Stephens, editors. *Goodness-of-fit Techniques*. Marcel Dekker, Inc., New York, NY, USA, 1986.

[13] D. J. Daley and T. Rolski. A light traffic approximation for a single-server queue. *Mathematics of Operations Research*, 9(4):624–628, 1984.

[14] D. J. Daley and T. Rolski. Light traffic approximations in queues. *Mathematics of Operations Research*, 16(1):57–71, 1991.

[15] H. Drees, L. de Haan, and S. Resnick. How to make a hill plot. *Annals of Statistics*, 28:2000, 2000.

[16] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*, volume 33 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997.

[17] P. Embrechts and G. Samorodnitsky. Ruin theory revisited: stochastic models for operational risk. Technical report, Cornell University Operations Research and Industrial Engineering, December 2002.

[18] P. Embrechts and N. Veraverbeke. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics & Economics*, 1(1):55–72, 1982.

[19] A. Feldmann and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31(3–4):245–279, January 1998.

[20] C. M. Harris, P. H. Brill, and M. J. Fischer. Internet-type queues with power-tailed interarrival times and computational methods for their analysis. *INFORMS Journal on Computing*, 12(4):257–260, 2000.

[21] B. M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics*, 3:1163–1174, 1975.

[22] A. Horváth and M. Telek. Approximating heavy tailed behavior with phase type distributions. In G. Latouche and P. Taylor, editors, *Advances in Algorithmic Methods for Stochastic Models, Proceedings of the Third International Conference on Matrix Analytic Methods*, pages 191–214, Leuven, Belgium, June 2000. Notable Publications.

[23] C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, June 1989.

[24] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*, volume 1 of *Wiley series in probability and mathematical statistics: Applied probability and statistics*. Wiley & Sons, second edition, 1995.

[25] R. E. A. Khayari, R. Sadre, and H. B. R. Fitting world-wide web request traces with the em-algorithm. *Performance Evaluation*, 52(2 - 3):175 – 191, 2003.

[26] J. F. C. Kingman. On queues in heavy traffic. *Journal of the Royal Statistical Society. Series B. Methodological*, 24:383–392, 1962.

[27] M. Olsson. The empht-programme. Technical report, Department of Mathematics, Chalmers University of Technology, June 1998.

[28] M. Olvera-Cravioto, J. Blanchet, and P. Glynn. On the transition from heavy traffic to heavy tails for the $M/G/1$ queue: the regularly varying case. *The Annals of Applied Probability*, 21(2):645–668, 2011.

[29] A. G. Pakes. On the tails of waiting-time distributions. *Journal of Applied Probability*, 12(3):555–564, 1975.

[30] R. Perline. Strong, weak and false inverse power laws. *Statistical Science*, 20(1):68–88, 2005.

[31] S. Resnick and C. Stărică. Smoothing the Hill Estimator. *Advances in Applied Probability*, 29(1), 1997.

[32] A. Riska, V. Diev, and E. Smirni. An EM-based technique for approximating long-tailed data sets with PH distributions. *Performance Evaluation*, 55(1–2):147–164, January 2004.

[33] J. F. Shortle, P. H. Brill, M. J. Fischer, D. Gross, and D. M. B. Masi. An algorithm to compute the waiting time distribution for the $M/G/1$ queue. *INFORMS Journal on Computing*, 16(2):152–161, 2004.

[34] K. Sigman. Light traffic for workload in queues. *Queueing Systems. Theory and Applications*, 11(4):429–442, 1992.

[35] K. Sigman. Appendix: A primer on heavy-tailed distributions. *Queueing Systems. Theory and Applications*, 33(1–3), March 1999. Queues with heavy-tailed distributions.

[36] D. Starobinski and M. Sidi. Modeling and analysis of power-tail distributions via classical teletraffic methods. *Queueing Systems. Theory and Applications*, 36(1-3):243–267, 2000.

[37] E. Vatamidou, I. J. B. F. Adan, M. Vlasiou, and B. Zwart. Corrected phase-type approximations of heavy-tailed risk models using perturbation analysis. *Insurance: Mathematics and Economics*, 53(2):366–378, September 2013.

[38] D. Wallace. Asymptotic approximations to distributions. *Annals of Mathematical Statistics*, 29:635–654, 1958.

[39] E. P. White and J. L. Green. On estimating the exponent of power-law frequency distributions. *Ecology*, 89:905–912, April 2008.