Statistical distributions in general insurance stochastic processes

by

Jan Hendrik Harm Steenkamp

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Department of Statistics
In the Faculty of Natural & Agricultural Sciences
University of Pretoria
Pretoria

January 2014

I, Jan Hendrik Harm Steenkamp declare that the dissertation, which I hereby submit for the degree Magister Scientiae in Mathematical Statistics at the University of Pretoria, is my own work and has not previously been submitted for a degree at this or any other tertiary institution.

SIGNATURE:

DATE: 31 January 2014

Summary

A general insurance risk model consists of in initial reserve, the premiums collected, the return on investment of these premiums, the claims frequency and the claims sizes. Except for the initial reserve, these components are all stochastic. The assumption of the distributions of the claims sizes is an integral part of the model and can greatly influence decisions on reinsurance agreements and ruin probabilities.

An array of parametric distributions are available for use in describing the distribution of claims. The study is focussed on parametric distributions that have positive skewness and are defined for positive real values. The main properties and parameterizations are studied for a number of distributions. Maximum likelihood estimation and method-of-moments estimation are considered as techniques for fitting these distributions. Multivariate numerical maximum likelihood estimation algorithms are proposed together with discussions on the efficiency of each of the estimation algorithms based on simulation exercises. These discussions are accompanied with programs developed in SAS PROC IML that can be used to simulate from the various parametric distributions and to fit these parametric distributions to observed data.

The presence of heavy upper tails in the context of general insurance claims size distributions indicates that there exists a high risk of observing very large and even extreme claims. This needs to be allowed for in the modeling of claims. Methods used to describe tail weight together with techniques that can be used to detect the presence of heavy upper tails are studied. These methods are then applied to the parametric distributions to classify their tails' heaviness.

The study is concluded with an application of the techniques developed to fit the parametric distributions and to evaluate the tail heaviness of real-life claims data. The goodness-of-fit of the various fitted distributions are discussed. Based on the final results further research topics are identified.

Acknowledgements

To Jesus Christ I am sincerely greatful for being blessed with the opportunity to have studied and also for the ability to have conducted this study. I received strength to persist despite setbacks and challenges.

To my family and in particular my mother I would like to express my thanks for your support and encouragement throughout the duration of my studies.

To my supervisor, Dr. Inger Fabris-Rotelli, thank you for your time, sharing of knowledge, valuable input and guidance during this research.

To my employer and colleagues I would like to express my thanks for your support and helping me to free up time to do this research.

Contents

List of Figures

List of Tables

Symbols and Notation

$\chi^2(\nu)$	Chi-square distribution with parameter ν.
$EXP(\theta)$	Exponential distribution with parameter θ.
$F_X(\cdot), F(\cdot), F$	Cumulative distribution function, where subscript X indicates that it is for a specific random variable X.
$f_X(\cdot), f(\cdot), f$	Probability density function, where subscript X indicates that it is for a specific random variable X.
$\overline{F}(\cdot)$	Tail function $1 - F(\cdot)$.
$F^{-1}(\cdot)$	Generalized inverse function of $F(\cdot)$.
$GAM(\theta, \kappa)$	Gamma distribution with parameters θ and κ.
$INV\chi^2(\nu)$	Inverse Chi-square distribution with parameter ν.
$INVGAM(\alpha, \theta)$	Inverse Gamma distribution with parameters α and θ.
$LN(\mu, \sigma)$	Lognormal distribution with parameters μ and σ.
$M_X(t)$	Moment generating function for a random variable X.
$N(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2.
$SN(\mu, \sigma, \lambda)$	Skew-normal distribution with parameters μ, σ and λ.
$\Psi(\cdot)$	Digamma function.
$\Psi'(\cdot)$	Trigamma function.

Chapter 1

Introduction

1.1 Problem Statement

Statistical distributions and their properties have been widely studied and applied in many fields in practice. One of the components of the risk models considered in general insurance is that of the claims. Two processes can be distinguished: the claim frequencies and the claim sizes:

- **Claim frequency**
  The number of claims that occur in a specific time period is a discrete random variable where such a variable can take on any nonnegative value. This is where a Poisson distribution is frequently used.

- **Claim size**
  Continuous distributions have in particular been studied in the context of general insurance to model claims sizes [75], [105]. The random variable for claim sizes is defined on $\mathbb{R}^+$. Well-known distributions, that are often used in practice to model claim sizes include the Gamma, Pareto and Weibull distributions.

In practice the assumption of normality is often used in problems across various industries as it has the advantages of being well known, well studied and easy to implement and use for inference. When the assumption of normality is satisfied, techniques such as Monte Carlo simulation can be used (while normality is not a prerequisite in order to be able to use this simulation technique). In some real-life problems the assumption of normality is reasonable and provides the opportunity for performing inference on this basis.
Luenberger [82] states that the stochastic process of stock prices can be described by a discrete-time multiplicative model that satisfies the condition that the natural logarithm of the return of a share of stock at a given time

(assuming stock prices are observed daily, for example) is a Normal random variable. Furthermore the series of log-returns are independent. This theory is given for a single stock price process. To illustrate this concept the log-returns based on daily historical closing values for the Johannesburg Stock Exchange All-share Index [1] (JSE ALSI) for the period January 2002 to July 2012 were considered. For this series were considered. The empirical distribution function is illustrated by means of the histogram in Figure 1.1.

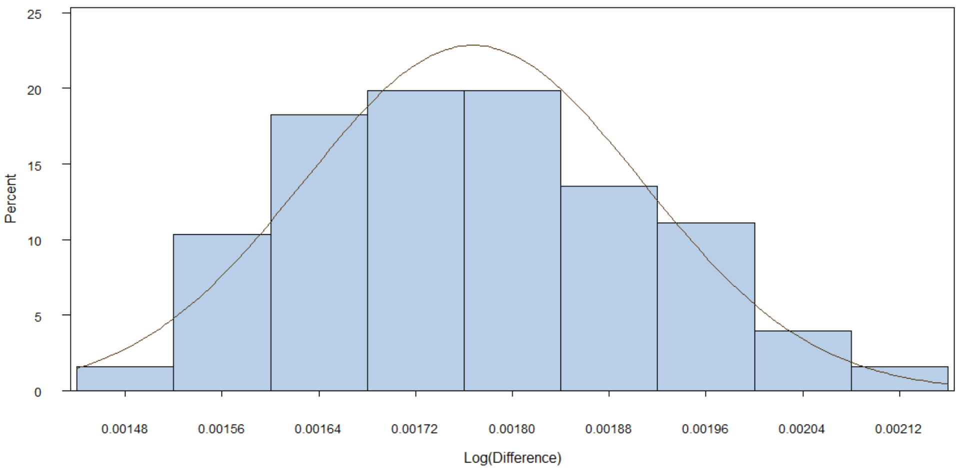

Figure 1.1: Distribution of log-returns for the JSE ALSI (2002-2012)

The Kolmogorov-Smirnov test statistic for testing the hypothesis of normality has an associated p-value of > 0.15 which means that the this hypothesis cannot be rejected, although it appears from the histogram in Figure 1.1 as if this distribution may be slightly skewed to the right. An assumption of normality therefore does seem to be reasonable for this example and it appears to be appropriate to model stock log-returns with a Normal distribution.

For the example of the stock price process the domain for log-returns is theoretically unbounded while the mean is non-zero. The Normal distribution is therefore suitable not only due to the shape of the empirical distribution being similar to the shape of the Normal distribution, but also in terms of the domain of the log-returns. These returns can also be centralised (in order to have a mean-adjusted series with mean zero and the variance unchanged) or standardized (in order to have a mean-adjusted series with mean zero and unit variance).
Often, in practice, the domain of a specific random variable is limited to a finite subset of the real numbers. If one considers a group of motor insurance policies, then the number of policyholders may be symmetrically distributed across the ages of the policyholders. Parametric distributions such as the

[1] Taken from http://markets.ft.com

Normal and Student-t distributions are not suitable to use, since these distributions will imply non-zero probabilities associated with values falling outside of the domain of the random variable for the policyholder's age.

When one consider general insurance claims, these claim sizes are often neither Normal nor symmetrically distributed. These dsitributions are also only defined for positive real values. An example of a typical claim size distribution as given in Figure 1.2, taken from Klugman et al [75], shows that this distribution is not symmetric.

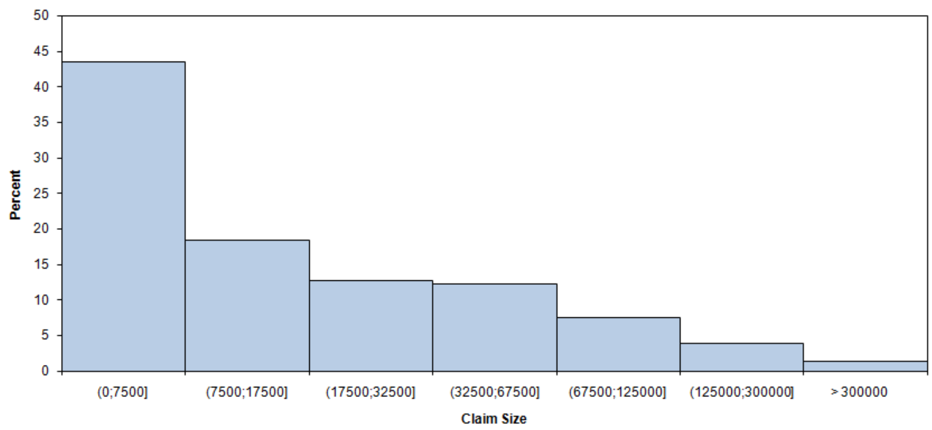

Figure 1.2: Example of a distribution of claim sizes

In cases where the distributions for claim sizes exhibit high degrees of positive skewness, it is important to consider the concept of upper tail heaviness of the distributions. A heavy upper tail in the context of general insurance claims implies that the probability of observing a very large claim in a specific range will become very small in relation to the probability of observing claims in that range and larger. This means that there will still be a significant weight in such an upper tail, despite having low densities on those larger values. Figure 1.3 gives an example of such a heavy-tailed distribution. Having an understanding of the upper tail weight of a distribution, can provide insight into the risk associated with the occurrence of very large claims.

This study is focussed on developing an understanding of the overall structure of a general insurance model as well as to identify and briefly describe each of the components within this model. One of these components deals specifically with claims sizes. The use of parametric distributions to model claims sizes is considered in this study by introducing an array of parametric distributions together with their properties. Techniques that can be used to identify heavy tails are studied and applied to the parametric distributions to classify each as heavy-tailed or not. The maximum likelihood estimation

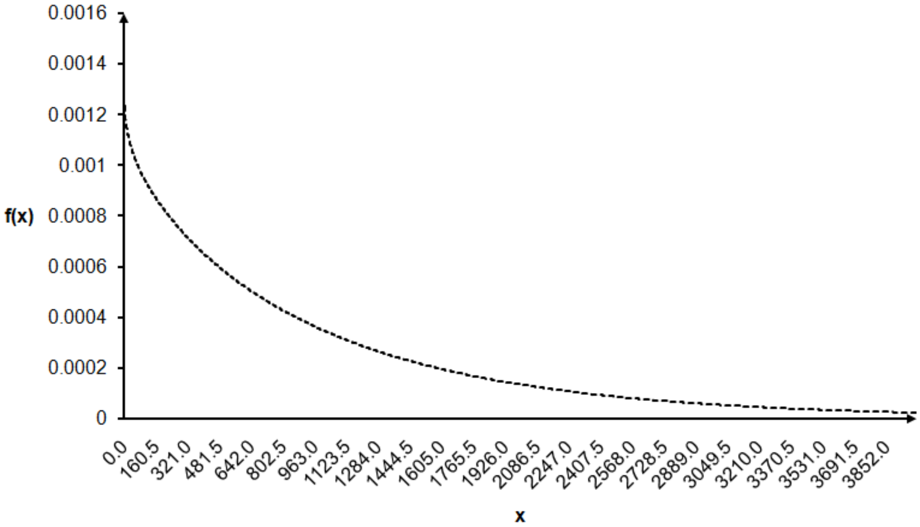

Figure 1.3: Example of a heavy-tailed distribution

process that can be followed to fit each distribution is derived after which it is applied to simulated datasets. The study is concluded by fitting the parametric distributions to a real-life set of claims after which the goodness-of-fit of each fitted distribution is briefly discussed together with describing the tail weight of these claims using the techniques introduced.

1.2 Literature Review

With the problem statement in mind, the literature review was conducted with a focus on skew distributions and the use of these distributions in general insurance. Furthermore, it was of interest to gain an overview of which topics in general insurance were studied more recently.

The topic of skew distributions is well studied across different industries, although it appears as if the application thereof in insurance (and specifically general or non-life insurance) had been studied to a lesser extent. Specific topics within general insurance (not necessarily related to the use of distributions) have been studied extensively more recently.
An overview of the observations made from the literature review is discussed below and is structured as follows:

- We start of with a brief overview given on the research conducted more recently on skew and heavy- tailed distributions.

- The focus then moves to research done in general insurance with the specific use of distributions to model claim sizes.
- A brief discussion is given on the main themes (not only limited to the use of statistical distributions) that have received the most attention in more recent research in general insurance.
- The discussion is concluded by consolidating these observations and discussing how these topics fit into the context of and serves as motivation for this dissertation.

Skew and Heavy-tailed Distributions

It is widely recognised in the literature that there is a need for considering skew and potentially heavy-tailed distributions. These parametric distributions have been used and studied extensively. It is argued by Önalan [93], for example, that asset returns are often skewed. If the distribution of asset returns are in fact skew, measures such as value at risk (VaR) and expected shortfall (which are important measures for measuring market risk) will be misstated if a Normal distribution is still assumed. In this context Önalan considers the Inverse Gaussian distribution as it has the flexibility to be represented as a skew distribution and has heavier tails than the Normal distribution.

An extensive amount of research has been conducted specifically around finding skew distributions by using generalizations of well-known parametric distributions. A generalization of the Normal distribution is compared with the known Skew-normal distributions while new forms of the skew Student-t and skew Cauchy distributions are studied by Abtahi et al [1].

Similar studies were conducted by Nadarajah and Kotz [89] [90] using the probability density and cumulative distribution functions of the Logistic, Laplace, Student-t, Uniform, Exponential Power, Bessel function and Generalized t distributions as well as Types II and VII of the Pearson distribution. In these studies the general form $2g(u)G(\lambda u)$ is used where $g(u)$ and $G(u)$ represent the probability density function and the cumulative distribution, respectively, coming from the same distribution. Generalizations of the Logistic distribution are also studied by Gupta and Kundu [61], specifically for unimodal data.

In more recent research Nadarajah and Kotz [92] use a similar form as considered in their earlier study [90]. In this research a skew distribution of the form $2f(u)G(\lambda u)$ is considered where $f(u)$ and $G(u)$ represent the probability function and the cumulative distribution, respectively, of any two distributions (not necessarily the same distribution). A combination of two distributions, typically chosen from the Normal, Student-t, Cauchy, Laplace, logistic or the uniform distribution, is used. When both the density function and the cumulative distribution function are from the Normal distribution the resulting distribution is the Skew-normal distribution as introduced by Azzalini and Capitanio [8], [9] and Pewsey [100].

In recognising that the topic of skew distributions had been widely studied, a general class of asymmetric, univariate distributions is considered by Arreleno-Valle et al [5] which contains the complete family of univariate, asymmetric distributions. Theoretical properties of this class of distributions are discussed while simulation methods, maximum likelihood estimation methods, method-of-moments estimation and extensions to multivariate skew distributions are discussed.

In a similar way a generalization is introduced by Goerg [58] in which a parametric nonlinear transformation of symmetric distributions is introduced in order to allow for asymmetry. In its most general form a random variable is modelled as the transformation of a standardized, symmetrically distributed random variable where the symmetric distribution of the underlying variables is well-known and has useful properties. Azzalini and Genton [10] consider a generalizations of a parametric class of distributions (of which the skew t distribution is a member) with parameters that can be used to regulate skewness and kurtosis as is reasonable to consider in some practical cases.

Different to finding skew distributions by means of generalizations of known parametric distributions, a class of Skew-normal distributions which is the result of a two-fold convolution of two independent random variables, one being a Normal and one being a Beta distributed random variable, is introduced by Barrera et al [14].

Non-parametric skew distributions have also been considered using techniques such as kernel density estimation. In one example of such a study, conducted by Lotti and Santarelli [81], kernel density estimation was used to obtain an estimated distribution which showed positive skewness and satisfied theory associated with industrial heterogeneity.

Skew Distributions in General Insurance

It is evident that the use of skew distributions in general insurance is more focussed on the use of parametric distributions such as the Weibull, Gamma, Lognormal, Pareto and Generalized Pareto for modeling claim sizes - see for example [39], [66], [20], [43], [105].

Space-time processes are studied by Davis and Mikosch [38] where the distributions thereof are not Gaussian, but of an extreme value type, such as the Pareto distribution. Such processes exhibit typical distributions at different locations at different points in time and therefore makes sense to be employed in the general insurance applications - for example where certain neighbourhoods are more likely to have burglaries and have them more frequently in the festive season.

To decide on an insurance deductible (the portion of a claim that will be retained by the policyholder) or the value of claim sizes (whether per individual policy or per aggregate claim size from all policies) from which reinsurance should be implemented, measures such as value at risk (VaR) can be used. Against this background Bali and Theodossiou [12] discuss positively skewed Extreme Value distributions, including the Generalized Pareto distribution, Generalized Extreme Value distribution and Box-Cox Generalized Extreme Value distribution. Skewed heavy-tailed distributions are also discussed which includes the Skewed generalized error distribution, Skewed Generalized Student-t distribution, Exponential Generalized Beta distribution of the Second Kind and the Inverse Hyperbolic Sign distribution.

Aspects relating to the upper right tail of claim frequency and claim size distributions are considered [95]. For this purpose quantile methods are used in order to develop statistical models for quantiles of these distributions to get better fitted tails. Issues regarding accurately capturing the tail behaviour by means of assessing the goodness-of-fit of these loss distributions with measures such as the Kolmogorov-Smirnov (KS) statistic, Chi-square statistic and the Akaike Information Criteria (AIC) are discussed. The number of near-extremes, which is the number of observed claim sizes being close to the m^{th} largest claim, is studied by Hashorva and Hüsler [64] and a consistent estimator for the upper tails of the largest claim size is derived.

Often claims are assumed to be independent or the underlying processes of claim frequency and claim severity are assumed to be independent. Guegan and Zhang [60] considers the fact that dependence does exist in multivariate financial and insurance data in practice and use copulas to describe influences from macro economic factors on the dependencies that exist among these processes. Furthermore research is focussed on taking the dependence among risks into account (for example the matrix-variate models discussed by Akdemir and Gupta [3] and the Koehler-Symanowski copula function discussed by Palmitesta and Provasi [97]), but not specifically for modeling dependent insurance risks that result in dependent claims which may affect claim frequencies and claim sizes.

General Insurance Themes

In general stochastic processes form part of the set of quantitative methods used for modeling components within the general insurance framework [47], [50]. These mostly relate to claim arrivals, reinsurance, ruin theory and experience rating.

Ruin theory had been researched often in the more recent years with a focus on the Lundberg inequality and the effect of implementing reinsurance [49], [40].

Research on ruin theory also includes studies on finding the finite time ruin probability under the assumption of stochastic interest rates (associated with invested portion of the surplus) as well as the derivation of precise estimates for finite time ruin probabilities. An example is where this is done for a discrete-time insurance risk model where losses are from a sub-exponential class of distributions [115]. Other examples include where bounds for ultimate ruin are given while approximations for the ruin probabilities under heavy-tailed claims occurrence are given, [120] [27] [28] and where asymptotic formulae are obtained for the ruin probabilities over finite and non-finite time horizons when claim sizes are assumed to be Pareto-distributed. This is done in [116] where the upper tail of stochastic present values of future aggregate claims are studied.

Reinsurance treaties had been studied with allowance made for investing premiums and recognising that investment returns are also of a stochastic nature [110].

Most of the literature on claims arrival makes allowance for single claims to arrive at a time. In practice a portfolio of policies introduces the reality of multiple claims within the same period arising from different insured risks (or even at the same time) where these risks may also be dependent in some instances [60]. By studying the bivariate random claims sizes, Hashorva [63] provides a way in which two claims sizes can be modelled jointly.

It was of interest to know what research has been done on factors that may affect claim sizes, specifically when it results in left-censoring. One such example is as a result of deductibles that are in place on insurance policies. Denuit et al [39] refers to this as the problem of unobserved heterogeneity when modeling claim counts. In addition to the known deductible per policy, there exists a pseudo-deductible that is defined by Braun et al [22] as a latent threshold above the policy deductible that governs the policyholder's claim behaviour and further censors the claim size distribution. This means that in spite of the policyholder incurring some loss(es) being covered by the insurance policy, the policy holder is willing to retain the loss instead of the premium of the policy going up due to the submitted claim. This reasoning is in light of the existence of Bonus-Malus systems which is a technique that helps in forming more homogeneous subgroups of policyholders to effectively charge certain subgroups lower or higher premiums based on recent claims experience, as is discussed by Boland [20].

Conclusion

From the in-depth literature review to determine the extent to which research has been conducted in the field of general insurance and what aspects of insurance are considered in a statistical context, the following was found:

- Skew and heavy-tailed parametric distributions have been studied extensively whilst generalizations of parametric distributions have received considerable attention in light finding distributions that are more flexible to the presence of skewness in observed data. Alternative methods such as finding skew distributions using convolutions as well as non-parametric skew distributions have been studied.

- A great amount of research had been done in methodologies to model claim frequencies, times between claim arrivals and claim sizes. It is widely recognised that the claim size distributions should be skewed and aspects such as these distributions being heavy-tailed are considered. While the skewness of these claim size distributions had been recognised, it was most frequently done using well-known parametric distributions. The effect of dependence among risks has also been considered by means of introducing copulas when modeling multiple claims.

- Research in the field of general insurance has been largely focussed on ruin theory and reinsurance in the presence of uncertainty related to returns on investments whilst the impact on claims behaviour due to the presence of bonus- malus systems has also received some attention.

This supports the need for investigating the possible uses of parametric and generalized families of skew distributions specifically in the context of claim sizes in general insurance. This conclusion serves as a motivation for the aim of this dissertation as outlined in Section 1.1.

1.3 Structure of the dissertation

The remainder of the dissertation is structured as follows:

- An overview of the stochastic elements the general insurance risk model consist of will be given in Chapter 2.

- Various parametric distributions that can be used to model skew and heavy-tailed claims size distributions together with their key properties. The main properties of these distributions in terms of the moments, parameterizations and links with other distributions. This is done in Chapter 4.

- Methods that can be used to describe tail weight and detect heavy-tails are studied in Chapter 3.

- In Chapter 5 the distributions studied in Chapter 4 are classified in terms of their tail behaviour using the techniques and underlying theory discussed in Chapter 3.

- Techniques that can be used to simulate from the distributions discussed in Chapter 4, methods to fit these distributions and to test the accuracy of the fitted distributions are discussed in Chapter 6.

- The theory and techniques discussed in Chapters 2 to 6 come together in Chapter 7.

Chapter 2

General Insurance Overview

Mathematical and statistical models are used in general insurance to perform calculations or make decisions in the following key areas:

- Premium Rating
- Reserving
- Reinsurance Arrangements
- Testing for Solvency

These models are based on historical claims experience in order to address the frequency of claims occurrence as well as their severity [39]. It is important to segment the business's underwritten risks into subgroups that are homogeneous. This means that the subgroups of the business should be similar in terms of the risks insured and the portion of the market it is aimed at delivering a service to. The insured risk refers to the actual asset being insured. In the case of vehicle insurance, for example, the business may be segmented into private car insurance, fleet car insurance and public transport insurance. Further segmentation might be in terms of the type of product that is made available which is related to the type of insurance product. Within the private car insurance one may find various types of risks insured like smaller sedans, large sedans and SUV's.
Upon setting up a new product in an insurance company or setting up an insurance company, the first aspect of concern in the initial reserve. The initial reserve is the amount of capital set aside for a specific subgroup or portfolio in order to meet claims that need to be paid out before receiving all the premiums or sufficient premiums to cover that risk.
Three main components in studying the key areas in general insurance are [71], [67]:

1. The initial reserve, denoted by u.

2. The premiums collected. These can either be collected up to a predetermined time, denoted by $(0, t]$ or over predetermined number of intervals, t.

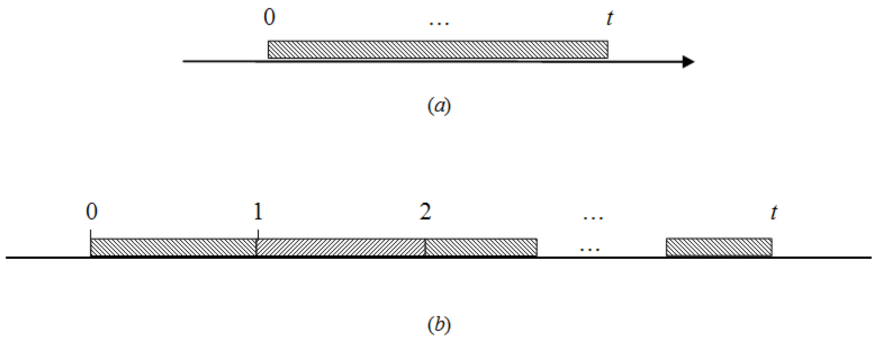

Figure 2.1: Illustration of discrete and continuous time intervals

   When working with a problem where one is interested in the total number of premiums collected up to a predetermined time, the premium collection process is defined on a continuous time space with $t \in \mathbb{R}^+$. This is illustrated in Figure 2.1.
   When working with a problem with outcomes or events associated with or states occupied at certain points in time or stages in a process, then the premium collection process is defined on a discrete time space $\{t : t = 0, 1, 2, \ldots\}$.

3. The claims made up to a predetermined time or over a predetermined number of time intervals, t, which is denoted by $X(t)$.

Various stochastic elements can be identified from these general insurance components.

2.1 Inter-occurrence times

The duration from inception of an investigation period up to the n^{th} claim can be denoted by τ_n. This means that from inception the time until the n^{th} claim arises is a random variable. The duration from the $(n-1)^{th}$ to the n^{th} claim is given by $T_n = \tau_n - \tau_{n-1}$, where T_n is therefore a random variable for $n \geq 1$. This also means that $\tau_0 = 0$ which implies that $T_1 = \tau_1$. This duration between successive claims is called the *inter-occurrence time* [105], [6].

2.2 Number of claims

Consider the number of claims from inception of the investigation period up to some time t where $t \geq 0$, which is denoted by $N(t)$ namely $N(t) = sup\{n : \tau_n \leq t\}$. A series of these outcomes for different values of t gives rise to a counting process defined as $\{N(t), t \geq 0\}$. There exists a relationship between the number of claims and the sequence of claim arrival times $\{\tau_1, \tau_2, \tau_3, \ldots, \tau_n\}$, since the event of $\{N(t) = n\}$ is equivalent to the event of $\{\tau_n \leq t \leq \tau_{n+1}\}$. The claim number process is assumed to be a counting process satisfying the following properties [39]:

- $N(0) = 0$
- $N(t) \in \mathbb{N}_0$
- For $h > 0$, $N(t) \leq N(t+h)$ where $N(t+h) - N(t) =$ number of claims in $(t, t+h]$.

Definition 1. ***Continuity from the right****. A function $f(x)$ is said to be continuous at a point c from the right, with c being an interior point of an interval $[a, b]$ in the sense that $a \leq c < b$, if for every $\varepsilon > 0$ there exists a $h = \delta(\varepsilon) > 0$ such that when it is true that $c \leq x < c + h$ and x is an element of the interval $[a, b]$ it is also true that $|f(x) - f(c)| < \varepsilon$, [15].*

Example 1. *The sample paths of the counting process are monotone non-decreasing and right continuous. This can be illustrated by means of an example: Suppose an insurance company has issued a group of one-year policies and the claims arising from losses on these insured risks are captured at the end of each month subsequent to inception of the policy. If the first claim occurred during the second month, then it will only be captured at the end of that month. If the year is partitioned such that t_i denotes the end of the i^{th} month, it means that at the end of the second month it is known that:*

- *$N(t_0) = 0$, $N(t_1) = 0$ and $N(t_2 - h) = 0$ for some arbitrary small $h > 0$.*
- *$N(t_2) = 1$.*
- *For $0 < t < t_2$, the function $N(t)$ will be differentiable, since one will always be able to find an arbitrary small $h > 0$ such that $N(t)$ is continuous on the interval $(h, t_2 - h)$.*

Similarly, if a second claim arises during the 5^{th} month since inception of the policy, then one will have that:

- *$N(t_0) = N(t_1) = 0$.*

- *$N(t_2) = N(t_3) = N(t_4) = 1$ and $N(t_5 - h) = 1$ for some arbitrary small $h > 0$.*

- *$N(t_5) = 2$.*

- *For $t_2 < t < t_5$, the function $N(t)$ will be differentiable, since one will always be able to find an arbitrary small $h > 0$ such that $N(t)$ is continuous on the interval $(t_2 + h, t_5 - h)$.*

The same principle will hold for any number of claims. It will also hold irrespective of how small the time intervals are made, say for instance daily. From this example, it can be seen that it satisfies the conditions to be met for a function to be right continuous. One can, for example, use the point t_2 and define an interval $J = [t_2, t_5]$. For any arbitrary small $\varepsilon > 0$, one can find an $h = \delta(\varepsilon)$, where $0 < \delta(\varepsilon) < (t_5 - t_2)$. Then t_2 is an interior point of J and $\delta(\varepsilon) \in J$. For t satisfying $t_2 \leq t \leq t_2 + h$, it is true that $|f(t) - f(t_2)| = 0 < \varepsilon$. Hence it is true that $N(t)$ is right continuous.

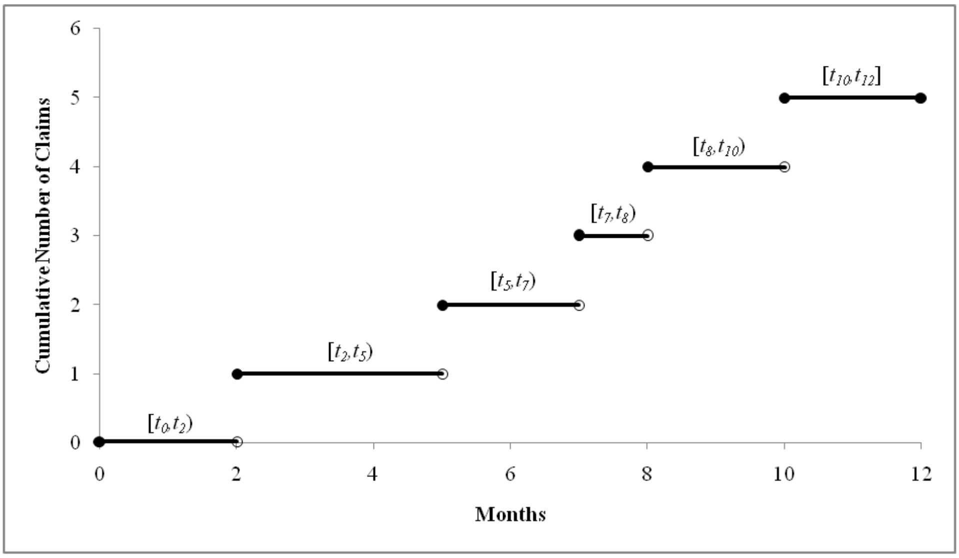

Figure 2.2: Example of a sample path of a claim number process

Figure 2.2 is a graph showing an example of a sample path for the claim number process as described above, with further claims arising over the course of the 12-month period.

Claims are generally assumed to be arriving one at a time (excluding the case of multiple arrivals of claims). Multiple arrivals are possible, but are not dealt with generally as they occur with low probability.

The probability density function of the number of claims is defined as:

$$p_k = \mathrm{P}\left(N(t) = k\right) = \mathrm{P}\left(\sum_{i=1}^{k} T_i \leq t < \sum_{i=1}^{k+1} T_i\right) \tag{2.1}$$

Definition 2. ***Indicator function***. *Define an indicator function as*

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true,} \\ 0 & \text{if } A \text{ is not true.} \end{cases} \tag{2.2}$$

Definition 3. ***Renewal Process***. *Let $T_1, T_2, \ldots$ be a sequence of independent and identically distributed random variables with $T_i \geq 0$ for $i = 1, 2, \ldots$. Consider the inter-occurrence times given by the sequence $\{T_n, n \in \mathbb{N}\}$. Define $\tau_n = T_1 + \ldots + T_n$ for $n = 1, 2, \ldots$ with $\tau_0 = 0$. The sequence $\{\tau_n, n \in \mathbb{N}\}$ is called the* renewal point process *with τ_n being the n^{th} (renewal) epoch which is simply the time of the arrival of the n^{th} claim. The* renewal counting process *is defined as $\{N(t), t \geq 0\}$ with $N(t) = \sum_{n=1}^{\infty} I(\tau_n \leq t)$. The renewal counting process is equivalent to the renewal epoch process, because $N(t) = n \Leftrightarrow \tau_n \leq t < \tau_{n+1}$ [105].*

Definition 4. ***Index of dispersion***. *Let X be a random variable, for an insurance risk for instance, with a probability mass/density function given by $f(x)$. Let the mean and variance of X be denoted by $E(X) = \mu$ and $var(X) = \sigma^2$, respectively. The index of dispersion of X is given by $I_X = I = \frac{\sigma^2}{\mu}$.*

The index of dispersion can be used to assess the homogeneity of data. Hoel mentions its use in the biological sciences for the Binomial and Poisson distributions. The index of dispersion as it is given here is also related to the coefficient of variation [57] which considers the ratio of the standard deviation to the mean as a relative measure of the spread of the data [112]. It therefore appears as if these two measures have a similar purpose. It is argued [112] that the spread of two datasets can't be compared purely based on their variances or standard deviations, but instead on relative measures of spread.

Different types of processes exist in order to model the number of claims. A few are discussed here.

2.2.1 Poisson Process

The Poisson process is a member of the class of renewal processes. In this case it is assumed that some renewal process of times of claim arrivals

$\{\tau_n, n \geq 1\}$ generates a series of inter-occurrence times, $\{T_n, n \geq 1\}$. The series of inter-occurrence times is assumed to consist of non-negative values being independent from each other for subsequent values of n and follows the same distribution as some generic variable T.

Definition 5. *A continuous time process $\{N(t), t \in [0, \infty)\}$ is said to have stationary increments if the distribution of $(N(t) - N(s))$ for $t > s$ depends only on the length of the interval $t - s$. Furthermore this stochastic process is said to have independent increments if increments for any set of disjoiny intervals are independent [75].*

The process is said to be a Poisson process with a parameter λ if it is a continuous time process $\{N(t), t \in [0, \infty)\}$ with state space $S = \{0, 1, 2, 3, \ldots\}$ satisfying the following properties:

- $N(0) = 0$
- $N(t)$ has independent increments [1] .
- $N(t)$ has stationary increments [2] , that is, the increments have a Poisson distribution with a parameter proportional to λ with a factor equal to the distance in time between two respective points in time. This means that for $\lambda > 0$

$$\mathrm{P}(N(t) - N(s) = k) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^k}{k!} \qquad (2.3)$$
$$\text{where } k = 0, 1, 2, \ldots \text{ and } \forall\, s < t.$$

From (2.3) it follows that:

$$\mathrm{P}(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \text{ where } k = 0, 1, 2, \ldots \text{ and } \forall\, t > 0.$$

This is exactly the probability mass function of a random variable from a Poisson distribution with parameter λt. It can be said that $N(t) \sim$POI(λt) and therefore E$(N(t))$ =var$(N(t)) = \lambda t$.
It is proven below that all inter-occurrence times are distributed similarly to a generic random variable, T, and that the time from one claim to the next possesses a no-memory property and is Exponentially distributed with

[1]For any partition $0 \leq t_0 < t_1 < \ldots < t_n < \infty$ and any $k \in \mathbb{N}$ the random variables $N(t_1) - N(t_0)$, $N(t_2) - N(t_1)$, $\ldots$, $N(t_k) - N(t_{k-1})$ are independent [39].

[2]For any $0 < s < t < \infty$ and any $k \in \mathbb{N}$ and an increment $h > 0$ it is true that $\mathrm{P}(N(t + h) - N(t) = k) = \mathrm{P}(N(s + h) - N(s) = k)$ [39].

parameter λ for which it is known that the expected time from the one claim arrival to the next is given by $\frac{1}{\lambda}$. The Exponential distribution is and its properties are defined in Section 4.1.2.
Proof:
Assuming that all inter-occurrence times are distributed to some generic random variable, T, defined as the time from the previous claim to the next in any stage of the process. Let T^* be the random variable for the time from the time s until the next claim occurs with s being any point in time after the previous claim, but before the next claim.

$$\mathrm{P}(T^* > t|N(s) = i) = \mathrm{P}(N(s+t) = i|N(s) = i) = e^{-\lambda t}.$$

This probability is independent of the value of i and therefore

$$\mathrm{P}(T^* > t) = \mathrm{P}(T^* > t|N(s) = i) = e^{-\lambda t}.$$

This then gives that

$$\mathrm{P}(T^* \leq t) = 1 - \mathrm{P}(T^* > t) = 1 - e^{-\lambda t},$$

meaning that $T^* \sim EXP(\lambda)$.
The distribution of the time from the previous claim until the next can be derived in a similar way. Suppose the i^{th} claim occurred at time τ_i. To find the distribution of the time from the i^{th} claim until the $(i+1)^{th}$ claim, T:

$$\mathrm{P}(T > t|N(\tau_i) = i) = \mathrm{P}(N(\tau_i + t) = i|N(\tau_i) = i) = e^{-\lambda t}.$$

Again this probability is independent of the value of i and therefore
$T \sim EXP(\lambda)$.

For the Poisson process the index of dispersion is 1, which is a special case, since

$$I(t) = \frac{\mathrm{var}(N(t))}{\mathrm{E}(N(t))} = 1. \tag{2.4}$$

This will always be true for the Poisson distribution irrespective of the value of its parameter. This implies that the homogeneity of the data under all Poisson distributions are the same. For other distributions this will generally be dependent on the values of the parameters.

2.2.2 Mixed Poisson Process

For a Poisson process the assumption is that the mean number of claims within a given period is a constant multiple, λ, of the length of that period. This period, t, can be considered to be, without loss of generality, 1 year. In many instances the mean number of claims may not necessarily be a fixed value. A mixed Poisson process involves the expectation of Poisson probabilities for which the parameter itself is random [39]. This will be true whenever the business is penetrating segments within the market that they did not previously have business with or when there is seasonality in the claims, for example when a particular year has seen more erratic weather conditions leading to policyholders incurring losses as a result thereof etc.

The distribution function of the claim numbers, $N(t)$, can be derived using arguments from Bayesian methods by treating the mean number of claims per period, λ, as a random variable itself. Having said that $N(t) \sim$POI(λt), λ is now an outcome of the random variable Λ with distribution function $F_\Lambda(\lambda)$ which is called the *mixing distribution*. The distribution of the claim size is therefore conditional on the outcome of λ [105]. Finding the marginal distribution of $N(t)$ yields

$$p_k(t) = \int_\Lambda \frac{e^{-\lambda t}(\lambda t)^k}{k!} f_\Lambda(\lambda)\,\mathrm{d}\lambda.$$

This gives the mixture or mixed distribution of $N(t)$ [11].
The mean and variance of $N(t)$ can be derived as follows:

$$\mathrm{E}(N(t)) = \sum_{k=0}^{\infty} k \times p_k(t) = \int_0^\infty \sum_{k=0}^{\infty} k \frac{e^{-\lambda t}(\lambda t)^k}{k!} f_\Lambda(\lambda)\,\mathrm{d}\lambda$$

This follows under assumption that the orders of integration and summation may interchange. The expression $\frac{e^{-\lambda t}(\lambda t)^k}{k!}$ is the probability mass function of a random variable, K, with a POI(λt) distribution. Furthermore we have that [11]

$$\mathrm{E}(K) = \sum_{k=0}^{\infty} k \frac{e^{-\lambda t}(\lambda t)^k}{k!} = \lambda t.$$

The expression $\sum_{k=0}^{\infty} k \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ in the integral can thus be substituted with λt to give

$$\mathrm{E}(N(t)) = \int_0^\infty \lambda t f_\Lambda(\lambda)\,\mathrm{d}\lambda = t\,\mathrm{E}(\Lambda)$$

and (from Bain and Engelhardt [11])

$$\begin{aligned}
\mathrm{E}(N(t)^2) &= \int_0^\infty \mathrm{E}((N(t))^2|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda \\
&= \int_0^\infty (\,\mathrm{var}(N(t)|\lambda) + \mathrm{E}(N(t)|\lambda)^2) f_\Lambda(\lambda)\,\mathrm{d}\lambda \\
&= t\,\mathrm{E}(\Lambda) + t^2\,\mathrm{E}(\Lambda^2).
\end{aligned}$$

From this follows that

$$\mathrm{var}(N(t)) = t\,\mathrm{E}(\Lambda) + t^2\,\mathrm{var}(\Lambda).$$

Also the index of dispersion is given by

$$I(t) = \frac{\mathrm{var}(N(t))}{\mathrm{E}(N(t))} = 1 + t\frac{\mathrm{var}(\Lambda)}{\mathrm{E}(\Lambda)}.$$

The index of dispersion is therefore not necessarily 1 (the Poisson process is a special case). It is however true that for the mixed Poisson process the index of dispersion is 1 if and only if the distribution of Λ is degenerate at some value λ_0 [105] .

Other models exist to model the number of claim arrivals at some point in time, t. Some of these methods include:

- The Sparre-Anderson model [4].
  When one considers a renewal counting process with inter-occurrence times being independent, identically distributed with distribution function F_T and also considers independent, identically distributed claim sizes with the distribution function F_Θ which is also independent of the inter-occurrence times, the model is called the Sparre-Anderson model [105].

- Compound Poisson process [43], [39], [75]

- Recursively defined claim number distributions which includes the Poisson, negative binomial and binomial distributions which satisfies Panjer's recurrence relation [39]

- Claim number processes as Markov processes which satisfies the Kolmogorov differential equation, such as the Pure Birth process [75], [29], [105].

2.3 Claim size

Let the claim arriving at time τ_n be of size Θ_n. The sequence of claim sizes, $\{\Theta_n, n = 1, 2, 3, \ldots\}$ is usually assumed to be made up of claims with sizes being independent and identically distributed, although this may not always be true.

Claim sizes are described by means of a distribution or more specifically a loss distribution. These risks taken on by an insurer are said to be dangerous whenever the claim size distribution has a heavy tail, which means that there exists relatively large probabilities of larger claims to arrive within the portfolio.

When large claims are possible, but with the chance of such an occurrence showing an exponential decline as the size of the claims increase, the risk or portfolio of risks are said to have a well-behaved distribution and generally satisfy the following relationship: $1 - \mathrm{F}_\Theta(\theta) \leq ce^{-a\theta}$ for some positive-valued c and a where Θ is a generic, nonnegative claim size random variable and $\mathrm{F}_\Theta(\theta)$ the distribution function of the claim sizes. To see why this bound makes sense, consider an arbitrary large claim of size θ^*. Then

$$\mathrm{P}(\Theta > \theta^*) = 1 - \mathrm{P}(\Theta \leq \theta^*) = 1 - \mathrm{F}_\Theta(\theta^*) \leq ce^{-a\theta^*}.$$

It is desired to have the above probability to tend to 0 as the claim size increase or becomes too large. If the loss distribution does satisfy this inequality, it is true that

$$\lim_{\theta^* \to \infty} \mathrm{P}(\Theta > \theta^*) = \lim_{\theta^* \to \infty} (1 - \mathrm{F}_\Theta(\theta^*)) \leq \lim_{\theta^* \to \infty} ce^{-a\theta^*} = 0. \tag{2.5}$$

For this reason it can be argued that a loss distribution is well-behaved if it is true that $1 - \mathrm{F}_\Theta(\theta) \leq ce^{-a\theta}$. To illustrate this concept, consider the following example.

Example 2. *Consider a random variable X that is Weibull distributed with parameters θ and β [75] with a cumulative dsitribution function given by (4.16). The Weibull distribution is introduced more formally in Section 4.1.8. It is argued in section 5.2.3 that the distribution has a heavy tail for values of β in the interval $(0, 1)$.*

If we suppose that $0 < \beta < 1$ and further suppose that there exists a value $c > 0$ such that

$$1 - F_X(x) \leq e^{-cx}, \tag{2.6}$$

which implies that

$$e^{-\left(\frac{x}{\theta}\right)^{\beta}} \leq e^{-cx}$$
$$\therefore \left(\frac{x^{\beta-1}}{\theta^{\beta}}\right) \geq c.$$

Hence if $0 < \beta < 1$, *then*

$$\lim_{x\to\infty} c \leq \lim_{x\to\infty} \frac{x^{\beta-1}}{\theta^{\beta}}$$
$$= 0$$

which forms a contradiction to the fact that there exists a value $c > 0$. *Hence there doesn't exist an* $a > 0$ *and* $c > 0$ *such that*

$$1 - F_X(x) \leq ae^{-cx}.$$

It can therefore be seen that the Weibull distribution with parameter values for $\beta \in (0, 1)$ *is not considered to be well-behaved.*

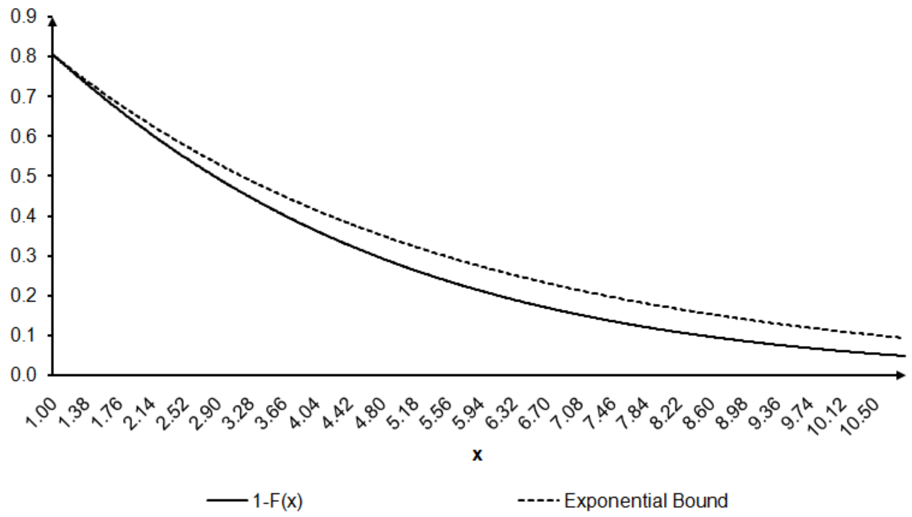

Figure 2.3: Comparison of Weibull ($\theta = 4$, $\beta = 1.1$) Upper Tail and Exponential Bound ($e^{-4^{-1.1}}$)

Conversely, if we consider the case where $\beta > 1$, then we can derive the following inqulity for values of $x > 1$:

$$\begin{aligned} 1 - F_X(x) &= e^{-\left(\frac{x}{\theta}\right)^{\beta}} \\ &< e^{-\left(\frac{1}{\theta}\right)^{\beta} x} \ \textit{if } x > 1 \end{aligned}$$

This then gives that there exists an $a = \theta^{-\beta}$ and $c = 1$ such that $(1 - F(x)) < ce^{-ax}$ for all $x > 1$. This is graphically illustrated in Figure 2.3.

In Chapter 3 the concept of an exponentially bounded upper tail is formally introduced. The condition stated here for a distribution to be considered as well-behaved is the same as the condition for a distribution to have and exponentially bounded upper tail [105].

In some instances insurance contracts are agreed for which the limit of insurance cover provided significantly exceeds the sizes of historically observed losses. In such a case one will have to find the probability of experiencing larger lossess than historically observed based on the actual historical losses. Wang [119] states that using extreme value theory can be useful in exptrapolating probabilities of observing these larger losses from the observed probabilities of the smaller losses.

2.4 Aggregate claim amount

The aggregate claim amount at time t depends on both the number of claims up to time t as well as the size of all those claims and is denoted $X(t)$ where

$$X(t) = \sum_{i=1}^{N(t)} \Theta_i. \tag{2.7}$$

Hence the cumulative distribution function is given by

$$\mathrm{F}_{X(t)}(x) = \mathrm{P}(X(t) \leq x) = \mathrm{P}\left(\sum_{i=1}^{N(t)} \Theta_i \leq x\right)$$

Define $X(t) = 0$ if $N(t) = 0$. Therefore $X(t)$ is a random sum (since $N(t)$ is stochastic) of random variables Θ_i for $i = 1, 2, \ldots, N(t)$.

Ideally the interdependence between $\{N(t); t \geq 0\}$ and $\{\Theta_n; n = 1, 2, \ldots, N(t)\}$ should be studied and made allowance for, but most often these two processes are treated as if they are stochastically independent.

Definition 6. ***k-fold convolution****. The distribution of the sum of two independent random variables, X and Y, can be calculated from their respective distributions, $\mathrm{F}_X(x)$ and $\mathrm{F}_Y(y)$. This distribution is called the convolution of F_X and F_Y, $\mathrm{F}_X * \mathrm{F}_Y$, which is defined [105] by*

$$\mathrm{F}_X * \mathrm{F}_Y(x) \int_{-\infty}^{\infty} \mathrm{F}_X(x-\phi) f_Y(\phi) d\phi. \tag{2.8}$$

Furthermore, the n-fold convolution of F_X is defined iteratively as follows:

$$\begin{aligned}
&\text{For } n=0: && \mathrm{F}_X^{*0}(x) = \delta_0(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases} \\
&\text{for } n=1: && \mathrm{F}_X^{*1}(x) = \mathrm{F}_X(x), \\
&\text{for } n=2: && \mathrm{F}_X^{*2}(x) = \mathrm{F}_X * \mathrm{F}_X(x), \\
&\text{for } n=3: && \mathrm{F}_X^{*3}(x) = \mathrm{F}_X^{*2} * \mathrm{F}_X(x), \\
&\quad\vdots && \\
&\text{and in general } && \mathrm{F}_X^{*n}(x) = \mathrm{F}_X^{*(n-1)} * \mathrm{F}_X(x).
\end{aligned}$$

Using the law of total probability and the property of a k-fold convolution being iteratively defined and determined,

$$\begin{aligned}
\mathrm{F}_{X(t)}(x) &= \mathrm{P}(X(t) \leq x) \\
&= \mathrm{P}\left(\bigcup_{k=1}^{\infty} \left\{ k \text{ claims } \ni \sum_{i=1}^{k} \Theta_i \leq x \right\} \right) \\
&= \mathrm{P}\left(\bigcup_{k=1}^{\infty} \left\{ \left\{ k \text{ claims} \right\} \bigcap \left\{ \sum_{i=1}^{k} \Theta_i \leq x \right\} \right\} \right) \\
&= \sum_{k=0}^{\infty} \mathrm{P}\left(\left\{ N(t) = k \right\} \bigcap \left\{ \sum_{i=1}^{k} \Theta_i \leq x \right\} \right)
\end{aligned}$$

since the events are disjoint and if $k=0$ then $\sum_{i=1}^{k} \Theta_i = 0$.

If it is further assumed that the claim number process, $\{N(t)\}_{t\geq 0}$, is independent from the claim sizes, $\{\Theta_i\}_{i\in\mathbb{N}}$, then the aggregate claim distribution can be written as

$$\begin{aligned}
\mathrm{F}_{X(t)}(x) &= \sum_{k=0}^{\infty} \mathrm{P}(N(t)=k) \times \mathrm{P}\left(\sum_{i=1}^{k} \Theta_i \leq x\right) \\
&= \sum_{k=0}^{\infty} \mathrm{P}(N(t)=k) \times \mathrm{P}(\Theta_1 + \Theta_2 + \ldots + \Theta_k \leq x) \\
&= \sum_{k=0}^{\infty} \mathrm{P}(N(t)=k) \times \mathrm{P}(\Theta + \Theta + \ldots + \Theta \leq x)
\end{aligned}$$

assuming that the claims are identically distributed. Thus

$$\mathrm{F}_{X(t)}(x) = \sum_{k=0}^{\infty} p_k(t) \times \mathrm{F}_{\Theta}^{*k}(x)$$

where $\mathrm{F}_{\Theta}^{*k}(x)$ is the k-fold convolution of F_{Θ}. Note that when the individual claims are assumed to be independent and identically Gamma distributed, the aggregate claim amount of a fixed number of claims, n, can be modelled using the Dirichlet distribution that is defined in Section 4.2.3.

Definition 7. ***Laplace-Stieltjes Transform****. For a probability function, $G(t)$, for a random variable, T, the Laplace-Stieltjes transform is defined as [73], [15]:*

$$\hat{\mathcal{L}}_T(s) = \int_0^{\infty} e^{-st} dG(t) = \int_0^{\infty} e^{-st} g(t) dt = \mathrm{E}(e^{-st}). \tag{2.9}$$

The properties of the aggregate claim amount can also be expressed in terms of a Laplace-Stieltjes transform which is defined as follows:

$$\hat{\mathcal{L}}_{X(t)}(s) = \mathrm{E}\left(e^{-sX(t)}\right) = \mathrm{E}\left(e^{-s\sum_{i=1}^{N(t)} \Theta_i}\right).$$

Defining the probability generating function of the number of claims, $N(t)$, as $\hat{g}_{N(t)}(z) = \sum_{i=0}^{\infty} p_i(t) z^i$ where $p_i(t) = \mathrm{P}(N(t) = i)$ allows one to decompose the Laplace-Stieltjes transform of the aggregate claim amount as the product of the probability generating function of the number of claims with the Laplace-Stieltjes transform of the individual claim sizes under the assumption that the claim sizes are independent and identically distributed:

$$
\begin{aligned}
\hat{\mathcal{L}}_{X(t)}(s) &= \mathrm{E}\left(e^{-sX(t)}\right) \\
&= \mathrm{E}_{N(t)}\left(\mathrm{E}\left(e^{-sX(t)}|N(t)\right)\right) \\
&= \mathrm{E}_{N(t)}\left(\mathrm{E}\left(e^{-(s\Theta_1+s\Theta_2+s\Theta_3+\ldots+s\Theta_{N(t)})}\right)\right) \\
&= \mathrm{E}_{N(t)}\left(\mathrm{E}\left(e^{-(s\Theta_1)}\right)\times \mathrm{E}\left(e^{-(s\Theta_2)}\right)\times\ldots\times \mathrm{E}\left(e^{-(s\Theta_{N(t)})}\right)\right) \\
&\quad \text{assuming claim sizes are independent} \\
&= \mathrm{E}_{N(t)}\left(\mathrm{E}\left(e^{-(s\Theta)}\right)\times \mathrm{E}\left(e^{-(s\Theta)}\right)\times\ldots\times \mathrm{E}\left(e^{-(s\Theta)}\right)\right) \\
&\quad \text{assuming claim sizes are identically distributed} \\
&= \mathrm{E}_{N(t)}\left(\mathrm{E}\left(e^{-(s\Theta)}\right)^{N(t)}\right).
\end{aligned}
$$

Using the definition of a Laplace-Stieltjes transform for the claim sizes, the Laplace-Stieltjes transform for the aggregate claim amount can now be written [75] as

$$
\hat{\mathcal{L}}_{X(t)}(s) = \mathrm{E}_{N(t)}\left(\left(\hat{\mathcal{L}}_{\Theta(s)}\right)^{N(t)}\right) = \sum_{i=0}^{\infty}\left(\hat{\mathcal{L}}_{\Theta}(s)\right)^{i}\times \mathrm{P}\left(N(t)=i\right) = \hat{g}_{N(t)}\left(\hat{\mathcal{L}}_{\Theta}(s)\right).
$$

Other alternatives for studying the claim size and its distribution include:

- Central limit theorem approximations [75] which may be unreliable when the true claim distribution is heavy-tailed.
- Fitting claim frequency and claim size distributions to data. This approach is suitable for modeling aggregate claims sizes as described above in which case the expression of the Laplace-Stieltjes transform given above can be used [75].
- Methods of risk comparison can also be used. If different risk classes have different distribution forms, one can compare aspects such as the tail heaviness, the index of dispersion to understand the spread of potential claims sizes and the stochastic order of these risks [105], [88], [75]. Concepts of the tail weight of distributions and stochastic orders are discussed in Chapter 3.

2.5 Premium Income

The premium charged to a policyholder is to provide cover to the policyholder against the estimated future losses resulting from the insured risks

[39]. From the insurer's point-of-view the premium income up to some time t is denoted by $\Pi(t)$. Premiums should be set up or calculated such that they accumulate quickly enough in order to meet claims as they arrive. In general, allowance should be made for covering administration costs, profit and other margins within the premium calculation. The accumulated premium up to time t is therefore given by:

$$\Pi(t) = (1 + \eta)\,\mathrm{E}\,(N(t))\,\,\mathrm{E}\,(\Theta)$$

where η is the safety loading to account for costs and profit and other margins [105] while the component given by $\mathrm{E}\,(N(t))\,\,\mathrm{E}\,(\Theta)$ is referred to as the pure premium [39]. This will be true if it can be assumed that the claim sizes are all independent and identically distributed like some generic claim size random variable, Θ, and if the process of the claim numbers and the claim size process are stochastically independent.

2.6 Risk Reserve

The risk reserve at a given time, t, is a function of the initial capital, u, the premiums received, $\Pi(t)$, and the realised claims up to time t, $X(t)$:

$$R(t) = u + \Pi(t) - X(t).$$

The premiums collected in the n^{th} collection period (usually months) is given by $\Pi_n = \Pi(n) - \Pi(n-1)$ while the total amount of claims reported in the n^{th} period is given by $X_n = X(n) - X(n-1)$. Therefore the reserve at the end of the n^{th} period is given by

$$R_n = R(n) = u + \Pi(n) - X(n) = u - \sum_{i=1}^{n}(X_i - \Pi_i).$$

Let $Y_n = X_n - \Pi_n$. Then Y_n is the surplus of the claims to the premiums accumulated during the n^{th} period. The process $\{Y_n\}_{n\geq 0}$ consists of random variables which are independent and identically distributed. Therefore the sum of these surpluses, $S_n = \sum_{i=1}^{n} Y_i$ yields a process $\{S_n\}_{n\geq 0}$ which is a random walk.

Assuming that the premiums are non-random, the probability that the reserve at some point in time t less than or equal to x is given by

$$\mathrm{P}(R(t) < x) = \mathrm{P}(u + \Pi(t) - X(t) \leq x) = \mathrm{F}_{X(t)}(\Pi(t) + u - x).$$

Understanding the underlying processes of the number of claims and claim sizes gives insight into the total claim amount that may arise within a particular period. Linking that with the premiums collected within that period allows one to get an idea of the solvability of the portfolio of risks underwritten. Historical experience can be used to develop models to estimate probabilities of policy lapses or of irregular payments, say where the policyholder skips a payment, but making up that payment in the following period. Since a portion of collected premiums are typically invested to generate profits for the business and to meet the claims on risks that are aligned with inflation, the macro-economic environment poses additional uncertainty and randomness to the premiums accumulated.

Let I_i be an index to denote the change in prices of an insured good relative to its price or cost at inception of the insurance policy or contract at times $i = 1, 2, 3, \ldots, n$. The average aggregate claim size can then be scaled using these indexes to represent an average aggregate claim amount at inception of the policy or contract [105]:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{I_i}.$$

Similarly for the variance of the aggregate claim amount:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i}{I_i} - \bar{X}\right)^2.$$

The indices can be estimated, say from economic factors with time-series analysis or interest rate models for example. Thus when the scaled average of the aggregate claims is known for the last n periods, an estimate for the aggregate claim amount can be found for the $(n+1)^{th}$ period as [105]

$$\hat{X}_{n+1} = \hat{I}_{n+1}\bar{X} \text{ and } \sigma^2_{n+1} = \hat{I}^2_{n+1}\sigma^2.$$

2.7 Reinsurance

Reasons for needing to cede claims liability or part thereof to a reinsurer include:

- To provide protection against excessively large claims
- To provide protection against a large number of claims close to each other in time or arriving almost simultaneously
- To provide protection against change, especially deterioration, in premium collection patterns

- Legal restriction in terms of reserving and holding provisions for solvency may limit the capacity of smaller insurers to offer competitive products at competitive premiums. Reinsurance can increase the business's capacity in order to offer and issue more products.

- It provides increased capacity for an insurance of any size insurer.

- Enables an insurer to enter a new branch of the business or introduce new products and be reasonably protected against risks that they have limited knowledge and no claims experience of.

Different types of reinsurance treaties exist. The key idea is for the insurer who issues the product or policy - called the direct insurer - to retain a portion of the claims and to cede the remaining portion to the reinsurer. The portion retained may be a fixed amount, called the deductible or else be a proportion of the claim amount. These two approaches may come into effect either per individual claim or on the aggregate claim amount. The retained portion is denoted by either $h(X(t))$ or $h(\Theta_i)$ for the aggregate and individual cases respectively. Hence the reinsured portion is simply $X(t) - h(X(t))$ or $\Theta_i - h(\Theta_i)$. The reinsurer can again enter a reinsurance treaty with a second reinsurer; therefore creating a reinsurance chain. The most common types of reinsurance treaties used are given below.

Proportional/Quota Share Reinsurance

A fixed proportion of the value of the risks, $1 - a$, is reinsured. This gives the reinsured portion of an aggregate claim amount [107], [114], [106] as

$$(1-a)X(t) = (1-a)\sum_{i=1}^{N(t)} \Theta_i = \sum_{i=1}^{N(t)} (1-a)\Theta_i.$$

which is the sum of proportions of the individual claims. From this it follows that

$$\mathrm{P}[(1-a)X(t) \le x] = F_{X(t)}\left(\frac{x}{1-a}\right). \tag{2.10}$$

This is a special case where a is constant for all risks [105]. Where a is not constant, the reinsurance agreement is referred to as *Surplus Reinsurance*.

Excess Loss Reinsurance

The effect of this treaty is a reduction in the mean amount and the variance of the amount paid in respect of the insurer's liability due to claims.
A fixed monetary value, b, is used as threshold - termed the retention level, [106]. The retention level is applied per individual claim and reinsurance will

therefore come into effect whenever an individual claim amount exceeds this retention level. An upper limit may be implemented from the reinsurer's side in order to limit the reinsurer's liability at some value c such that the liability is limited to $c-b$ where claims above the second retention level may be ceded to a second-in-line reinsurer or may again be the liability of the direct insurer [75], [105].
The reinsured amount of the direct insurer is therefore given by

$$\sum_{i=1}^{N(t)} (\Theta_i - b)_+ \quad \text{where } x_+ = \max\{x, 0\}.$$

The retention level may also be linked to some inflation index.

Stop-Loss Reinsurance

This arrangement is similar to that of Excess Loss Reinsurance by having a retention level, b, but with the difference that it is applied to the aggregate claim amount. Therefore it provides an advantage in that smaller claims can now be aggregated to contribute in order for the reinsurance to come into effect [105], [75]. The reinsured amount of the direct insurer is therefore given by

$$\left(\sum_{i=1}^{N(t)} \Theta_i - b\right)_+ \quad \text{where } x_+ = \max\{x, 0\} \tag{2.11}$$

ECOMOR Reinsurance

The acronym refers to the Latin phrase describing this agreement, *excédent du coût moyen relative*. If the individual claims are arranged from the smallest to the largest claim size and then being denoted as order statistics $(\Theta_{(1)}, \Theta_{(2)}, \Theta_{(3)}, \ldots, \Theta_{(N(t)-r-1)}, \Theta_{(N(t)-r)}, \Theta_{(N(t)-r+1)}, \ldots, \Theta_{(N(t))})$. The r largest claims, $\Theta_{(N(t))}, \Theta_{(N(t)-1)}, \Theta_{(N(t)-2)}, \ldots, \Theta_{(N(t)-r+1)}$, are reinsured with a random retention level being equal to the $(r+1)^{th}$ largest claim , $\Theta_{(N(t)-r)}$ [105], [48]. Thus the reinsured amount is given by

$$\begin{aligned} Z(t) &= \sum_{i=1}^{r} \left(\Theta_{(N(t)-i+1)} - \Theta_{(N(t)-r)}\right) \\ &= \sum_{i=1}^{N(t)} \left(\Theta_i - \Theta_{(N(t)-r)}\right)_+ \quad \text{where } x_+ = \max\{x, 0\}. \end{aligned}$$

2.8 Ruin Problems

When starting a new branch of insurance products, an initial amount of capital, u, is put at risk. The aim will then be to minimize the probability that the claims surplus, given by $X(t) - \Pi(t)$, exceeds the initial capital amount. If the claims surplus exceeds the initial capital amount then it means that $R(t) = u + \Pi(t) - X(t) < 0$. This occurrence is called *ruin*.
The ruin time, denoted by ζ, is defined [7] as

$$\zeta = \inf\{t \geq 0 : R(t) < 0\} = \zeta(u). \tag{2.12}$$

The ruin time depends on the stochastic elements making up $R(t)$.
For a finite time horizon the probability of no ruin or survival probability [42], [75] is given by

$$\psi(u, x) = \mathrm{P}\left(\inf_{0 \leq t \leq x} R(t) \geq 0\right) = \mathrm{P}(\zeta(u) > x) \tag{2.13}$$

and for an infinite time horizon

$$\psi(u) = \mathrm{P}\left(\inf_{0 \leq t} R(t) \geq 0\right) = \mathrm{P}(\zeta(u) = \infty). \tag{2.14}$$

Finding estimates of ruin probabilities and determining distributions for the event of ruin is difficult. In the literature bounds for ultimate ruin are given by Chen et al [27], [28]. Ruin probabilities are largely influenced by the claim size distribution, especially when the distribution is heavy-tailed. Wei and Hu [120] studied approximations for the ruin probabilities under the occurrence of heavy-tailed claims. The probability of ruin is obviously also dependent on the level of initial capital.

Chapter 3

Theory and Properties of Distributions

3.1 Introduction

The aim of this chapter is to gather and introduce measures and methods that can be used to describe parametric distributions in terms of the location, spread and skewness together with tail heaviness. We introduce an argument below, based on our own reasoning, on how to detect and compare tail heaviness after which we present formal theoretical results from the literature that can be used to evaluate tail heaviness in Sections 3.2 to 3.7.

Let the probability density function and the cumulative distribution function for a random variable X be given by $f_X(x)$ and $F_X(x)$, respectively, with $X \in \mathbb{R}^+$. We would like to evaluate the behaviour of $f_X(x)$ and $F_X(x)$ in particular for large values of X; that is evaluating the upper tail of the function $\overline{F}_X(x) = (1 - F_X(x))$.

When a unimodal distribution exhibits skewness, the widths of the ranges on both sides of the mode, say X_{mode}, are not the same. If a distribution exhibits right skewness, the range of values larger than the mode is wider than the range of values smaller than the mode. Furthermore the probability densities attached to the values in this upper range are typically lower due to the fact that a larger range of values is now present than compared to the lower range. Lastly the probability density function will be monotone decreasing over the range (X_{mode}, ∞).

Consider the comparison of the probability density functions of an Exponential (with $\theta = 2$) and Weibull (with $\beta = 0.95, \theta = 3$) distribution as given in Figure 3.1 - the Exponential and Weibull distributions are formally

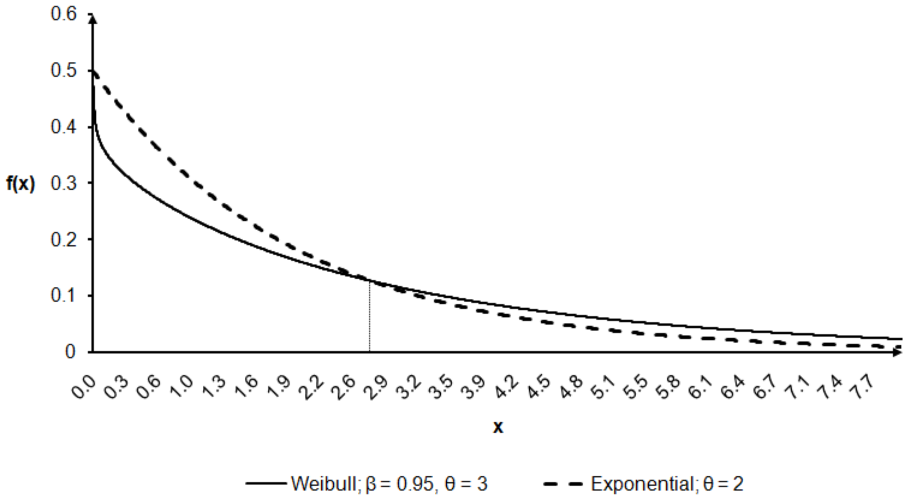

Figure 3.1: Comparison of Exponential ($\theta = 2$) and Weibull ($\beta = 0.95, \theta = 3$) probability density functions

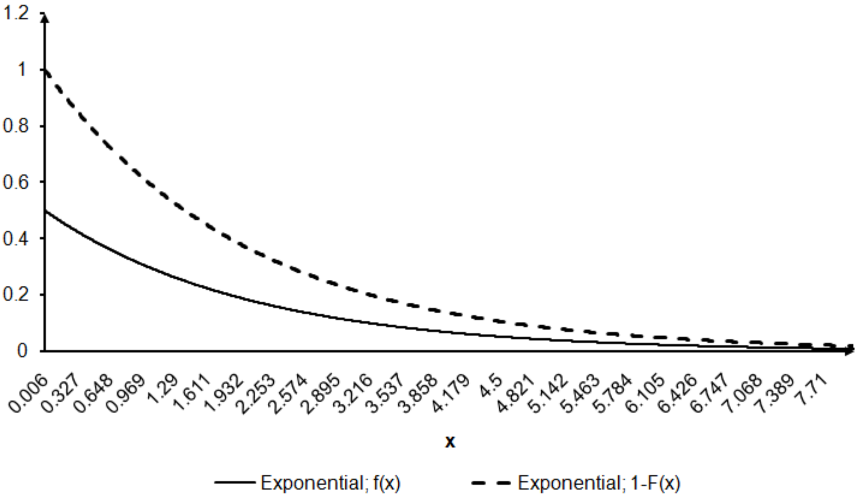

Figure 3.2: Comparison of the probability density and cumulative distribution functions for Exponential distribution with $\theta = 2$

defined in Sections in 4.1.2 and 4.1.8. From this figure it can be seen that for a value of x of approximately 2.8 the two probability density functions are equal (as indicated by the vertical reference line on the graphical comparison) whereafter the probability density function of the Weibull distribution remains larger than that of the Exponential distribution. It also shows that

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

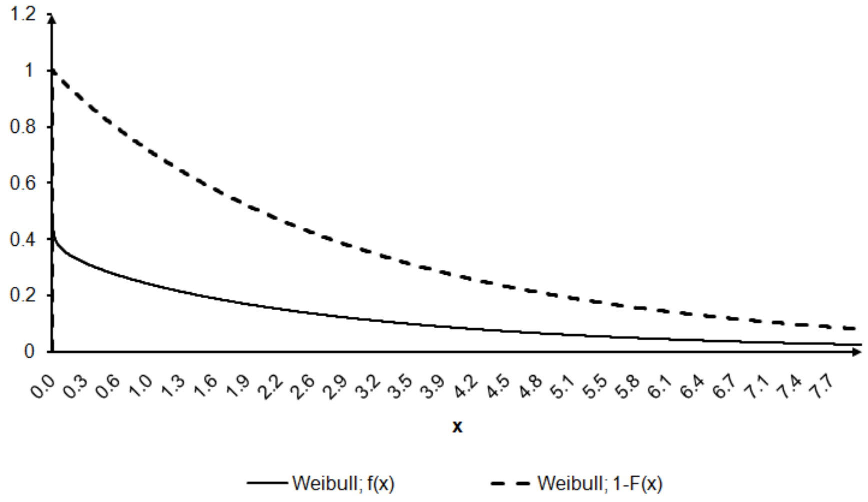

Figure 3.3: Comparison of the probability density and cumulative distribution functions for Weibull distribution with $\beta = 0.95$ and $\theta = 3$

the upper tail weight at this point is larger for the Weibull distribution than for the Exponential distribution.

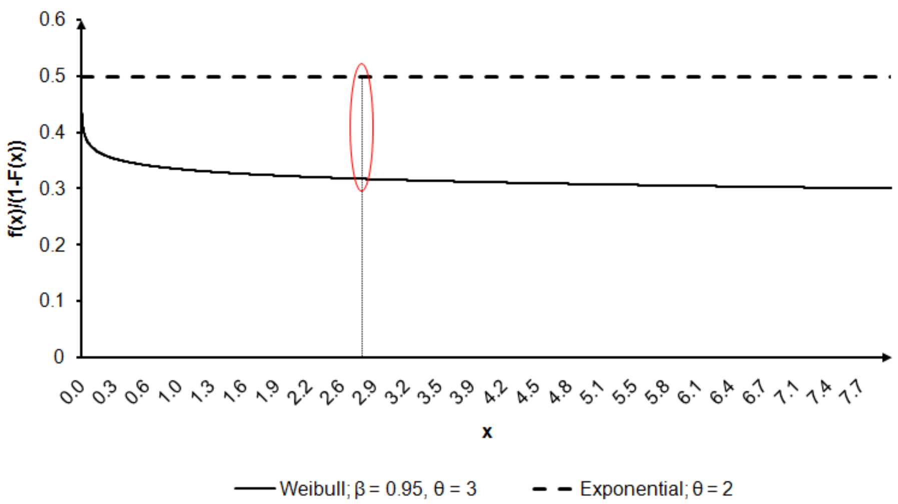

Figure 3.4: Comparison of ratios of Exponential ($\theta = 2$) and Weibull ($\beta = 0.95, \theta = 3$) probability density functions to upper tail weights

If a distribution has a heavy tail, it means that as x gets larger the probability density becomes smaller while the upper tail weight $1 - F_X(x) = \overline{F}_X(x)$ tends to not go towards a value of 0 as quickly. This suggests that after some point x^* the probability density function $f_X(x^* + t)$ goes to 0 much faster than $1 - F_X(x^* + t)$ as t gets larger. This suggests that the ratio of the probability density function to $\overline{F}_X(x)$ will be a decreasing function of t. This is illustrated in Figures 3.2 and 3.3 where it can be seen that for the Exponential distribution that both functions move towards 0 at a similar rate whilst it is evident that for the Weibull distribution the upper tail does not move to 0 as quickly for large values of x.

In Figure 3.4 the ratios of the probability density functions to the upper tail weights of the Exponential ($\theta = 2$) and Weibull($\beta = 0.95, \theta = 3$) distributions are graphically compared for various values of x:

From (B.34) and (4.1)

$$\frac{f_X(x)}{\overline{F}_X(x)} = \frac{1}{\theta} \text{ for } x \geq 0 \text{ (Exponential) and} \tag{3.1}$$

from (B.179) and (4.16)

$$\frac{f - X(x)}{\overline{F}_X(x)} = \frac{\beta}{\theta^\beta} x^{\beta - 1} \text{ for } x \geq 0 \text{ (Weibull),}$$

From this comparison it can be seen that for all values of x the ratio for the Weibull distribution remains smaller than the ratio for the Exponential distribution. This clearly indicates at the point where the densities of the distributions are equal that the upper tail weight of the Weibull distribution is larger than for the Exponential distribution. The ratio is constant for the Exponential distribution while it is consistently lower for the Weibull distribution and is consistently decreasing.

Figure 3.4 suggests that the Weibull distribution has a heavier tail than the Exponential distribution and can also be classified as being heavy-tailed. In Chapter 5 it will be shown mathematically that the Exponential distribution doesn't have a heavy tail, but that the Weibull distribution does have for values of β between 0 and 1.

Continuing our reasoning, suppose now that in general we consider any two different distributions, $F_X(\cdot)$ and $G_X(\cdot)$ and suppose one can find a value situated in the upper tail of the distribution, say x^* such that $f_X(x^*) = g_X(x^*)$. Consider the case where $\overline{F}_X(x^*) > \overline{G}_X(x^*)$ while $f_X(x^*) = g_X(x^*)$. This implies that, given the value of x^*, the weight in the upper tail of $F_X(\cdot)$ is larger than the weight in the upper tail of $G_X(\cdot)$. It follows then that:

$$\frac{g_X(x^*)}{\overline{G}_X(x^*)} > \frac{f_X(x^*)}{F_X(x^*)}$$

This comparison is, however, not necessarily true for all values of $x \geq x^*$, but instead only proven to hold at the point x^*.

One may need to consider whether the upper tail weight of the one distribution is in general larger than the weight of the other across all values in the upper tail; that is for any arbitrary value of x^*. If we have identified such a point where the tail weight of the one distribution $F(\cdot)$ exceeds that of the other distribution $G(\cdot)$, one can consider the following ratios:

$$h_F(a) = \frac{f_X(x^*+a)}{\overline{F}_X(x^*+a)} \text{ for any } a \geq 0,$$
$$h_G(a) = \frac{g_X(x^*+a)}{\overline{G}_X(x^*+a)} \text{ for any } a \geq 0.$$

If the slope of $h_F(a)$ is consistently less than the slope of $h_G(a)$, distribution $F_X(\cdot)$ will always have greater upper tail weight than distribution $G_X(\cdot)$. Mathematically it means that for any $a \geq 0$

$$\frac{f_X(x^*+a)}{F_X(x^*+a)} < \frac{g_X(x^*+a)}{G_X(x^*+a)}.$$

If this condition can be confirmed, then distribution $F_X(\cdot)$ has a heavier tail than distribution $G_X(\cdot)$ irrespective of which value x^* we consider.

Example 3. *Consider two Exponential distributions with parameters θ_1 and θ_2, respectively, where $0 < \theta_2 < \theta_1 < \infty$. The two pairs of probability density and cumulative distribution functions follow from (B.34) and (4.1) with $\theta = \theta_1$ and $\theta = \theta_2$, resepctively.*

In Figure B.34, shown in Chapter 4, it can be seen that for increasing values of parameter θ, the weight towards the upper tail of the distribution is increasing. For the purpose of this example consider Figure 3.5 where we compare the probability density functions for the cases where $\theta = 0.5$ and $\theta = 0.25$.

We can find the value x^* such that $f_X(x^*) = g_X(x^*)$; that is the value:

$$x^* = -\frac{\theta_1\theta_2 \ln\left(\frac{\theta_2}{\theta_1}\right)}{\theta_1 - \theta_2}.$$

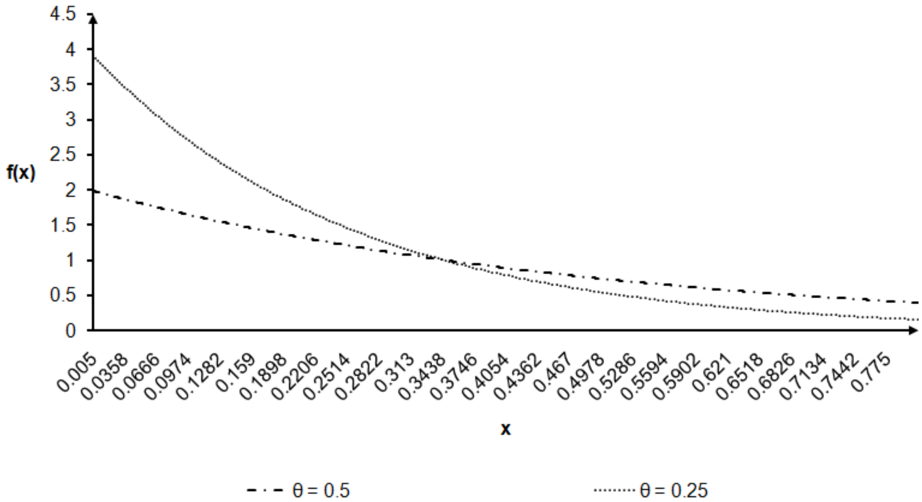

Figure 3.5: Comparison Exponential density functions for different values of θ

In the context of the values of θ considered in Figure 3.5, $x^* = 0.35$. Consider the ratio of $\overline{F}_X(x^*)$ to $\overline{G}_X(x^*)$ taken from (4.1):

$$
\begin{aligned}
\frac{\overline{F}_X(x^*)}{\overline{G}_X(x^*)} &= \frac{e^{-\frac{x^*}{\theta_1}}}{e^{-\frac{x^*}{\theta_2}}} \\
&= \frac{\theta_1}{\theta_2} > 1 \quad \forall x > 0
\end{aligned}
$$

This means that $\overline{F}_X(x^*) > \overline{G}_X(x^*)$. Furthermore

$$\frac{f_X(x)}{\overline{F}_X(x)} = \theta_1^{-1} \quad \forall\ x > 0$$

and

$$\frac{g_X(x)}{\overline{G}_X(x)} = \theta_2^{-1} \quad \forall\ x > 0$$

Therefore

$$\frac{f_X(x)}{\overline{F}_X(x)} < \frac{g_X(x)}{\overline{G}_X(x)} \quad \forall\ x \geq x^* > 0$$

which means that $F_X(\cdot)$ has a heavier tail than $G_X(\cdot)$.

Conversely, one can consider this evaluation of the tail weights of two distributions for a given upper tail weight α. We can find x_A and x_B such that

$\overline{F}_X(x_A) = \overline{G}_X(x_B) = \alpha$. One can then evaluate $f_X(x_A)$ and $g_X(x_B)$. If $f_X(x_A) > f_X(x_B)$, it suggests that for a given upper tail weight α distribution $G_X(\cdot)$ has a heavier tail than $F_X(\cdot)$. It means that the upper tail of $G_X(\cdot)$ is spread across a wider range of values on $\mathbb{R}^+$ than the upper tail of $F_X(\cdot)$.

Example 4. *As a continuation of example 3, consider Exponential distributions $F_X(\cdot)$ and $G_X(\cdot)$ with parameters θ_1 and θ_2, where $0 < \theta_2 < \theta_1 < \infty$. To find x_A and x_B such that $\overline{F}_X(x_A) = \alpha$ and $\overline{G}_X(x_B) = \alpha$. Solving these equations yields the following solutions:*

$$x_A = -\theta_1 \ln(\alpha)$$
$$x_B = -\theta_2 \ln(\alpha)$$

from which the probability densities can be evaluated:

$$f_X(x_A) = \frac{\ln(\alpha)}{\theta_1}$$
$$g_X(x_B) = \frac{\ln(\alpha)}{\theta_2}$$

From this follows that for the values associated with the specified upper tail weight α, x_A and x_B, we have that $f_X(x_A) < g_X(x_B)$ which suggests that $F_X(\cdot)$ has a heavier tail than $G_X(\cdot)$ at this point of the distribution.

This is testing one point only and does not yet prove heavier tail weight of $G_X(\cdot)$ for all values in its domain. One will have to consider a range of values for α and evaluate densities at various values of α to determine whether the conclusion holds over the full domain in order to conclude that distribution $F_X(\cdot)$ does in fact have a heavier tail than $G_X(\cdot)$.

In the remainder of this chapter theoretical statistical measures will be given that will be used in the chapter 4 in order to describe and analyse the characteristics of statistical distributions. Distributions of insurance claims typically exhibits skewness, as was discussed in Section 1.1. For this reason one needs to identify rules that can assist in identifying skewness and also help in ascertaining whether the heavy tails exist.

When considering claim size distributions large claim sizes are represented by the upper (right) tail. Klugman et al [75] highlight the importance of studying tails of distributions of losses (or claims in this study) due to the fact that the losses (claims) in the tail of the distribution will have the biggest impact on total loss (aggregate claim size). They further suggest that having an understanding of the tail weight of some distribution can help to choose or confirm a potential distribution to model the losses (claim sizes). It is therefore required to have an understanding of:

- whether a distribution exhibits skewness,
- whether the upper tail of the distributions is heavy and
- if multiple distributions are considered, then it is usefull to have a measure of comparing upper tail heaviness.

The remainder of this chapter is structured as follows:

- Definitions and formulae for calculating moments from parametric distributions are given in Section 3.2.
- Properties of hazard rates and hazard rate functions are introduced in Section 3.3. Based on these results the concept of stochastic order is introduced.
- Criteria that can be used to assess whether a distribution has a heavy tail are introduced in Section 3.4. This is done by using the properties of hazard rates and hazard rate functions as presented in Section 3.3 as well as considering the use of the moment generating function, the concept of distributions being exponentially bounded and comparitive limiting tail behavior of pairs of distributions.
- In Section 3.5 aspects relating to data where left censoring is present are discussed. In this context measures that can be used to test whether the assumption of heavy-tailedness holds, are given.
- A class of distributions, the subexponential class, is introduced in Section 3.6. It is then also showed that these distributions have heavy tails.
- The chapter is concluded in Section 3.7 with practical methods that can be used to detect whether the distributions of observed data are heavy-tailed.

Skewness and kurtosis properties of parametric distributions are given in Chapter 4 in the context of the coefficient of skewness and coefficient of kurtosis. Coefficients of skewness, kurtosis and variation are unitless measures and can therefore be used to compare distributions' skewness, kurtosis and spread (where spread is measured in terms of the coefficient of variation) [57], [75].

3.2 Moments of parametric distributions

In this section definitions are presented that can be used to calculate the moments from parametric distributions either directly or by using the moment generating function. The use of the formulae presented in these definitions are dependent on whether these moments exist and whether the moment generating function for a specific distribution exists.

Definition 8. *The k^{th} moment [11],[112] about the origin is defined as*

$$\mathrm{E}(X^k) = \begin{cases} \displaystyle\int_A x^k f_X(x)dx & \text{if } X \text{ is continuous,} \\ \displaystyle\sum_A x^k P(X = x) & \text{if } X \text{ is discrete.} \end{cases}$$

where A is the domain of X for $k \in \mathbb{Z}$.

Definition 9. *The k^{th} moment [11],[112] about the mean is defined as*

$$\mathrm{E}\left((X - \mathrm{E}(X))^k\right) = \begin{cases} \displaystyle\int_A (x-\mu)^k f_X(x)dx & \text{if } X \text{ is continuous,} \\ \displaystyle\sum_A (x-\mu)^k P(X = x) & \text{if } X \text{ is discrete.} \end{cases}$$

where A is the domain of X for $k \in \mathbb{Z}$.

Definition 10. *For a random variable X the moment generating function [11],[112] is given by*

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = \begin{cases} \displaystyle\int_A (e^{tx})^k f_X(x)dx & \text{if } X \text{ is continuous,} \\ \displaystyle\sum_A (e^{tx})^k P(X = x) & \text{if } X \text{ is discrete.} \end{cases}$$

The variance, skewness and kurtosis (second, third and fourth moments about the mean) can be calculated from the first four moments about the origin:

$$\mathrm{var}(X) = \mathrm{E}\left((X-\mu_X)^2\right) = \sigma_X^2 \text{ where } \mu_X = \mathrm{E}(X) \tag{3.2}$$

$$\mathrm{skewness}(X) = \frac{\mathrm{E}\left((X-\mu_X)^3\right)}{\sigma_X^3} \text{ where } \sigma_X = \sqrt{\mathrm{var}(x)} \tag{3.3}$$

$$\mathrm{kurtosis}(X) = \frac{\mathrm{E}\left((X-\mu)^4\right)}{\sigma_X^4} \tag{3.4}$$

3.3 Distributions with monotone hazard rates

In this section the hazard function and hazard rate function are introduced. The associated residual hazard distribution and mean residual hazard function are given. The section is concluded by deriving a relationship between the hazard rate and stochastic order of distributions.

Let X be a nonnegative, absolutely continuous [1] random variable .

Definition 11. ***Hazard function***
The hazard function of a random variable, X, with continuous density given by $F(\cdot)$ is defined as

$$h_X(x) = -\ln(1 - F_X(x)) = -\ln(\overline{F}_X(x)). \tag{3.5}$$

Definition 12. ***Hazard rate function***
The hazard rate function of X defined for $F_X(t) < 1$ [13] is given by

$$h_X^*(t) = \frac{f_X(t)}{1 - F_X(t)}. \tag{3.6}$$

If $h_X^*(t)$ is an increasing function of t, then the random variable is said to have an increasing hazard rate (IHR), while the random variable is said to have a decreasing hazard rate (DHR) if it is a decreasing function of t [105].

It follows from [105] that if the probability density function $f_X(x)$ is continuous, then $h_X(x)$ is differentiable and hence that $\frac{d}{dx}h_X(x) = h_X^*(x)$.

Definition 13. ***Residual hazard distribution***
The residual hazard rate distribution [105], $F_t(x)$, is given by

$$\begin{aligned} F_t(x) &= \frac{F_X(t+x) - F_X(t)}{1 - F_X(t)} \qquad (3.7)\\ &= \mathrm{P}(X - t \leq x | X > t) \ \textit{if}\ F_X(t) < 1 \qquad (3.8) \end{aligned}$$

In Theorem 1 it is given that an increasing hazard rate function is associated with a decreasing residual hazard rate function.

Definition 14. ***Mean residual hazard function***
The mean residual hazard function [105], [48] is given by

$$\begin{aligned} \mu_{F_t} &= \mathrm{E}(X - t | X > t)\\ &= \frac{\int_t^\infty \overline{F}_X(x)dx}{1 - F_X(t)} \ \textit{if}\ F_X(t) < 1 \qquad (3.9) \end{aligned}$$

[1] A random variable X is absolutely continuous if a measurable function [80] $f : \mathbb{R} \to \mathbb{R}^+$ exists such that $\int f(x)dx = 1$ and for each $M \in M(\mathbb{R})$ it is true that $\mathrm{P}(X \in M) = \int_M f(x)dx$. The distribution of X is called absolutely continuous [105].

The mean residual function essentially gives the average amount by a which the random variable X exceeds the value t. It therefore follows that if the density function moves to 0 at a very slow rate (which occurs for heavy-tailed distributions) that this function will increase with increasing values of t [24].

Definition 15. ***Stochastic Order*** *A distribution $F_X(\cdot)$ is said to be stochastically larger than a distribution $G_X(\cdot)$ if $\overline{F}_X(x) \geq \overline{G}_X(x)\ \forall\ x \in \mathbb{R}$. This can alternatively be written as $F \geq_{st} G$ [105].*

This definition implies that if the upper tail of the one distribution is larger than or equal to the upper tail of the other for all values of x, then the distribution is stochastically larger. It also means that the distribution has a heavier tail. Using the results of definitions 12 and 15, Rolski et al [105] give the following theorem. An alternative proof is given here.

Theorem 1. *The distribution F is said to have an increasing hazard rate if and only if $\forall\ t_1 \leq t_2$ it is true that $F_{t_1} \geq_{st} F_{t_2}$.*
Proof:

$$\begin{aligned} h_X^*(t) &= \frac{f(t)}{1-F(t)} \\ &= \frac{-\frac{d}{dt}(1-F(t))}{1-F(t)} \\ &= -\frac{d}{dt} ln(1-F(t)). \end{aligned}$$

Hence

$$1 - F_X(t) = e^{-\int_0^t h_X^*(s)ds} \tag{3.10}$$

From (3.10) and Definition 13 we have that

$$\begin{aligned} \overline{F}_t(x) &= \frac{\overline{F}_X(t+x)}{\overline{F}_X(t)} \\ &= \frac{e^{-\int_0^{t+x} h_X^*(s)ds}}{e^{-\int_0^{t} h_X^*(s)ds}} \\ &= e^{-\int_t^{t+x} h_X^*(s)ds} \end{aligned} \tag{3.11}$$

If $h_X^*(s)$ is increasing we have that $\overline{F}_t(x)$ is decreasing. Also for $t_1 \leq t_2$ we have that $\overline{F}_{t_1} \geq \overline{F}_{t_2} \forall x \in \mathbb{R}$. This implies that $F_{t_1} \geq_{st} F_{t_2}$.
Conversely, if for $t_1 \leq t_2$ we have that $\overline{F}_{t_1} \geq \overline{F}_{t_2} \forall x \in \mathbb{R}$, then $\overline{F}_t(x)$ is decreasing. From 3.11 follows that $h_X^*(s)$ is increasing.

3.4 Heavy-tailed distributions

Klugman et al [75] argue that knowing something about the tail weight of an underlying distribution may assist in narrowing down the choices of potential theoretical distributions that can be used to model the particular claims data. In this section techniques will be discussed that can be used to determine whether parametric distributions are heavy-tailed.

Definition 16. *A distribution $F(\cdot)$ for a random variable X is considered to be heavy-tailed [56] if and only if*

$$\int_{\mathbb{R}} e^{\lambda x} F(x)dx = \infty \text{ for all } \lambda > 0.$$

Definition 17. *A distribution $F(\cdot)$ for a random variable X is considered to be light-tailed [56] if and only if*

$$\int_{\mathbb{R}} e^{\lambda x} F(x)dx < \infty \text{ for some } \lambda > 0.$$

These two definitions make sense in that if a distribution has a heavy tail, there is a high likelihood to observe large values of x in which case the quantity $e^{\lambda x}$ can become very large and can tend to infinity. This further implies that the quantity $e^{\lambda x}F(x)$ can also tend towards infinity in which case the integral will be evaluated on a function that is not finite and does not tend to 0 as x tends to infinity which will imply that the integral of the expression will also not be finite.

3.4.1 Classification of tail heaviness based on the moment generating function

From Klugman et al [75] it follows that one can classify a distribution as having a light or heavy tail based on whether all of its moments exist. If limited cases or none of the moments exist, then it is indicative of a distribution with a heavy tail whilst if all moments exist, the distribution is classified as having a relatively light tail. This statement is more formally defined by Rolski et al [105] as follows:

Definition 18. *An absolutely continuous distribution with cumulative distribution function, $F_X(x)$, is said to have a heavy tail if $\forall\ s > 0$ the moment generating function $M_X(s)$ is infinite.*

Example 5. *Consider a random variable X that is Gamma distributed with parameters θ and κ. The r^{th} order moment for X is given in (B.63). Also consider a random variable Y that is Pareto Type II distributed with parameters κ and θ. The r^{th} order moment for Y is given in (B.140), [75]. The* *moments of Y exist only if $-2 < r < \kappa + 1$ - i.e. not all of the moments of Y exist. There is no restriction on the value of r in terms of calculating moments for X.*

We can conclude in saying that the Gamma distribution does not have a heavy tail; on the other hand the Pareto distribution does have a heavy tail. Note that the Gamma and Pareto Type II distributions are discussed in Chapter 4 in Sections 4.1.1 and 4.1.11, respectively.

It is important to note that if a distribution doesn't have all its moments , it will not have a moment generating function in which case $\mathrm{E}(e^{tX}) = \infty\ \forall\ t > 0$. The converse is, however, not true.

3.4.2 Classification of tail heaviness based on exponential boundedness

Definition 19. *An absolutely continuous distribution with cumulative distribution function, $F_X(x)$, is said to have an exponentially bounded tail if there exists an $a > 0$ and $b > 0$ such that*

$$\overline{F}_X(x) = 1 - F_X(x) \leq ae^{-bx} \quad \forall \quad x > 0 \tag{3.12}$$

A distribution with an exponentially bounded tail is said to have a light tail [105].

It was stated and argued in Section 2.3 that if we denote the random variable for the claim size by Θ, then the distribution of the claim size distribution $F_\Theta(\theta)$ is well behaved if $1 - \mathrm{F}_\Theta(\theta) \leq ce^{-a\theta}$. In the context of Definition 19 it means that such a claim size distribution is exponentially bounded and hence has a light tail.

Example 6. *If X is Exponentially distributed with parameter θ, then it follows from (4.1) that*

$$\overline{F}_X(x) = e^{-\frac{x}{\theta}} < e^{-\frac{x}{\theta+\varepsilon}}$$

Thus there exist $a, b > 0$ (where $a = 1$ and $b = \frac{1}{\theta+\varepsilon}$ where $\varepsilon < \theta$) such that $\overline{F}_X(x) < ae^{-bx}$ which implies that $F_X(x)$ is lightly tailed. In Figure 3.6 the Exponential distribution's upper tail is displayed for a parameter value θ of 0.9. *It also shows an exponential bound calculated using $a = 1$, $b = \frac{-x}{\theta+0.3}$.*

Example 7. *If X is Weibull distributed with parameters β and θ where $\beta > 1$, then from (4.16) and (4.1) follow that $\overline{F}(x) = e^{-\left(\frac{x}{\theta}\right)^\beta} \leq e^{-\frac{x}{\theta}}$ which is the upper tail of the Exponential dsitribution which from Example 6 follows that it is exponentially bounded. This indicates that for $\beta > 1$ the Weibull distribution is exponentially bounded and does not exhibit a heavy tail.*

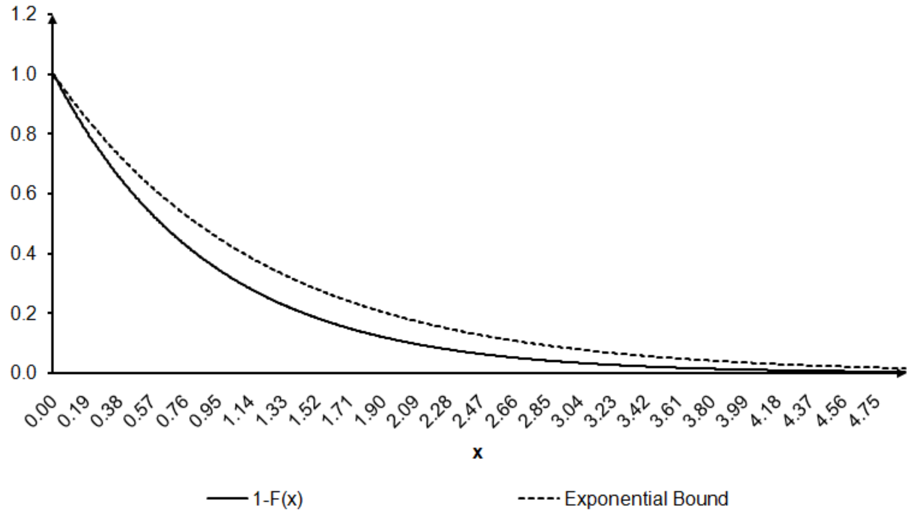

Figure 3.6: Comparison of upper tail of Exponential distribution and exponential bound

3.4.3 Comparison of tail weights using limiting tail behaviour

Klugman et al [75] discuss the use of the limiting behaviour of the ratio of the tails of two distributions to determine whether the tail of the one distribution is heavier than the tail of the other. Suppose we consider any two distributions with cumulative distribution functions $F_1(x)$ and $F_2(x)$. Now consider the limit as x tends to infinity of the ratio of $\overline{F}_1(x)$ to $\overline{F}_2(x)$. Using L'Hospital's Rule [111]

$$\lim_{x\to\infty} \frac{\overline{F}_1(x)}{\overline{F}_2(x)} = \lim_{x\to\infty} \frac{f_1(x)}{f_2(x)}. \tag{3.13}$$

If the ratio tends to infinity as x tends to infinity, it means that there is a divergence in the probabilities of the two distributions with respect to larger values; more specifically, there is higher probabilities for larger values for the distribution in the numerator of the ratio.

Often the limit of the ratio of the tails of the two distributions under consideration does not tend to either 0 or infinity which suggests that neither one of the two distributions has a dominantly heavier tail. The concept of tail equivalence as introduced by Rolski et al [105] may come in useful in this instance. Two distributions $F_X(\cdot)$ and $G_X(\cdot)$ are tail-equivalent if the ratio $\frac{\overline{G}_X(x)}{\overline{F}_X(x)}$ tends to some constant c as x tends to infinity where $0 < c < \infty$.

The following example is taken from Klugman et al [75].

Example 8. *Consider two random variables; X that is Gamma distributed with parameters $\theta = \theta_1$ and $\kappa = \kappa_1$ and Y that is Pareto Type II distributed with parameters $\kappa = \kappa_2$ and $\theta = \theta_2$.*

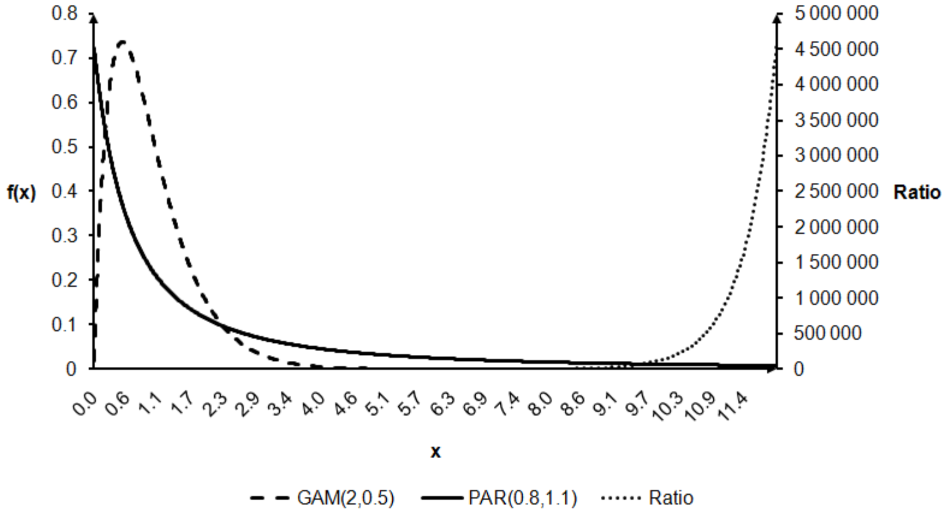

Figure 3.7: Comparison of the limiting tail behaviour of the Gamma and Pareto Type II distributions

From (B.61) and (B.139) follow that:

$$\begin{aligned}
\lim_{x\to\infty} \frac{f_Y(x)}{f_X(x)} &= \lim_{x\to\infty} \frac{\kappa_2 \theta_2^{\kappa_2} (x+\theta_2)^{-\kappa_2-1}}{x^{\kappa_1-1}\theta_1^{-\kappa_1} e^{-\frac{x}{\theta_1}} \Gamma(\kappa_1)^{-1}} \\
&> c \lim_{x\to\infty} \frac{e^{\frac{x}{\theta_1}}}{(x+\theta_2)^{\kappa_2+\kappa_1}} \quad \textit{for } \kappa_1 > 1 \textit{ and } \theta_2 > 0 \\
&= \infty.
\end{aligned}$$

This means that the distribution of Y has a heavier tail than X. The density functions for a Gamma and Pareto Type II distribution are given in Figure 3.7 together with a ratio of the Pareto density to the Gamma density up to a reasonably high value of x to indicate that this ratio is tending to infinity as x tends to infinity. It therefore indicates that the Pareto distribution does have a heavier tail than the Gamma distribution.

The concept of tail equivalence is illustrated in the next example:

Example 9. *Consider the Exponential and Two-parameter Exponential distributions. These distributions are formally introduced in Chapter 4. The cumulative distribution function for the Exponential distribution is given in (4.1) and the cumulative distribution function for the Two-parameter Exponential distribution is given by [11]:*

$$G_X(x) = 1 - e^{-\frac{x-\eta}{\theta}} \text{ for } x > \eta.$$

The Two-parameter Exponential distribution is essentially the Exponential distribution shifted η units to the right on the real line. Alternatively the Exponential distribution can be seen as a special case of the Two-parameter Exponential distributions with $\eta = 0$. This means that the tail weights of these two distributions are the same, but if we consider at any point x^* the area under the distribution to the right (i.e. the upper tail weight), then the weight for the Two-parameter Exponential distribution will be larger than that of the Exponential distributions because of the shift to the right of η units. Hence the tails are expected to be equivalent:

$$\begin{aligned}\frac{\overline{G}_X(x)}{\overline{F}_X(x)} &= \frac{e^{-\frac{x-\eta}{\theta}}}{e^{-\frac{x}{\theta}}} \\ &= e^{\frac{\eta}{\theta}}\end{aligned}$$

Hence there exists a value $c = e^{\frac{\eta}{\theta}}$ where $0 < c < \infty$ such that $\lim_{x\to\infty} \frac{\overline{G}_X(x)}{\overline{F}_X(x)} = c$. Furthermore this value of c is larger than 1 which indicates that the the upper tail weight at any point x^* will be larger for the Two-parameter Exponential distribution than for the Exponential distribution.

3.4.4 Classification of tail heaviness based on the hazard function

In this section a result will be given which is based on the hazard function that can be used to determine whether a distribution has a heavy tail.

The following result is presented in [105] and is quite useful in deriving a relationship between the hazard function, $h_X(x)$, and the heavy-tailedness of the distribution with cumulative distribution function $F_X(x)$.

Result 1. *For a random variable X with a distribution function $F_X(x)$*

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f(x)dx \; \forall \; s \in \mathbb{R}$$

Furthermore

$$\begin{aligned}M_X(s) - 1 &= \int_{-\infty}^{\infty} (e^{sx} - 1) f_X(x)dx \\ &= -s\int_{-\infty}^{0} F_X(x)e^{sx}dx + s\int_{0}^{\infty} \overline{F}_X(x)e^{sx}dx\end{aligned}$$

Theorem 2. *Let* $\alpha_F = \limsup_{x\to\infty} \frac{h_X(x)}{x}$. *If* $\alpha_F = 0$, *then the distribution of* X *is heavy-tailed* [2] *[56].*
Proof:
We will prove this result for X *being a nonnegative random variable. This is sufficient to prove in the context of claim sizes as claim sizes are nonnegative. The proof as provided by Rolski et al [105] is given here, but additional steps are included in order to make the derivation clearer and easier for the reader to follow.*

Suppose that $\alpha_F = 0$, *then* $\limsup_{x\to\infty} \frac{h(x)}{x} = 0$. *Hence for each* $\varepsilon > 0$ *there exists an* x^* *such that* $\forall\ x \geq x^*$ *we have that* $h(x) \leq \varepsilon x$. *Hence*

$$\begin{aligned} -\ln(\overline{F}_X(x)) &\leq \varepsilon x \\ \overline{F}_X(x) &\geq e^{-\varepsilon x} \end{aligned}$$

Therefore, for $s > 0$

$$\begin{aligned} \int_0^\infty e^{sx}\overline{F}_X(x)dx &\geq \int_0^\infty e^{sx-\varepsilon x}dx \\ &= \infty\ \forall\ s \geq \varepsilon \end{aligned}$$

ε *is an arbitrary number and therefore*

$$\int_0^\infty e^{sx}\overline{F}_X(x)dx = \infty\ \forall\ s > 0$$

From Result 1 follows that

$$\int_0^\infty e^{sx}\overline{F}_X(x)dx = \frac{1}{s}(M_X(s) - 1 + s\int_{-\infty}^0 F_X(x)e^{sx}dx)$$

If X *is a nonnegative random variable, then* $F_X(x) = 0$ *for* $x < 0$. *Hence*

$$\int_0^\infty e^{sx}\overline{F}(x)dx = \frac{M_X(s) - 1}{s} = \infty\ \forall\ s > 0$$

Thus

$$M_X(s) = \infty\ \forall\ s > 0 \tag{3.14}$$

Therefore, if for a nonnegative random variable X *with a distribution function* $F_X(\cdot)$ *we have that* $\alpha_F = 0$, *it follows that the distribution is heavy-tailed.*

[2] Let $x = (x_n)$ be a bounded sequence in $\mathbb{R}$. The limit superior of x, denoted by $\limsup x$ is the infimum of the set $\{v \in \mathbb{R}\}$ such that there are at most a finite number of $n \in \mathbb{N}$ such that $v < x_n$ [15].

In practice the Normal distribution is often used if the distribution of the data appears to be close to symmetric. When considering a problem where observed data is strictly positive, the Normal distribution may still be useful given that the chosen values of the parameters are such that the $\mathrm{P}(X \leq 0)$ is very close to 0. In such an event the question may be asked whether this Normal distribution does have a heavy tail. Since the Theorem 2 is proved only for distributions associated with nonnegative domains, an alternative method needs to be considered to ascertain whether a Normal distribution does have a heavy tail. The easiest way to show that it doesn't have a heavy tail is by using (3.13) in which the limiting tail behaviour of a Normal distribution (for any parameter values) is compared to the limiting tail behaviour of an Exponential distribution (for any parameter value). It follows from (B.34) and (B.133) that the Exponential distribution has a heavier tail than the Normal distribution in that:

$$\lim_{x \to \infty} \frac{f_Y(x)}{f_X(x)} = 0$$

Figure 3.6 shows that the Exponential distribution has an exponentially bounded upper tail and consequently is not heavy-tailed. Hence the Normal distribution does not have a heavy tail.

3.4.5 Classification of tail heaviness based on the hazard rate function

Klugman et al [75] argue that a distribution with a decreasing hazard rate function, $h_X^*(t)$, has a heavy tail. In this case the term *decreasing* should be interpreted as *non-increasing* in that the function may be flat over some ranges. The hazard rate function expresses the probability of a certain value within the range of the distribution with respect to the tail to the right of that value. Having a decreasing hazard rate function therefore suggests that the probability of observing specific large values decreases at a faster rate than the probability of observing large values beyond these specific values. As such, this suggests that the right tail weights for these distributions are not light. In conclusion Klugman et al state that one distribution has a heavier tail than another if the hazard rate of the distribution decreases at a faster rate than the hazard rate of the other distribution.

The following mathematical reasoning can be followed to understand why a decreasing hazard rate function is indicative of a heavy tail:

$$\begin{aligned}\frac{f_X(t)}{1-F_X(t)} &\approx \frac{\mathrm{P}(X \in (t+dt))}{P(X>t)}\\ &\approx \frac{\mathrm{P}(t-dt<X<t+dt)}{\mathrm{P}(X>t-dt)}\\ &= \frac{\mathrm{P}(X>t-dt)-\ \mathrm{P}(X>t+dt)}{\mathrm{P}(X>t-dt)}\\ &= 1-\ \mathrm{P}(X>t+dt|X>t-dt) \qquad (3.15)\end{aligned}$$

From the expression in (3.15) it follows that if the hazard rate is a decreasing function, that the ratio

$$\frac{\mathrm{P}(X>t+dt)}{\mathrm{P}(X>t-dt)}$$

will tend to 1 as t tends to ∞. This means that over a small region of various values of t the probability to the right of that value (i.e. in the right tail) is not changing very quickly - suggesting that there is still significant mass in the right tail. This is indicative of a distribution having a heavy tail.

Alternatively the argument of Klugman et al [75] that a decreasing hazard rate function is indicative of a heavy-tailed distribution can be supported by the fact that $\frac{d}{dx}h_X(x) = h_X^*(x)$. A decreasing hazard rate function will therefore yield that

$$\begin{aligned}\alpha_F &= \limsup_{x\to\infty} \frac{h_X(x)}{x}\\ &= \limsup_{x\to\infty} \frac{\frac{d}{dx}h_X(x)}{1} \text{ using L'Hospital's Rule [111]}\\ &= \limsup_{x\to\infty} h_X^*(x)\\ &= 0 \text{ since } h_X^* \text{ is a decreasing function.}\end{aligned}$$

This suggests that if the hazard rate function is a decreasing function, $\alpha_F = 0$ which, by Theorem 2, is indicative of a heavy-tailed distribution.

3.4.6 Classification of tail heaviness based on the mean residual hazard rate function

The classification of a distribution as being heavy-tailed can also be done using the mean residual hazard rate function. From (3.10) follows that if we consider the ratio of right tail probabilities

$$\begin{aligned}\frac{\overline{F}_X(t+y)}{\overline{F}_X(t)} &= \frac{e^{-\int_0^{t+y} h_X^*(x)dx}}{e^{-\int_0^{t} h_X^*(x)dx}} \\ &= e^{-\int_0^{y} h_X^*(x+t)dx}. \qquad (3.16)\end{aligned}$$

Therefore it follows from (3.16) that the ratio of $\overline{F}_X(x+t)$ to $\overline{F}_X(t)$ is increasing if the hazard rate function is decreasing. The mean residual hazard function, as defined in Definition 14, can be written as:

$$\begin{aligned}\mu_{F_t} &= \mathrm{E}(X-t|X>t) \\ &= \frac{\int_t^{\infty} \overline{F}(x)dx}{\overline{F}(t)} \\ &= \int_0^{\infty} \frac{\overline{F}(x+t)}{\overline{F}(t)}dx \\ &= \int_0^{\infty} e^{-\int_0^{x} h_X^*(y+t)dy} dx. \qquad (3.17)\end{aligned}$$

This means that if the hazard rate function is decreasing, the mean residual hazard function is increasing, [23]. Klugman et al [75] and Brown [23] highlight that the converse does not necessarily hold. An increasing mean residual hazard rate function is therefore associated with distribution with a heavy tail.

From Davis [37] it follows that if the hazard rate function $h_X^*(t)$ is a constant function (i.e. equal to some constant $k>0$) it yields a differential equation of the form:

$$\frac{d}{dt}\left(-\ln(\overline{F}(t))\right) = k$$

for which a solution exist of the form $\overline{F}(t) = e^{-kt}$. In a similar fashion Bryson [24] considers testing the null hypothesis of the mean residual hazard rate function being constant against the alternative that it is increasing. Under the alternative hypothesis the set of linear functions of the form $a+bx$ is considered which results in a differential equation for which a solution is given by

$$\overline{F}(x) = \left(\frac{a}{a+bx}\right)^{1+\frac{1}{b}}.$$

The solution under the null hypothesis is therefore obtained as the limiting case of where $b \to 0$ which is given by

$$\overline{F}(x) = e^{-\frac{x}{a}}$$

which corresponds to the solution that Davis obtained when considering a constant hazard rate function. In this context it can be seen that when the alternative hypothesis is true where the mean residual hazard function is increasing, an exponential bound cannot be obtained. It therefore follows that if the hazard rate function is decreasing, the mean residual hazard rate function for which does not have an exponentially bounded tail.

The mean residual hazard function can also be used to compare the tail heaviness of two distributions where a distribtion with a heavier tail than another distribution will have a mean residual hazard function that increases at a higher rate than the mean residual hazard function of the other distribution [75]. This concept is illustrated in Example 10 below.

Example 10. *Consider two random variables X_1 and X_2 that are both Pareto Type II distributed with parameters θ and κ where $\theta = 1$ for both random variables and κ has values of 2 and 4 for X_1 and X_2, respectively. The distributions are illustrated in Figure 3.8.*

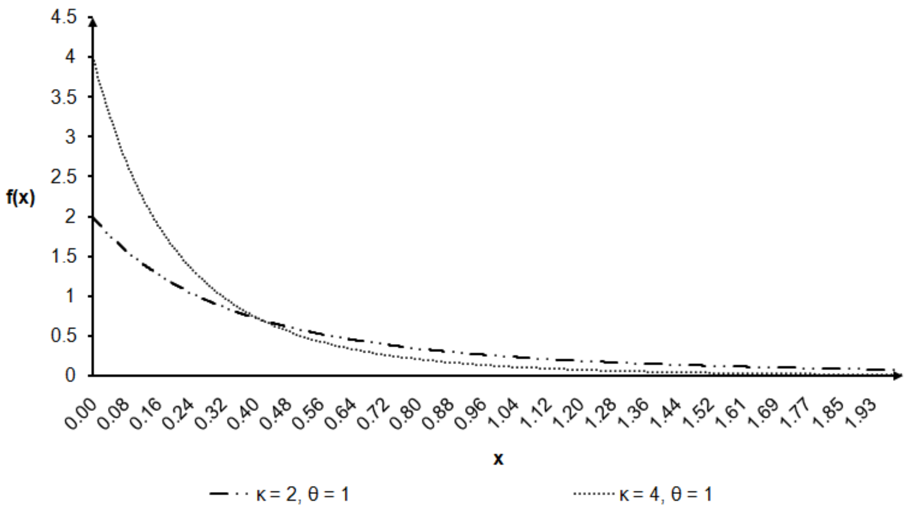

Figure 3.8: Comparison of two Pareto Type II distributions

In Figure 3.8 it can be seen that for $\kappa = 4$ the upper tail weight is lower than for $\kappa = 2$. The general expression for the mean residual hazard rate function can be derived as follows:

$$\mu_{F_t} = \frac{\int_t^\infty \overline{F}(x)dx}{\overline{F}(t)} = \frac{\int_t^\infty \left(\frac{x}{\theta}+1\right)^{-\kappa} dx}{\left(\frac{t}{\theta}+1\right)^{-\kappa}} = \frac{t+\theta}{\kappa-1}.$$

The rate of increase for the mean residual hazard rate function is therefore given by

$$\frac{d}{dt}\mu_{F_t} = (\kappa-1)^{-1}.$$

This means that if we consider the two cases where $\kappa = \kappa_1 = 2$ and $\kappa = \kappa_2 = 4$, we have that $(\kappa_1-1)^{-1} > (\kappa_2-1)^{-1}$ which indicates that the mean residual hazard rate function for X_1 increases at a higher rate than for X_2 which suggests that the distribution for X_1 should have a heavier tail than X_2. This is also supported by the graphical representation of the two distributions in Figure 3.8.

3.5 Left Censored Variables and Integrated Tail Distributions

Consider a random variable, X, with distribution function $F_X(x)$ defined on $\mathbb{R}^+$. If X is a random variable for measuring the loss (or any similar type of exposure), then one can and should clearly distinguish between the following associated random variables:

- Excess loss variable
- Left censored variable
- Limited loss variable

Klugman et al [75] give formal definitions of each of these associated random variables. These are discussed in Sections 3.5.1 to 3.5.3.

3.5.1 Excess Loss Variable

Definition 20. ***Excess Loss Variable***
The excess loss variable for a random variable X with a threshold t is defined as

$$Y = X - t \text{ where } X > t.$$

The expected value of this variable is given by $\mathrm{E}(X-t|X>t)$ which is the mean residual hazard function in (3.9), as was defined in definition 14.

In the context of life contingencies [99] the expectation as given in Definition 14 is referred to as the complete (remaining) life expectation of a person given that the person is t years old. The variable Y will have a 0 density for all values less than t.

3.5.2 Left Censored

Definition 21. ***Left Censored Variable***
The left censored variable for a random variable X with a threshold, t, is defined as

$$Y = (X-t)_+ = \begin{cases} 0 & \text{if } X \le t, \\ X-t & \text{if } X > t. \end{cases}$$

The expectation of the left censored variable can be derived as follows:

$$\begin{aligned} \mathrm{E}((X-t)_+) &= \int_0^\infty (x-t)_+ f_X(x)dx \\ &= \int_0^t 0 \times f_X(x)dx + \int_t^\infty (x-t)f_X(x)dx \\ &= (1-F_X(t)) \times \mathrm{E}(X-t|X>t) \\ &= \int_t^\infty (1-F_X(x))dx \text{ from (14).} \end{aligned} \tag{3.18}$$

The random variable Y has a value of 0 for values of $X \le t$, but has non-zero density on the range $[0, t]$. This variable and its distribution is of paricular interest when insurance claims are modelled where an excess or retention level is in place with respect to the insured party. In reinsurance this will be of interest to the reinsurer (second line insurer) with respect to claims from the cedant (first line insurer).

3.5.3 Limited Loss Variable

Definition 22. ***Limited Loss Variable***
The limited loss variable for a random variable X with an upper threshold, ν, is defined as

$$Y = \begin{cases} X & \text{if } X < \nu, \\ \nu & \text{if } X \ge \nu. \end{cases} \tag{3.19}$$

The expectation of the limited loss can be derived as follows:

$$\begin{aligned}
\mathrm{E}(Y) &= \int_0^{\nu} x f_X(x)dx + \int_{\nu}^{\infty} \nu f_X(x)dx \\
&= \int_0^{\infty} x f_X(x)dx - \int_{\nu}^{\infty} x f_X(x)dx + \int_{\nu}^{\infty} \nu f_X(x)dx \\
&= 1 - \int_{\nu}^{\infty} (x-\nu) f_X(x)dx \\
&= 1 - \int_{\nu}^{\infty} (1 - F_X(x))dx \text{ from (14).}
\end{aligned} \tag{3.20}$$

The random variable Y and its distribution is of particular interest when insurance claims are modelled from the point of view of the first line insurer when reinsurance treaties are in place. The upper threshold is the retention level. The random variable is commonly referred to in statistics theory as a right censored variable.

3.5.4 Coefficient of variation when the distribution is heavy-tailed

Another useful approach introduced by Klugman et al [75] is by considering the coefficient of variation of a nonnegative random variable. Knowing whether the hazard rate is increasing or decreasing can be indicative of whether the coefficient of variation is limited to certain values. This limitation may be given either by a lower limit or an upper limit. The converse isn't necessarily implied; one can, however, prove the assertion that a distribution is heavy or not heavy-tailed wrong by means of finding a counter-argument in the coefficient of variation.

Definition 23. ***Integrated Tail Distribution***
Consider the excess of loss and left censored variables as defined in Definitions 20 and 21. If $t = 0$, then the expectation of the excess loss variable equals the expectation of the left censored variable:

$$\mathrm{E}(X) = \int_0^{\infty} \overline{F}_X(x)dx$$

hence

$$\int_0^{\infty} \frac{\overline{F}_X(x)}{\mathrm{E}(X)} dx = 1 \tag{3.21}$$

This is referred to as the integrated tail distribution [105]. The ratio in this integral behaves like a probability density function in that the integral over the full range of the values that X can take on is equal to 1.

Let $f_e(x) = \dfrac{\overline{F}_X(x)}{\mathrm{E}(X)}$ and therefore

$$F_e(x) = \int_0^x f_e(u)du = \int_0^x \frac{\overline{F}_X(u)}{\mathrm{E}(X)}du = 1 - \int_x^\infty \frac{\overline{F}_X(u)}{\mathrm{E}(X)}du. \tag{3.22}$$

If the mean residual hazard rate function is increasing, one will have that for positive, non-zero values of X [75] that

$$\begin{aligned}
\mu_{F_x} &\geq \mu_{F_0} \\
\frac{\int_x^\infty (1-F_X(u))du}{1-F_X(x)} &\geq \frac{\int_0^\infty (1-F_X(u))du}{(1-F_X(0))} \\
\frac{\int_x^\infty (1-F_X(u))du}{1-F_X(x)} &\geq \mathrm{E}(X), \text{ hence} \\
\frac{\int_x^\infty (1-F_X(u))du}{\mathrm{E}(X)} &\geq (1-F_X(x)) \\
\therefore (1-F_e(x)) &\geq (1-F_X(x)) \\
\therefore \int_0^\infty (1-F_e(x))dx &\geq \int_0^\infty (1-F_X(x))dx = \mathrm{E}(X).
\end{aligned} \tag{3.23}$$

but

$$\begin{aligned}
\int_0^\infty (1-F_e(x))dx &= \mathrm{E}_e(X) \\
&= \int_0^\infty x f_e(x)dx \\
&= \int_0^\infty \frac{x\overline{F}_X(x)}{\mathrm{E}(X)}dx \\
&= \frac{1}{\mathrm{E}(X)} \int_0^\infty x\overline{F}_X(x)dx \\
&= \frac{1}{\mathrm{E}(X)} \left\{ \frac{x^2}{2}\overline{F}_X(x)\Big|_0^\infty + \int_0^\infty \frac{x^2}{2} f_X(x)dx \right\}.
\end{aligned}$$

Hence

$$\begin{aligned}
\int_0^\infty (1-F_e(x))dx &= \frac{1}{2\,\mathrm{E}(X)} \int_0^\infty x^2 f_X(x)dx \\
&= \frac{\mathrm{E}(X^2)}{2\,\mathrm{E}(X)} \\
&\geq \int_0^\infty (1-F_X(x))dx \\
&= \mathrm{E}(X) \text{ following from the expression in (3.23).}
\end{aligned} \tag{3.24}$$

Using the expression in (3.24) and the fact that $\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$, it can be concluded that if the mean residual hazard rate function is increasing (which is when the distribution has a heavy tail), that

$$\frac{\mathrm{var}(X)}{(\mathrm{E}(X))^2} \geq 1. \tag{3.25}$$

This means that if the distribution has a heavy tail, the coefficient of variation has a value of at least 1. Furthermore if a distribution has a decreasing mean residual hazard rate function, the coefficient of variation will be at most 1 [75].

The relationship between heavy-tailed distributions, their mean residual hazard rate functions and coeffients of variation can therefore be summarised as follows:

- Case 1:
  Distribution with cumulative distribution function $F_X(\cdot)$ has a heavy tail
  $\Rightarrow \mu_{F_t}$ is increasing (from (3.17)
  $\Rightarrow \dfrac{\mathrm{var}(X)}{(\mathrm{E}(X))^2} \geq 1$ (from (3.25))
- Case 2:
  Distribution with cumulative distribution function $F(\cdot)$ doesn't have a heavy tail with μ_{F_t} that is decreasing
  $\Rightarrow \dfrac{\mathrm{var}(X)}{(\mathrm{E}(X))^2} \leq 1$
- Case 3:
  Distribution with cumulative distribution function $F(\cdot)$ doesn't have a heavy tail with μ_{F_t} that is increasing
  $\Rightarrow \dfrac{\mathrm{var}(X)}{(\mathrm{E}(X))^2} \geq 1$

The relationship in (3.25) can be therefore only be used to test a proposition of heavy tailedness in that if we, for example, assume the distribution has a heavy tail, but the coefficient of variation is less than 1, it forms a contradiction to our assumption. In this case the proposition will be proven wrong. If we observe a coefficient of variation that is greater than or equal to 1 it does not necessarily imply that the distribution under consideration has a heavy tail. This can be used in practice where we don't know the distribution, but can calculate a sample coefficient of variation.

3.6 Subexponential Distributions

The aim of this section is to introduce the subexponential class of distributions. Examples and key properties of this class of distributions are also given.

Definition 24. ***Subexponential Class***
If a distribution function $F_X(x)$, defined on $\mathbb{R}^+$, satisfies the following property

$$\lim_{x\to\infty} \frac{1-F^{*2}(x)}{1-F(x)} = 2 \tag{3.26}$$

the distribution is said to be a member of the subexponential class [3]*, denoted by S [118], [56], [7].*

Examples of distributions belonging to the subexponential class of distributions include the Lognormal distribution, Pareto distribution and Weibull distribution with shape parameter $\beta \in (0,1)$ [105]. These distributions are formally introduced in Chapter 4.

Example 11. *The Exponential distribution is not a member of the subexponential class, since*

$$\lim_{x\to\infty} \frac{1-F^{*2}(x)}{1-F(x)} = \lim_{x\to\infty} \frac{e^{-\lambda x}(1+\lambda x)}{e^{-\lambda x}} = \lim_{x\to\infty} (1+\lambda x) = \infty$$

Theorem 3. *Each distribution $F \in S$ is heavy-tailed [105].*

Once it is known whether a distribution is a member of the subexponential class, it follows imediatiely that the distribution has a heavy tail. It requires of one to know which parametric distribution is under consideration together with the cumulative distribution function of the distribution. If these are all known, one can determine algebraically whether the distribution satisfies the condition as set out in Definition 24.

Often it is algebraically not possible or difficult to calculate the convolution or to integrate the convolution of the parametric cumulative distribution function. In some instances an explicit formula for the cumulative distribution function does not exist in which case it will not be possible to derive the convolution. In practice the parametric distribution is not always known in advance and there might not even be a suitable parametric distribution that can be fitted to the observed data. In other cases where the parametric distribution is known, the parameter values are not always known. For the Weibull distribution, for instance, it is true that it belongs to the subexponential class, but only for certain values of its parameters - as it is discussed

[3] F^{*2} refers to the two-fold convolution of $F_X(x)$.

earlier in this section.

Although it might be useful to determine whether a distribution is a member of the subexponential class, it provides a technique that is potentially a more difficult way of trying to determine whether a particular distribution has a heavy tail. There may, however, be cases available in the literature in which these results are already provided and can be leveraged off.

3.7 Practical Methods to Detect Heavy Tails

3.7.1 Overview

To classify claims as large is difficult in that claims can be regarded as small or large relative to a class of insured risks. Firstly some additional notation needs to be introduced in the context of insurance claims.

Let $\{\Theta_i, 1 \leq i \leq n\}$ denote a series of n successive observed claims with the aggregate claim size as given in (2.7) with $X(t) = X_n$ and $N(t) = n$.

The value of X_n will be large if at least $\Theta_{(n)}$ is large, where $\Theta_{(n)}$ is the n^{th} order statistic from the series of n claims.

If exessively large claims are causing X_n to be large, then $\mathrm{P}(X_n > x) \approx \mathrm{P}(\Theta_{(n)} > x)$ as $x \to \infty$.

Rolski et al [105] propose a few mathematical arguments that can be used to classify claims as being large. These arguments will be discussed in the following sections.

3.7.2 Method 1: Using definition of the subexponential class

Under this argument the aim is to use the empirical cumumlative distribution function $F_n(x)$ to determine whether the condition for the distrbution being a member of the subexponential class as given in Definition 24 needs to be tested. This requires one to show that for a large value of x it is true that $\dfrac{1 - F_n^{*2}(x)}{1 - F_n(x)}$ tends to 2 as $x \to \infty$. One can evaluate this expression for a large (or the largest) order statistic $\Theta_{(n)}$. It is, however not easy to show that this expression is close to or tends to 2 from using an empirical cumulative distribution function. Note that:

$$F_n(x) = \frac{1}{n} \max\{i; \Theta_i \leq x\} \quad \forall\, x \in \mathbb{R}.$$

and that the following events are equivalent:

$$\left\{F_n(x) = \frac{k}{n}\right\} \equiv \left\{\Theta_{(k)} \leq x \leq \Theta_{(k+1)}\right\}.$$

3.7.3 Method 2: Using the mean residual hazard rate function

From (3.9) in Definition 14 follows that the aim here is to determine whether the mean residual hazard rate function is increasing and therefore if $\mu_{F_t} \to \infty$ as $t \to \infty$. If this can be confirmed, it will imply that $\alpha_F = 0$ and that the distribution is heavy-tailed.

The empirical analogue of μ_{F_t} is given by

$$\mu_n(\Theta_{(n-k)}) = \int_{\Theta_{(n-k)}}^{\infty} \frac{1 - F_n(y)}{1 - F_n(\Theta_{(n-k)})} dy$$

where n should be large and k be chosen such that $\frac{k}{n} \to 0$.

From the fact that $\left\{F_n(x) = \frac{k}{n}\right\} \equiv \left\{\Theta_{(k)} \leq x \leq \Theta_{(k+1)}\right\}$ follows that

$$\begin{aligned}
\mu_n(\Theta_{(n-k)}) &= \int_{\Theta_{(n-k)}}^{\infty} \frac{1 - F_n(y)}{1 - F_n(\Theta_{(n-k)})} dy \\
&= \sum_{i=n-k}^{n-1} \int_{\Theta_{(i)}}^{\Theta_{(i+1)}} \frac{1 - F_n(y)}{1 - F_n(\Theta_{(n-k)})} dy \\
&= \frac{1}{k} \sum_{i=n-k}^{n-1} (n-i) \left(\Theta_{(i+1)} - \Theta_{(i)}\right) \\
&= \frac{1}{k} \sum_{j=n-k+1}^{n} \left(\Theta_{(j)} - \Theta_{(n-k)}\right).
\end{aligned}$$

The idea here is to choose a series of values for k and to then evaluate the empirical mean hazard rate function values for each of these k's. One can then assess whether this series of empirical values are increasing or decreasing. Despite being able to derive these results algebraically, it is not clear how to choose suitable values for k in order to show that if $\Theta_{(n-k)} \to \infty$ then $\mu_n(\Theta_{(n-k)}) \to \infty$.

In practice the empirical analogue, $\mu_n(\Theta_{(n-k)})$, can be used. If the claims distribution has a heavier tail than the proposed distribution, then $\mu_n(\Theta_{(n-k)})$ for $0 < k \leq n$ will be consistently larger than μ_{F_t}.

One should plot the empirical mean residual hazard function against the theoretical function to ascertain whether the claims distributions has a lighter or a heavier tail.

For both this method and the previous method it follows that while it is theoretically possible to indentify heavy-tailedness, it is difficult to use these theoretical results to establish heavy-tailedness empirically.

3.7.4 Method 3: Testing for an exponentially bounded tail using quantile plots

Using Definition 19 one can argue from the concept of an exponentially bounded tail that a distribution with a lighter tail than an Exponential distribution will satisfy the inequality in (3.12). If this inequality holds we can derive for $x \geq 0$

$$\ln(\overline{F}_X(x)) \leq \ln(ce^{-ax})$$
$$\limsup_{x\to\infty} \frac{-\ln(\overline{F}_X(x))}{x} > \limsup_{x\to\infty} \frac{-\ln(ce^{-ax})}{x} = 0$$

$\therefore \alpha_F > 0$ for $\overline{F}_X(x) \leq ce^{-ax}$. This means that the distribution doesn't have a heavy tail. Conversely, if $F_X(\cdot)$ satisfies the condition

$$\overline{F}_X(x) > ce^{-ax} \ \forall \ a > 0$$

then $\alpha_F = 0$.

Methods for detecting light tails will involve comparing samples with Exponential distributions using techniques proposed by Rolski et al [105]:

- Quantile plots
- Mean residual hazard functions

Using quantile plots

Quantile plots are reliant on a sample of observed claim sizes $\Theta_1 = \theta_1$, $\Theta_2 = \theta_2$, ..., $\Theta_n = \theta_n$, where $\theta_1, \theta_2, \ldots, \theta_n$ are the observed values of the random variables $\Theta_1, \Theta_2, \ldots, \Theta_n$.

Definition 25. ***Generalized Inverse Function***
For a monotone increasing cumulative distribution function $F_X(x)$, the generalized inverse function $F_X^{-1}(y)$ [105] is given by

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}. \tag{3.27}$$

Definition 26. ***Quantile Function***
The quantile function is defined as

$$Q_F(y) = F^{-1}(y) \ \textit{as given in}(3.27). \tag{3.28}$$

Given the order statistics of claim sizes $\Theta_{(1)} \leq \Theta_{(2)} \leq \ldots \leq \Theta_{(n)}$ the following events are equivalent

$$\{Q_n(y) = \Theta_{(k)}\} \equiv \left\{ \frac{k-1}{n} < y \leq \frac{k}{n} \right\}$$

Example 12. *For the Exponential the cumulative distribution function in (4.1) can be used to obtain the associated quantile function*

$$Q_F(y) = \frac{-\ln(1-y)}{\theta} \text{ for } 0 < y < 1.$$

One can then compare $Q_F(y)$ and $Q_n(y)$ in order to identify whether the sample is from a distribution close to an Exponential distribution. The slope of the line given by plotting $Q_F(y)$ on the x-axis and $Q_n(y)$ on the y-axis will give an approximation to the value of $-\frac{1}{\theta}$ if the plot is close to linear and the underlying distribution therefore close to an Exponential distribution. If the true underlying distribution has a tail that is heavier than the tail of an Exponential distribution, then the quantile plot is expected to increase faster than a straight line with a slope of 1 in which case the slope of the line will be decreasing.

The next example is based on simulated data and illustrates how a quantile plot can be used to assess whether the observed data has a distribution with a tail being heavier than the Exponential distribution.

Example 13. *Consider four sets of simulated observations of size n:*

- *From an Exponential distribution with parameter $\theta = 3$.*
- *From a Birnbaum-Saunders distribution with parameters $\alpha = 2$ and $\beta = 1$.*
- *From a Folded Normal distribution with parameters $\mu = 3$ and $\sigma = 2$.*
- *From a Pareto Type II distribution with parameters $\theta = 2$ and $\kappa = 3$.*

These distributions are formally introduced in Chapter 4.

For each of the observed samples the following was done:

- *Construct the empirical cumulative distribution function based on the observed values. These values relate to n quantiles.*
- *Fit an Exponential distribution to the data.*

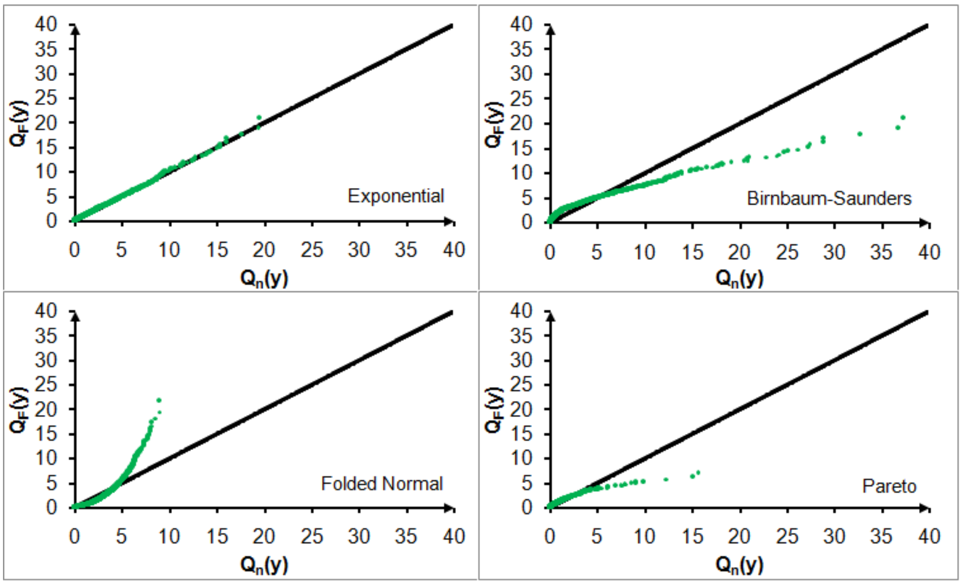

Figure 3.9: Quantile plots to assess fitted Exponential distributions

- *Calculate the* n *quantiles as implied by the fitted Exponential distribution.*

- *Plot the quantiles from the fitted distribution* $Q_F(y)$ *against the observed quantiles* $Q_n(y)$.

It can be seen that for the data simulated from the Exponential distribution the fit is good and the tail is therefore considered to be exponentially bounded. For the data simulated from the Birnbaum-Saunders distribution it can be seen from the quantile plot that the distribution has a heavier tail than the fitted Exponential distribution suggesting that an exponential bound may perhaps not exist for this distribution. For the data simulated from the Folded Normal distribution it is evident that the tail is lighter than that of the fitted Exponential which clearly indicates that the tail is exponentially bounded. For the data simulated from the Pareto Type II distribution it is clear that the tail is heavier than that of the fitted Exponential distribution and it is therefore suggesting that an exponential bound may not exist. It is argued in Chapter 5 that the Pareto Type distribution is heavy-tailed, while the Exponential, Birnbaum-Saunders and Folded Normal distributions are not.

Chapter 4

Parametric Loss Distributions

In this chapter we consider parametric distributions that can be used to model the severity of claims (losses associated with claims). Since claims can only take on nonnegative values, only distributions with domain in $\mathbb{R}^+$ are considered. Each of the distributions is discussed in terms of its first four moments:

- Mean - as a measure of centrality
- Variance/Standard Deviation - as a measure of spread
- Skewness
- Kurtosis

Because of the fact that claims distributions exhibit skewness, the skewness of a particular distribution is of interest. Furthermore, the shape of distributions for different parameter values are investigated graphically. Links among variables are also indicated.

The Normal distribution is popular to be used due to its shape being suitable for particular problems and studies whilst its properties are well-known, easy to use and often built-in functions are available in statistical and mathematical software packages. In cases where the observed data appears to be close to symmetric and even close to Normal, one might be interested in using the Normal distribution. The only problem is that the domain of this distribution includes negative values.
Montgomery [87] considers 3-sigma limits (that is three standard deviations from the mean) on a Normal distribution for process quality control where this implies that for any process that is operating in an in-control state, the

probability of observing a non-conforming good produced is 0.27%. This 0.27% is a two-sided probability, meaning that a lower tail probability is 0.135%. This means that if we have data suggesting normality and the mean is greater than or equal to three times the standard deviation, a fitted distribution will imply a probability of observing a negative value of at most 0.135%. In such an instance, the Normal distribution may prove to still be useful. Alternatively, variations on the Normal distribution such as the Folded Normal, Skew-normal and Birnbaum-Saunders distributions may be useful. These distributions are discussed in sections 4.1.24, 4.1.28 and 4.1.15, respectively.

They provide details with respect to the moments, techniques for the estimation of parameters in cases of censored and non-censored data, related distributions and techniques for simulating frome these distributions are provided. Some of these distributions are discussed in Section 4.1 below for which the estimation of the parameters are discussed in Chapter 6.

4.1 Continuous Distributions for Claim Sizes

4.1.1 Gamma Distribution

The probability density function for the Gamma distribution, with parameters θ and κ, is defined in (B.61), [11], [105].

The espressions for the r^{th} moment about the origin, mean, variance, skewness and kurtosis, as given by Bain and Engelhardt [11], are given in (B.61) to (B.67). It can be seen that the skewness and kurtosis are dependent on the value of κ. Depending on the value of κ, the distribution may be negatively or positively skewed or even symmetric if $\kappa = 4$.

Figure 4.1 shows the effect of varying the values for θ from which it can be seen that θ is a scale parameter.

Figure 4.2 shows the effect of varying the values for κ from which it can be seen that κ is a shape parameter.

4.1.2 Exponential Distribution

The probability density function of a random variable X that is Exponentially distributed [105],[11], as given in (B.34), is a special case of the probability density function of the Gamma distribution on the same domain with $\kappa = 1$ and using the fact that

$$\Gamma(r+1) = r\Gamma(r) = \ldots = r!$$

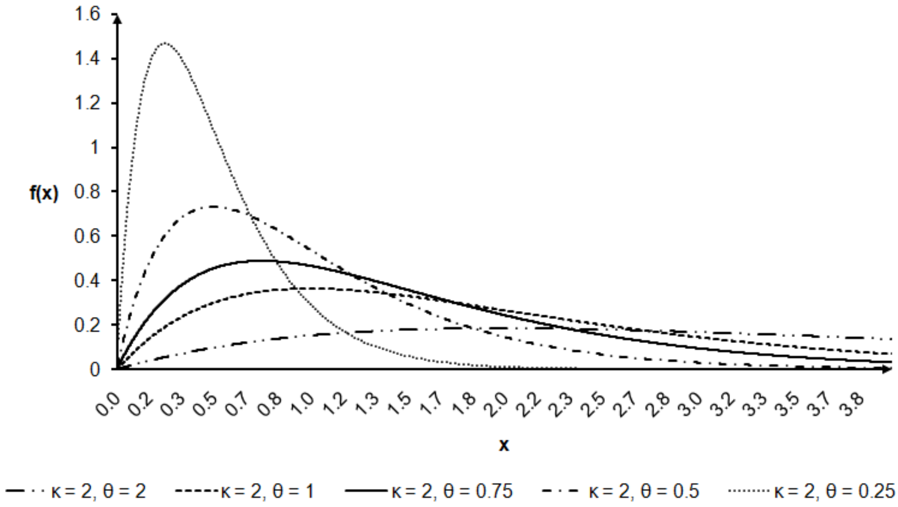

Figure 4.1: Gamma Density Function with varying values of θ

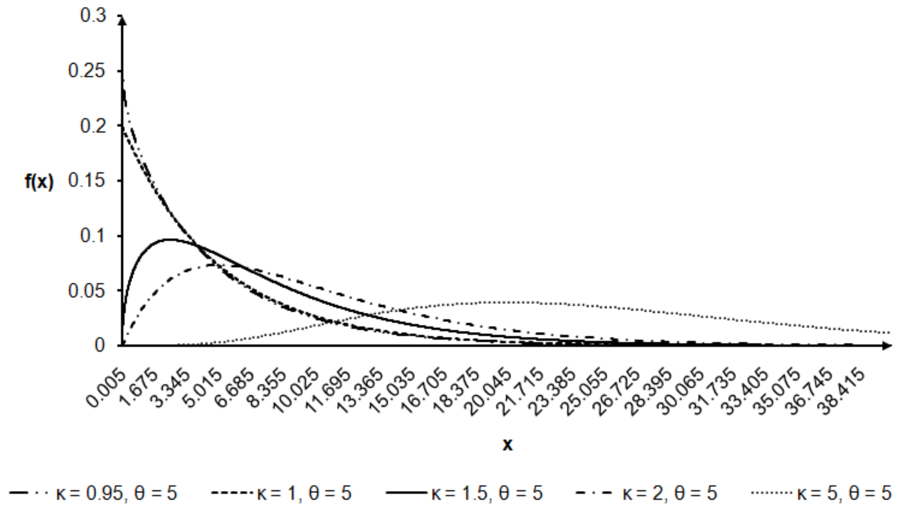

Figure 4.2: Gamma Density Function with varying values of κ

The cumulative distribution function is given by:

$$F_X(x) = 1 - e^{-\frac{x}{\theta}} \text{ for } x \geq 0 \tag{4.1}$$

The expression for r^{th} moment, which can be obtained using (B.63) with $\kappa = 1$, is given in (B.36).

One can leverage of the moments derived for the Gamma distribution to find the expressions for the moments of the Exponential distribution. These expressions are given in (B.37), (B.38), (B.39) and (B.40). Similarly the moment generating function can be found using the moment generating function of the Gamma distribution (B.62) with $\kappa = 1$.

Since the Exponential distribution is a special case of the Gamma distribution where $\kappa = 1$, the shape will not be affected by varying values of θ, but the scale will vary. In Figure B.34 the probability density functions for various values of θ can be seen.

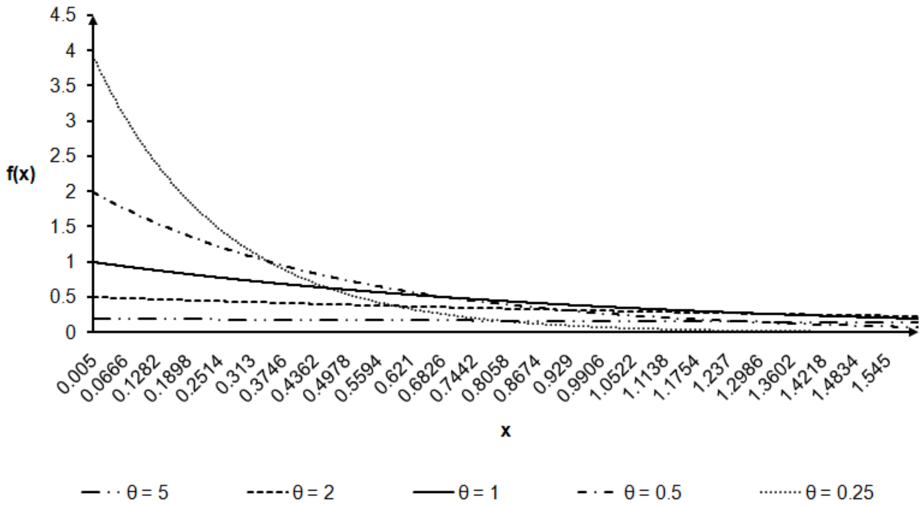

Figure 4.3: Exponential Density Function with varying values of θ

From Figure 4.3 it can be seen that probability density starts at 0 at a maximum level equal to $\frac{1}{\theta}$.

4.1.3 Chi-square Distribution

Consider a random variable Y from a Gamma distribution with $\theta = 2$ and $\kappa = \frac{\nu}{2}$. From expression (B.61) the probability density function of Y follows directly, as given in (B.13).

From Bain and Engelhardt [11] follows that Y is Chi-square distributed with ν degrees of freedom (denoted $Y \sim \chi^2(\nu)$). The r^{th} moment can be obtained using (B.63) with $\theta = 2$ and $\kappa = \frac{\nu}{2}$. This expression is given in (B.15).

Using the expressions for the Gamma distribution given in (B.65), (B.66) and (B.67), one can find the other moments for this distribution (where $X \sim \chi^2(\nu)$).

The relationship can be stated more generally using the moment generating functions of the Chi-square and Gamma distributions, respectively (see Bain and Engelhardt [11]). The moment generating function for a random variable $Y \sim \chi^2(\nu)$ is given in (B.14). Furthermore, if X is gamma distributed with parameters θ and κ, then $Z = \frac{2X}{\theta} \sim \chi^2(2\kappa)$.

Hence, the Chi-square distribution, as a special case of the Gamma distribution, is not dependent on the value on θ as this special case is where $\theta = 2$. Figure 4.4 shows how the density functions for different values of ν.

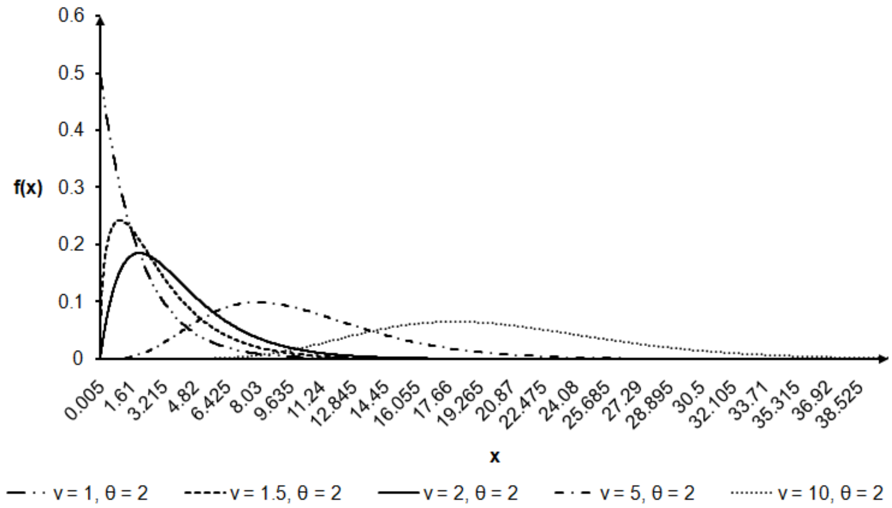

Figure 4.4: Chi-square Density Function with varying values of ν

4.1.4 Two-parameter Exponential Distribution

The probability density function of a random variable X that is said to have a Two-parameter Exponential distribution [11] is given in (B.171).

It can be seen that the Exponential distribution is a special case of the Two-parameter Exponential distribution with $\eta = 0$. For this reason the variance, skewness and kurtosis of the Two-parameter Exponential are the same as for the Exponential distribution, since the η parameter is purely a location parameter that shifts the distributions η units from the origin. For completeness, expressions for the mean, variance, skewness and kurtosis are given in Section B.29 in Appendix B.

The moment generating function in (B.172) can be found from first principles as follows:

$$\begin{aligned}
\mathrm{E}\left(e^{tX}\right) &= \int_{\eta}^{\infty} \frac{1}{\theta} e^{\left(-\frac{x-\eta}{\theta}\right)} dx \\
&= e^{t\eta} \int_{0}^{\infty} e^{-y(1-t\theta)} dy \text{ by letting } y = \frac{x-\eta}{\theta} \\
&= \frac{e^{t\eta}}{1-t\theta} \qquad (4.2)
\end{aligned}$$

Consider the transformation $Y = X - \eta$.

$$\begin{aligned}
F_Y(y) &= P(X \leq y + \eta) \\
&= 1 - e^{-\frac{y}{\theta}} \qquad (4.3)
\end{aligned}$$

From (4.3) follows that $Y \sim EXP(\theta)$. To derive the moments of X, one can therefore leverage of the fact that $\mathrm{E}(Y^r) = \theta^r r!$ from (B.36). Hence $\mathrm{E}(X^r) = \mathrm{E}((Y + \eta)^r)$.

Figure 4.5 shows how the density of X varies with different values of θ while Figure 4.6 shows how it varies with different values of η.

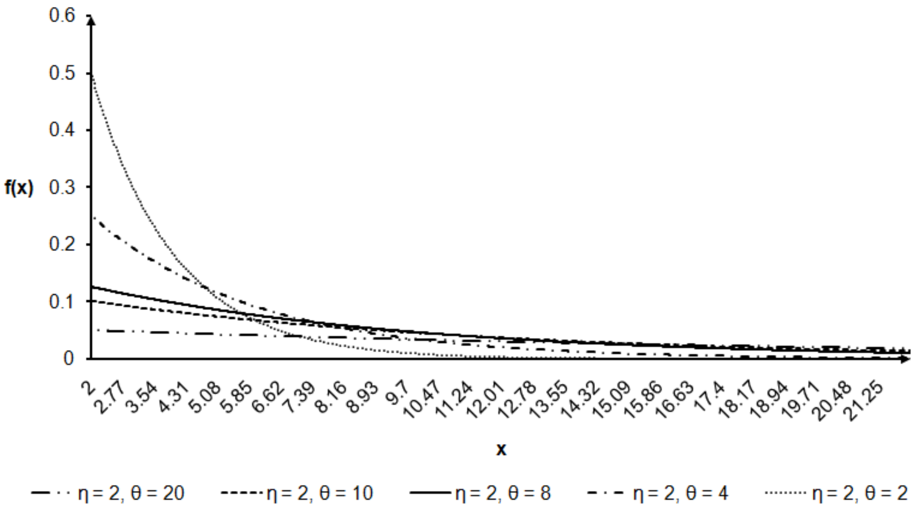

Figure 4.5: Double Exponential Density Function with varying values of θ

4.1.5 Erlang Distribution

A random variable X that is said to have an Erlang distribution with parameters n and λ, denoted as $X \sim \mathrm{ERL}(n, \lambda)$, has a probability density function as given by (B.25), [105].

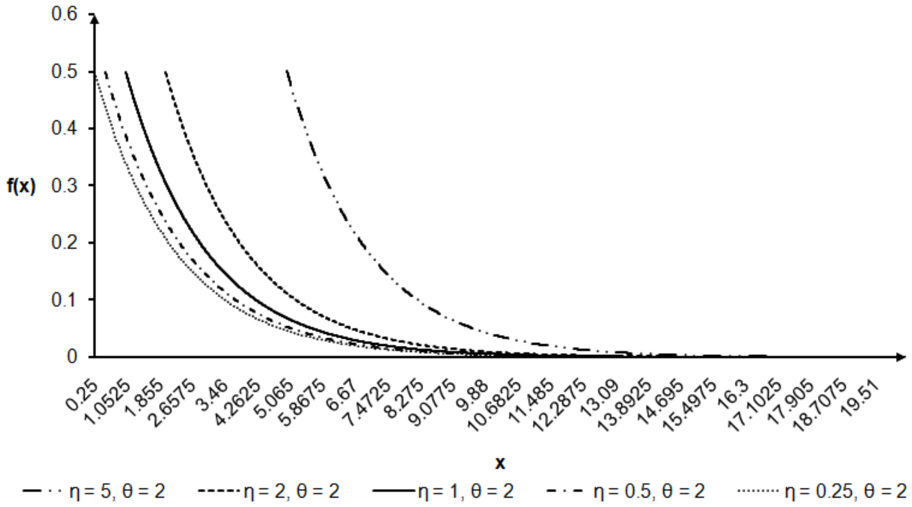

Figure 4.6: Double Exponential Density Function with varying values of η

If one consider a set of n random variables $(X_1, X_2, ..., X_n)$ where the individual random variables are independent, identically Exponential distributed with parameter θ and let

$$Z = \sum_{i=1}^{n} X_i,$$

then the distribution of Z can be derived using the moment generating function of the Exponential distribution as given in (B.35). From Bain and Engelhardt [11] it follows that the moment generating function for Z is exactly the same as the moment generating function for a random variable that exhibits an Erlang distribution with parameters λ and n. It can therefore be concluded that an Erlang distribution is the resultant of a sum of n independent and identically EXP$\left(\frac{1}{\lambda}\right)$ distributed random variables. Klugman et al [75] considers aggregate severity distributions which are representative of a portfolio of losses. In such cases the overall severity is an aggregation of the individual losses. The Erlang distribution is a specific case where the severity distribution is a discrete (since n is an integer value) mixture of Exponential distributions.

The moments for the Erlang distribution can be found using (B.27) which is a general formula for its r^{th} moment. The mean, variance, skewness and kurtosis are given in (B.28), (B.29), (B.30) and (B.31), respectively.

Rolski et al [105] also state that the sum of two independent Erlang distributed random variables with the same parameter λ will again be Erlang distributed; that is if $X_1 \sim \text{ERL}(n_1, \lambda)$ and $X_2 \sim \text{ERL}(n_2, \lambda)$, then $Y = (X_1 + X_2) \sim \text{ERL}(n_1 + n_2, \lambda)$. This can follow from general reasoning in that the relationship between the sum of independent, identically distributed exponential random variables and the Erlang distribution holds for any $n \in \mathbb{N}$. Since the result holds for any n, it automatically follows that it holds for n_1 and n_2, but it should therefore also hold for $n_1 + n_2$.

Figure 4.7 shows how the density of X varies with different values of λ while Figure 4.8 shows how it varies for different sample sizes.

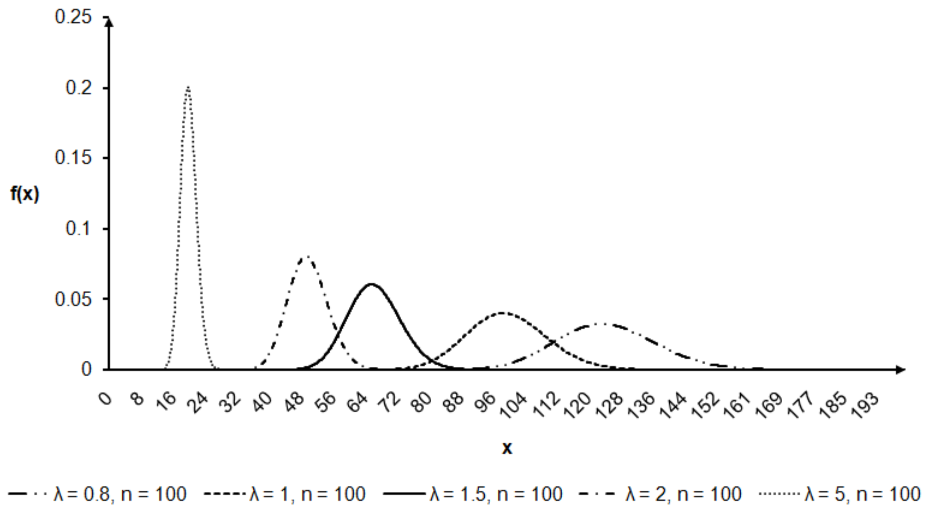

Figure 4.7: Erlang Density Function with varying values of λ

4.1.6 Extreme Value Distribution

Hosking et al [68] uses the Generalized Extreme Value distribution as introduced by Jenkinson [70] to derive probability weighted moments for this distribution.

$$F_X(x) = \begin{cases} e^{-(1-k\frac{x-\xi}{\alpha})^{1/k}} & k \neq 0, \\ e^{-e^{-\frac{x-\xi}{\alpha}}} & k = 0. \end{cases} \tag{4.4}$$

X is bounded by $(\xi + \alpha)/k$ from above if $k > 0$ and from below if $k < 0$ where $\xi \in \mathbb{R}$ and $\alpha > 0$. Usually the shape parameter k lies in $(-1/2; 1/2)$. This form of the distribution is a combination of three possible types of limiting distributions used for extreme values originally derived by Fisher and Tippet [54],[76], [31]. There are essentially three distinct forms of the Extreme Value distributions as encapsulated by the generalized form. These forms stem from the resulting distributions for different values of k [98].

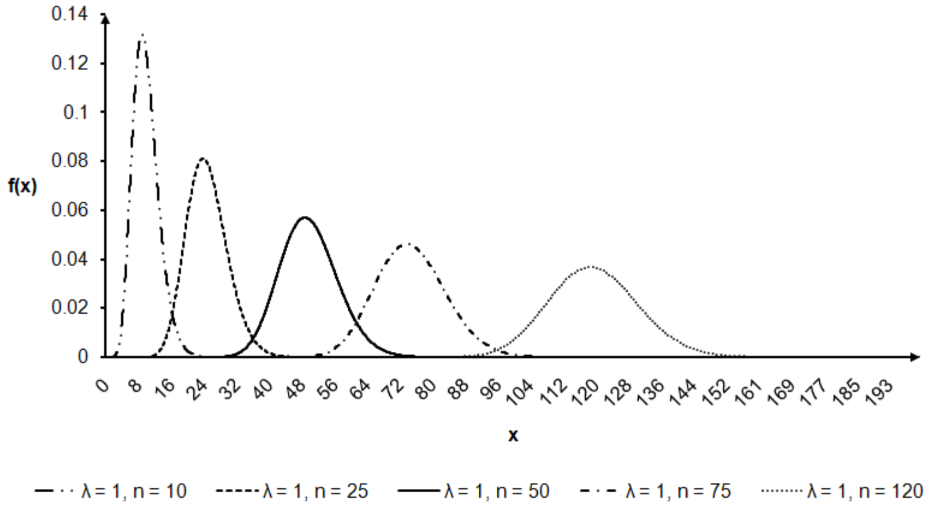

Figure 4.8: Erlang Density Function with varying values of n

- Frechet, when $k < 0$
- Weibull, when $k > 0$
- Gumbel, when $k = 0$

These distributions are particularly useful in accurately modeling large claims in non-life insurance policies for which other distributions potentially lack goodness-of-fit for these claims occurring in the upper tails [39].These three distributions are discussed in the Sections 4.1.7, 4.1.8 and 4.1.10.

A concept of probability weighted moments was introduced by Greenwood [59]. This is a generalization of the usual moments and provides a way in which the parameters of the three extremal distributions can be estimated.

Definition 27. ***Probability Weighted Moments***
The probability weighted moments of a random variable X with a distribution function $F(x) = \mathrm{P}(X \leq x)$ are given by

$$M_{p,r,s} = \mathrm{E}(X^p(F(x))^r(1 - F(x))^s). \tag{4.5}$$

The usual p^{th} order non-central moment of X is given by $M_{p,0,0}$.

The inverse distribution function of $F(x)$ in (4.4) is given by

$$x = x(F) = \begin{cases} \frac{\alpha}{k}(1 - (-\ln(F))^k) + \xi & k \neq 0, \\ \xi - \alpha \ln(-\ln(F)) & k = 0. \end{cases}$$

Deriving moments for X where $k \neq 0$
To derive the moments for X for the case where $k \neq 0$, one can follow Greenwood's approach using probability weighted moments.

For $p = 1$:

$$\begin{aligned}
M_{1,r,0} &= \mathrm{E}(X(F(x))^r(1-F(x))^0) \\
&= \int_0^1 x(F(x))^r dF \\
&= \int_0^1 (F(x))^r \left(\xi + \frac{\alpha}{k}(1-(-\ln(F))^k)\right) dF \\
&= -\int_{-\infty}^0 (\xi + \frac{\alpha}{k}(1-w^k))e^{-w(r+1)}dw \text{ by letting } w = \ln(F) \\
&= -\left(\int_{-\infty}^0 \xi e^{-w(r+1)}dw + \int_{-\infty}^0 \frac{\alpha}{k}(1-w^k)e^{-w(r+1)}dw\right) \\
&= \int_0^\infty (\xi + \frac{\alpha}{k})e^{-w(r+1)}dw - \frac{\alpha}{k}\int_0^\infty (r+1)^{-(k+1)}t^k e^{-t}dt \\
&\quad \text{by letting } w(r+1) = t \\
&= \frac{1}{r+1}\left((\xi + \frac{\alpha}{k}) - \frac{\alpha}{k}\frac{\Gamma(k+1)}{(r+1)^k}\right)
\end{aligned}$$

For $p = 2$:

$$\begin{aligned}
M_{2,r,0} &= \mathrm{E}(X^2(F(x))^r(1-F(x))^0) \\
&= \int_0^1 x^2(F(x))^r dF \\
&= \int_0^1 (F(x))^r \left(\xi + \frac{\alpha}{k}(1-(-\ln(F))^k)\right)^2 dF
\end{aligned}$$

Hence

$$
\begin{aligned}
M_{2,r,0} &= \int_0^1 \left((\frac{\alpha}{k}+\xi)^2 - 2((\frac{\alpha}{k})^2 + \frac{\alpha\xi}{k})(-\ln(F))^k + (-\ln(F))^{2k} \right) (F(X))^r dF \\
&= \frac{\left(\frac{\alpha}{k}+\xi\right)^2 (F(x))^{r+1}}{r+1}\bigg|_0^1 - 2\left((\frac{\alpha}{k})^2 + \frac{\alpha\xi}{k}\right) \int_0^1 (-\ln(F))^k F^r dF + \\
&\int_0^1 (-\ln(F))^{2k} F^r dF \\
&= \frac{\left(\frac{\alpha}{k}+\xi\right)^2}{r+1} - 2\left(\left(\frac{\alpha}{k}\right)^2 + \frac{\alpha\xi}{k}\right) \int_0^\infty w^k e^{-w(r+1)} dw + \\
&\int_0^\infty w^{2k} e^{-w(r+1)} dw \text{ by letting } w = -\ln(F) \\
&= \frac{\left(\frac{\alpha}{k}+\xi\right)^2}{r+1} - 2\left(\left(\frac{\alpha}{k}\right)^2 + \frac{\alpha\xi}{k}\right) \int_0^\infty \frac{t^k}{(r+1)^{k+1}} e^{-t} dt + \\
&\int_0^\infty \frac{t^{2k}}{(r+1)^{2k+1}} e^{-t} dt \text{ by letting } w = \frac{t}{r+1} \\
&= \frac{\left(\frac{\alpha}{k}+\xi\right)^2}{r+1} - 2\left(\left(\frac{\alpha}{k}\right)^2 + \frac{\alpha\xi}{k}\right) \frac{\Gamma(k+1)}{(r+1)^{k+1}} + \frac{\Gamma(2k+1)}{(r+1)^{k+2}} \qquad (4.6)
\end{aligned}
$$

The derivation of $M_{3,r,0}$ follows similarly to that of $M_{1,r,0}$ and $M_{2,r,0}$, but with the algebra being a bit more involved:

$$
\begin{aligned}
M_{3,r,0} &= \mathrm{E}(X^3(F(X))^r(1-F(X))^0) \\
&= \int_0^1 (x(F))^3 (F(x))^r dF \\
&= \left(\left(\frac{\alpha}{k}\right)(1-(-\ln(F))^k) + \xi\right)^2 (F(x))^r dF \\
&= \int_0^1 \left(\frac{\alpha}{k} + \frac{\xi\alpha}{k}\right)\left(\frac{\alpha}{k}+\xi\right)^2 F^r dF \\
&\quad - \left(2\left(\frac{\alpha}{k} + \frac{\xi\alpha}{k}\right)\left(\left(\frac{\alpha}{k}\right)^2 + \frac{\xi\alpha}{k}\right) - \frac{\alpha}{k}\left(\frac{\alpha}{k}+\xi\right)^2\right) \int_0^1 (-\ln(F))^k F^r dF \\
&\quad + \left(\left(\frac{\alpha}{k} + \frac{\xi\alpha}{k}\right) + 2\frac{\alpha}{k}\left(\left(\frac{\alpha}{k}\right)^2 + \frac{\alpha\xi}{k}\right)\right) \int_0^1 (-\ln(F))^{2k} F^r dF \\
&\quad - \frac{\alpha}{k}\int_0^1 (-\ln(F))^{3k} F^r dF \qquad (4.7)
\end{aligned}
$$

In (4.7) two distinct integrals can be identified with simplifications as given below:

$$
\int_0^1 F^r dF = \frac{1}{r+1} \qquad (4.8)
$$

and

$$
\begin{aligned}
\int_0^1 (-\ln(F))^{jk} F^r dF &= \int_0^\infty \left(\frac{w}{r+1}\right)^{jk} \frac{\left(e^{-\frac{w}{r+1}}\right)^r \left(e^{-\frac{w}{r+1}}\right)}{r+1} dw \\
&= \frac{1}{(r+1)^{jk+1}} \int_0^\infty w^{jk} e^{-w} dw \\
&= \frac{\Gamma(jk+1)}{(r+1)^{jk+1}} \qquad (4.9)
\end{aligned}
$$

where $j = 1, 2, 3$.

Using the results in (4.8) and (4.9) one can find the final expression for $M_{3,r,0}$:

$$
\begin{aligned}
M_{3,r,0} = {} & \frac{\left(\frac{\alpha}{k} + \frac{\xi\alpha}{k}\right)\left(\frac{\alpha}{k} + \xi\right)^2}{r+1} - \frac{\left(2\left(\frac{\alpha}{k} + \frac{\xi\alpha}{k}\right)\left(\frac{\alpha^2}{k^2} + \frac{\alpha\xi}{k}\right)\right)\Gamma(k+1)}{(r+1)^{k+1}} \\
& + \frac{\left(\frac{\alpha}{k}\left(\frac{\alpha}{k} + \xi\right)^2\right)\Gamma(k+1)}{(r+1)^{k+1}} \\
& + \left(\left(\frac{\alpha}{k} + \frac{\xi\alpha}{k}\right) + 2\frac{\alpha}{k}\left(\frac{\alpha^2}{k^2} + \frac{\alpha\xi}{k}\right)\right)\frac{\Gamma(2k+1)}{(r+1)^{2k+1}} - \frac{\alpha\Gamma(3k+1)}{k(r+1)^{3k+1}} \qquad (4.10)
\end{aligned}
$$

Hence if $k \neq 0$, the first three moments about the origin can be found from $M_{1,r,0}$, $M_{2,r,0}$ and $M_{3,r,0}$ where $r = 0$:

$$
\begin{aligned}
\mathrm{E}(X) &= M_{1,0,0} \\
&= \left(\xi + \frac{\alpha}{k}\right) - \frac{\alpha}{k}\Gamma(k+1) \qquad (4.11)
\end{aligned}
$$

and consequently the expression for the variance can be found:

$$
\mathrm{var}(X) = \Gamma(2k+1) - \frac{\alpha^2}{k^2}\left(\Gamma(k+1)\right)^2 \qquad (4.12)
$$

Deriving moments for X where $k = 0$

The derivations of these moments can be done similar to how Fisher and Tippett did it [54]. They consider each of the three possible cases that the Generalized Extreme Value distribution can take on. For the case where $k = 0$ (which rleates to the Gumbel distribution) they derive the moments that have the same form as the expressions for the mean, variance, skewness and kurtosis as given in (B.80) to (B.83) with $\xi = 0$ and $\alpha = 1$.

These moments and the derivation thereof will be studied in more detail in Section 4.1.10.

4.1.7 Frechet Distribution

The generalized form as proposed by Jenkinson [70] for maxima [98] is given below:

$$F(x) = e^{-\left(1-\frac{k(x-\xi)}{\alpha}\right)^{1/k}} \text{ with } k < 0,\ \alpha > 0 \text{ and } \xi \in \mathbb{R} \tag{4.13}$$

The form as given in (4.13) is for maxima. The form for minima is also given in [98]. Essentially this means that for n observed independent, identically distributed random variables, $X_1, X_2, ..., X_n$, the maximum will exhibit an Extreme Value distribution. Consider the order statistics $X_{(1)}, X_{(2)}, ..., X_{(n)}$, where $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$. The cumulative distribution function for the $X_{(n)}$ is given by

$$\begin{aligned}
F(x) &= \mathrm{P}(X_{(n)} \leq x) \\
&= \mathrm{P}(\{X_{(1)} \leq x\} \cup \{X_{(2)} \leq x\} \cup ... \cup \{X_{(n)} \leq x\}) \\
&= \mathrm{P}(\{X_1 \leq x\} \cup \{X_2 \leq x\} \cup ... \cup \{X_n \leq x\}) \\
&= \mathrm{P}(\{X_1 \leq x\})\, \mathrm{P}(\{X_2 \leq x\}) ... \mathrm{P}(\{X_n \leq x\}) \\
&\text{by independence of } X_1, X_2, ..., X_n \\
&= F_X(x) F_X(x) ... F_X(x) \\
&\text{since } X_1, X_2, ..., X_n \text{ are identically distributed} \\
&= (F_X(x))^n
\end{aligned}$$

where $F_X(x)$ is the cumulative distribution function of the X_i's.

If one considers the maxima of a set of independent, identically distributed random variables (with distribution function given by $F_X(x)$, the maxima is said to exhibit an Extreme Value distribution. If this distribution is assumed to be for the case where $k < 0$, then the maxima is said to exhibit a Frechet distribution. In this case it is then true that:

$$\begin{aligned}
F(x) = (F_X(x))^n = e^{-\left(\frac{\delta}{x-\lambda}\right)^{\beta}} \\
\text{by letting } \lambda = \left(\frac{\alpha}{k} + \xi\right), \delta = -\frac{\alpha}{k} \text{ and } \beta = -k^{-1}
\end{aligned} \tag{4.14}$$

The form of the cumulative distribution function in (4.14) is for the Frechet distribution as given by Park and Sohn [98] and is valid $\forall\ x \geq \lambda$.

If one considers an observed sample $\{Y_1, Y_2, ...Y_n\}$ of independent, identically distributed random variables, the order statistics of the observed sample are given by $\{Y_{(1)}, Y_{(2)}, ...Y_{(n)}\}$. The distribution of $Y_{(1)}$ can be found by using (4.14) if it is true that $Y_{(i)} \leq \lambda$. If this is true, one can find a series

$$\{-X_{(n)} + 2\lambda, -X_{(n-1)} + 2\lambda, ..., -X_{(1)} + 2\lambda\}$$

where

$$-X_{(n)} \leq -X_{(n-1)} \leq ... \leq -X_{(1)} \text{ and } Y_{(i)} = -X_{(n-i+1)} + 2\lambda \text{ for } i = 1, 2, ..., n$$

$$\begin{aligned}
\therefore \ \mathrm{P}(Y_{(1)} \leq y) &= \ \mathrm{P}(-X_{(n)} + 2\lambda \leq y) \\
&= 1 - \ \mathrm{P}(X_{(n)} < 2\lambda - y) \\
&= \begin{cases} 1 - e^{-\left(\frac{\delta}{2\lambda - y - \lambda}\right)^{\beta}} & \text{if } 2\lambda - y \geq \lambda, \\ 1 - 0 & \text{otherwise.} \end{cases} \quad \text{using (4.14)} \\
&= \begin{cases} 1 - e^{-\left(\frac{\delta}{\lambda - y}\right)^{\beta}} & \text{if } y \leq \lambda, \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}$$

The focus of the study is on distributions that can be used to model claim sizes and therefore the focus is on the form of the Frechet distribution for maxima. The probability density function , where $X \geq \lambda$ can be fund by considering the first order derivative of (4.14). This expression is given in (B.56).

The probability density function can now be used to obtain the moments abouth the origin for $X_{(n)}$. The general expression as given in (B.57).

From (B.57) and using the properties of the Gamma function the expressions for the mean, variance and skewness can be found. These expressions are presented in (B.58), (B.59) and (B.60).

Graphs can be constructed to study the behaviour of the shape and scale of this distribution for various sets of parameters. Consider Figures 4.9, 4.10 and 4.11 showing the probability density functions for various values of β, δ and λ, respectively.

It is clear from figure 4.9 that the scale of the density function is affected by the value of β which suggests that β is a scale parameter. Similar to β it can be seen that the value of δ is also affecting the scale of the density function (see figure 4.10). Lastly, it can be seen from figure 4.11 that λ only affects the location of the distribution, hence being a location parameter.

4.1.8 Weibull Distribution

Consider the expression for the Generalized Extreme Value distribution (4.4) for values of k larger than 0 as given in section 4.1.6.

$$F(y) = e^{-\left(-\frac{y-\lambda}{\delta}\right)^{\beta}} \text{ where } \delta = \frac{\alpha}{k}, \lambda = \left(\frac{\alpha}{k} + \xi\right) \text{ and } \beta = \frac{1}{k}. \tag{4.15}$$

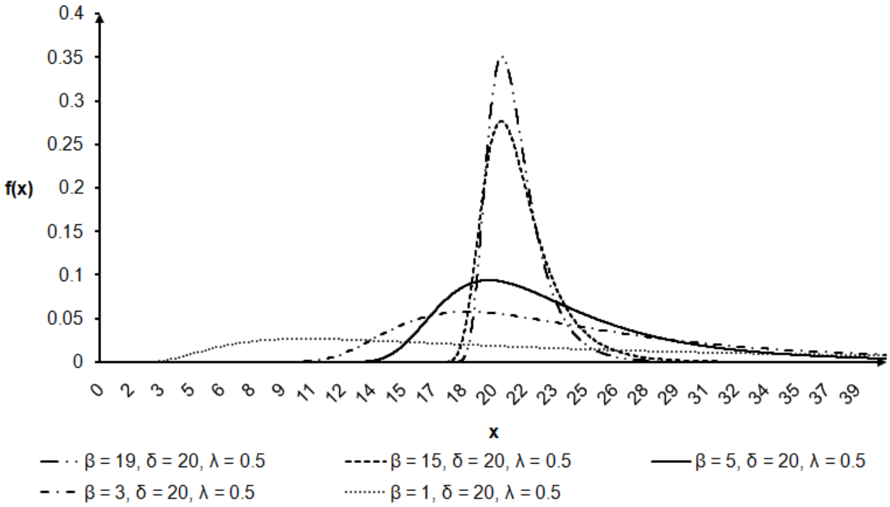

Figure 4.9: Frechet Density Function with varying values of β

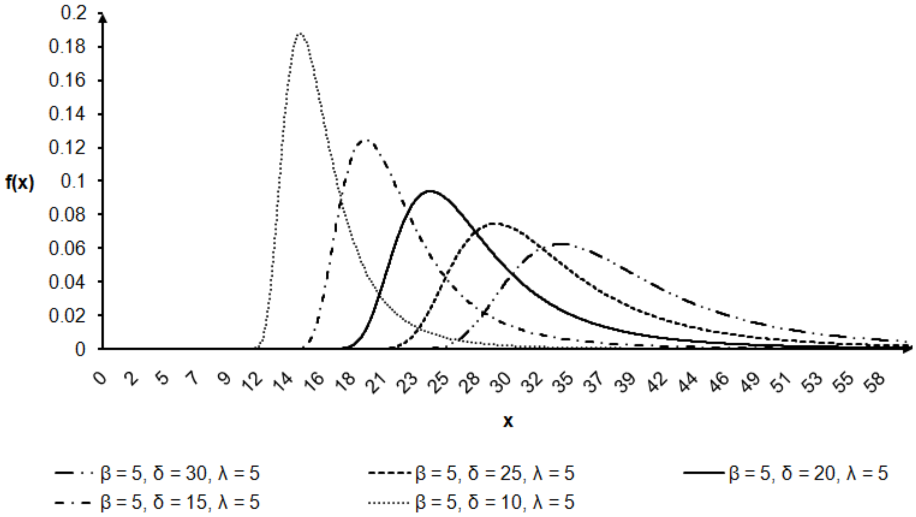

Figure 4.10: Frechet Density Function with varying values of δ

The cumulative distribution function given in (4.15) is referred to as the standardized Weibull distribution and is only valid for values of $Y \leq \lambda$. Also note that since this is for the case where $k > 0$, it means that $\beta > 0$. Klugman et al [75] state that this standardized form of the Weibull distribution is associated with the distribution of minimum values of distributions. This form is different from the well-known form given in (4.16). The standardized Weibull is also further associated with distributions that have a

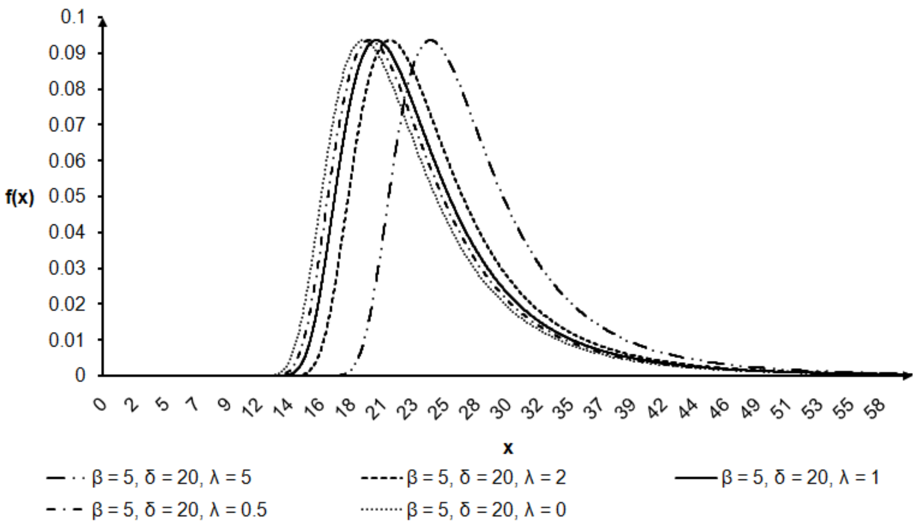

Figure 4.11: Frechet Density Function with varying values of λ

finite right-hand endpoint (in which case it is definitely not suitable for modelling) large claims.

The random variable Y associated with the standardized Weibull has a domain defined for values less than or equal to λ and that has a distribution with a finite upper tail. If we consider the converse to this where a variable is defined for values of λ or more and doesn't have a finite upper tail, it makes it useful for modelling of large claims. Consider the transformation $X = \lambda - Y$ and let $\delta = \theta$, then

$$\begin{aligned} F(x) &= 1 - F_Y(\lambda - x) \\ &= \begin{cases} 1 - exp\left(-\left(\frac{x}{\theta}\right)^{\beta}\right) & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The form as given in (4.16) is the expression often used in the literature [11], [105] and [75]. The probability density function in (B.179) is given by [75] The r^{th} moment, given in (B.180), follows directly from the probability density function by using the properties of the Gamma function:

$$\begin{aligned} \mathrm{E}(X^r) &= \int_0^{\infty} x^r \frac{\beta}{\theta^{\beta}} x^{\beta-1} e^{-\left(\frac{x}{\theta}\right)^{\beta}} dx \\ &= \theta^r \Gamma\left(\frac{r}{\beta} + 1\right) \text{ with } r > -\beta. \end{aligned} \quad (4.16)$$

Using (B.180), the expressions in (B.181), (B.182) and (B.183) for the mean, variance and skewness can be obtained. The moment generating function for the Weibull distribution does not exist [75], [105], [11].

Figures 4.12 and 4.13 shed some light on how the probability density function is varying in terms of its shape and scale for different values of θ and β. It should be noted that figure 4.12 shows various density functions for different values of θ while keeping the value β constant at 2. The special case where $\beta = 2$ is referred to as the Rayleigh Distribution and is discussed in section 4.1.9.

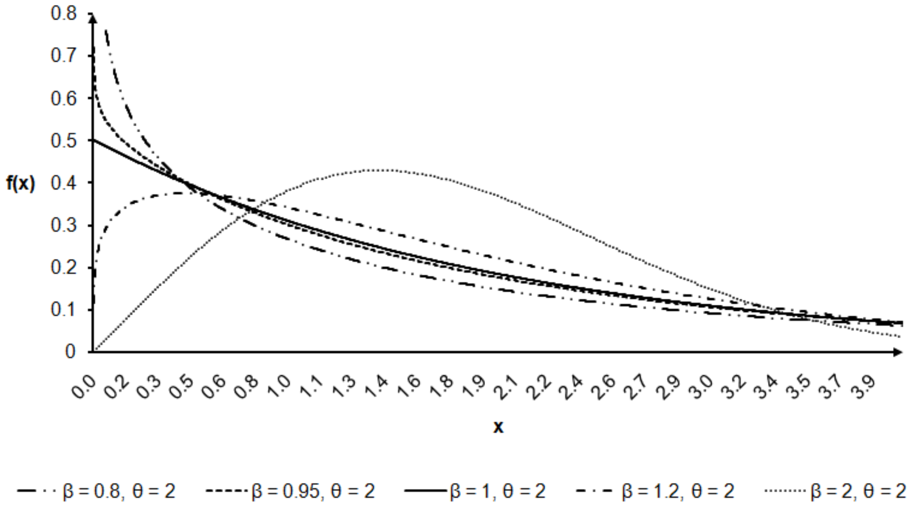

Figure 4.12: Weibull Density Function with varying values of θ

4.1.9 Rayleigh Distribution

The Rayleigh distribution is a special case of the Weibull, as given in (B.179), with $\beta = 2$ [11]. Using (B.179), the probability density function that is given in (B.146) can be obtained. The mean, variance and skewness follow directly from (B.180), (B.181), (B.182) and (B.183). Since the moment generating function for the Weibull distribution does not exist, the moment generating function also doesn't exist for the Rayleigh distribution.

4.1.10 Gumbel Distribution

Again, we start off with the generalized extreme value representation in (4.4) of Jenkinson [70]. The case where $k = 0$ relates to the Gumbel distribution [98], [76], [75]. This form is suitable for the distribution of the maxima from

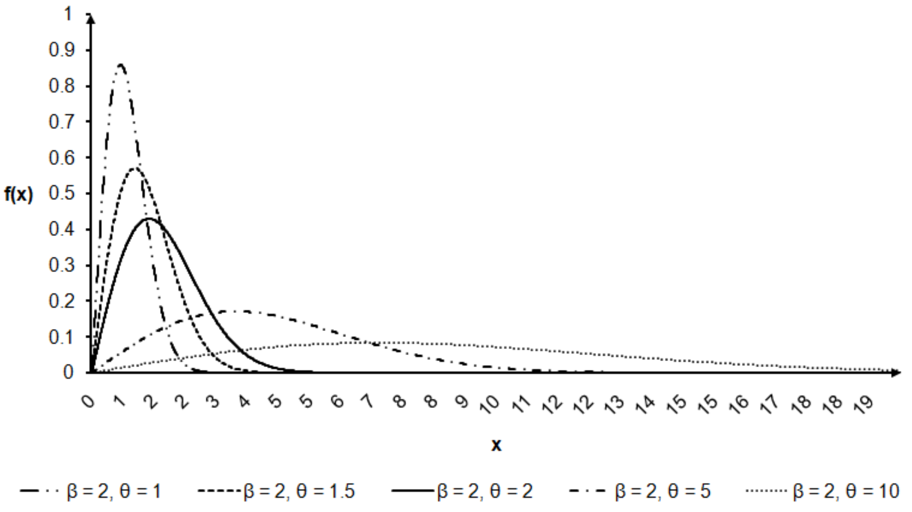

Figure 4.13: Weibull Density Function with varying values of β

vector samples [98]. This means that if we were to observe multiple sets of claims, we can identify the largest claim from each set of claims (maximum). This will give us a series of maxima. Let the random variables for any one set of claims be denoted by $(X_1, X_2, \ldots, X_n)$ with $X_{(n)}$ the random variable for the largest order statistic of the set of claims. The largest claim will be the observed value for the maximum denoted by $X_{(n)}$. The Gumbel distribution is then a potential distribution that can be used to model the maxima of the sets of claims.

The associated distribution for the minima can be obtained by defining a series $(Y_1, Y_2, ...Y_n)$, where $Y_i = (2\xi - X_i)$ and $\xi > 0$. Since we also have a series $(X_1, X_2, ...X_n)$, we can obtain two series of order statistics $(Y_{(1)}, Y_{(2)}, ...Y_{(n)})$ and $(X_{(1)}, X_{(2)}, ...X_{(n)})$ where $Y_{(i)} = 2\xi - X_{(n-i+1)}$. In particular $Y_{(1)} = 2\xi - X_{(n)}$. The cumulative distribution function can then be derived for the minima:

$$F(y) = \mathrm{P}(Y_{(1)} \leq y) = 1 - e^{-e^{-\left(\frac{\xi - y}{\alpha}\right)}}$$

The probability density function of the Gumbel distribution can be found by taking the first order derivative of (4.4) for the case where $k = 0$. A generalization of the Gumbel distribution is given by Adeyemi and Ojo [2] in terms of the probability density function:

$$f(u) = \frac{\lambda^\lambda}{\Gamma(\lambda)} e^{-\lambda u - \lambda e^{-u}}, \text{ where } -\infty < u < \infty,\ \lambda > 0 \qquad (4.17)$$

If one let $u = \dfrac{x-\xi}{\alpha}$ and $\lambda = 1$, then the expression in (4.17) reduces to the probability density function of the Gumbel distribution as given in (B.79). To find the moments of X, consider the relationship between the moment generating function of X with the moment generating function of U:

$$\begin{aligned} M_X(t) &= \mathrm{E}(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\alpha}e^{-\left(e^{-\left(\frac{x-\xi}{\alpha}\right)}+\left(\frac{x-\xi}{\alpha}\right)\right)}dx \\ &= e^{t\xi}\int_{-\infty}^{\infty} e^{tu\alpha}e^{-\left(e^{-u}+u\right)}du \text{ by lettting } u = \frac{x-\xi}{\alpha} \\ &= e^{t\xi}M_U(\tau) \text{ where } \tau = t\alpha \end{aligned} \tag{4.18}$$

The moment generating function for U is given by Adeyemi and Ojo:

$$M_U(\tau) = \frac{\lambda^{\tau}}{\Gamma(\lambda)}\Gamma(\lambda-\tau) \tag{4.19}$$

They define the cumulant generating function in terms of the moment generating function as:

$$c_U(\tau) = ln(M_U(\tau)) = \tau ln(\lambda) + ln(\Gamma(\lambda-\tau)) - ln(\Gamma(\lambda)) \tag{4.20}$$

From (4.20) the cumulant generating function for X can be derived as

$$\begin{aligned} c_X(t) &= \ln(M_X(t)) \\ &= \ln(e^{t\xi}M_U(\tau)) \text{ from (4.18)} \\ &= t\xi + c_U(\tau) \\ &= t\xi + \ln(\Gamma(1-\alpha t)) \text{ if } \lambda = 1 \end{aligned} \tag{4.21}$$

The Maclaurin series expansion of (4.21) at the point 0 [111] yields

$$\begin{aligned} c_X(t) &= c_X(0) + \frac{c_X^{'}(0)}{1!}t + \frac{c_X^{''}(0)}{2!}t^2 + \frac{c_X^{'''}(0)}{3!}t^3 + \dots \\ &= 0 + \zeta_1\frac{t}{1!} + \zeta_2\frac{t^2}{2!} + \zeta 3\frac{t^3}{3!} + \dots \text{ where } \zeta_r = \frac{d^r}{dt^r}c_X(t) \end{aligned} \tag{4.22}$$

From Adeyemi and Ojo follows that

$$\begin{aligned} \frac{d^r}{dt^r}\ln\Gamma(\lambda-t)|_{t=0} &= (r-1)!\left(\sum_{j=1}^{\infty} j^{-r} - \sum_{j=1}^{\lambda-1} j^{-r}\right) \\ &= (r-1)!\sum_{j=1}^{\infty} j^{-r} \text{ since } \lambda = 1 \end{aligned}$$

Hence, the first three cumulants can be evaluated using (4.22) and the expressions given by Adeyemi and Ojo [2] for the first four cumulants:

$$\begin{aligned}
\zeta_1 &= \xi + \frac{d}{dt}\ln(\Gamma(1-\alpha t))|_{t=0} \\
&= \xi + \alpha(\gamma - \sum_{j=1}^{\lambda-1} j^{-1}) \\
&= \xi + \alpha\gamma \\
&\text{where } \gamma \equiv \text{Euler-Mascheroni constant,}([2],[75],[76],[111]) \\
&= \xi + 0.57721566\alpha
\end{aligned}$$

$$\begin{aligned}
\zeta_2 &= \frac{d^2}{dt^2}(\ln(\Gamma(\lambda-\alpha t)))|_{t=0} \\
&= \alpha^2 \sum_{j=1}^{\infty} j^{-2} \\
&= \alpha^2 \frac{\pi^2}{6} \text{ from } [111],[2],[76] \\
&= 1.64493407\alpha^2
\end{aligned} \tag{4.23}$$

$$\begin{aligned}
\zeta_3 &= \frac{d^3}{dt^3}(\ln(\Gamma(\lambda-\alpha t)))|_{t=0} \\
&= 2\alpha^3 \sum_{j=1}^{\infty} j^{-3} \\
&\approx 2(1.2021)\alpha^3 \text{ from } [111],[2] \\
&= 2.4042\alpha^3
\end{aligned}$$

$$\begin{aligned}
\zeta_4 &= \frac{d^4}{dt^4}(\ln(\Gamma(\lambda-\alpha t)))|_{t=0} \\
&= 3\alpha^4 \sum_{j=1}^{\infty} j^{-4} \\
&= \frac{\pi^4}{15}\alpha^3 \text{ from } [2]
\end{aligned}$$

If we consider the form of the probability density function as given in (B.79), the effects of varying the values of α and ξ can be seen in Figures 4.14 and 4.15. It is clear from these figures that α is the scale parameter while ξ affects the location.

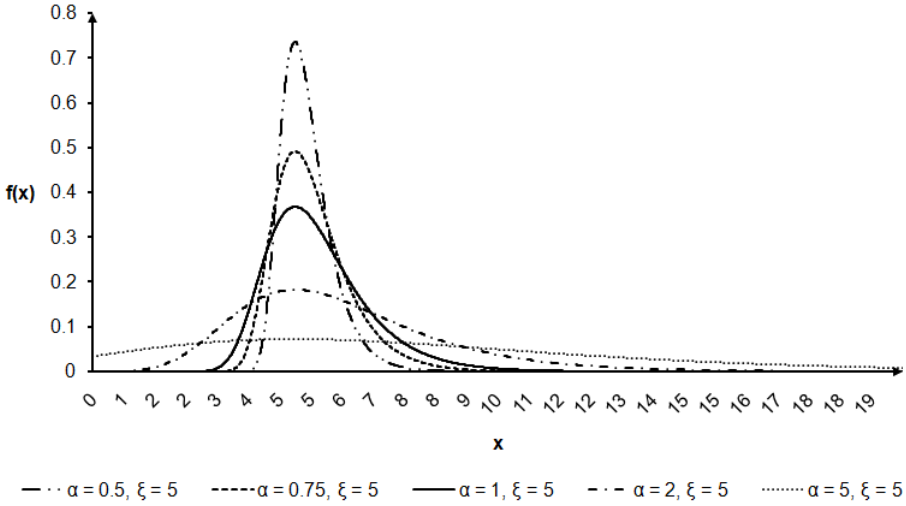

Figure 4.14: Gumbel Density Function with varying values of α

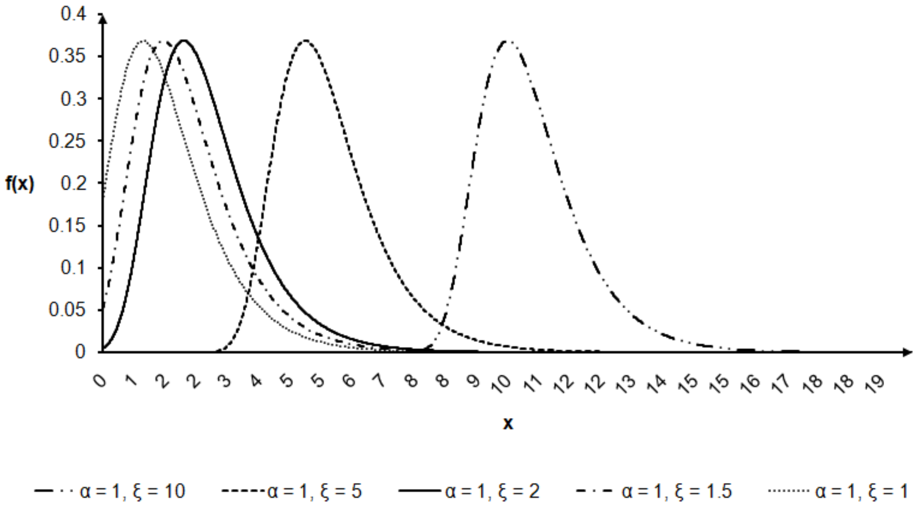

Figure 4.15: Gumbel Density Function with varying values of ξ

4.1.11 Pareto (Lomax) Distribution

The probability density function for a random variable X that is said to have Pareto Type II distribution [11] is given in (B.139). The probability density function can be rewritten as follows [75]:

$$f_X(x) = \frac{\kappa\theta^\kappa}{(x+\theta)^{\kappa+1}} \text{ for } x > 0 \tag{4.24}$$

In this form it is also known as the Lomax distribution [75]. Lomax originally used the following functional form in 1954 as a cumulative distribution function [113]:

$$F_X(x) = 1 - \left(\frac{a}{x+a}\right)^k \tag{4.25}$$

in order to model business failure data. The associated probability density function (that can be found by taking the first order derivative of (4.25) with respect to x) is the same as expression (B.139) with $\kappa = k$ and $\theta = a$.

The general expression to calculate the r^{th} moment (B.140) can be obtained using the properties of the beta function (as defined in Definition 28).

Definition 28. ***Beta Function*** *Let $B(x, y)$ be defined for x, y in $P = \{x \in \mathbb{R} : x > 0\}$ by*

$$\begin{aligned} B(x,y) &= \int_{0^+}^{1^-} t^{x-1}(1-t)^{y-1}dt \\ &= \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \end{aligned} \tag{4.26}$$

This integral will be proper if $x \geq 1$ and $y \geq 1$. If $0 < x < 1$ or $0 < y < 1$, the integral is improper [15].

$$\begin{aligned} \mathrm{E}(X^r) &= \int_0^\infty x^r \frac{\kappa}{\theta}\left(1+\frac{x}{\theta}\right)^{-(\kappa+1)} dx \\ &= \theta^r \kappa \int_0^1 (1-y)^{(r+1)-1} y^{(\kappa-r)-1} dy \text{ by letting } y = \frac{\theta}{\theta+x} \\ &= \theta^r \frac{\Gamma(r+1)\Gamma(\kappa-r)}{\Gamma(\kappa)} \text{ by using definition 28} \end{aligned}$$

The first four moments about the origin follows directly from (B.140). These moments can then be used to find the expressions for the mean, variance and skewness given in (B.141),(B.142) and (B.143).

In Section 4.1.12 it will be shown that the Pareto Type II distribution is a special case of the Generalized Pareto distribution. For the purpose of calculating the skewness of a random variable X from a Pareto Type II distribution, one can also leverage of the expression as derived for the Generalized Pareto given in (B.78).

The moment generating function for the Pareto Type II distribution does not exist. It is therefore not possible to study the heaviness of the tail of this distribution by means of using a moment generating function. The cumulative distribution function can, however, be used and is given below:

$$F_X(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\kappa} \quad \text{for } x > 0 \tag{4.27}$$

The effects of varying the parameter values (that is for κ and θ) can be seen in Figures 4.16 and 4.17.

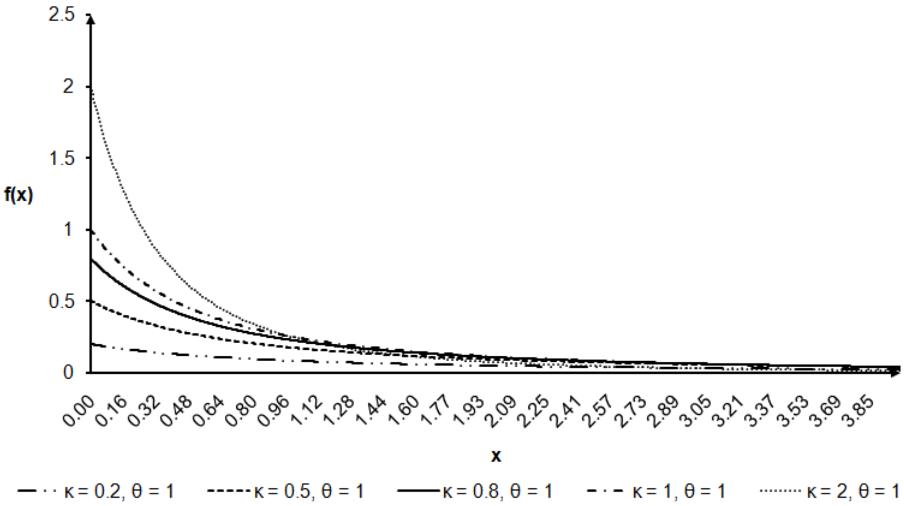

Figure 4.16: Pareto Type II Density Function with varying values of κ

4.1.12 Generalized Pareto Distribution

Denuit et al [39] indicate the use of the Generalized Pareto distribution in accurately modeling large claims in non-life insurance policies for which other distributions potentially lack goodness-of-fit for these claims occurring in the upper tails. An expression for the probability density function, as presented by Klugman et al [75], for a random variable defined on $\mathbb{R}^+$ that is said to have a Generalized Pareto distribution is given in (B.74).

An expression to calculate the r^{th} moment is given by Klugman et al [75]. This expression, as given in (B.75), can be used to find expressions for the moments about the origin directly from which expressions for the mean, variance and skewness follows - see (B.76) to (B.78).

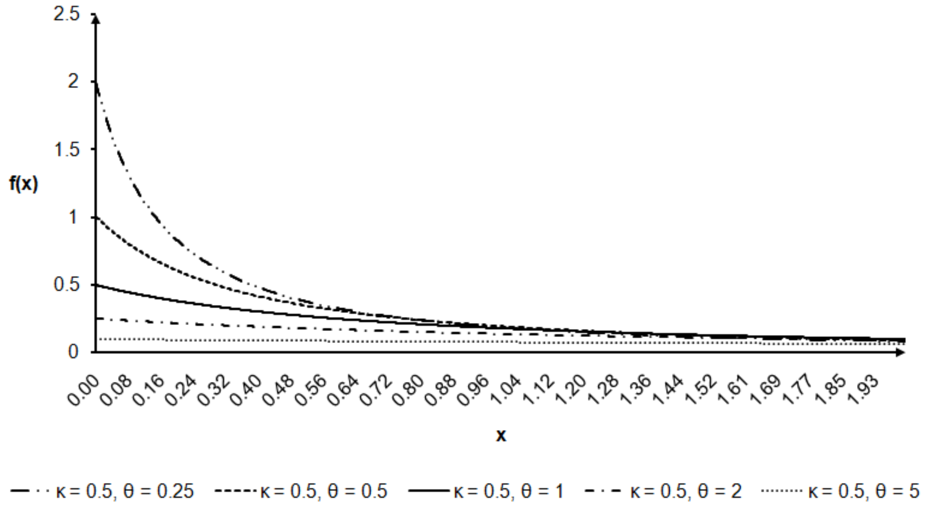

Figure 4.17: Pareto Type II Density Function with varying values of θ

If we let $\tau = 1$ in the probability density function (B.74), then it reduces to:

$$f_X(x) = \frac{\kappa\theta^\kappa}{(x+\theta)^{\kappa+1}} \text{ which is the same as the expression in (B.139)}$$

The Pareto Type II (Lomax) distribution is therefore a special case of the Generalized Pareto distribution. Consequently the mean, variance and skewness of the Pareto Type II disribution can be calculated directly from (B.76), (B.77) and (B.78).

The effect of using different values for the parameters κ, τ and θ can be seen in figures 4.18, 4.19 and 4.20.

4.1.13 Lognormal Distribution

A random variable X that has a Lognormal distribution (denoted $X \sim LN(\mu, \sigma^2)$) has a probability density function given by [11], [75]

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(\ln(x)-\mu)^2}{\sigma^2}\right)} \text{ for } x > 0 \text{ with } -\infty < \mu < \infty \text{ and } \sigma > 0.$$

There exists a relationship, given in Result 2 between the Lognormal and Normal distributions that is quite useful. We will use this relationship to derive moments for the Lognormal distribution.

Result 2. *A random variable* $Y \sim LN(\mu, \sigma^2) \iff$ *the random variable* $X = \ln(Y) \sim N(\mu, \sigma^2)$ *[82], [11].*

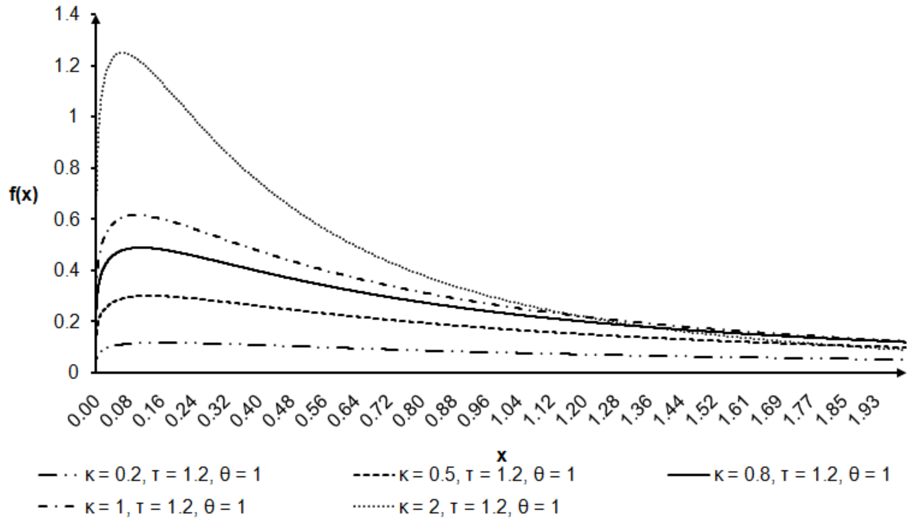

Figure 4.18: Generalized Pareto Density Function with varying values of κ

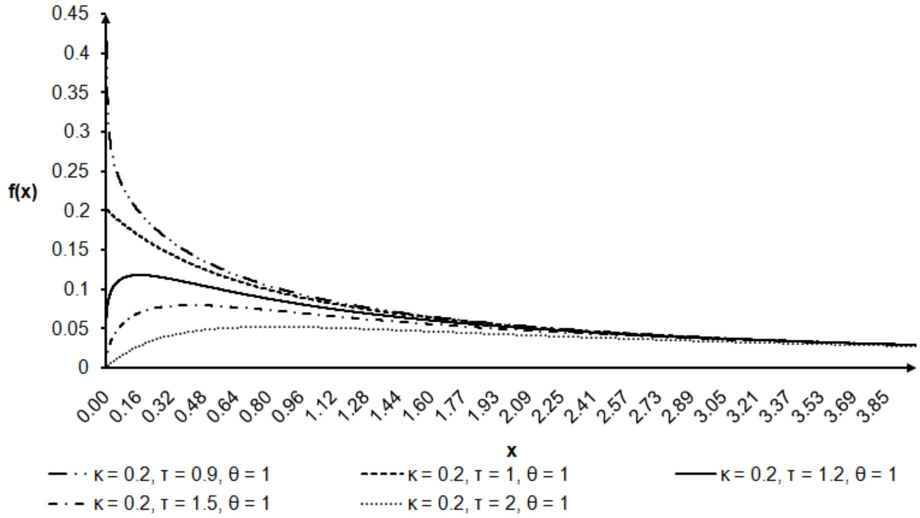

Figure 4.19: Generalized Pareto Density Function with varying values of τ

Expression (B.127) for the r^{th} moment of a random variable X that is $LN(\mu, \sigma^2)$ distributed can be obtained using Result 2:

$$\mathrm{E}(X^r) = \mathrm{E}((e^Y)^r) = M_Y(r),$$

where $M_Y(r)$ is the moment generating function of Y which is Normal distributed with parameters μ and σ^2. From (B.127) the mean as well as the rest of the first four moments about the origin can be obtained from which

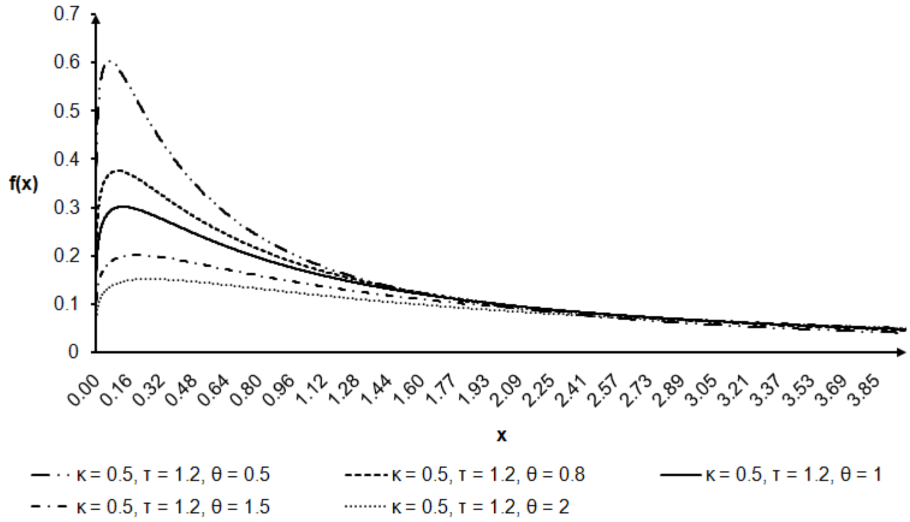

Figure 4.20: Generalized Pareto Density Function with varying values of θ

expressions (B.129) and (B.130) for the variance and skewness follow.

It is clear from (B.130) that the distribution is positively skewed for all values σ which makes it suitable for modeling insurance claims. Figures 4.21 and 4.22 illustrates this skewness as well as how the scale and shape of the distribution is affected by varying the values of δ_1 and δ_2.

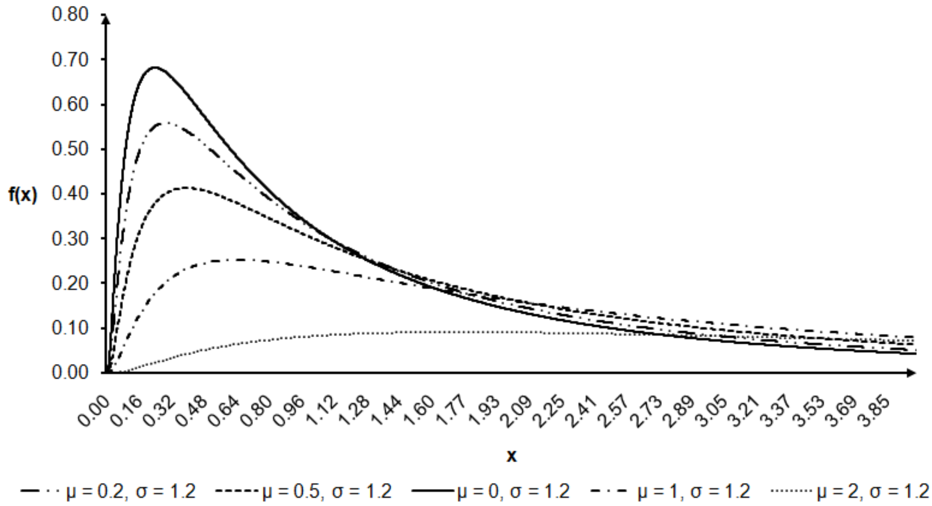

Figure 4.21: Lognormal Density Function with varying values of μ

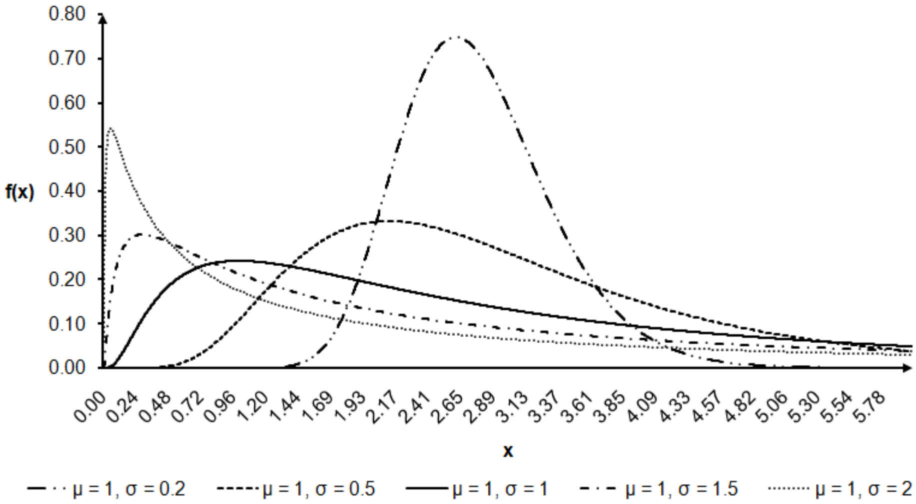

Figure 4.22: Lognormal Density Function with varying values of σ

In the literature is is argued that the Lognormal distribution is popular for modeling of non-life insurance claims sizes [41], for example modeling of motor insurance claim data [104]. While there may be instances in which the data appears to be close to normal, the Normal distribution may appear to lack realism in the context of modeling claims data in that the Normal random variables can also be negative. Luenberger specifically highlights this shortcoming when considering distributions for modeling stock prices [82]. He concludes that the log of the change in stock prices appears to be normal, but mentions that real-life stock prices suggests heavier tails, higher concentration around the mean with somewhat of a larger weight attached to larger changes which suggests some degree of skewness. It is in this context that Goerg [58] argues for the use of skew distributions even when data appears to be close to normal, but with a slight skewness and therefore supports the view of considering the Lognormal distribution in the context of modeling claims size data.

4.1.14 Beta-prime Distribution (Pearson Type VI)

Fabián [51] considers the Beta-prime distribution function in the context of finding an appropriate measure of central tendency. In this context the following form is considered:

$$f_{p,q}(x) = \frac{1}{xB(p,q)} \frac{x^p}{(x+1)^{p+q}} \text{ for } x \geq 0 \text{ , with } p, q > 0 \tag{4.28}$$

This form of the distribution is also known as the Pearson Type VI distribution.

Bradlow et al [21] consider this distribution as prior distribution of the parameter of the negative binomial distribution used to perform Bayesian inference on the negative binomial distribution. The form they consider is a reparameterization of the form considered by Fabián which is given in (B.1). The r^{th} moment can be derived by recognising that (B.1) can be rewritten in a format that is similar to the integrand of the beta function (as given in Definition 28) by letting $\frac{x}{x+1} = z$:

$$\begin{aligned} \mathrm{E}(X^r) &= \frac{1}{B(\delta_1, \delta_2)} \int_0^1 z^{(\delta_1+r)-1}(1-z)^{(\delta_2-r)-1} dz \\ &= \frac{B(\delta_1 + r, \delta_2 - r)}{B(\delta_1, \delta_2)} \end{aligned}$$

From (B.2) the first three moments can be derived in order to obtain expressions for the mean, variance and skewness coefficient which are given in (B.3), (B.4) and (B.5).

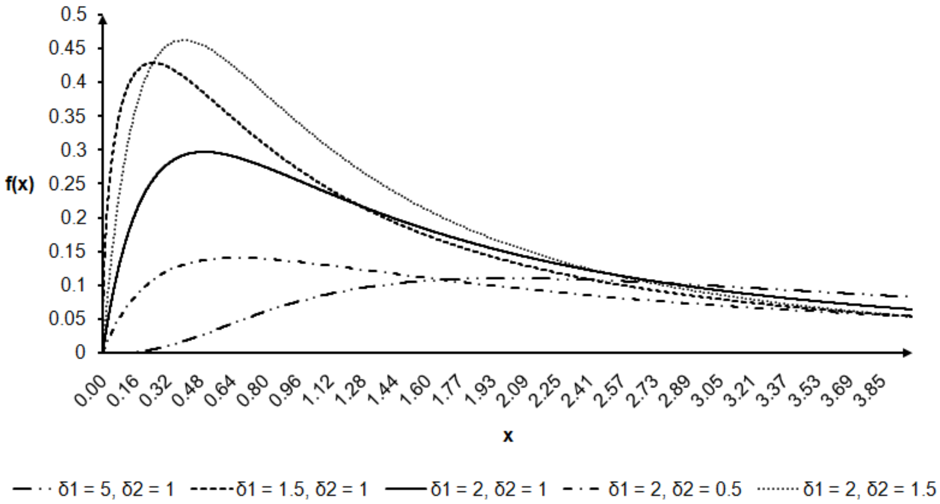

Figure 4.23: Pearson Type VI Density Function with varying values of δ_1 and δ_2

The scale of this distribution is affected by both δ_1 and δ_2. This is illustrated in Figure 4.23.

4.1.15 Birnbaum-Saunders Distribution

Rieck and Nedelman [103] defines the Birnbaum-Saunders distribution for a random variable T having a cumulative distribution function given by

$$F_T(t) = \Phi\left(\alpha^{-1}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right) \text{ for } t \geq 0 \text{ with } \alpha, \beta > 0 \tag{4.29}$$

where $\Phi(x)$ is the standard Normal cumulative distribution function, and α is the shape parameter and β is the scale parameter as well as the median of T.

The distribution was originally introduced by Birnbaum and Saunders [19]. The distribution, as it is presented here, follows from renewal theory for the number of time frames needed in order to observe a specific event. One can therefore use this distribution to model the number of months until the first claim arise for a specific insured party.

From the cumulative distribution function follows that [19]

$$F_T(t) = \Phi\left(\alpha^{-1}h\left(\frac{t}{\beta}\right)\right) \text{ where } h(x) = \sqrt{x} - \frac{1}{\sqrt{x}}$$

from which it follows that

$$\begin{aligned} \alpha^{-1}h\left(\frac{T}{\beta}\right) &\sim N(0,1) \\ \therefore h\left(\frac{T}{\beta}\right) &\sim N(0,\alpha^2) \end{aligned}$$

and

$$\therefore \text{ if } X \sim N\left(0, \frac{\alpha^2}{4}\right).$$

Therefore it means that

$$2X = h\left(\frac{T}{\beta}\right).$$

If one let $\varphi(x) = h^{-1}(2x)$, $T = \beta\varphi(X)$. It is further shown that

$$\begin{aligned} \varphi(X) &= \left(1 + 2X^2 + 2X\sqrt{(X^2+1)}\right) \\ \therefore T &= \beta\left(1 + 2X^2 + 2X\sqrt{(X^2+1)}\right) \text{ with } X \sim N\left(0, \frac{\alpha^2}{4}\right). \end{aligned} \tag{4.30}$$

The mean and variance as given by Birnbaum and Saunders [19] are given in (B.7) and (B.8) in Appendix B. It is also shown that the density function of T is a maximum at $t = \beta$, which means that the median and the mode of T are equal to the same value β. From (B.7) the mean of T is strictly larger than the median and mode which indicates that the distribution possesses positive skewness.

In order to obtain a graphical presentation of the Birnbaum-Saunders distribution, one must first obtain an expression of the probability density function. This expression, as given in (B.6), can be derived from (4.29) using the chain rule and the Fundamental Theorem of Calculus [111].

The resulting expression, as given in (B.6), is also given by Kundu et al [77] when studying the shape of hazard function of the Birnbaum-Saunders distribution.

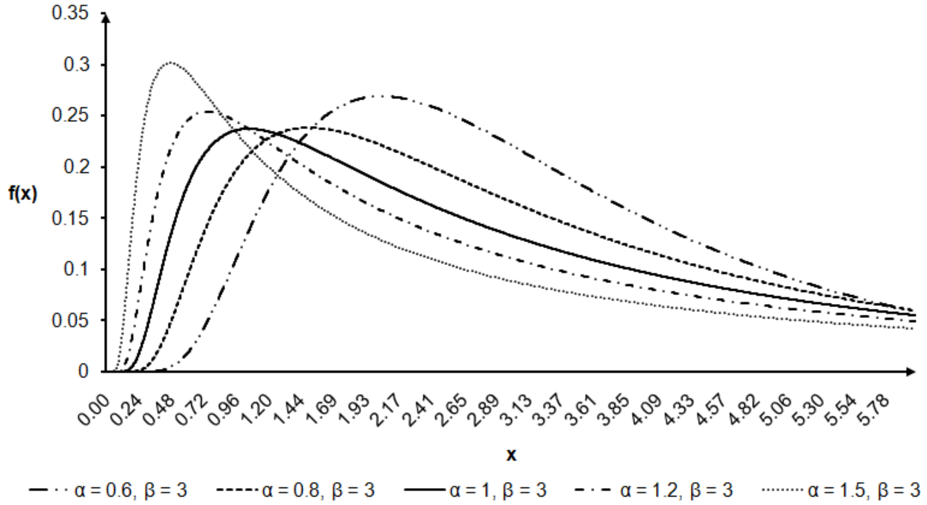

Figure 4.24: Birnbaum-Saunders Density Function with varying values of α

Expression (B.6) was used to generate graphs for the density of T for varying values of α and β as shown in Figures 4.24 and 4.25.

4.1.16 Burr Distribution

A number of forms of cumulative distribution functions were suggested by Burr [26] yielding a wide range of values for skewness and kurtosis [113]. These functions all have either a similar algebraic form or are reparameterizarions of one another. The form that Burr considers and further analyses

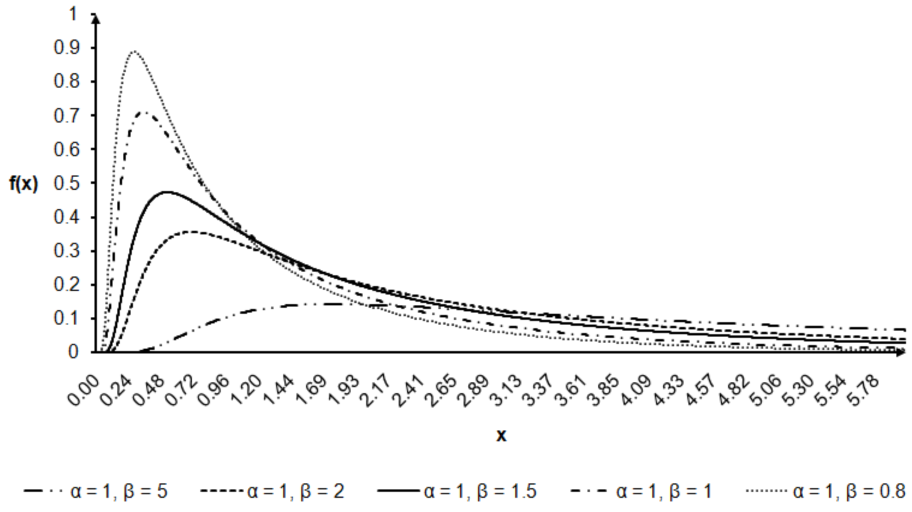

Figure 4.25: Birnbaum-Saunders Density Function with varying values of β

in his paper [26] is the form known as Type XII of the Burr distribution which has a cumulative distribution given by

$$F_X(x) = 1 - (1 + x^c)^{-k} \text{ for } x > 0, \text{ with } c, k > 0 \tag{4.31}$$

Both c and k are shape parameters.

Tadikamalla [113] argues that the Burr distribution can be fitted to a given dataset by matching the mean, variance, skewness and kurtosis - which essentially refers to the method-of-moments [11].

A scale parameter α can be introduced as follows [113]:

$$F_X(x) = 1 - \left(1 + \left(\frac{x}{\alpha}\right)^c\right)^{-k} \tag{4.32}$$

that can be rewritten as

$$F_X(x) = 1 - \left(1 + \frac{x^c}{\alpha^*}\right)^{-k}, \text{ where } \alpha^* = \alpha^c.$$

The probability density function, given in (B.9), can be derived for the Type XII distribution by taking the derivative with respect to x of the cumulative distribution function (4.31). Taking the first derivative of the probability density function and setting it equal to 0 gives that if $c > 1$, the distribution is unimodal at the value of $x = \sqrt[c]{\frac{c-1}{kc+1}}$.

The r^{th} moment of X, as given in (B.10), can be derived as follows:

$$\begin{aligned}
\mathrm{E}(X^r) &= \int_0^\infty x^r kcx^{c-1}\left(1+x^c\right)^{-(k+1)}dx \\
&= k\int_0^1 y^{\frac{r+c}{c}-1}(1-y)^{k-\frac{r}{c}-1}dy \\
&\qquad \text{by letting } y = \frac{x^c}{1+x^c} \\
&= kB\left(\frac{r}{c}+1, k-\frac{r}{c}\right) \text{ using definition 28} \\
&= k\frac{\Gamma(\frac{r}{c}+1)\Gamma(k-\frac{r}{c})}{\Gamma(k+1)}
\end{aligned}$$

The first four moments follows directly from B.10. From these moments expressions for the mean and variance, given in (B.11) and (B.12), can be calculated. To obtain expressions for for the skewness and kurtosis the following general expressions can be used:

$$\text{skewness}(X) = \frac{\mathrm{E}(X^3) - 3\,\mathrm{E}(X)\,\mathrm{E}(X^2) + 2(\,\mathrm{E}(X))^3}{\sqrt{\text{var}(X)}} \tag{4.33}$$

$$\text{kurtosis}(X) = \frac{\mathrm{E}(X^4) - 4\,\mathrm{E}(X)\,\mathrm{E}(X^3) + 6\,\mathrm{E}(X^2)(\,\mathrm{E}(X))^2 - 3(\,\mathrm{E}(X))^4}{(\,\text{var}(X))^2} \tag{4.34}$$

Figures 4.26 and 4.27 shows how the shape and scale of the probability density function for the Burr Type XII distribution varies with varying values of k and c.

Let $Y = \frac{1}{X}$, then

$$\begin{aligned}
G_Y(y) &= 1 - F_X\left(\frac{1}{y}\right) \\
&= (1+y^{-c})^{-k} \text{ for } y > 0 \text{ with } c, k > 0
\end{aligned} \tag{4.35}$$

The distribution as given in (4.35) for random variable Y is referred to as the Burr Type III distribution [113]. The probability density function can be obtained by taking the first derivative of (4.35) with respect to y:

$$g_Y(y) = kcy^{-(c+1)}(1+y^{-c})^{-(k+1)} \text{ for } y > 0 \text{ with } c, k > 0$$

Furthermore, the r^{th} moment can be derived from the relationship between random variables X and Y using (B.10):

$$\mathrm{E}(Y^r) = \,\mathrm{E}(X^{-r}) = k\frac{\Gamma\left(1-\frac{r}{c}\right)\Gamma\left(k+\frac{r}{c}\right)}{\Gamma(k+1)} \tag{4.36}$$

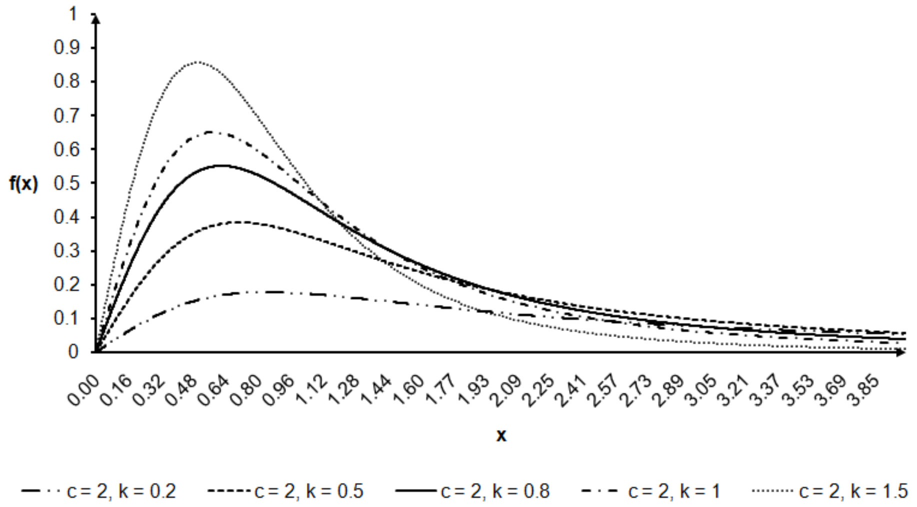

Figure 4.26: Burr Type XII Density Function with varying values of k

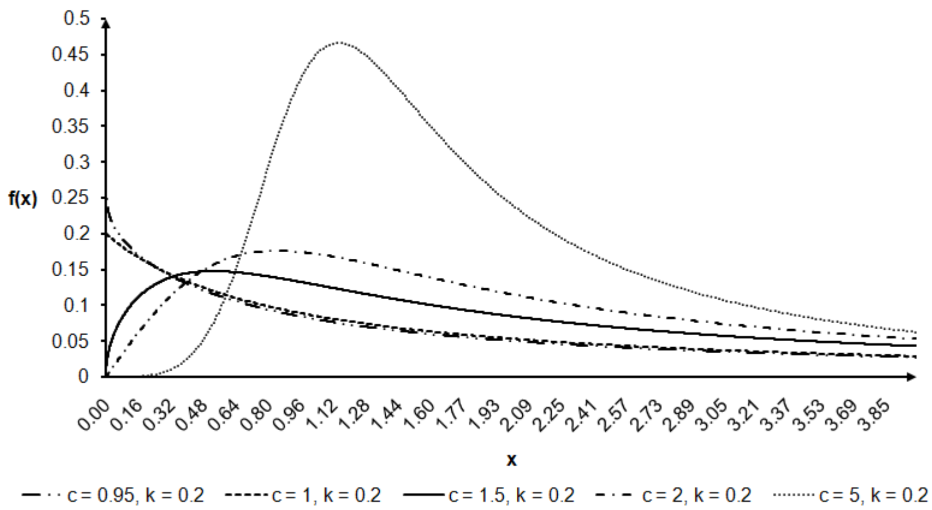

Figure 4.27: Burr Type XII Density Function with varying values of c

The first three moments and the variance follows directly from (4.36) and from the moments and variance derived for Type XII.

$$\mathrm{E}(Y) = k\frac{\Gamma\left(k+\frac{1}{c}\right)\Gamma\left(1-\frac{1}{c}\right)}{\Gamma(k+1)}$$

and

$$\text{var}(Y) = kB\left(\frac{2}{c}+k, 1-\frac{2}{c}\right) - k^2\left(B\left(\frac{1}{c}+k, 1-\frac{1}{c}\right)\right)^2$$

Thus far we have that X is Burr Type XII distributed. Then $Y = \frac{1}{X}$ is Burr Type III distributed. If we now consider $Z = c\ln(Y)$, then Z is again Burr distributed Type II [113].

$$\begin{aligned} F_Z(z) &= G_Y\left(e^{z/c}\right) \\ &= \left(1+e^{-z}\right)^{-k} \text{ for } -\infty < z < \infty \text{ with } k > 0 \end{aligned} \tag{4.37}$$

Consequently the probability density function can be obtained by taking the first order derivative of (4.37) to get:

$$f_Z(z) = k\frac{e^{zk}}{(1+e^z)^{k+1}}$$

Furthermore, the moment generating function can be obtained as follows:

$$\begin{aligned} M_Z(t) &= \mathrm{E}\left(e^{tZ}\right) \\ &= \mathrm{E}\left(Y^{tc}\right) \\ &= k\frac{\Gamma(1-t)\Gamma(k+t)}{\Gamma(k+1)} \text{ using (4.36)} \end{aligned}$$

from which the moments of the Burr Type II distribution can be derived.

Figures 4.29 and 4.28 shows how the densities of the Burr Type III distribution varies with varying values of c and k.

4.1.17 Dagum Distribution

This distribution was originally introduced by Dagum in 1977 with a distribution function [74]:

$$F_X(x) = \left(1+\left(\frac{b}{x}\right)^a\right)^{-p} \text{ for } x > 0 \text{ with } a, b, p > 0 \tag{4.38}$$

This is the Dagum Type I distribution. Generalizations of this distributions were introduced by Dagum in 1977 and 1980 that were used to model the distribution of personal income [34], [35]. It is therefore expected that the distribution should possess positive skewness.

An important link with the Burr distribution can be found when (4.38)

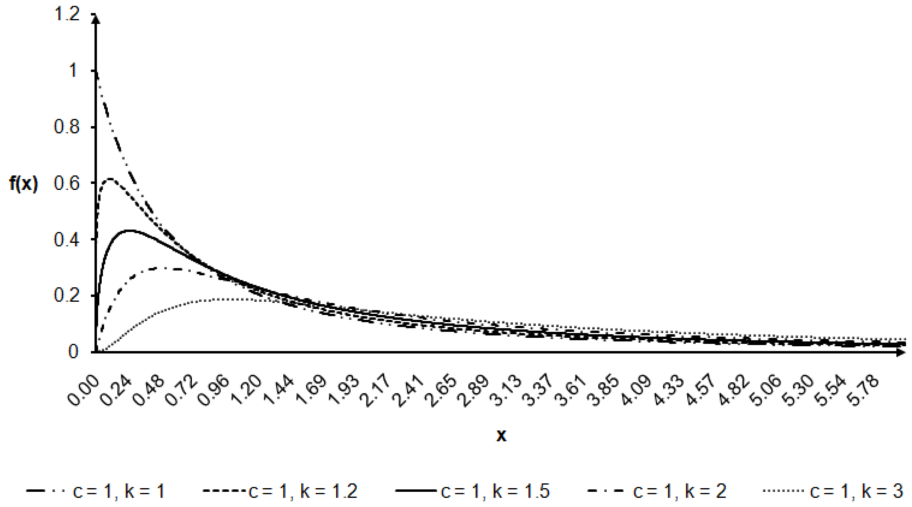

Figure 4.28: Burr Type III Density Function with varying values of k

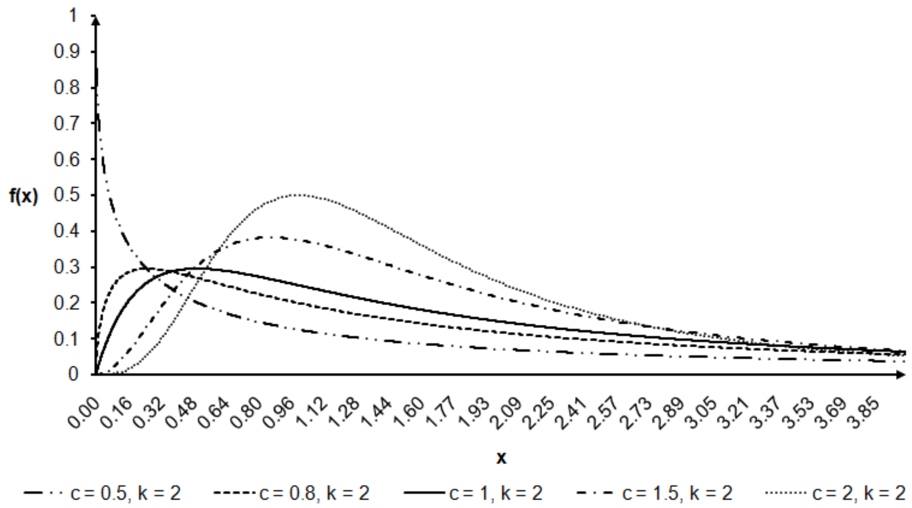

Figure 4.29: Burr Type III Density Function with varying values of c

is compared with the expression for the cumulative distribution function of a random variable with a Burr Type XII distribution that incorporates a scale parameter - which is given in (4.32). Suppose Y is Burr distributed with parameters b^{-1}, a and p, then let $X = Y^{-1}$ in which case it follows from (4.32) that

$$F_X(x) = 1 - F_Y\left(\frac{1}{x}\right) = \left(1 + \left(\frac{b}{x}\right)^a\right)^{-p}$$

which is the cumulative distribution function of a random variable being Dagum(b, a, p) distributed.

The probability density function of X, as given in (B.21), can be obtained by taking the first derivative of (4.38) with respect to x.

An expression for the r^{th}, as given in (B.22), moment can be derived as follows:

$$\begin{aligned}
\mathrm{E}(X^r) &= \int_0^\infty x^r ab^a p x^{-(a+1)} \left(1 + b^a x^{-a}\right)^{-(p+1)} dx \\
&= b^r p \int_0^1 y^{\left(\frac{r}{a}+p\right)-1}(1-y)^{\left(1-\frac{r}{a}\right)-1} \quad \text{by letting } y = \frac{x^a}{x^a + b^a} \\
&= b^r p B\left(\frac{r}{a} + p, 1 - \frac{r}{a}\right) \qquad (4.39)
\end{aligned}$$

The moments follow directly from (B.22) and are given in (B.23) and (B.24).

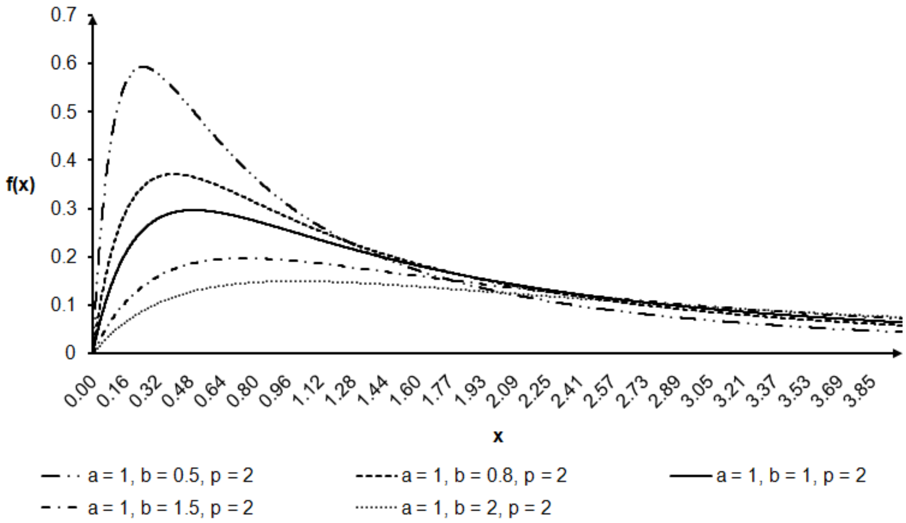

Figure 4.30: Dagum Density Function with varying values of b

For varying values of a, the shape of the density function will be affected in a similar way to how the shape of the Burr Type III distribution is affected by varying values of c as illustrated in Figure 4.29. Figures 4.30 and 4.31 illustrates how the scale and location is affected by varying values of b and p.

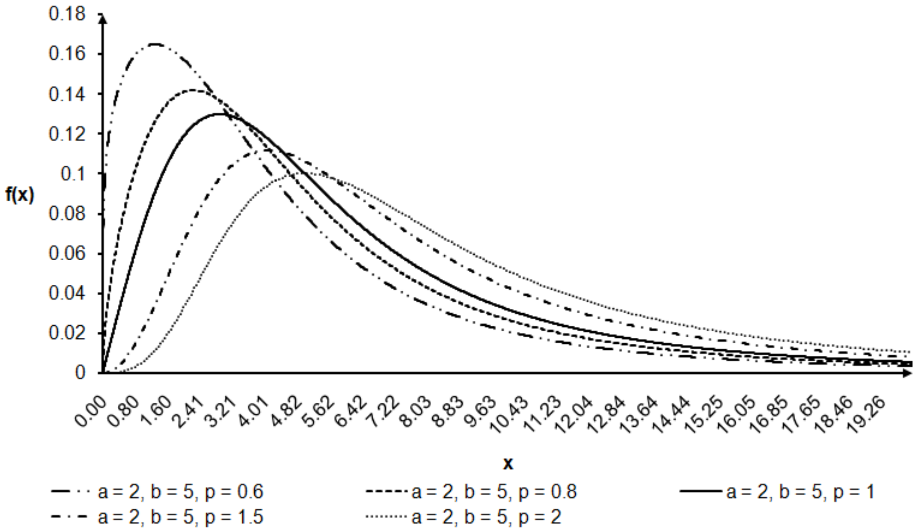

Figure 4.31: Dagum Density Function with varying values of p

4.1.18 Generalized Beta Distribution of the Second Kind

A distribution that was first introduced in 1984, Kleiber [74] discusses similarities of this distribution with respect to the Dagum distribution (as discussed in Section 4.1.17) and the Singh-Maddala distribution (as discussed in Section 4.1.19). The probability density function for a random variable X that is said to be Generalized Beta distributed ($X \sim GB2(a, b, p, q)$) is given by (B.70).

It appears as if the probability density function is very similar to the probability density function of the Generalized Pareto distribution given in (B.74). Klugman et al [75] also specifically refer to the Generalized Pareto distribution as the beta of the second kind, which is not exactly the same as the Generalized Beta of the Second Kind that they present. It is therefore not possible to obtain a direct link between the Generalized Pareto distribution and the Generalized Beta of the Second Kind.

Klugman et al [75] refer to the Generalized Beta of the Second Kind also by *transformed beta* and the *Pearson Type VI* distributions. If we consider a special case of the probability density function of the Generalized Beta (of the Second Kind) distribution with $a, b = 1$, then we get the following function:

$$f_X(x)|_{a=1,b=1} = \frac{x^{p-1}}{B(p,q)(1+x)^{p+1}} \text{ for } x \geq 0 \text{ with } p, q > 0$$

which is exactly the same as the expression for the probability density function of the beta prime distribution (Pearson Type VI) as given by Fabián [51] - see (4.28).

A general expression can be derived using (B.70) that can be used to obtain the r^{th} moment. Key steps in deriving this expression, as given in (B.71), are shown here:

$$\begin{aligned}
\mathrm{E}(X^r) &= \int_0^\infty \frac{x^{r-aq-1}ab^{aq}}{B(p,q)} \left(\frac{\left(\frac{x}{b}\right)^a}{1+\left(\frac{x}{b}\right)^a} \right)^{p+q} dx \\
&= \frac{b^r}{B(p,q)} \int_0^1 y^{\left(\frac{r}{\alpha}\right)-1}(1-y)^{\left(q-\frac{r}{a}\right)-1} dy \qquad \text{by letting } y = \frac{\left(\frac{x}{b}\right)^a}{1+\left(\frac{x}{b}\right)^a} \\
&= b^r \frac{B\left(\frac{r}{a}+p, q-\frac{r}{a}\right)}{B(p,q)} \text{ from the Definition 28 of the Beta function}
\end{aligned}$$

From (B.71) the first four moments about the origin can be derived, from which the expressions for the mean and variance, as given in (B.72) and (B.73), follow directly.

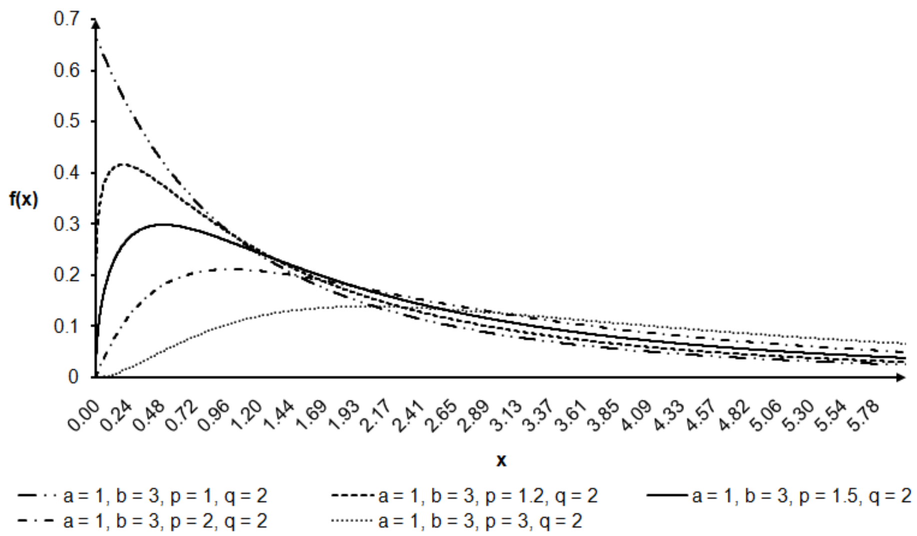

Figure 4.32: Generalized Beta (of the Second Kind) Density Function with varying values of p

The effect of varying the values of parameter a is similar to the effect of varying parameter c in the Burr Type III distribution - which is illustrated in Figure 4.29. Also the effect of varying the values of parameters b is similar to the effect of varying parameter b for the Dagum distribution - which is illustrated in Figure 4.30. The effects of varying the values of parameters p

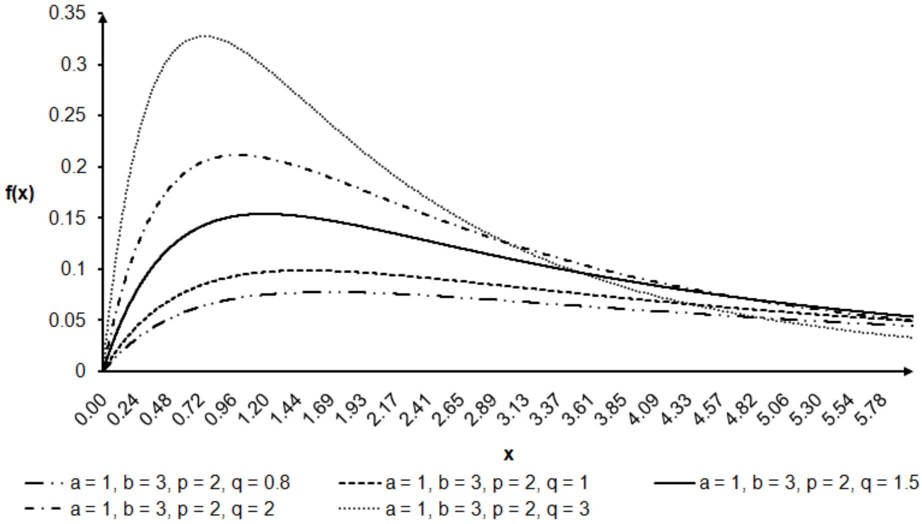

Figure 4.33: Generalized Beta (of the Second Kind) Density Function with varying values of q

and q can be seen in Figures 4.32 and 4.33. It is evident from these figures that the value of parameter p has an effect on the shape and scale of the distribution while the value of parameter q has an effect on the scale of the distribution.

4.1.19 Singh-Maddala Distribution

Kleiber [74] states in a theorem that a random variable X is Singh-Maddala distributed with parameters a, b and q if and only if $Y = \frac{1}{X}$ is Dagum$(a, \frac{1}{b}, q)$ distributed. Using this theorem, we derive the distribution and density functions for random variable X from the expression for the cumulative distribution function of the Dagum distribution - given by (4.38).

$$\begin{aligned} F_X(x) &= 1 - F_Y\left(\frac{1}{x}\right) \\ &= 1 - \left(1 + \left(\frac{x}{b}\right)^q\right)^{-q} \quad \text{with } a, b, q > 0 \end{aligned} \tag{4.40}$$

It can be seen that the cumulative distribution function (4.40) for X is a special case of the Burr Type XII distribution with a scale parameter, as given in (4.32), with $c = k(= q)$.

By taking the first derivative of the cumulative distribution function with respect to x the probability density function, given in (B.153), can be obtained. The r^{th} moment can now be derived by using (B.153):

$$\begin{aligned}
\mathrm{E}(X^r) &= \int_0^\infty x^r \frac{qa}{x}\left(1+\left(\frac{x}{b}\right)^a\right)^{-(q+1)}\left(\frac{x}{b}\right)^a dx \\
&= qb^r \int_0^1 y^{\left(\frac{r}{a}+1\right)-1}(1-y)^{\left(q-\frac{r}{a}\right)-1} dy \text{ by letting } y=\left(\frac{x}{b}\right)^a \\
&= qb^r B\left(\frac{r}{a}+1, q-\frac{r}{a}\right) \text{ from definition 28.}
\end{aligned}$$

The first four moments of X follows directly from (B.154) from which expressions for the variance, skewness and kurtosis can be found. The expressions for the mean and the variance are given in (B.155) and (B.156).

If a random variable X is Lomax(a, k) distributed, then the probability density function is exactly the same as (B.139) given in Section 4.1.11 with $\kappa = k$ and $\theta = a$. If we now consider the probability density function for the Singh-Maddala distribution, given in (B.153) and let $a = 1, b = a$ and $q = k$, we get an expression which is exactly the same as the probability density function for the Lomax distribution with parameters a and k. Furthermore consider the cumulative distribution function for the Lomax distribution:

$$F_X(x) = 1 - \left(1 + \frac{x}{a}\right)^{-k}.$$

This is a special case of the Burr Type XII distribution (with a scale parameter), since it is the same as expression in (4.32) with $c = 1$ and $\alpha = a$.

In conclusion, if a random variable is Lomax(a, k) distributed, it is also Singh-Maddala$(1, a, k)$ distributed as well as Burr Type XII distributed (with a scale parameter) and with parameter $c = 1$ [113]. Moments for this special case of the Singh-Maddala distribution can be derived from (B.154):

$$\begin{aligned}
\mathrm{E}(X) &= \frac{a}{k-1} \\
\mathrm{var}(X) &= \frac{a^2 k}{(k-1)^2(k-2)} \\
\mathrm{skewness}(X) &= 2\left(\frac{k+1}{k-3}\right)\sqrt{\frac{k-2}{k}} \text{ for } k \geq 2
\end{aligned}$$

Note that for $k = 2$ the skewness$(X) = 0$ in which case the distribution is considered to be symmetric.

Because of the similarity between the Dagum distribution and the Singh-Maddala the shape, scale and location of the Singh-Maddala distribution are affected by varying values of parameters a,b and q in a similar manner as to how the shape, scale and location of the Dagum distribution is affected by varying values of parameters a, b and p. The distribution's shape is primarily affected by the choice of parameter a as illustrated in Figure 4.27 for varying values of parameter c in the Burr Type XII distribution (recall that we have established a correspondence between the Dagum and Burr Type XII distributions in Section 4.1.17).

The scale and location of the Singh-Maddala distribution is affected by parameter b in a similar manner as to how parameter b affects the scale of the Dagum distribution. Since the Singh-Maddala distribution is a special case of the Dagum distribution with parameter $b = \frac{1}{b}$ it is expected that the scale of the Singh-Maddala distribution when $b = 1.25$ (for example) would be comparable with the scale if the Dagum distribution when $b = 0.8$ (if the other parameters are kept the same). The effect of varying values for b is illustrated in Figure 4.30. Note that the effect of increasing b for the Singh-Maddala distribution will be illustrated by the effect of decreasing b for the Dagum distribution due to the inverse relationship between the b parameters for these distributions.

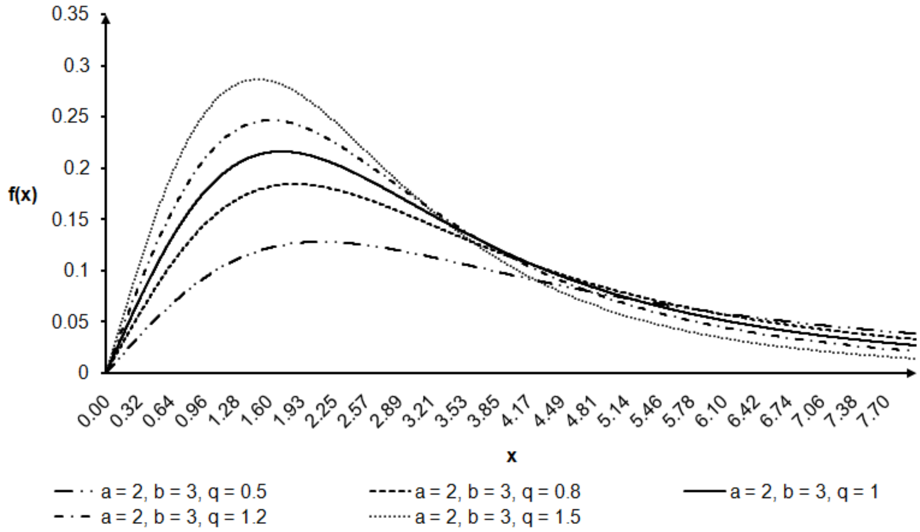

Figure 4.34: Singh-Maddala Density Function with varying values of q

Lastly the effect of varying parameter q for the Singh-Maddala distribution appears to be somewhat different from the effect that varying values of p have on the Dagum distribution. Figure 4.34 illustrates the effect of varying values of q on the Singh-Maddala distribution.

4.1.20 Kappa Family of Distributions

There are two variations considered by Tadikamalla [113]; the two-parameter and the three-parameter case.

Two-parameter Case

In this instance the two parameters relate to:

Shape parameter - β

Scale parameter - α

This two-parameter version of the distribution was introduced in 1973 by Mielke [85] and is a special case of the Burr Type III distribution as given in (4.35). Consider a random variable X being Burr Type III distributed with cumulative distribution function given by (4.35). Let $Y = \beta X$ and let the parameters of the Burr Type III distribution be taken as $c = \alpha$ and $k = \frac{1}{\alpha}$:

$$F_Y(y) = \left(1 + \left(\frac{y}{\beta}\right)^{-\alpha}\right)^{-\frac{1}{\alpha}} \quad \text{for } y \geq 0 \text{ with } \alpha > 0 \text{ and } \beta > 0. \tag{4.41}$$

The random variable Y is said to have a Kappa (α, β) distribution [113]. The probability density function, given in (B.104), can be found by taking the first order derivative of (4.41).

To derive the r^{th} moment of Y, one can leverage of (4.36) derived for the Burr Type III distribution, recognising that $\mathrm{E}(Y^r) = \beta^r \, \mathrm{E}(X^r)$ where X is Burr Type III distributed. From this approach the expression in (B.105) can be obtained. Expressions for the mean, variance and skewness as given in (B.106) to (B.108) can be found from the general expression for the r^{th} moment of Y in a similar way that the moments for the Burr Type III distribution was found in Section 4.1.16.

Graphs were generated for varying values of α and β to study the effect that these parameters have on the scale, shape and location of the density functions of the Kappa Two-Parameter distribution.

Figure 4.35 suggests that for smaller values of α the scale would be larger (seen in larger variances with larger values of α from (B.107)) while the location is also affected with increased values as α decreases. The shape is also affected by the value of α.

Figure 4.36 suggests that for larger values of β the scale is increased (corresponding with (B.107)) while the location is also increased. The shape does, however appear not be affected by the value of β.

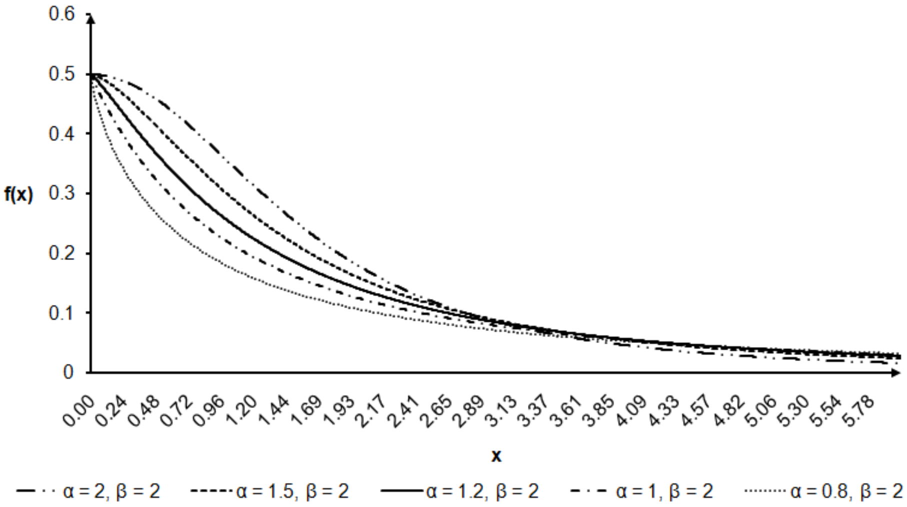

Figure 4.35: Kappa Two-Parameter Density Function with varying values of α

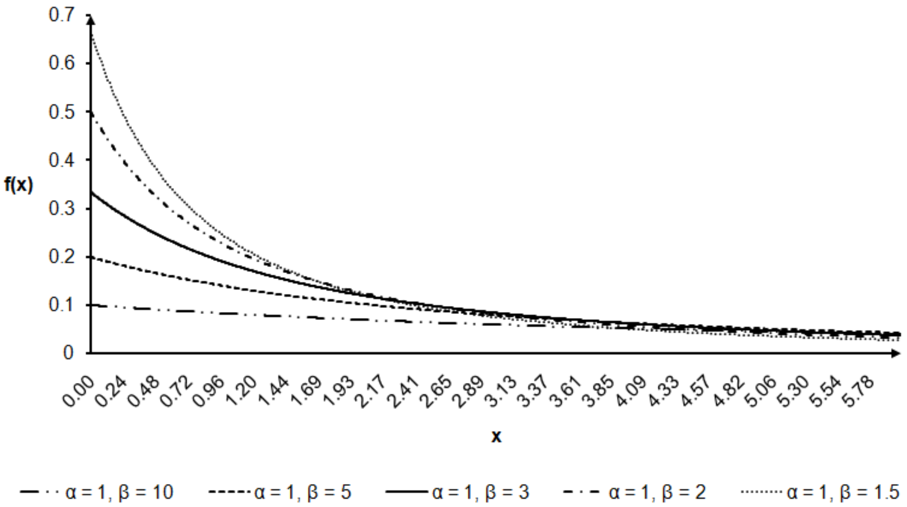

Figure 4.36: Kappa Two-Parameter Density Function with varying values of β

Three-parameter Case

Mielke and Johnson [86] introduced a version of the three-parameter distribution in 1973 [113], [24]:

$$F_X(x) = \left(\frac{\left(\frac{x}{\beta}\right)^{\alpha\theta}}{\alpha + \left(\frac{x}{\beta}\right)^{\alpha\theta}} \right)^{\frac{1}{\alpha}}$$

An alternative parameterization was given by Mielke and Johnson in 1974. Both the original and alternative forms have been used in hydrology and meteorology problems as alternatives to the Gamma and Lognormal distributions [113]. The alternative form is a special case of the Generalized Beta distribution of the Second Kind, as given in (B.70), with $a = \theta$, $b = \beta$, $p = \alpha$ and $q = 1$. The probability density function of this alternative form is given in (B.109).

This can can also be seen as a special case of the Burr Type III distribution with distributin function as given in (4.35). If we consider a random variable Y that has a Burr Type distribution and let $X = \beta Y$ and further let $c = \theta$ and $k = \alpha$, then

$$F_X(x) = F_Y\left(\frac{x}{\beta}\right) = \left(\frac{\left(\frac{x}{\beta}\right)^\theta}{1 + \left(\frac{x}{\beta}\right)^\theta}\right)^\alpha \quad \text{for } x \geq 0 \text{ with } \alpha, \beta, \theta > 0 \tag{4.42}$$

One can use the cumulative distribution as derived in (4.42) to obtain the associated probability density function by differentiating with respect to x. The resulting expression has exactly the same form as the probability density function given in (B.109). Therefore, it can be concluded that the three-parameter Kappa distribution is a special case of the Generalized Beta distribution (of the the Second Kind), as given in (B.70), with $a = \theta$, $p = \alpha$, $b = \beta$ and q=1 as well as being a special case of the Burr Type III distribution.

The general expression for the r^{th} moment, given in (B.110), can therefore be derived by using (B.71), as derived for the Generalized Beta distribution (of the Second Kind), by making suitable substitutions for a, b, p and q:

Using (B.110), the first four moments about the origin can be derived.

These values can be evaluated from which the mean and variance follows - see (B.111) and (B.112). Furthermore, the skewness and kurtosis can be calculated using (4.33) and (4.34).

Figures 4.37 and 4.38 show that the values of α and β affect the scale and location of the Kappa Three-Parameter distribution. The shape of the distribution is affected by the value of θ as can be seen in Figure 4.39.

4.1.21 Loggamma Distribution

If a random variable X is Loggamma distributed with parameters a and λ then its probability density function is given by [113], [105]:

It is given by Tadikamalla [113] that if a random variable Y is Gamma distributed with parameters a and λ, then $X = e^Y$ is Loggamma distributed

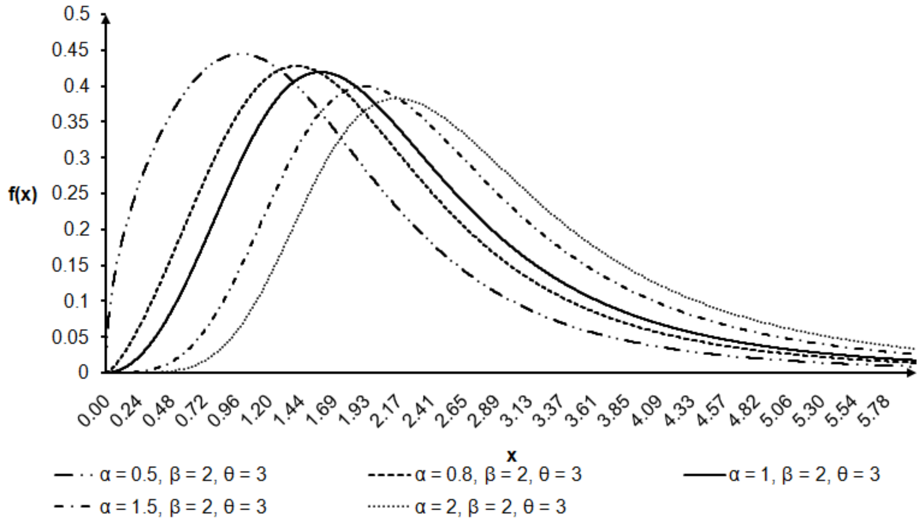

Figure 4.37: Kappa Three-Parameter Density Function with varying values of α

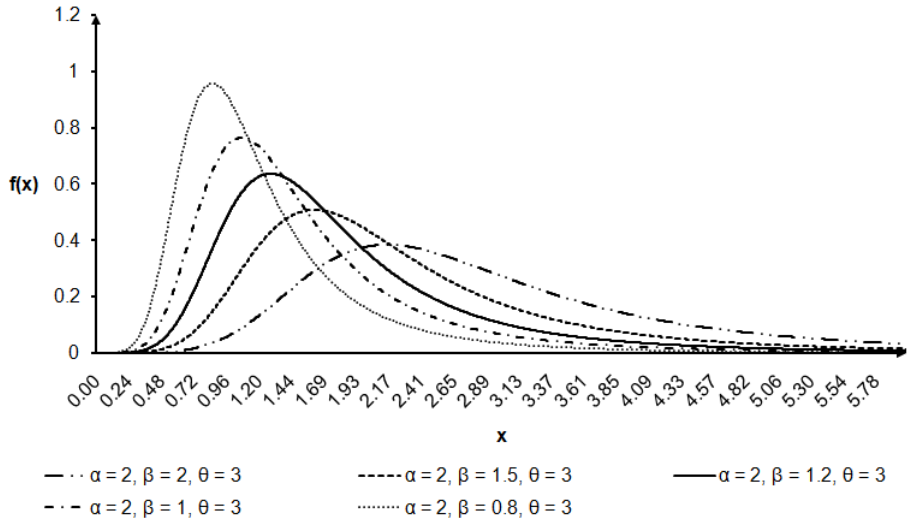

Figure 4.38: Kappa Three-Parameter Density Function with varying values of β

with parameters a and λ.
Consider the probability density function of random variable Y using the expression as given in (B.61) with $\kappa = a$ and $\theta = \frac{1}{\lambda}$ and let $X = e^Y$. The resulting probability density function for X, which is also given by Tadikamalla [113] and Rolski et al [105], is given in (B.113).

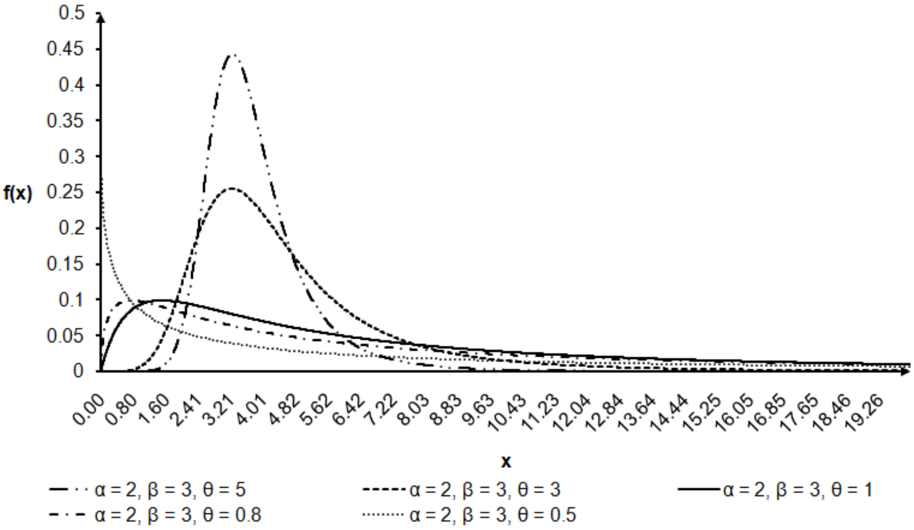

Figure 4.39: Kappa Three-Parameter Density Function with varying values of θ

Note that since the probability density function for Y is only valid for values where $y > 0$, the probability density function for X is only valid for values of x such that $x = e^y > 1$.

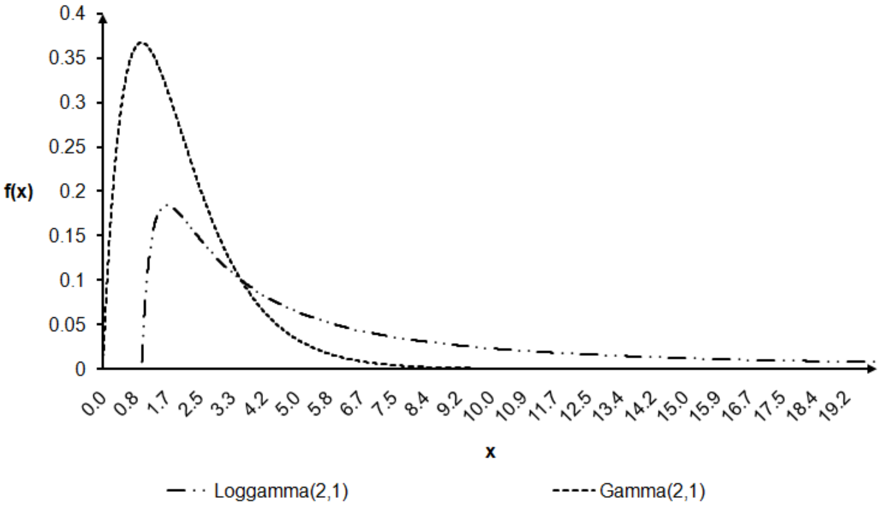

Figure 4.40: Comparison of the Loggamma and Gamma distributions

Figure 4.40 shows a comparison of the Gamma and Loggamma distributions with the same parameter values from which it can be seen that the Loggamma distribution has more weight in the upper tail with lower density for the lower values of x.

Now, to obtain an expression for the r^{th} moment of X, as given in (B.114), the relationship between the Gamma and Loggamma distributions can be used, in which case it follows that: The first four moments follow directly from expression (B.114), from which the expressions for the mean, variance and skewness that are given in (B.115), (B.116) and (B.117) follow.

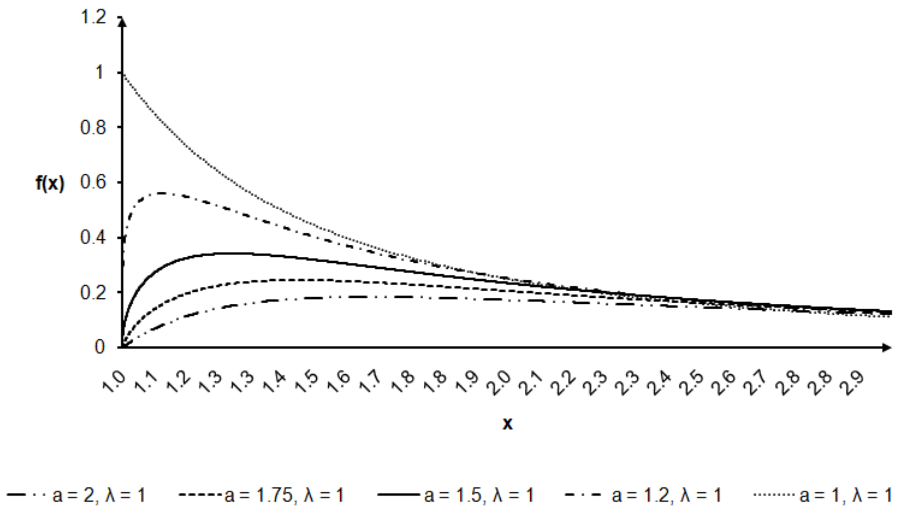

Figure 4.41: Loggamma Density Function with varying values of a

The parameter a affects both the scale and shape of the probability density function. This is illustrated in Figure 4.41. Figure 4.42 shows how the λ parameter affects the scale of the distribution.

4.1.22 Snedecor's F Distribution

Consider twon independent random variables, V_1 and V_2, where $V_1 \sim \chi^2(\nu_1)$ and $V_2 \sim \chi^2(\nu_2)$. Let $X = \frac{V_1/\nu_1}{V_2/\nu_2}$, then $X \sim F(\nu_1, \nu_2)$ [11].

In order to derive the r^{th} moment of X that is given in (B.165), we can use the fact that V_1 and V_2 are independent and thus us the expressions derived for the r^{th} moment of the Chi-square distribution as given in (B.15).This expression can be used to obtain the moments about the origin from which expressions for the mean, variance and skewness follows. These expressions, as provided by Bain and Engelhardt [11], are given in Section B.28 in Appendix B.

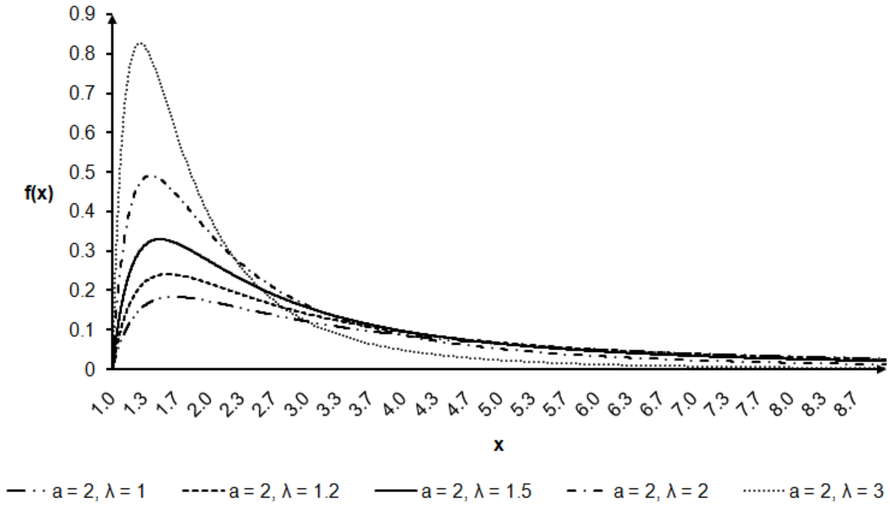

Figure 4.42: Loggamma Density Function with varying values of λ

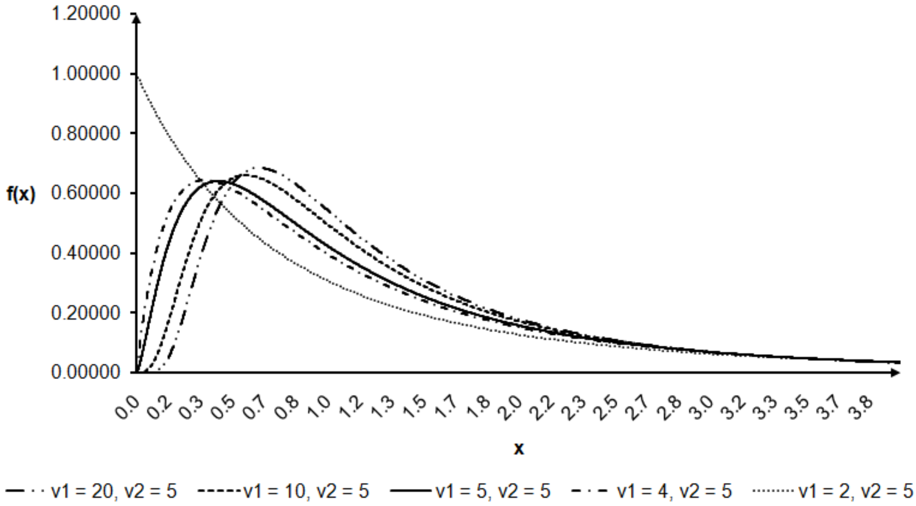

Figure 4.43: F Density Function with varying values of ν_1

Figures 4.43 and 4.44 show the effects of varying values for ν_1 and ν_2. From these figures it is evident that the value of ν_1 affects the shape, location and scale of the distribution. For varying values ν_2 the scale and location appears to be affected.

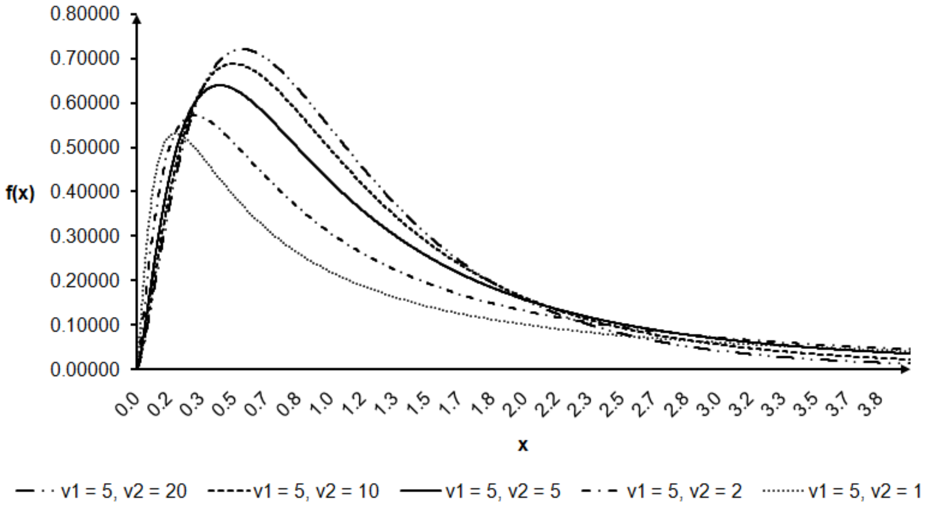

Figure 4.44: F Density Function with varying values of ν_2

4.1.23 Log-logistic Distribution

Consider a random variable Y that has a logistic distribution (as given in Section 4.2.2). Now let $X = e^Y$, $\theta = \frac{1}{\beta}$ and $\xi = \ln(\alpha)$. The cumulative distribution function can be derived using the cumulative distribution function of the logistic distribution in (4.55):

$$\begin{aligned} F_X(x) &= F_Y(\ln(x))|_{\theta=\frac{1}{\beta},\xi=\ln(\alpha)} \\ &= \frac{\left(\frac{x}{\alpha}\right)^\beta}{1+\left(\frac{x}{\alpha}\right)^\beta} \end{aligned} \tag{4.43}$$

This is the expression for the cumulative distribution function for a random variable X that has a Log-logistic distribution which is also referred to as the Fisk distribution in some literature [113], [75].

We can find the probability density function by taking the first order derivative of (4.43) with respect to x to find the expression in (B.122). This form of the proabability density function is also given by Klugman et al [75].

The probability density function of X as given in (B.122) will be used to derive a general expression for the r^{th} moment of X as is given in (B.123):

$$\begin{aligned}
\mathrm{E}(X^r) &= \int_0^\infty x^r \frac{\beta\left(\frac{x}{\alpha}\right)^\beta}{x\left(1+\left(\frac{x}{\alpha}\right)^\beta\right)^2 dx} \\
&= \alpha^r \int_0^1 y^{\left(\frac{r}{\beta}+1\right)-1}(1-y)^{\left(1-\frac{r}{\beta}\right)-1} dy \text{ by letting } y = \frac{x^\beta}{\alpha^\beta + x^\beta} \\
&= \alpha^r \Gamma\left(1+\frac{r}{\beta}\right)\Gamma\left(1-\frac{r}{\beta}\right) \qquad (4.44)
\end{aligned}$$

The first four moments about the origin of X follows directly from (B.123) from which the expression for the variance in (B.125) can be obtained.

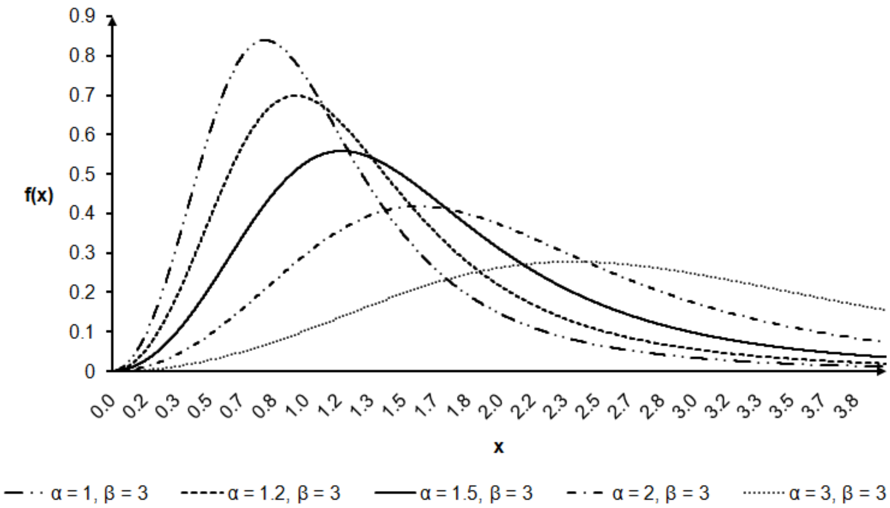

Figure 4.45: Log-logistic Density Function with varying values of α

The scale, location and shape of the Log-logistic distribution is affected by the value of β while varying values of α only affects the location and scale. Figures 4.45 and 4.46 support these statements.

4.1.24 Folded and Half Normal Distribution

Consider a random variable W that is Normally distributed (as described in section 4.2.1) with mean μ and variance σ^2. The random variable $X = |W|$ is Folded Normally distributed with probability density function given by [46], [79] - see (B.43).

Figure 4.47 compares the Normal and Folded Normal distributions with the same values for μ and σ. These graphs shows exactly how the part of the

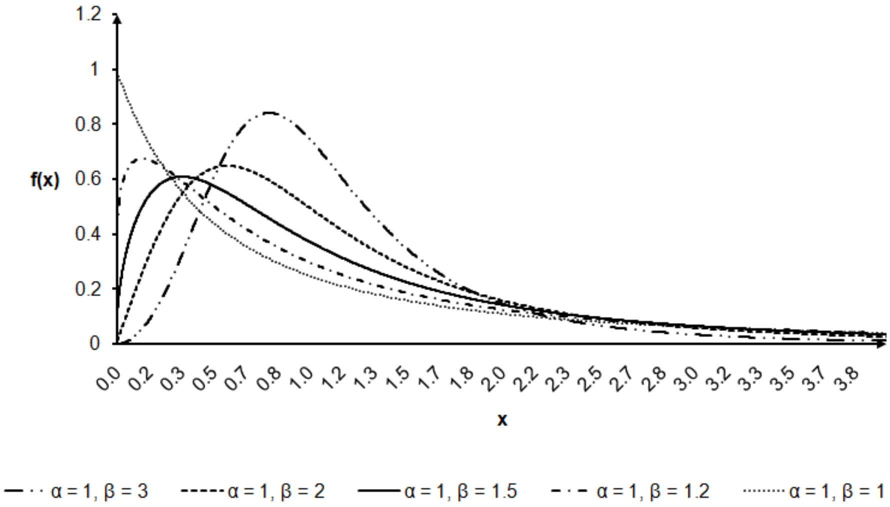

Figure 4.46: Log-logistic Density Function with varying values of α

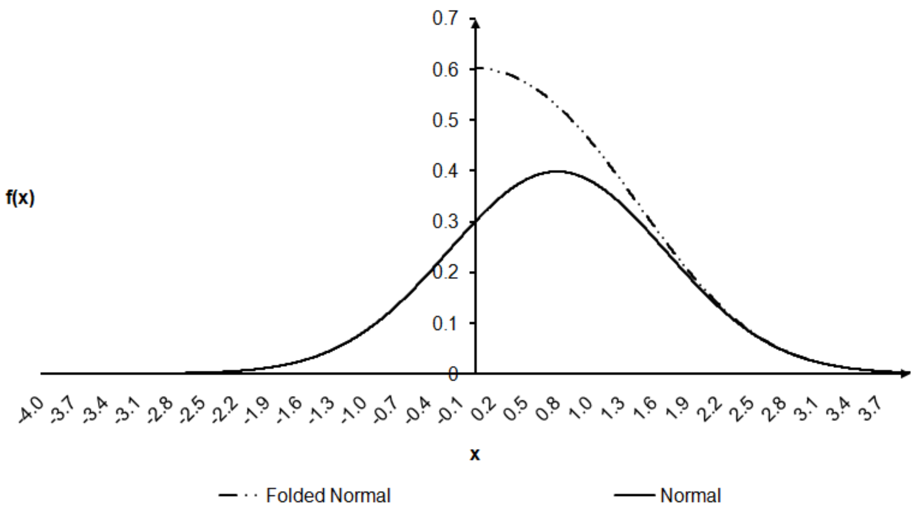

Figure 4.47: Comparison of the Normal and Folded Normal Distributions for $\mu \neq 0$

Normal distribution left of the point where $x = 0$ is folded over to the part where x is nonnegative and is basically adding to the density of the lower nonnegative values of x for the Folded Normal distribution.

A general expression for the r^{th} moment is derived by Elandt [46]:

$$\mathrm{E}(X^r) = \sigma^r \sum_{j=0}^{r} \begin{pmatrix} r \\ j \end{pmatrix} \theta^{r-j} \left(I_j(-\theta) + (-1)^{r-j} I_j(\theta) \right) \tag{4.45}$$

where

$$I_j(a) = \frac{1}{\sqrt{2\pi}} \int_a^{\infty} y^j e^{-\frac{1}{2}y^2} dy \text{ for } j = 1, 2, ...$$

$$I_0(a) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}y^2} dy$$

and

$$\theta = \frac{\mu}{\sigma}$$

For the purpose of deriving the first four moments of this distribution, we have to calculate $I_0\left(\frac{\mu}{\sigma}\right), I_1\left(\frac{\mu}{\sigma}\right), ..., I_4\left(\frac{\mu}{\sigma}\right)$ and $I_0\left(-\frac{\mu}{\sigma}\right), I_1\left(-\frac{\mu}{\sigma}\right), ..., I_4\left(-\frac{\mu}{\sigma}\right)$. One can consider deriving the terms of the form $I_j(\theta) + (-1)^{r-j} I_j(\theta)$ for $r = 1, 2, 3, ...$ and $j \leq r$. The expressions required to calculate the first four moments are given in (B.45) to (B.53). The expressions for the mean and variance, which can be derived using (B.45) to (B.53), are given in (B.54) and (B.55).

One can calculate the the moments about the origin first after which the skewness and kurtosis can be calculated using (4.33) and (4.34).

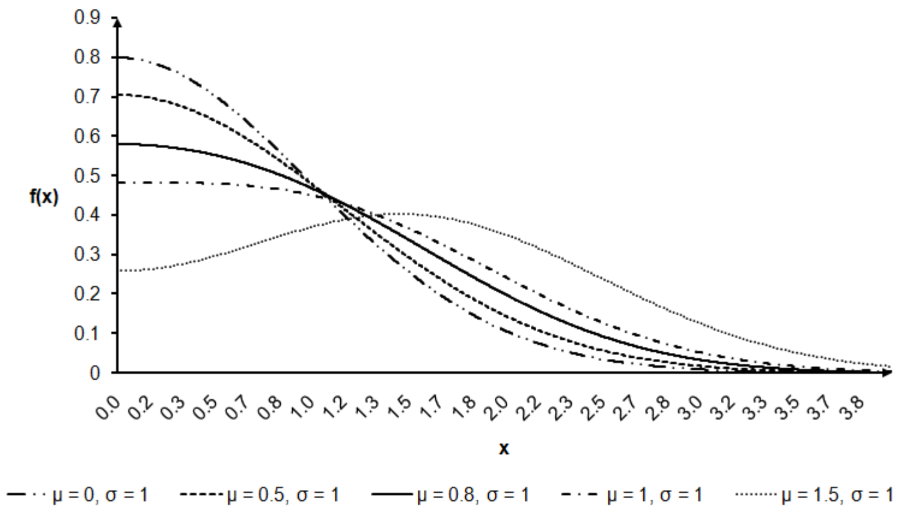

Figure 4.48: Folded Normal Density Function with varying values of μ

Similar to the Normal distribution (as discussed in section 4.2.1) the Folded Normal distribution's scale is affected by the value of σ. It also affects the

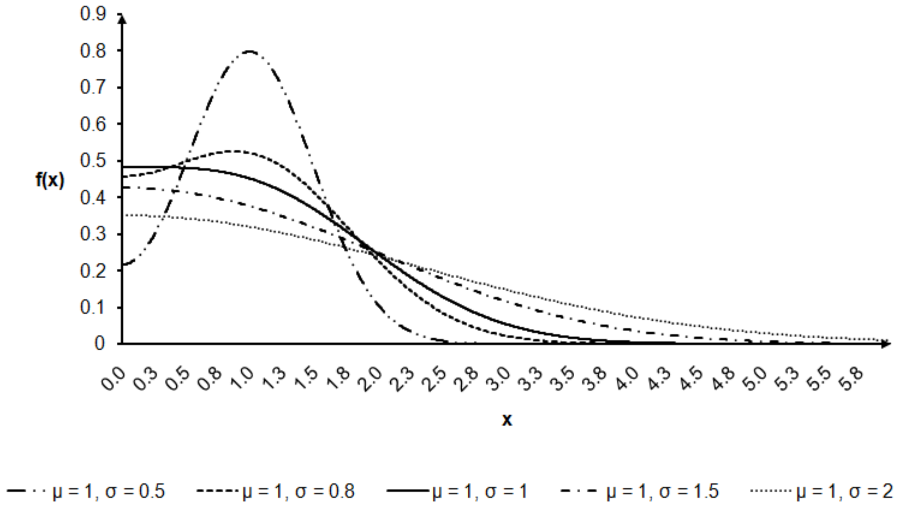

Figure 4.49: Folded Normal Density Function with varying values of σ

shape (especially for values less than μ). This is illustrated in Figure 4.49. In contrast to the μ purely being a location parameter for the Normal distribution, it affects both the location as well the shape of the Folded Normal distribution. The reason for the value of μ affecting the shape is because the larger the value μ is, the smaller the area of the Normal distribution that lies to the left of 0 and hence the smaller the area that is folded over to the nonnegative values of x. Also, the larger the value of μ, the closer the Folded Normal will be to the Normal distribution. This is evident from Figure 4.48.

The Half Normal distribution follows directly from the Folded Normal as a special case where $\mu = 0$ [36]:

$$f_X(x) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \text{ for } x \geq 0$$

This form is also stated by Birnbaum [18] and is used by Daniel [36] to interpret factorial two-level experiments. The case where $\mu = 0$ in Figure 4.48 provides a graphical presentation of the probability density function of the Half Normal distribution.

Maximum likelihood estimation and method-of-moments estimation are used by Leone et al [79] and Elandt [46] to obtain parameter estimates.

4.1.25 Inverse Gamma Distribution

Result 3. *If a random variable Y has a $GAM(\theta, \kappa)$ distribution, then $X = \frac{1}{Y}$ has an $INV\ GAM(\alpha, \kappa)$ distribution with $\alpha = \frac{1}{\theta}$.*
Proof: *From (B.61) we have that if $X \sim GAM(\theta, \kappa)$, then*

$$f_Y(y) = \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa - 1} e^{-\frac{x}{\theta}} \text{ for } x > 0$$

If $Y = \frac{1}{X}$, then

$$\begin{aligned} f_X(x) &= f_Y\left(\frac{1}{x}\right) \left|\frac{d}{dx} h(x)\right| \text{ where } h(x) = \frac{1}{x} \\ &= \frac{\left(\frac{\theta}{x}\right)^\alpha e^{-\left(\frac{\theta}{x}\right)}}{x\Gamma(\alpha)} \text{ by letting } \theta = \frac{1}{\theta} \text{ and } \kappa = \theta, \end{aligned}$$

which is of the form of the expression given by Klugman et al [75] for the probability density function for the Inverse Gamma distribution - also given in B.89.

Using the relationship between $X \sim INVGAM(\alpha, \theta)$ and $Y \sim GAM(\theta, \kappa)$, an expression for the r^{th} moment of X can be derived from the moments of Y. The resulting expression is given in (B.90), from which expressions for the mean, variance and skewness follow. These expressions are given in (B.91) to (B.93).

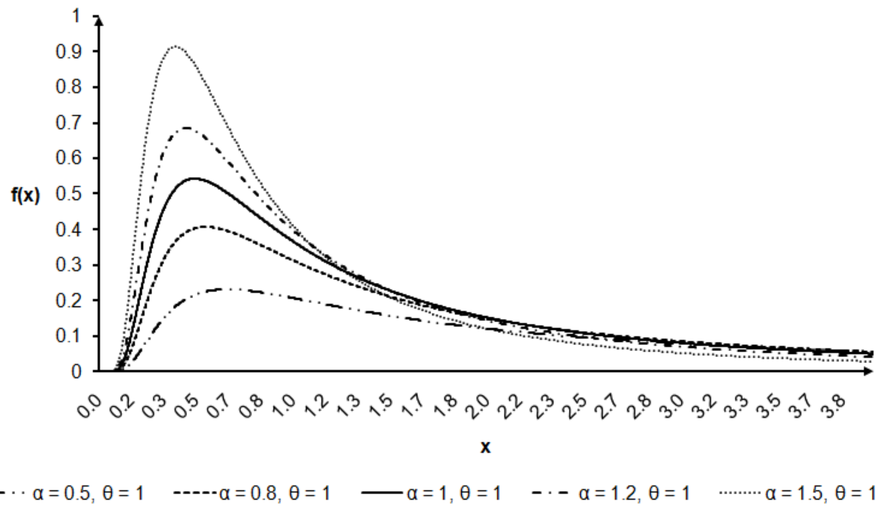

Figure 4.50: Inverse Gamma Density Function with varying values of α

Both the scale and location of the distribution are affected by varying the values of α or varying the values of β. This is illustrated by the graphs in

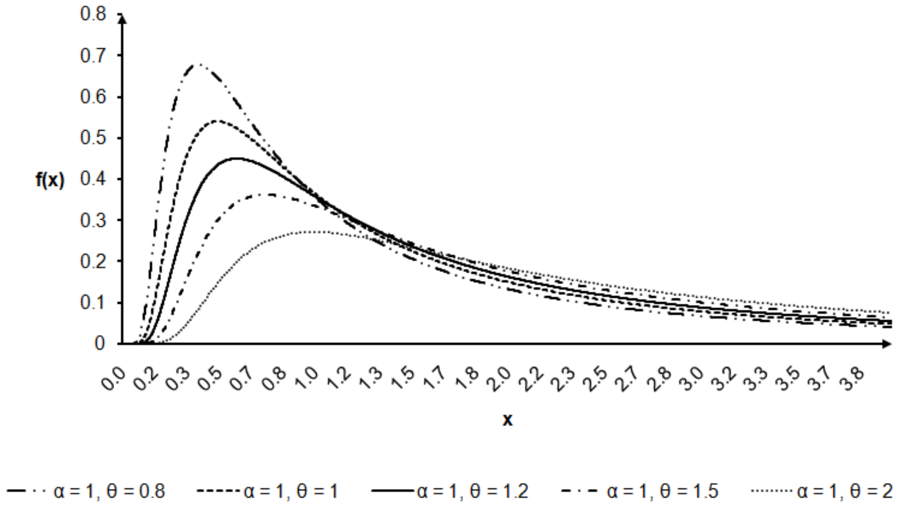

Figure 4.51: Inverse Gamma Density Function with varying values of β

Figures 4.50 and 4.51 and can be deduced from the fact that both these parameters affect the values of the mean and variance (as per (B.91) and (B.92)).

If the gamma and Inverse Gamma distributions are compared for equal parameter values for κ and $\alpha = \theta^{-1}$, the Inverse Gamma distribution will have a heavier tail. This is graphically displayed in Figure

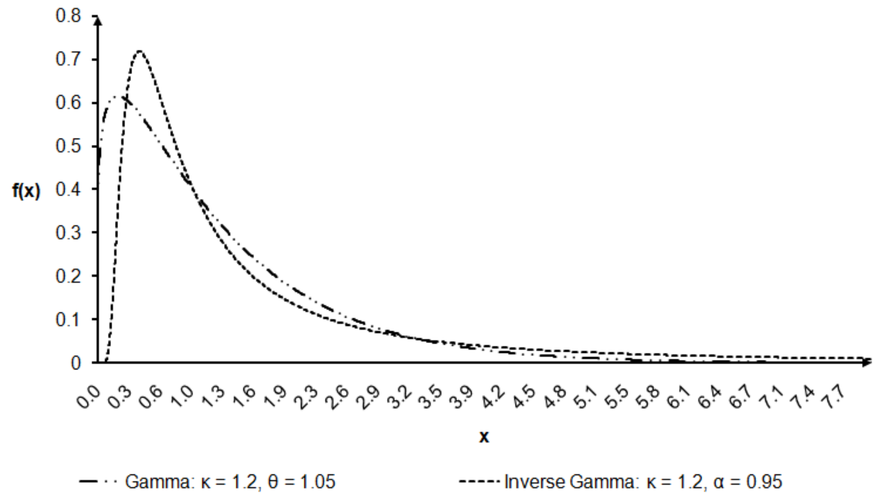

Figure 4.52: Comparison of Gamma and Inverse Gamma Distributions

This can theoretically also be proven by considering the ratio of the two distributions' limiting tail behaviour as using (3.13). Let the random variable for from the Inverse Gamma be denoted by Y and the random variable from the Gamma distribution be denoted by X:

$$\begin{aligned}\lim_{x\to\infty}\frac{f_Y(x)}{f_X(x)} &= \lim_{x\to\infty}\frac{\frac{\left(\frac{\alpha}{x}\right)^{\kappa}e^{-\frac{\alpha}{x}}}{x\Gamma(\kappa)}}{\frac{x^{\kappa-1}e^{\frac{x}{\theta}}}{\theta^{\kappa}\Gamma(\kappa)}} \text{ using } (B.61) \text{ and } (B.89)\\ &= \lim_{x\to\infty}\frac{(\alpha\theta)^{\kappa}e^{-\left(\frac{\alpha}{x}-\frac{x}{\theta}\right)}}{x^{2\kappa}}\\ &= \infty\end{aligned}$$

This indicates that the Inverse Gamma distribution has a heavier tail than the Gamma distribution if the parameter values for κ are the same and $\alpha = \theta^{-1}$.

4.1.26 Inverse Chi-square Distribution

From Bernardo and Smith [17] it follows that if a random variable $Y \sim \chi^2(\nu)$, then $X = \frac{1}{Y} \sim INV\chi^2(\nu)$. The probability density function of X, as given in (B.84) can be found from the probability density function of Y as given in (B.13) using the following relationship:

$$f_X(x) = f_Y(h(x))\left|\frac{d}{dx}h(x)\right| \text{ where } h(x) = \frac{1}{x}$$

It is shown in Section 4.1.3 that if $Z \sim GAM(\theta,\kappa)$, then $X = \frac{2Z}{\theta} \sim \chi^2(2\kappa)$. Similarly, it can be shown that if $W \sim INVGAM\left(\frac{\nu}{2},\theta\right)$, then $X = \frac{W}{2\theta} \sim INV\chi^2(\nu)$ using (B.89) with $\alpha = \nu$ and the following relationship:

$$f_X(x) = f_W(h(x))\left|\frac{d}{dx}h(x)\right| \text{ where } h(x) = 2\theta x$$

Now, using the relationship between a random variable W that is Inverse Gamma distributed with parameters $\frac{\nu}{2}$ and θ and a random variable X that is Inverse Chi-square distributed with parameter ν, the general expression (B.85) for the r^{th} moment of X can be found from (B.90) (which is the general expression for the r^{th} moment of W). The first four moments about the origin of X follow directly from (B.85) from which the variance and skewness can be calculated. Refer to (B.86) to (B.88) for expressions for the mean, variance and skewness.

Since the Inverse Chi-square distribution is a special case of the Inverse Gamma distribution with parameters $\alpha = \frac{\nu}{2}$ and $\theta = 2$, the probability density function will be affected by varying values of ν in a similar manner as the density function for the Inverse Gamma distribution is affected by varying values of α which is illustrated in Figure 4.50.

4.1.27 Inverse Gaussian Distribution

A detailed review of the Inverse Gaussian Distribution is given by Folks and Chhikara [55] with its properties and applications.

The distribution was originally presented by Schrödinger in 1915 as a probability density function of the first passage time in a Brownian motion (also referred to as a Weiner Process) [105]. A formal definition of the Weiner process is given by given by Luenberger [82].

Schrödinger's study considers a Weiner process $\{Z_t\}$ with a drift ν and variance σ^2. It is then argued that the time T (which is a random variable) until the value α is reached by Z_t for the first time (referred to as the first passage time [105]) is a random variable that is Inverse Gaussian distributed with probability density function given by:

$$f_T(t) = \frac{\alpha}{\sigma\sqrt{2\pi t^3}} e^{-\frac{(\alpha-\nu t)^2}{2\sigma^2 t}} \text{ for } t > 0, \nu > 0.$$

A reparameterization of this function form was presented by Tweedie with parameters μ and λ [55]. This reparameterization can be obtained by letting $\alpha = \nu\mu$ and $\frac{\alpha}{\sigma} = \sqrt{\lambda}$ and is given by Rolski et al [105]. This form is given in (B.96).

It also follows from the review of Folks and Chhikara that if $\mu = 1$ the probability density function given in (B.96) reduces to the density function of the Wald distribution.

The expression for the moment generating function (B.97) as presented by Rolski et al[105] can be used to find the moments about the origin, from which expressions for the variance and skewness follows.. Expressions for the mean, variance and a general expression for the r^{th} moment of T are also given by Folks and Chhikara [55] - see expressions (B.98) to (B.101).

Figures 4.53 and 4.54 show how the scale and location of the distribution is affected by varying the values of μ and λ, respectively.

The cumulative distribution function for T is given by Folks and Chhikara

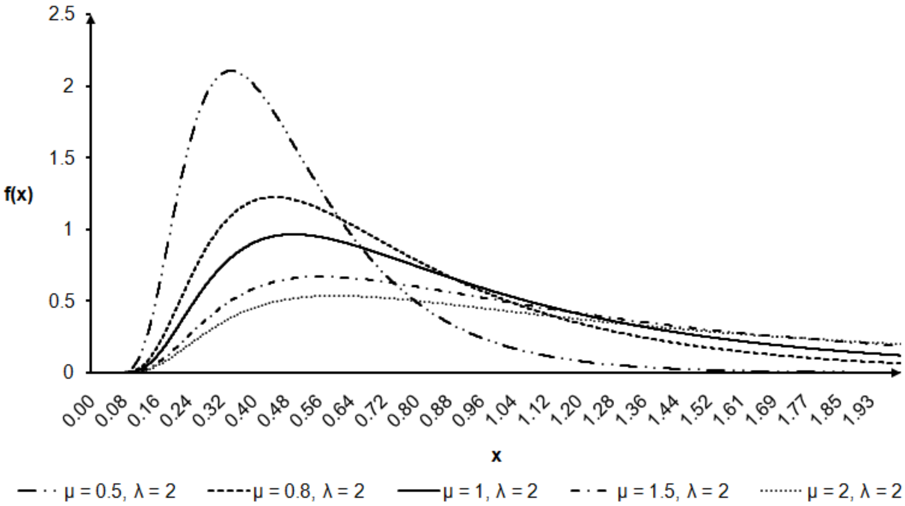

Figure 4.53: Inverse Gaussian Density Function with varying values of μ

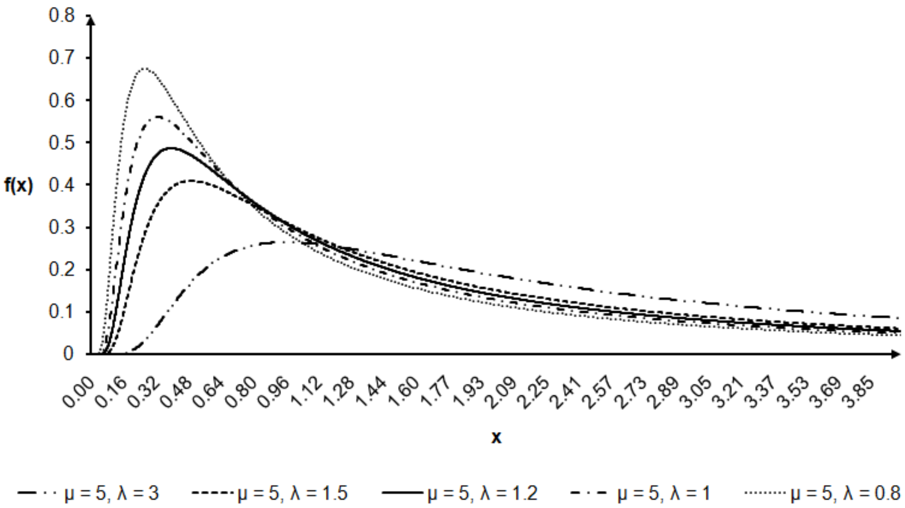

Figure 4.54: Inverse Gaussian Density Function with varying values of λ

[55] as follows:

$$F_T(t) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right) + e^{2\frac{\lambda}{\mu}}\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right) \tag{4.46}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal distribution. This form is also derived by Shuster [109].

4.1.28 Skew-normal Distribution

Azzalini introduced a general family of distributions [8] as follows: For a probability density function, f, that is symmetric about 0 and a cumulative density function, G, that is an absolute continuous function for which G' is symmetric about 0, the function

$$h(y) = 2G(\lambda y)f(y) \text{ for } -\infty < y < \infty$$

is a valid probability density function for any real-valued λ.

If f and G are probability density and cumulative distributions, respectively, of a standard Normal random variable, then

$$\begin{aligned} h(y) &= 2G(\lambda y)f(y) \\ &= 2\Phi(\lambda y)\phi(y) \text{ for } -\infty < y < \infty \end{aligned} \tag{4.47}$$

is the probability density function of the Skew-normal distribution, denoted as $Y \sim SN(\lambda)$.

Some of the most important properties of the Skew-normal distribution given by Azzalini [8] are as follows:

- If $\lambda = 0$, then $Y \sim N(0, 1)$
- $h(y)$ tends to the half-normal distribution as $\lambda \to \infty$
- If $Y \sim SN(\lambda)$, then $-Y \sim SN(-\lambda)$
- $h(y)$ is unimodal
- If $Y \sim SN(\lambda)$, then $Y^2 \sim \chi^2(1)$ (i.e. similar to the case of a standard Normal variable Z for which it is true that $Z^2 \sim \chi^2(1)$).
- $\ln(h(y))$ is a concave function of y.

Denote the probability density and cumulative distribution functions of Y as follows:

$$\begin{aligned} \phi(y; \lambda) &= h(y) = 2\Phi(\lambda y)\phi(y) \text{ for } -\infty < y < \infty \\ \Phi(y; \lambda) &= 2\int_{-\infty}^{y}\int_{-\infty}^{\lambda t} \phi(t)\phi(u)dudt \end{aligned} \tag{4.48}$$

The probability density function, as given in (4.47), is illustrated for various values of λ in Figure 4.55.

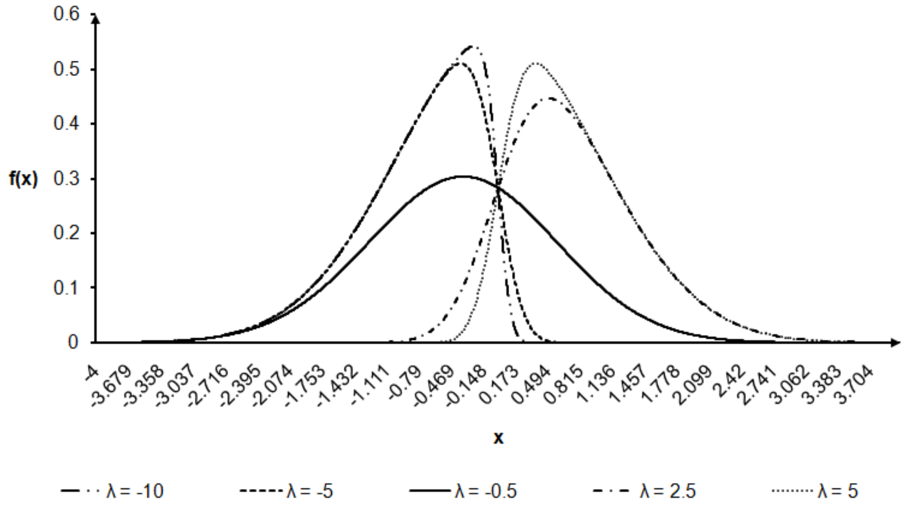

Figure 4.55: Skew-normal Density Function with varying values of λ

The moment generating function is given in (4.49). This is the form as proposed by Azzalini [8]. Expressions for the mean, variance, skewness and kurtosis are also given by Azzalini - see (4.50) to (4.53).

$$M_Y(t) = 2e^{\frac{t^2}{2}}\Phi(\delta t) \text{ where } \delta = \frac{\lambda}{\sqrt{(1+\lambda^2)}} \tag{4.49}$$

and

$$\mathrm{E}(Y) = b\delta \tag{4.50}$$
$$\mathrm{var}(Y) = 1 - b^2\delta^2 \tag{4.51}$$

where

$$b = \sqrt{\frac{2}{\pi}} \text{ and } \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$$

while the skewness and kurtosis are given by:

$$\mathrm{skewness}(Y) = \frac{1}{2}(4-\pi)\mathrm{sign}(\lambda)\left(\frac{(\mathrm{E}(Y))^2}{\mathrm{var}(X)}\right)^{\frac{3}{2}} \tag{4.52}$$

and

$$\text{kurtosis}(Y) = 2(\pi - 3)\left(\frac{(\mathrm{E}(Y))^2}{\text{var}(X)}\right)^2 \tag{4.53}$$

Alternative parameterizations can also be considered where one can introduce location and scale parameters [8], [100]. Let ξ and η denote the location and scale parameters, respectively. If $X \sim SN(\lambda)$ then we can have

$$Y = \xi + \eta X$$

in which case $Y \sim SN_D(\xi, \eta, \lambda)$, with $\xi, \eta \in \mathbb{R}$,which is referred to as the direct parameterization.

The expressions for the mean and variance of Y under the direct parameterization follows directly from (4.50) and (4.51):

$$\mathrm{E}(Y) = \xi + \eta b\delta$$

$$\text{var}(Y) = \eta^2(1 - b^2\delta^2)$$

while the skewness coefficient is given by:

$$\gamma_1 = \frac{b\delta^3(2b^2 - 1)}{(1 - b^2\delta^2)^{\frac{3}{2}}}$$

where

$$b = \sqrt{\frac{2}{\pi}} \text{ and}$$

$$\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$$

The skewness coefficient can take on values in the interval $(-0.99527, 0.99527)$ [100].

It is further argued that the information matrix associated with the maximum likelihood estimation process for the ξ, η and λ is singular. In this light an alternative parameterization based on a centred random variable X is introduced:

$$Y = \mu + \sigma\frac{X - \mathrm{E}(X)}{\sqrt{\text{var}(X)}} \text{ with } -\infty < \mu < \infty \text{ and } \sigma > 0$$

where $X \sim SN(\lambda)$ with skewness coefficient given by γ_1. This skewness coefficient is preserved in this location-scale, centred parameterization. The random variable Y is then said to be $SN_C(\mu, \sigma, \gamma_1)$, which is referred to as the centered parameterization [100]. Expressions for the mean, variance and skewness of this form are given in (B.158) to (B.160).

Recently published work by Azzalini and Capitanio [9] provides a detailed formulation of the Skew-normal distribution in a univariate and multivariate context and details on extensions and generalizations of this distribution. In this context they provide the properties and likelihood estimation details for these distributions.

4.1.29 Exponential-Gamma Distribution

Consider a random variable X with an Exponential distribution that is conditional on its parameter θ, then we have that $X|\theta$ has a probability density function as given in (B.34). Now suppose $\beta = \frac{1}{\theta}$ is $GAM(\alpha, k)$ distributed for $\beta \geq 0$ with $\alpha, k > 0$, then β has a probability density function of the form as given in (B.61). Hence, the unconditional distribution of X can be found as follows:

$$\begin{aligned}
f_X(x) &= \int_0^\infty f(x|\beta) f_B(\beta) d\beta \\
&= \int_0^\infty \beta e^{-\beta x} \frac{1}{\alpha^k \Gamma(k)} e^{-\beta/\alpha} \beta^{k-1} d\beta \\
&= \left(x + \alpha^{-1}\right)^{-(k+1)} \frac{\Gamma(k+1)}{\alpha) k \Gamma(k)} \text{ by letting } y = \beta\left(x + \frac{1}{\alpha}\right) \\
&= k\alpha \left(\frac{1}{x\alpha + 1}\right)^{k+1} \qquad (4.54)
\end{aligned}$$

This marginal probability density function of X is referred to as the Exponential-Gamma distribution [113].

If one now let $\alpha^{-1} = a$, then we can rewrite the probability density function in (4.54) in the exact same form as given in (B.139) for the Pareto Type II (Lomax) distribution. It was also shown in Section 4.1.19 that the Lomax distribution is a special case of the Singh-Maddala distribution.

If one let $Y = \dfrac{X}{a}$ then

$$\begin{aligned}
f_Y(y) &= f_X(h(y)) \left|\frac{d}{dy} h(y)\right| \text{ where } h(y) = ay \\
&= k(1+y)^{-(k+1)}.
\end{aligned}$$

This probability density function has the exact same form as (B.9), but with $c = 1$, which is the probability function for a random variable that is Burr

Type XII distributed. One can then conclude that Y is Burr Type XII distributed with parameters c and k where $c = 1$.

Expression (B.10) derived for the r^{th} moment of the Burr Type XII distribution can then be used to derive a similar expression for X which has an Exponential-Gamma distribution. Alternatively, we can use the fact that X is also Lomax distributed with parameters a and k where $a = \frac{1}{\alpha}$ and use (B.141), (B.142) and (B.143) for the mean, variance and skewness as derived in Section 4.1.11.

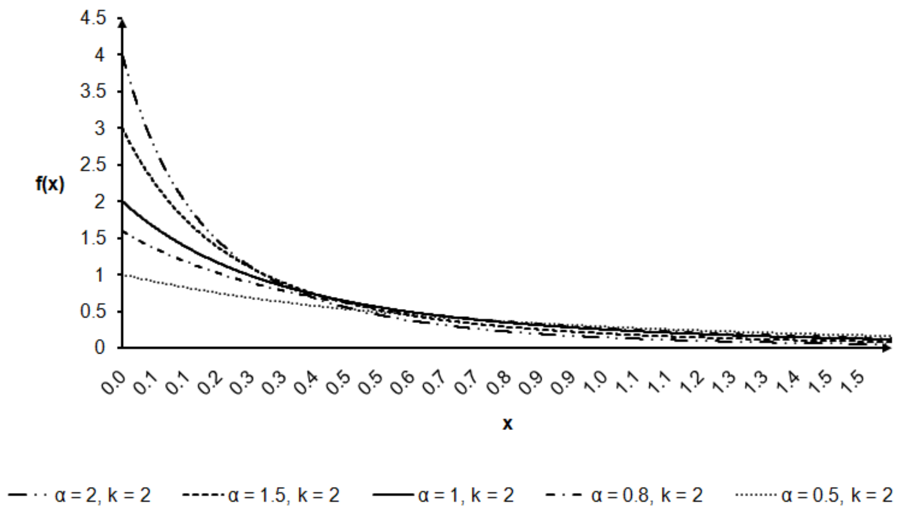

Figure 4.56: Exponential-Gamma Density Function with varying values of α

Figures 4.56 and 4.57 shows how the shape of the distribution changes with varying values of α and k, respectively. As exepeted these compare well with trends seen from varying the parameters for the Pareto Type II (Lomax) distribution as shown in Figures 4.16 and 4.17. It should be noted that for increasing values of θ the weight of the upper tail of the Pareto Type II distribution increases while the density on the lower values decreases. Similarly for decreasing values of α the weight of the upper tail of the Exponential-Gamma distribution increases with an associated decrease on the density for the lower values. This makes sense from the fact that we know that the Exponential-Gamma distribution is a special case of the Lomax distribution with parameters $\kappa = k$ and $\theta = \alpha^{-1}$.

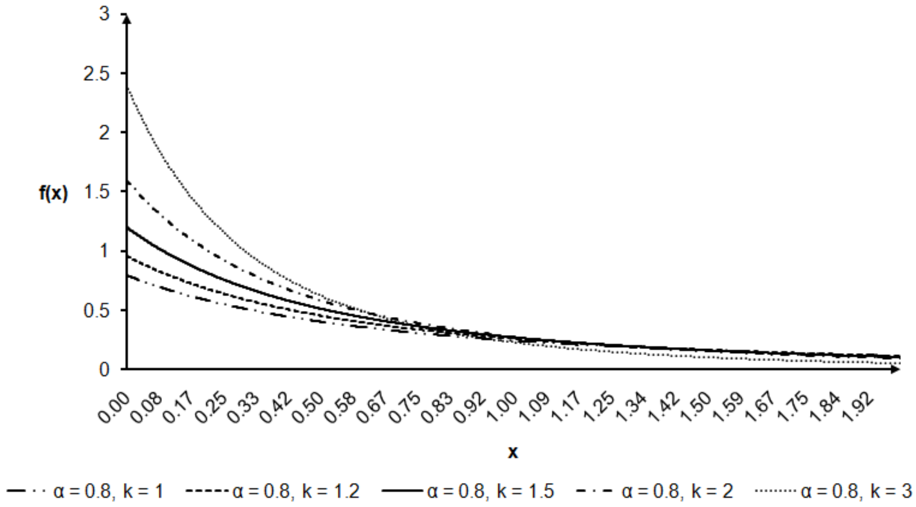

Figure 4.57: Exponential-Gamma Density Function with varying values of k

4.2 Other Useful Distributions

In this section overviews of the Normal and Logistic distributions are given. These distributions are symmetric and are defined for all values in $\mathbb{R}$. As such that do not possess the properties of being skewed and defined on $\mathbb{R}^+$ as one would ideally require in order to model claims size data. There are distributions suitable for modeling claims size data which are related to these distributions, though. An understanding of these two distributions enable one to obtain a better understanding of their related distributions.

4.2.1 Normal Distribution

The Normal distribution is symmetric around its mean value (hence has skewness of 0) and is defined for all values in $\mathbb{R}$. Because of its symmetry it is not really suitable for modelling claim sizes, especially since the underlying distributions of claims are skewed and often heavy-tailed. The probability density function, moment generating function and expressions for the mean and variance, as presented by Bain and Engelgardt [11], are given in (B.133) to (B.136).

The Normal distribution does have useful properties, important relationships with other non-central distributions, is well-studied, easy to simulate from whilst certain transformations of the Normal distribution are useful for modeling skew claims data. Knowing and understanding the relationship between the skew distributions and the Normal distribution can enable one to utilise the useful properties of the Normal distribution. Good examples of such cases include the Skew-normal, Birnbaum-Saunders and Folded-Normal distributions. All three of these distributions can be simulated using the Normal distributions, see for example the SAS codes given for these simulations in appendices A.26, A.14 and A.20.

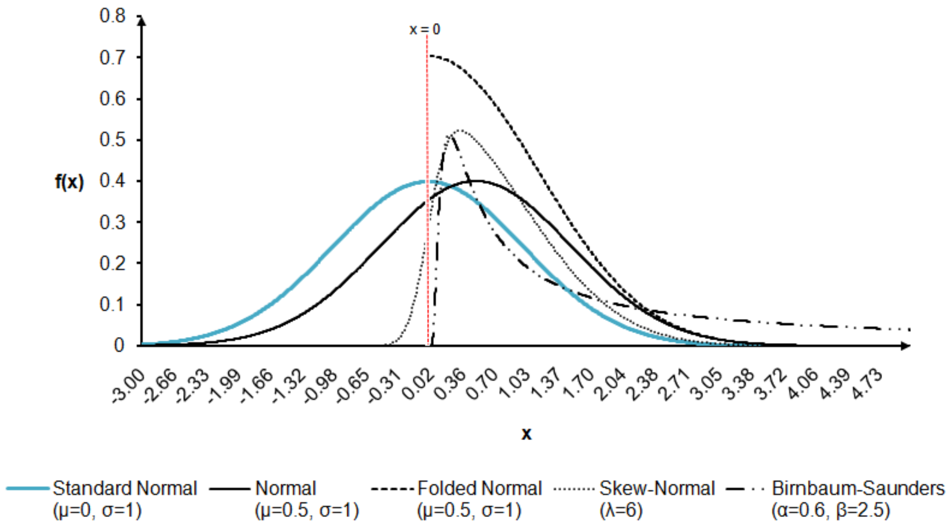

Figure 4.58: Comparison of Normal and related distributions

Consider Figure 4.58 which illustrates the following in terms of distributions related to the Normal distribution:

- Firstly consider the standard Normal distribution which is symmetric around 0. This distribution is well studied and has useful properties. For large samples approximations to statistics, such as the sample mean [31], can be obtained using the central limit theorem [105], [75]. Under the central limit theorem the limiting behaviour of the stanardized statistic [1] conforms to a standard Normal distribution. This is a very important use of the standard Normal distribution for various problems irrespective what the domain of the problem is.

- A non-central version of the standard Normal distribution may be used as is given in Figure 4.58 with the mean of 0.5. One can also adjust

[1]For a statistic T_n, where n denotes the sample size, the $\lim_{n\to\infty} \frac{T_n - \mathrm{E}(T_n)}{\sqrt{\mathrm{var}(T_n)}}$ has a Normal distribution with mean 0 and variance 1 [75], [31].

the the variance. In this example it is clear, though, that a large portion of the distribution still has a density on the negative values. One can work around this, by choosing values for μ and σ such that the density on the negative domain becomes very small. Alternatively one can resort to one of the related distributions.

- A possible choice is to consider the folded-normal, as discussed in Section 4.1.24 where the density associated with the negative values is folded over to their positive counterparts which significantly increase the density on the lower positive values. Depending on how far the value μ is from 0 this folding over may have a small impact on the upper tail, but inflate the density on the lower positive values too much.

- An alternative related distribution may therefore be to consider a skewed Normal distribution, as discussed in Section 4.1.28. From Figure 4.58 how this distribution shows a higher density on the lower positive values while it is not as high as for the Folded Normal distribution. It shows some density on the negative values, but this density may be reduced by varying the value for the paramter λ. It is also evident that the upper right tail has a lower weight than the Normal and Folded Normal distributions.

- In many practical problems related to financial and operational risk as well as insurance tail heaviness play a role and should as such be accounted for in the choice of distribution [82], [75]. As an alternative to the Skew-normal in our example in Figure 4.58 it can be seen that the choice of parameters for the Birnbaum-Saunders distribution gives a distribution that is somewhat similar in shape to the Skew-normal, but with more weight on the upper tail.

- In conclusion it can be seen from Figure 4.58 how the distributions related to the Normal distributions can be used to better capture the behaviour of particular problems in practice. The importance of the Normal distribution and its use are also highlighted, especially in cases where large sample properties such as the central limit theorem can be used. Since the Normal distribution forms the base for these related distributions, it is useful in:

  Understanding the properties of these distributions,

  Obtain a sense of the heaviness of the upper tails of these distributions, and

  Simulating from these distributions using the standard Normal distribution.

4.2.2 Logistic Distribution

A random variable Y that is said to have a logistic distribution with location parameter ξ and scale parameter θ, has a probability density function given by (B.118), [11]. Expressions for the mean and variance, as gvien in (B.119) and (B.120), are given by Gupta et al [62].

If we let $Z = \frac{Y-\xi}{\theta}$, then we have that

$$\begin{aligned} f_Z(z) &= f_Y(h(z))\left|\frac{d}{dz}h(z)\right| \text{ where } h(z) = \theta z + \xi \\ &= \frac{e^z}{(1+e^z)^2} \text{ using expression B.118} \\ &\text{and similarly} \\ f_Z(-z) &= \frac{e^z}{(1+e^z)^2}. \end{aligned}$$

Hence we can conclude that $f_Z(z) = f_Z(-z)$ that shows that the logistic distribution is symmetric around the value of ξ and therefore its skewness is 0.

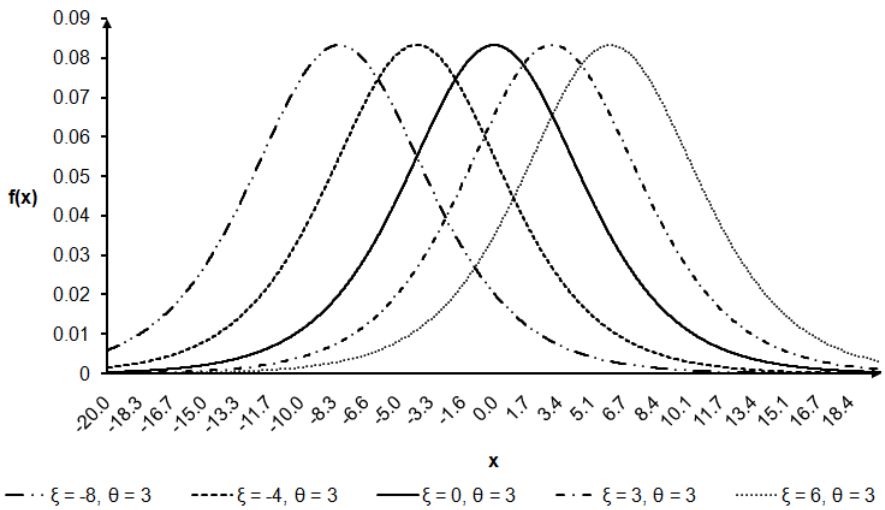

Figure 4.59: Logistic Distribution with varying values of ξ

From Figures 4.59 and 4.60 it follows that ξ affects the location and θ the scale of the distribution.

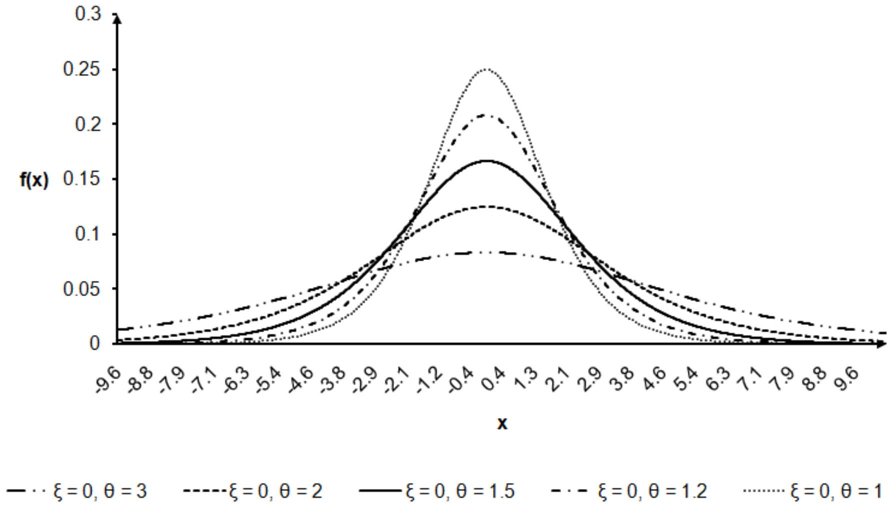

Figure 4.60: Logistic Distribution with varying values of θ

The cumulative distribution function is given by [11]:

$$F_Y(y) = \frac{e^{\frac{y-\xi}{\theta}}}{1+e^{\frac{y-\xi}{\theta}}} \text{ for } y \in \mathbb{R} \tag{4.55}$$

The Log-logistic distribution, as discussed in Section 4.1.23, is related to this distribution in that if Y is logistic distributed then e^Y is Log-logistic distributed. Similar to the Normal distribution, the logistic distribution is defined for all values on $\mathbb{R}$ whilst the focus for the purpose of modeling claims sizes is on distributions with domain on $\mathbb{R}^+$. Having an understanding of this distribution does, however, provide insight to the properties of the associated Log-logistic distribution.

4.2.3 Dirichlet Distribution

Consider k independent $GAM(1,\kappa_i)$ distributed random variables $Z_1, Z_2, ..., Z_k$ [11], [76] such that the joint probability density function is given by

$$f_{Z_1,Z_2,...,Z_k}(Z_1, Z_2, ..., Z_k) = \prod_{i=1}^{k} \frac{1}{\Gamma(\kappa_i)} z_i^{\kappa_i - 1} e^{-z_i} \text{ for } 0 < z_i < \infty$$

$$\text{where } \kappa_j > 0 \text{ for some } j = 1, 2, ..., k$$

Let $Y_i = Z_i / \sum_{j=1}^{k} Z_j$ for $j = 1, 2, ..., k-1$ and $Y_k = \sum_{j=1}^{k} Z_j$. Bain and Engelhardt [11] derive the joint probability density function for the Y_i random variables for a special case where $k = 3$. The value of Y_i can be interpreted as the size of the i^{th} random variable relative to the sum of the sizes of all random variables. The joint probability density function of $Y_1, Y_2, ..., Y_{k-1}$ is given by:

$$f_{Y_1,Y_2,...,Y_{k-1}}(y_1, y_2, ..., y_{k-1}) = \frac{\Gamma(\kappa_1 + \kappa_2 + ... + \kappa_k)}{\Gamma(\kappa_1)\Gamma(\kappa_2)...\Gamma(\kappa_k)} \left(1 - \sum_{j=1}^{k-1} y_j\right)^{\kappa_k - 1} \prod_{j=1}^{k-1} y_j^{\kappa_j - 1} \tag{4.56}$$

This joint density function of $Y_1, Y_2, ..., Y_{k-1}$ given in (4.56) is referred to as the Dirichlet distribution [11], [52].

In the context of non-life insurance this distribution may be used to model aggregate losses where the individual losses are independent and gamma distributed. The loss from each claim is assumed to have parameters $\theta = 1$ and κ_j, where κ_j is the parameters associated with the j^{th} claim.

Chapter 5

Tail Behaviour of Parametric Distributions

5.1 Introduction

In Chapter 4 a comprehensive collection of distributions were studied in terms of the distribution functions, their shape, scale and location together with derivations of the moments. In Chapter 3 seful properties of heavy tailed distributions were discussed in detail. The aim of this chapter is to consider the distributions introduced in Chapter 4 and to apply the theory developed in Chapter 3 to ascertain which of these distributions have heavy tails and how these tails compare amongst one another.

In chapter 3 the following concepts that describe the properties of distributions in terms of tail weight were introduced:

- Hazard function - given in (3.5) Definition 11.
- Hazard rate function - given in (3.6) in Definition 12.
- Residual hazard (rate) function - given in (3.7) in Definition 13.
- Mean residual hazard function - given in (3.9) in Definition 14.
- Stochastic order - as per Definition 15.
- Exponentially bounded tails - inequality given in (3.12) as per Definition 19.

From Chapter 3 one can identify criteria that can be utilized in order to classify distributions as heavy-tailed or not. These criteria are summarised in the techniques below:

Technique 1: As discussed in Section 3.4.1, if only some or if none of the moments of a particular distribution exist, then it is indicative of a distribution that has a heavy tail.

Technique 2: From Definition 18 it follows that a distribution of a random variable X is said to have a heavy tail if $M_X(s) = \infty \ \forall \ s > 0$.

Technique 3: If $\alpha_F = \limsup_{x\to\infty} \frac{h_X(x)}{x} = 0$, then it follows from Theorem 2 that the distribution has a heavy tail.

Technique 4: In Section 3.4.5 it is argued that if $h_X^*(t)$ is a decreasing function for increasing values of t, then the distribution has a heavy tail. It is also given that this condition implies the condition as given as Technique 3.

Technique 5: A distribution is said to be contained in the subexponential class of distributions if

$$\lim_{x\to\infty} \frac{\overline{F}_X^{*2}(x)}{\overline{F}_X(x)} = 2 \text{ from Definition 24}$$

From Theorem 3 it follows that if $F_X(\cdot) \in S$ then the distribution is heavy-tailed.

Technique 6: If a cumulative distribution function $F_X(\cdot)$ for a random variable X has a heavy tail, then it follows from (3.25) in Section 3.5.4

$$\frac{\text{var}(X)}{(\text{E}(X))^2} \geq 1$$

It is important to note that the converse doesn't necessarily hold.

Technique 7: A distribution is said to have an exponentially bounded tail if there exists an $a > 0$ and a $b > 0$ such that

$$1 - F_X(x) \leq ae^{-bx} \ \forall \ x \geq 0.$$

The distribution $F(\cdot)$ is said to have a light tail. This follows from Definition 19.

5.2 Classification of Tail Heaviness for Parametric Continuous Distributions

In the following sections each of the distributions as introduced in Chapter 4 will be assessed, based on the criteria as set out in Techniques 1 to 7, to determine whether the distribution has a heavy tail or not.

5.2.1 Gamma and related distributions

In order to derive the hazard function and hazard rate function, the cumulative distribution function is required. An explicit expression for the cumulative distribution function for the Gamma distribution is not available. It is therefore also not possible assess whether the Gamma distribution is a member of the subexponential class as defined in Defintion 24. For this reason, techniques 1, 2 and 6 will be considered here.

Assessment using Techniques 1 and 2

Consider the moment generating function for a random variable X that is gamma distributed with parameters θ and κ as given in (B.62). The r^{th} moment of X can be derived by evaluating the r^{th} order derivative of the moment generating function where $t = 0$:

$$M_X^{(r)}(t) = \theta^r (1 - \theta t)^{\kappa - r} \prod_{j=0}^{r-1} (\kappa - j) \text{ for } r = 1, 2, 3, ... \tag{5.1}$$

From (5.1) it appears as if all the moments of X exist (although they will tend to infinity). Based on these criteria it implies that the Gamma distribution doesn't have a heavy tail.

Assessment using Technique 6

An alternative way to show that the distribution doesn't have a heavy tail, is by supposing that it does in fact have a heavy tail. Then

$$\begin{aligned}
\frac{\text{var}(X)}{(\text{E}(X))^2} &= \frac{\kappa\theta^2}{\kappa^2\theta^2} \text{ from } (B.64) \text{ and } (B.65) \\
&= \frac{1}{\kappa} \text{ where } \kappa > 0
\end{aligned}$$

If $\kappa > 1$, then $\frac{\text{var}(X)}{(\text{E}(X))^2} < 1$ which contradicts the argument that a distribution that has a heavy tail will have a value for the ratio $\frac{\text{var}(X)}{(\text{E}(X))^2}$ of at least 1. If $\kappa = 1$, this ratio is equal to 1 which is not a contradiction to an assertion that a distribution has a heavy tail. This technique does not provide any evidence against the assertion of a heavy tail.

If $\kappa < 1$ the ratio will be larger than 1. This does not provide a contradiction to an assertion of a heavy tail, but does not serve as a proof of the Gamma distribution having a heavy tail for the case where $\kappa < 1$.

Conclusion

It can therefore be concluded, based on assessment using techniques 1 and 2 that the Gamma distribution doesn't have a heavy tail. Because the Exponential and Chi-square distributions are special cases of the Gamma distribution, it follows that neither of these two distributions has a heavy tail.

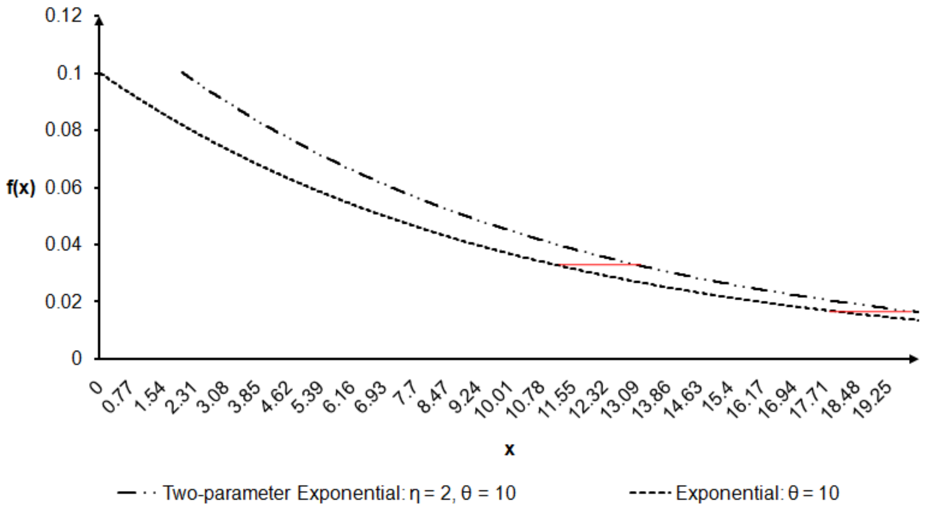

Figure 5.1: Comparison of the Exponential and Two-parameter Exponential Densities

From Figure 5.1 it can be seen that the Two-parameter Exponential density function is exactly the same as the Exponential density function just shifted η units to the right. This means that the tail weight of the Two-parameter Exponential distribution will be the same as for the Exponential distribution, implying that it does not have a heavy tail. Alternatively one can consider the ratio of the variance to square of the expected value from (B.173) and (B.174):

$$\frac{\text{var}(X)}{(\text{E}(X))^2} = \frac{\theta^2}{(\theta + \eta)^2} < 1,$$

since $\theta, \eta > 0$ which implies from Technique 6 that the Two-parameter Exponential distribution does not have a heavy tail.

5.2.2 Erlang Distribution

In order to derive the hazard function and hazard rate function, the cumulative distribution function is required. An explicit expression for the cumulative distribution function for the Erlang distribution is not available. It is therefore also not possible assess whether the Erlang distribution is a member of the subexponential class as defined in Defintion 24. For this reason, techniques 1, 2 and 6 will be considered here.

Assessment using Techniques 1 and 2

For a random variable, X, that is Erlang distributed the moment generating function is given in (B.26) from which it follows that

$$M_X^{(r)}(t) = \lambda^n \left(\prod_{j=0}^{r-1} (n+j) \right) (\lambda - t)^{-(n+r)},$$

which exists for all values of $r \in \mathbb{N}$. It can therefore be concluded that all moments of the Erlang distribution exists which suggests that the distribution doesn't have a heavy tail.

Assessment using Technique 6

Alternatively the argument that the distribution does not have a heavy tail can be supported by considering the fact that:

$$\frac{\operatorname{var}(X)}{(\operatorname{E}(X))^2} = \frac{1}{n^2} \leq 1 \quad \forall n \in \mathbb{N}$$

where $\operatorname{E}(X)$ and $\operatorname{var}(X)$ are as given in (B.28) and (B.29). This ratio is therefore never larger than 1 and for all values of $n > 1$ it is always less than 1. If $n = 1$, the it follows from Section 4.1.5 that the distribution reduces to the Exponential distribution which does not have a heavy tail.

Conclusion

Based on assessments using Techniques 1, 2 and 6, the Erlang distribution does not have a heavy tail.

5.2.3 Generalized Extreme Value Distribution

The Generalized Extreme Value distribution as it is given in (4.4) incorporates the Frechet, Gumbel and Weibull distributions.

Working with these distributions can be algebraically cumbersome. As such, the focus in this section is on using the techniques to assess whether distributions have heavy tails that are algebraically easier to apply. The techniques used are 3, 4 and 7.

Assessment using Technique 7

Given the form of the Generalized Extreme Value distribution as given by Jenkinson (see (4.4)) it has an exponentially bounded tail for $k = 0$ (i.e. for the Gumbel distribution) and for $k > 0$ (i.e. for the Weibull distribution) [105]. For the case where $k < 0$ in (i.e. for the Frechet distribution) this form doesn't have an exponentially bounded tail.

It can therefore be concluded at this stage that the Frechet distribution potentially has a heavy tail (since it doesn't have an exponentially bounded tail) while the Gumbel distribution doesn't, based on Technique 7.

Assessment using Techniques 3 and 4

Consider the hazard rate function of the Frechet that is stated here in terms of the parameterizations as given in Sections (4.1.7):

$$h_X^*(t) = \beta \left(\frac{\left(\frac{\delta}{t-\lambda}\right)^{\beta}}{(t-\lambda)\left(e^{\left(\frac{\delta}{t-\lambda}\right)^{\beta}} - 1\right)} \right) \text{ for } t \geq \lambda$$

This hazard rate function is a decreasing function of t. This then suggests that, based on Technique 4 (and consequently based on Technique 3), that the distribution does have a heavy tail.

Klugman et al [75] argued that the Weibull distribution, as a special case where $k > 0$ in (4.4), is the form for minima. We are interested in knowing whether the Weibull distribution for maxima has a heavy tail. A form for the maxima which is related to the version given in (4.16) for minima as follows:

$$F_X(x) = \begin{cases} 1 - e^{-\left(\frac{x+\lambda}{\delta}\right)^{\beta}} & \text{for } x \geq \lambda \\ 0 & \text{otherwise} \end{cases}$$

We then have that the hazard rate function is given by:

$$\begin{aligned} h_X^*(t) &= \frac{d}{dx}\left(-\ln(1 - \overline{F}_X(x))\right)\Big|_{x=t} \text{ for } T \geq \lambda \\ &= \frac{\beta}{\delta}\left(\frac{t+\lambda}{\delta}\right)^{\beta-1} \end{aligned}$$

From this hazard rate function it follows that for any t_1 and t_2 for which it is true that $\lambda < t_1 < t_2$ we have that:

* If $\delta > 0$ and $\beta > 1$ or $\beta < 0$ then

$$\frac{\beta}{\delta}\left(\frac{t_1+\lambda}{\delta}\right)^{\beta-1} < \frac{\beta}{\delta}\left(\frac{t_2+\lambda}{\delta}\right)^{\beta-1}$$

  in which case $h_X^*(t)$ is an increasing function.

* If $\delta > 0$ and $0 < \beta < 1$ then

$$\frac{\beta}{\delta}\left(\frac{t_1+\lambda}{\delta}\right)^{\beta-1} > \frac{\beta}{\delta}\left(\frac{t_2+\lambda}{\delta}\right)^{\beta-1}$$

  in which case $h_X^*(t)$ is an decreasing function and hence, the distribution has a heavy tail.

Similarly, for the two-parameter form of the Weibull distribution for maxima (as given in (4.16)) the distribution has a heavy tail for values of β between 0 and 1.

Consequently we have that the Rayleigh distribution doesn't have a heavy tail, since it is a special case of the two-parameter form of the Weibull distribution with $\beta = 2$. This consequence can be stated more formally in terms of the hazard rate function of the Rayleigh distribution which is given by:

$$h_X^*(t) = \frac{2t}{\theta^2} \text{ for } t \geq 0$$

from which it follows that for any t_1 and t_2 for which it is true that $t_1 < t_2$ we have that $h_X^*(t_1) < h_X^*(t_2)$.

Conclusion

The Frechet distribution does have a heavy tail. The Weibull distribution (for maxima) does have a heavy tail if $\beta \in (0, 1)$. Neither the Gumbel distribution nor the Rayleigh distribution has a heavy tail.

5.2.4 Pareto Distribution

For the Pareto Type II distribution we will consider Techniques 3, 4 and 6 to assess whether the distribution has a heavy tail. Techniques 1 and 2 requires the moment generating function, which is not available and Techniques 5 and 7 require extensive algebra.

Assessment using Techniques 3 and 4

Using (B.139) and (4.25), with $a = \theta$, from Section 4.1.11 we have for $x > 0$ that the hazard rate function is given by

$$\begin{aligned} h_X^*(t) &= \frac{f_X(t)}{1 - F_X(t)} \\ &= \frac{\kappa}{t + \theta} \end{aligned} \tag{5.2}$$

The hazard rate function is a decreasing function of t for positive values of θ and κ which suggests that the Pareto distribution does have a heavy tail.

Assessment using Technique 6

Using the mean and variance in (B.141) and (B.142), it follows that

$$\frac{\text{var}(X)}{(\,\text{E}(X))^2} = \frac{\kappa}{\kappa - 2} \geq 1 \;\forall\; \kappa \geq 2.$$

This suggests that the condition for heavy-tailedness is violated for values of $\kappa < 2$.

Conclusion

The Pareto distribution does have a heavy tail, but it appears to be only true for values of $\kappa \geq 2$.

5.2.5 Generalized Pareto Distribution

An explicit expression for the cumulative distribution function for the Generalized Pareto distribution is not available. The moment generating function also does not exist, but an general expression for the r^{th} moment is given in (B.75). Hence we consider Techniques 1 and 6.

Assessment using Technique 1

From (B.75) follows that

$$f_X(x) = \frac{\Gamma(\kappa + \tau)\theta^\kappa x^{\tau - 1}}{\Gamma(\kappa)\Gamma(\tau)(x + \theta)^{\kappa + \tau}} \text{ for } x \geq 0$$

and

$$\mathrm{E}\left(X^r\right) = \theta^r \frac{\Gamma(\tau + r)\Gamma(\kappa - r)}{\Gamma(\kappa)\Gamma(\tau)} \text{ for } -\tau < r < \kappa$$

which suggests that the moments of X only exist for values of r in the interval (τ, κ). Based on this, it appears as if this distribution may have a heavy tail.

Assessment using Technique 6

Consider the following (using the expressions for the mean and variance given in (B.76) and (B.77)):

$$\frac{\operatorname{var}(X)}{\left(\mathrm{E}(X)\right)^2} = \frac{\kappa + \tau - 1}{\tau(\kappa - 2)}$$

Evaluating the value of this ratio for various coordinates of (κ, τ) showed that this ratio is not always greater than or equal to 1, which suggests that this distribution is not necessarily heavy-tailed.

Limiting tail behaviour relative to the Pareto distribution

Alternatively consider the limit of the ratio of the probability density function of the Pareto distribution to that of the Generalized Pareto distribution.

$$\lim_{x\to\infty} \frac{\frac{\Gamma(\kappa+\tau)}{\Gamma(\kappa)\Gamma(\tau)} \frac{\theta^\kappa x^{\tau-1}}{(x+\theta)^{\kappa+\tau}}}{\frac{\kappa\theta^\kappa}{(x+\theta)^{\kappa+1}}} \text{ from (B.139) and (B.74)}$$
$$= \frac{\Gamma(\kappa + \tau)}{\kappa\Gamma(\kappa)\Gamma(\tau)} > 0$$

The limit of the ratio of these distributions do not tend to infinity. If we consider the reciprocal of this ratio it will also not tend to infinity, but to some constant. From Section 3.4.3 follows that the Pareto distribution will not necessarily have a heavier tail than the Generalized Pareto distribution, instead the density in the upper tail will be some multiple (as suggested by the ratio tending to some constant) of the density of the Generalized Pareto distribution. The same will apply when considering the Generalized Pareto distribution relative to the Pareto distribution.

Conclusion

It is true that the Pareto distribution (as a special case of the Generalized Pareto distribution) has a heavy tail (as shown in the previous section. Also see Rolski et al [105]). It is also true that there is no evidence that the either one of the Pareto or Generalized Pareto distributions has a heavier tail. It can therefore be concluded that, depending on the values of $\kappa$and τ, the Generalized Pareto has a heavy tail.

5.2.6 Lognormal Distribution

An explicit expression for the cumulative distribution function for the Lognormal distribution does not exist, neither does the moment generating function. Since the moment generating function of X does not exist, it is therefore argued that the distribution does have a heavy tail [105] . As an alternative we will consider an assessment using Technique 6.

Assessment using Technique 6

Consider the ratio of the variance to the square of the expected value using (B.128) and (B.129):

$$\frac{\text{var}(X)}{(\text{E}(X))^2} = e^{\sigma^2} - 1 \tag{5.3}$$

This ratio will only be larger than 1 if $|\sigma| > 0.83255$.

Conclusion

The Lognormal distribution does have a heavy tail. This does depend on the values for σ, since it was shown above that if $|\sigma| < 0.83255$, then the condition for heavy-tailedness is is violated.

5.2.7 Beta-prime Distribution

Explicit expressions for the cumulative distribution function and the moment generating function do not exist. For this reason we will consider an assessment using Technique 6.

Assessment using Technique 6

The probability density function for the Beta-prime distribution is given by:

Consider the ratio of the mean and variance given in (B.3) and (B.4):

$$\begin{aligned}\frac{\operatorname{var}(X)}{(\mathrm{E}(X))^2} &= \frac{(\delta_1+1)(\delta_2-1)}{\delta_1(\delta_2-2)}\\ &> \frac{(\delta_2-1)}{(\delta_2-2)}\\ &> 1 \end{aligned} \tag{5.4}$$

The ratio is greater than 1 which doesn't violate the consequent result of a distribution having a heavy tail, but this does not prove that this distribution does in fact have a heavy tail.

If we consider the probability density funtion of the Generalized Pareto distribution as given in (B.74) with $\theta = 1$, $\kappa = \delta_1$ and $\tau = \delta_2$, then the resulting probability density function is that of the Beta-prime distribution as given in (4.28). This means that the Beta-prime distribution is a special case of the Generalized Pareto distribution. Based on the arguments given in respect of the tail weight of the Generalized Pareto distribution, it is expected that the tail weight of the Beta-prime distribution will be dependent on the values of δ_1 and δ_2.

Limiting tail behaviour relative to the Pareto distribution

One can also consider the tail behaviour of the Beta-prime distribution relative to the tail behaviour of the Pareto distribution for which we confirmed tail heaviness in Section 5.2.4.

Let $f_X(\cdot)$ denote the Beta-prime density function as given in (4.28) and $f_Y(\cdot)$ the Pareto Type II density function as given in (B.139). The ratio of these two density functions are given by:

$$\frac{f_X(x)}{f_Y(x)} = \frac{\Gamma(\delta_1+\delta_2)}{\kappa\theta^\kappa\Gamma(\delta_1)\Gamma(\delta_2)}\frac{x^{\delta_1-1}}{(x+1)^{\delta_1+\delta_2}(x+\theta)^{\kappa+1}} \tag{5.5}$$

If $\delta_1 > 0$, $\delta_2 > 0$, $\kappa > 0$ and $\theta > 0$, then

$$\begin{aligned}x^{\delta_1-1} &< x^{\delta_1} < (x+1)^{\delta_1+\delta_2} \ \forall\ x > 0\\ &\leq (x+1)^{\delta_1+\delta_2}(x+\theta)^{\kappa+1} \ \forall\ x \geq 1.\end{aligned}$$

Hence for $x \geq 1$ we have that

$$\frac{x^{\delta_1 - 1}}{(x+1)^{\delta_1+\delta_2}(x+\theta)^{\kappa+1}} < 1$$

from which follows that for $x \geq 1$

$$\lim_{x\to\infty} \frac{f_X(x)}{f_Y(x)} = \frac{\Gamma(\delta_1+\delta_2)}{\kappa\theta^\kappa \Gamma(\delta_1)\Gamma(\delta_2)} \lim_{x\to\infty} \frac{x^{\delta_1-1}}{(x+1)^{\delta_1+\delta_2}(x+\theta)^{\kappa+1}} = 0.$$

This implies that the Beta-prime distribution has a lighter tail than the Pareto distribution.

Conclusion

From the assessment using Technique 6 no contradiction can be found to the assertion that the Beta-prime distribution has a heavy tail. The limiting tail behaviour of the Beta-prime distribution relative to the Pareto distribution suggests that the Beta-prime distribution has a lighter tail than the Pareto distribution. It can therefore not be confirmed that the Beta-prime has a heavy tail, neither can any contradiction to an assertion of heavy-tailedness be found.

5.2.8 Birnbaum-Saunders Distribution

The cumulative distribution function of the Birnbaum-Saunders is defined in terms of the cumulative distribution function of the standard Normal distribution - as given in (4.29). An explicit expression for the cumulative distribution function does not exist, hence we will use Technique 6 to show that the ratio of the variance to the square of the expected value forms a contradiction to an assertion that the Birnbaum-Saunders distribution has a heavy tail.

Assessment using Technique 6

The mean and variance, as given in (B.7) and (B.8), can be used to evaluate the ratio of the variance to the square of the expected value:

$$\begin{aligned}\frac{\operatorname{var}(T)}{(\mathrm{E}(T))^2} &= \frac{\alpha^2\beta^2\left(1+\frac{5}{4}\alpha^2\right)}{\left(\beta+\left(\frac{\alpha^2\beta}{2}\right)\right)^2} \\ &= \frac{4\alpha^2+5\alpha^4}{(\alpha^2+2)^2} \\ &= 1+4\frac{\alpha^4-1}{(\alpha^2+2)^2}\end{aligned}$$

This ratio can take on values greater than or less than 1 as is graphically shown in Figure 5.2 for varying values of α.

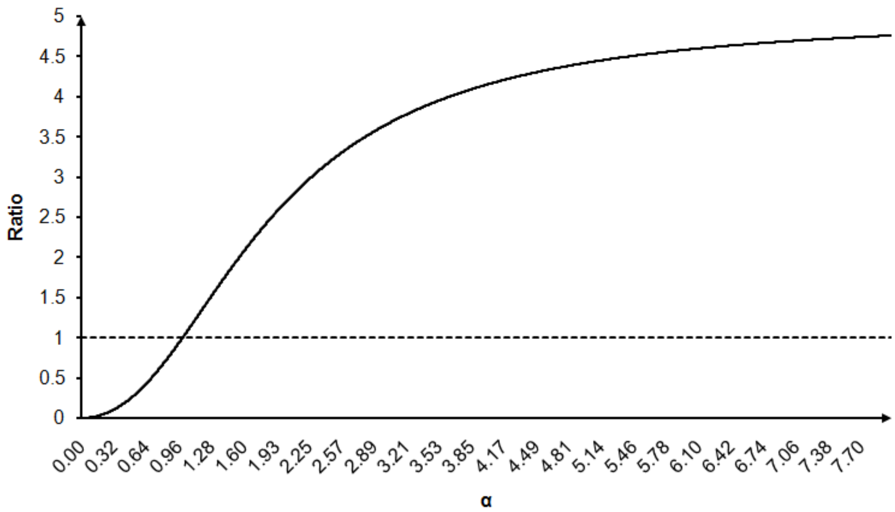

Figure 5.2: Ratio of Variance to Square of Expected value for the Birnbaum-Saunders distribution for varying values of α

Conclusion

For the Birnbaum-Saunders distribution there are instances, depending on the value of α, for which the ratio of the variance to the square of the expected value is less than 1 in which case it indicates that the distribution cannot have a heavy tail.

5.2.9 Burr and related distributions

The cumulative distribution function for the Burr distribution is available in an explicit form, but finding the two-fold convolution of the distribution function in order to determine if the distribution is a member of the subexponential class is algebraically cumbersome. An expression for the moment generating function does not exist. We will consider Techniques 1, 3 and 4 to determine whether the Burr distribution has a heavy tail.

Assessment using Technique 1

An expression for the r^{th} moment exist and is given in (B.10). From this expression it follows that the moments only exists for values of $-cr < ck$. This is indicative of the distribution potentially having a heavy tail.

Assessment using Techniques 3 **and** 4

Using the cumulative distribution and probability density functions as given in Section 4.1.16 for the Burr Type XII distribution, the hazard rate function can be derived using (B.9) and (4.31):

$$h_X^*(t) = \frac{kct^{c-1}}{\alpha^c + t^c}$$

It is not possible to determine algebraically whether this function is decreasing or increasing. We can, however, determine where this function will reach a stationary point. Consider the first order derivative with respect to t:

$$\frac{d}{dt}h_X^*(t) = \frac{kc\left(\alpha^c ct^{c-2} - \alpha^c t^{c-2} - t^{2c-2}\right)}{\left(\alpha^c + t^c\right)^2}$$

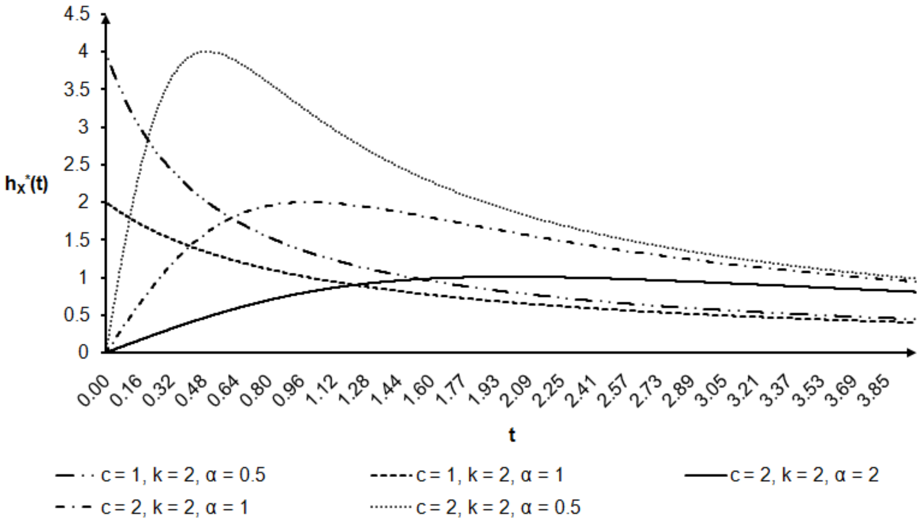

Figure 5.3: Hazard Rate Function for the Burr Type XII distribution for varying parameter values

Setting $\frac{d}{dt}h_X^*(t) = 0$ yields that the function will reach a stationary point at $t = \alpha(c-1)^{\frac{1}{c}}$. From numerical examples (as shown in Figure 5.3) it can be found that for values smaller than the stationary point, the hazard rate function is increasing and for values larger than this stationary point it will be decreasing and hence have a heavy tail.

Conclusion

Based on the numerical analysis of the hazard rate function, it was seen that after the point $t = \alpha(c-1)^{\frac{1}{c}}$ the hazard rate function is a monotone decreasing function. A decreasing hazard rate function also implies that $\alpha_F = 0$ - see Section 3.4.5. The Burr distribution therefore has a heavy tail.

5.2.10 Dagum Distribution

An expression for the moment generating function does not exist, hence we will consider Techniques 1, 3 and 4 to determine whether the Dagum distribution has a heavy tail.

Assessment using Technique 1

An expression for the r^{th} moment exist and is given in (B.22). From this expression it follows that the moments only exists for values of $-pa < r < a$. This is indicative of the distribution potentially having a heavy tail.

Assessment using Techniques 3 and 4

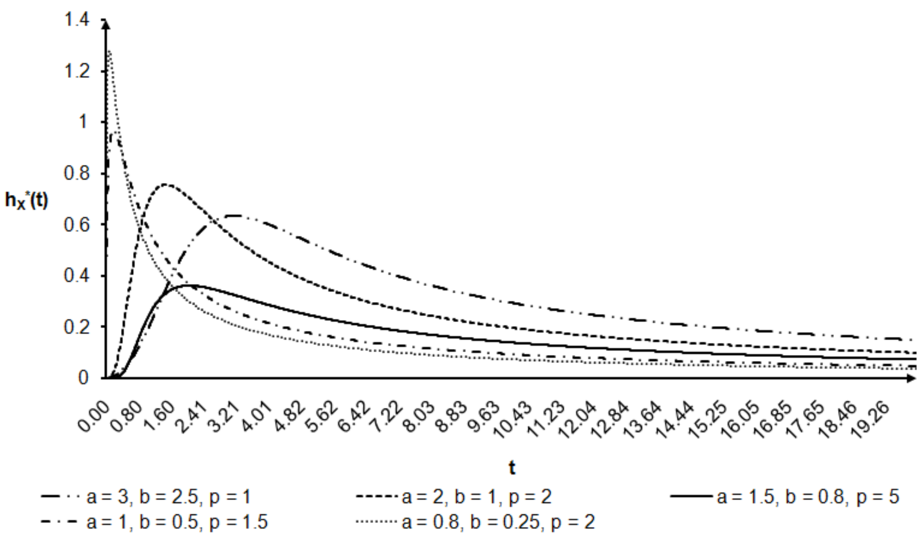

Figure 5.4: Hazard Rate Function for the Dagum distribution for varying parameter values

The hazard rate function for a random variable X from the Dagum distribution is given by:

$$h_X^*(t) = \frac{ab^a pt^{-(a+1)}\left(1 + b^a t^{-a}\right)^{-(p+1)}}{1 - \left(1 + \left(\frac{b}{t}\right)^a\right)^{-p}} \tag{5.6}$$

This function cannot really be simplified algebraically in order to obtain a form from which it is clear whether this function is increasing or decreasing and for which values of t it is increasing or decreasing. It is therefore recommended to consider a sketch of this function for various values of a, b and p as shown in Figure 5.6. From this figure it can be seen that for t sufficiently large the hazard rate function is strictly decreasing and hence suggests that the Dagum distribution has a heavy tail.

Conclusion

Based on the numerical analysis of the hazard rate function, the Dagum distribution has a heavy tail.

5.2.11 Singh-Maddala Distribution

In Section 4.1.19 it is shown that if a random variable Y follows a Dagum distribution with parameters a, $\frac{1}{b}$ and q, then $X = \frac{1}{Y}$ follows a Singh-Maddala distribution with parameters a, b and q. Furthermore this parameterization of the Singh-Maddala distribution is a special case of the Burr Type XII distribution with parameters $c = a, k = q$ and $\alpha = b$. Based on the fact that the Singh-Maddala distribution is a special case of the Burr distribution, it follows directly that the distribution has a heavy tail. We will show that the condition for heavy-tailedness still holds for Singh-Maddala distribution.

Assessment using Technique 1

An expression for the r^{th} moment exist and is given in (B.154). From this expression it follows that the moments only exists for values of $-qa < r < a$, which is indicative of the distribution potentially having a heavy tail.

Assessment using Techniques 3 **and** 4

The hazard rate function can be obtained using the expression for the Dagum distribution in (5.6) by letting $t = t^{-1}$ and $b = b^{-1}$:

$$h_X^*(t) = \frac{qat^{a-1}}{b^a + t^a}$$

From this expression it can be seen that for $t > 1$ the numerator will always be less than the denominator. Furthermore the function will reach a stationary point at $t = b(a-1)^{\frac{1}{a}}$. This is similar to what one can see for both the Burr and Dagum distributions as discussed in Sections 5.2.9 and 5.2.10. It was argued that for the Burr Type XII distribution its hazard rate function is decreasing for values of t larger than the stationary point; hence the same argument holds for the hazard rate function of the Singh-Maddala distribution.

Conclusion

The Singh-Maddala distribution has a heavy tail.

5.2.12 Kappa Family of Distributions

In Section 4.1.20 the two- and three-parameter cases are discussed.

Assessment using Technique 1

For the two-parameter case the expression for the r^{th} moment is given in (B.105). From this expression it is evident that the moments exist only for values of r where $-1 < r < \alpha$. Similarly for the three-parameter case the expression for the r^{th} moment is given in (B.110) from which it follows that the moments exist only for values of r such that $-\alpha\theta < r < \theta$. Since the moments for these distributions only exist for limited values of r, it is suggested that these distributions are heavy-tailed.

Assessment using Techniques 3 and 4

From (B.104), (4.41), (B.109) and (4.42) the hazard rate function follows:

$$h_X^*(t) = \begin{cases} \dfrac{\frac{\beta^\alpha}{(t^\alpha+\beta^\alpha)}}{(t^\alpha+\beta^\alpha)^{\frac{1}{\alpha}}-t} & \text{for the two-parameter case} \\[2ex] \dfrac{\left(\frac{\alpha\theta\frac{t^{\alpha\theta-1}}{\beta^{\alpha\theta}}}{1+\left(\frac{t}{\beta}\right)^\theta}\right)}{\left(1+\left(\frac{t}{\beta}\right)^\theta\right)^\alpha-\left(\frac{t}{\beta}\right)^{\alpha\theta}} & \text{for the three-parameter case} \end{cases}$$

It is not clear, using algebraic manipulation, whether these hazard rate functions are decreasing or increasing and over which parts of the domain these are increasing or decreasing. To assess this, we can use the graphs

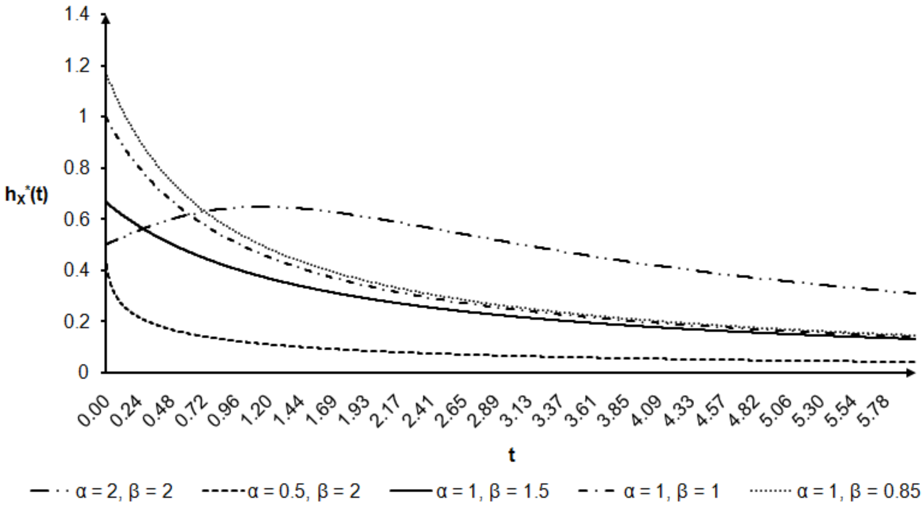

Figure 5.5: Hazard Rate Function for the Two-Parameter Kappa distribution for varying parameter values

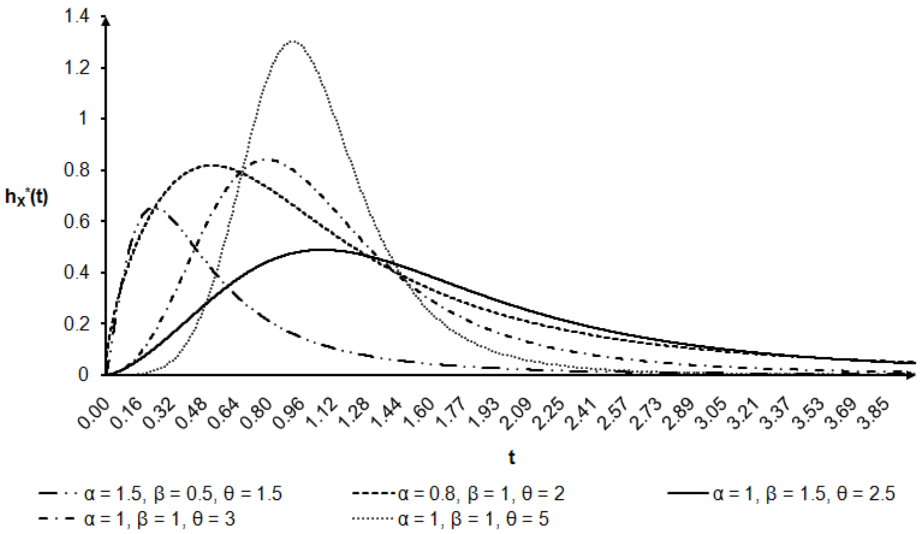

Figure 5.6: Hazard Rate Function for the Three-Parameter Kappa distribution for varying parameter values

shown in Figures 5.5 and 5.6, respecitvely. From these figures it can be seen that for large values of t the hazard rate functions are decreasing. A decreasing hazard rate function also implies that $\alpha_F = 0$ - see Section 3.4.5. This suggests that the two- and three-parameter Kappa distributions exhibit heavy tails.

Conclusion

It follows from the assessment results given above that the Kappa family of distributions are heavy-tailed distributions.

5.2.13 Generalized Beta Distribution of the Second Kind

We will use Techniques 1 and 6 to determine whether the distribution has a heavy tail.

Assessment using Technique 1

The expression for the r^{th} moment of X as given in (B.71) is valid for values of r where $-pa < r < qa$. This means that not all of the moments exist which further indicates that the distribution may potentially have a heavy tail.

Assessment using Technique 6

The variance and the expected value are known and therefore one can consider the ratio of the variance to square of the expected value:

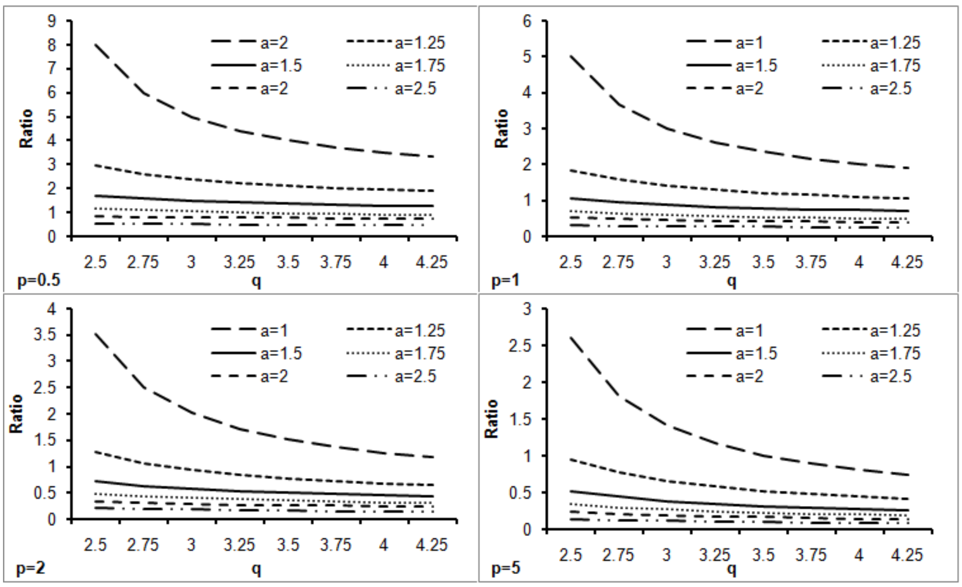

Figure 5.7: Ratio of Variance to Square of Expected value for the Generalized Beta distribution for varying parameter values

$$\frac{\text{var}(X)}{(\text{E}(X))^2} = \frac{\Gamma\left(\frac{2}{a}+p\right)\Gamma\left(q-\frac{2}{a}\right)}{\Gamma\left(\frac{1}{a}+p\right)^2\Gamma\left(q-\frac{1}{a}\right)^2}\Gamma(p)\Gamma(q) - 1$$

Algebraically it is not possible to determine whether this ratio is less than or greater than 1 and for which parameter values it is. Consider Figure 5.7 where this ratio is presented for varying values of a, p and q. From the graphs in this figure it can be seen that:

- As the values of a increase, with the values of p and q kept fixed, this ratio is increasing.
- As the values of q increases, with the values of a and p kept fixed, this ratio is decreasing.
- As the values of p increases, with the values of a and q kept fixed, this ratio is decreasing.

It can be seen that depending on the combination of the values for a, p and q, this ratio can be less than 1 which contradicts an assertion of heavy-tailedness in those instances.

Conclusion

Based on the assessment using Technique 1 it appears as if the distribution might exhibit tail heaviness. From the assessment using Technique 6 it is evident,though, that there are certain combinations of values for a, p and q for which the distribution definitely does not have a heavy tail.

5.2.14 Log-logistic Distribution

We will use Techniques 1, 3 and 4 to determine whether the Log-logistic distribution has a heavy tail.

Assessment using Technique 1

The expression for the r^{th} moment as given in (B.123) suggests that the moments exists only for values of r satisfying $-\beta < r < \beta$. This suggests that the distribution potentially has a heavy tail.

Assessment using Techniques 3 and 4

The hazard rate function for the Log-logistic distribution can be found using the probability density and cumulative distribution functions as given in (B.122) and (4.43):

$$h_X^*(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta + t^\beta}$$

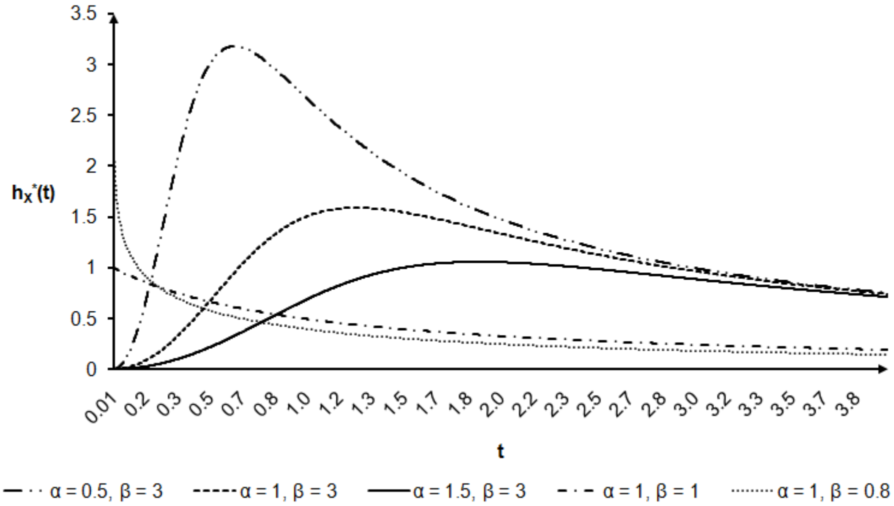

Figure 5.8: Hazard rate function of Log-logistic distribution for various parameter values

From this expression we can find the point t at which the function will reach a stationary point, that is where $t = \alpha(\beta-1)^{\frac{1}{\beta}}$ for values of $\beta \geq 1$. When $0 < \beta < 1$, the hazard rate function is monotone decreasing for all values of $t \geq 0$. As can be seen from Figure 5.8, the shape of the function is dependant on the values of α and β. When considering these graphs, especially for values of $t > \alpha(\beta-1)^{\frac{1}{\beta}}$, the function is decreasing and therefore indicative of the distribution having a heavy tail.A decreasing hazard rate function also implies that $\alpha_F = 0$ - see Section 3.4.5.

Conclusion

From the grpahical analysis of the hazard rate function it is evident that the fucntion is decreasing for values of t greater than the stationary point of the hazard rate function for values of $\beta \geq 1$. When $0 < \beta < 1$, the function is monotone decreasing for all positive values of t. This suggests that the disribution is heavy-tailed.

5.2.15 Folded Normal Distribution

The Folded Normal distribution is a transformation of the Normal distribution (as discussed in Section 4.1.24) with a probability density function as given in (B.43). The Half Normal distribution is a special case of the Folded Normal with a value of $\mu = 0$. Essentially the Folded Normal is taking a Normal distribution and maps the area of the distribution which corresponds to values of $X < 0$ onto the positive domain ($\mathbb{R}^+$). This means that generally the density for lower postive values will be higher than for an associated Normal distribution. The extreme case in which the most will be added to the density associated with the upper right tail will be when $\mu = 0$, which is the Half Normal distribution in which case the tail density will be double the density of an associated Normal distribution on the same domain. This means that if the Half Normal distribution doesn't have a heavy tail, the Folded Normal will also not have a heavy tail.

We will use Technique 6 to show that the distribution does not have a heavy tail.

Assessment using Technique 6

We will proceed in finding a contradiction to this condition for the Half Normal to argue that the Half Normal doesn't have a heavy tail and consequently the Folded Normal in general doesn't have a heavy tail. Following from (B.54) and (B.55):

$$\frac{\text{var}(X)}{(\text{E}(X))^2} = \frac{\pi}{2} - 1 < 1 \quad \forall \quad \mu \geq 0, \sigma > 0$$

which indicates that the Half Normal doesn't have a heavy tail.

Conclusion

It was argued above that the Half Normal distribution is the special case of the Folded Normal with the heaviest tail. Since the Half Normal distribution doesn't have a heavy tail, it can be concluded that the Folded Normal, in general, does not have a heavy tail.

5.2.16 Inverse Gamma and related distributions

Techniques 1 and 6 will be used to ascertain whether this distribution has a heavy tail.

Assessment using Technique 1

The r^{th} moment of an Inverse Gamma random variable X is given in (B.90) which is only valid for $r < \alpha$. Since the moments exist only for limited values of r, it is indicative of the distribution potentially being heavy-tailed.

Assessment using Technique 6

Using (B.91) and (B.92), the ratio of the variance to the square of the expected value results in $(\alpha - 2)^{-1}$ which is

- less than or equal to 0 if $\alpha \leq 2$,
- between 0 and 1 if $\alpha \geq 3$ and
- larger than 1 if $2 < \alpha < 3$.

We have identified values for α that leads to a contradiction of the assertion that the distribution does have a heavy tail.

It was shown in Section 4.1.26 that if W is Inverse Gamma distributed with parameters $\frac{\nu}{2}$ and θ, then $X = \frac{W}{2\theta}$ is Inverse Chi-square distributed with parameter ν. Hence the ratio of the variance to the square of the expected value results in $\frac{2}{\nu - 4}$ which is

- less than or equal to 0 if $\nu \leq 4$,
- between 0 and 1 if $\nu \geq 6$ and
- larger than 1 if $4 < \nu < 6$.

Conclusion

From the assessment using Technique 1 it appears as if the distribution may have a heavy tailed. This cannot be confirmed, though. Technique 6's assessment indicated that we can indentify values for α for which the assertion of heavy-tailedness is contradicted. It can therefore be concluded that we know that for values of $\alpha \notin (2, 3)$ the Inverse Gamma distribution does not have a heavy tail.

Similarly for the Inverse Chi-square distribution we know that for values of $\nu \notin (4, 6)$ the distribution does not have a heavy tail.

5.2.17 Loggamma Distribution

We will consider Techniques 1 and 6 to determine whether the distribution has a heavy tail.

Assessment using Technique 1

The expression for the r^{th} moment as given in (B.114) suggests that the moments exists for all values of r. Therefore there is no indication based on this Technique of the distribution having a heavy tail.

Assessment using Technique 6

Using (B.116) and (B.115), the ratio of the variance to the square of the expected value can be evaluated:

$$\frac{\operatorname{var}(X)}{(\operatorname{E}(X))^2} = \left(\frac{(\lambda-1)^2}{\lambda(\lambda-2)}\right)^a - 1$$

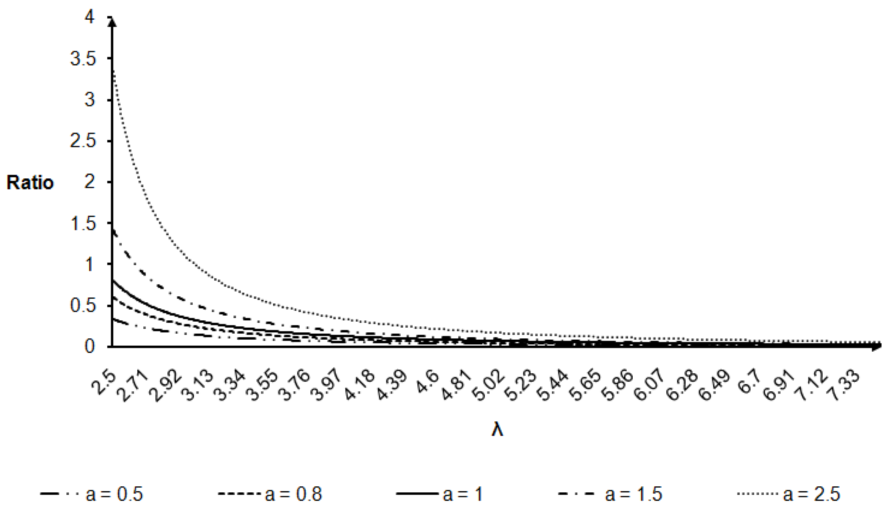

Figure 5.9: Evaluation of $\frac{\operatorname{var}(X)}{(\operatorname{E}(X))^2}$ for varying values of λ and a

Figure 5.9 presents this ratio for varying values of λ for various values of a. The graph only includes values for $\lambda \geq 2.5$. It can be seen that for larger values of λ this ratio becomes smaller than 1 which indicates that there exists values of λ and a for which the distribution does not have a heavy tail.

Conclusion

When using Technique 6, evaluation of the ratio for various values of a and λ shows that this ratio can have values that are greater than or less than 1 which suggests that there exists parameter values for which this distribution doesn't have a heavy tail.

5.2.18 Inverse Gaussian Distribution

Techniques 1, 2 and 6 will be used here to determine whether this distribution has a heavy tail.

Assessment using Technique 1

It is stated by Rolski et al [105] that all the positive and negative moments of this distribution exists which suggests that the distribution does not have a heavy tail.

Assessment using Technique 2

The expression for the moment generating function given in (B.97) is finite which also suggest that the distribution does not have a heavy tail.

Assessment using Technique 6

The ratio of the variance to the square of the expected values is obtained using (B.99) and (B.100) given by:

$$\frac{\mathrm{var}(T)}{(\mathrm{E}(T))^2} = \frac{\mu}{\lambda}.$$

which is larger than 1 if $\mu > \lambda$, less than 1 if $\mu < \lambda$ and equal to 1 if $\mu = \lambda$ which suggests that the distribution doesn't have a heavy tail.

Conclusion

From the assessment using Technique 6 we could identify values for μ and λ for which an assertion of heavy-tailedness is contradicted but the assessments from Techniques 1 and 2 indicate that, irrespective of the values of μ and σ, the distribution does not have a heavy tail.

5.2.19 Snedecor's F Distribution

Techniques 1 and Techniques 6 will be considered here to determine if the F distribtion has a heavy tail.

Assessment using Technique 1

The expression for r^{th} moment is given in (B.165) from which it follows that the moments exist only for values of $r \in (-\frac{\nu_1}{2}, \frac{\nu_2}{2})$. This suggests that the moments exist only for limited values of r in which case the distribution potentially has a heavy tail.

Assessment using Technique 6

Consider the ratio of the variance to the square of the expected value:

$$\frac{\mathrm{var}(X)}{(\,\mathrm{E}(X))^2} = \frac{\nu_1 + 2}{\nu_1(\nu_2 + 2)}. \text{ from (B.166) and (B.167)}$$

This ratio will be less than 1 for values of ν_1 and ν_2 satisfying the inequality

$$\nu_1 > \frac{2}{\nu_2 + 1}.$$

This means that for the case where the parameter values satisfy this inequality the distribution does not have a heavy tail.

Conclusion

When using Technique 6, algebraic analysis of the ratio of the variance to the square of the expected value indicates that this ratio can have values that are greater than or less than 1 which suggests that there exists parameter values for which this distribution doesn't have a heavy tail.

5.2.20 Skew-normal Distribution

Consider the parameterization of the Skew-normal as given in Section 4.1.28 in (4.47). We will use Techniques 3 together with some of the results discussed in Section 4.1.28 to argue why the Skew-normal distribution does not have a heavy tail.

Assessment using Technique 3

Using the expression for the cumulative distribution as given in (4.48), the hazard rate function for a random variable that is $SN(\lambda)$ distributed can be obtained as follows:

$$\begin{aligned} h_Y(t) &= \frac{\phi(t;\lambda)}{1-\Phi(t;\lambda)} \\ &= \frac{2\phi(t)\Phi(\lambda t)}{1-2\int_{-\infty}^{t}\int_{-\infty}^{\lambda s}\phi(s)\phi(u)duds} \\ &= \frac{2\phi(t)\Phi(\lambda t)}{1-2\int_{-\infty}^{t}\phi(s)\Phi(\lambda s)ds} \end{aligned}$$

This function has an integral in the denominator which makes it difficult to assess algebraically whether the function is increasing or decreasing. For this assessment we consider sketching this function whilst making use of numerical integration to calculate the denominator at the various values for t. This is shown in Figure 5.10.

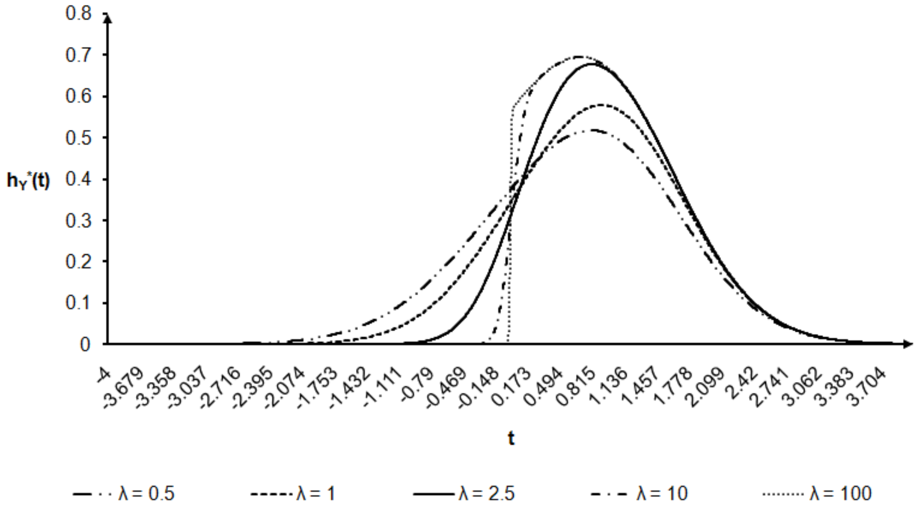

Figure 5.10: Hazard rate function of Skew-normal distribution for various parameter values

From the graphs of the hazard rate function for various values of λ and the evaluation of the ratio above, it is clear that there is evidence that at least for some values of λ the distribution doesn't have a heavy tail whilst it follows that the distribution's skewness and heaviness in tail increases with the absolute value increase in the value of λ.

Conclusion

Whether the distribution has a heavy tail for some values of λ is not clear from our investigation thus far. From Section 4.1.28 follows that as λ tends to infinity, the Skew-normal distribution will tend to the half-normal distribution [8]. In Section 5.2.15 it is argued that the half-normal distribution is a special case of the Folded Normal and that the Half Normal doesn't have a heavy tail. Since it can be seen from Figure 4.55 that for increasing values of λ the tail weight of the Skew-normal increases, it follows that the Skew-normal also doesn't have a heavy tail as the half-normal is considered to be the extreme case.

Chapter 6

Fitting Parametric Loss Distributions

6.1 Introduction

In this chapter methods will be introduced that can be used to fit the parametric distributions that were introduced in Chapter 4 to observed data. Algorithms for the distribution fitting are proposed for each distribution. Methods that can be used to simulate from each of the distributions are also discussed. These simulations were used to assess the performance of the fitting algorithms.

The process of fitting parametric distributions to observed data requires that the parameters need to be estimated. Two frequently used techniques include:

- Method-of-moments [11], [75]
- Maximum likelihood estimation [11], [75], [76], [31].

These techqniques are defined below.

Definition 29. ***Method-of-moments Estimation***
Consider a distribution with parameters $\underline{\theta}$. The method-of-moments estimate of $\underline{\theta}$ is the solution to the p equations [75]

$$\mu_k^{'}(\underline{\theta}) = \hat{\mu}_k^{'}, k = 1, 2, \ldots, p$$

where $\mu_k^{'}$ denotes the k^{th} moment about the origin as given in Definition 8 and $\hat{\mu}_k^{'}$ denotes the k^{th} observed moment about the origin given by $\frac{1}{n}\sum_{j=1}^{n} x_j^k$.

Definition 30. ***Maximum Likelihood Estimation***
Maximum likelihood estimation requires one to find the values of the parameters that maximises the likelihood function [31], where the likelihood function is given by

$$L(\underline{\beta}) = \prod_{j=1}^{n} f_{X_i}(x_i; \underline{\beta}) \tag{6.1}$$

where $\underline{\beta}$ denotes the vector of distribution parameters that need to be estimated. The log-likelihood function is given by

$$\begin{aligned} l(\underline{\beta}) &= \ln(L(\underline{\beta})) \\ &= \sum_{j=1}^{n} \ln(f_{X_i}(x_i; \underline{\beta})). \end{aligned} \tag{6.2}$$

The maximum likelihood estimate is given by

$$\hat{\underline{\beta}} = \arg\max_{\underline{\beta}} l(\underline{\beta}) \tag{6.3}$$

Various statistical software packages are available that can be used to fit parametric distributions and to simulate from these distributions. Examples include:

- **SAS**
  This is a well-known statistical package with an offering of built-in statistical distributions that can be fitted to observed data. Built-in functions also exist from which can be simulated. In this study we have used these functions only for the purpose of simulating. The built-in functions were not used to fit distributions, since we considered fitting these distributions from first principles using numerical techniques as described in Section 6.4 below.

- **R**
  Becoming increasingly popular amongst practitioners, R provides a distribution fitting package which performs the parameter estimation and goodness-of-fit evaluation statistics and graphics. Empirical density estimation and smoothing can also be performed. In terms of the variety of parametric distributions for which functionality is available, R compares well with SAS. In general distributions can be fitted using the built-in function *fitdistr()* in the *MASS* package using various methods, including method-of-moments and maximum likelihood estimation.

  Alternatively, the *fitdistrplus* package provides an integrated solution which facilitates the choice of parametric distribution by means of assessing a skewness-kurtosis plot of distributions which can then be compared with the skewness and kurtosis of the observed data to assess which of the possible distributions are viable. This package is flexible in terms of various estimation techniques and also provides statistics and graphical results that can be used to evaluate the goodness-of-fit of fitted distributions.

- **Mathematica**
  This software package provides possibly the most extensive set of parametric distributions that can be fitted to the data making use of maximum likelihood estimation, method-of-moments estimation and other similar techniques. The functions in the package can be used for simulation. Furthermore the package provides comprehensive graphical presentations of probability density functions, cumulative distribution functions, hazard rate functions. In comparison to other packages (such as SAS and R mentioned above), the analysis in Mathematica is much more interactive in which settings and options can be adjusted without having to code any syntax whilsy the results are immediately obtained.

6.2 Numerical Optimisation

To maximise the log-likelihood partial derivatives with respect to each of the parameters should be derived which can then all be set equal to 0 and the solution to these expressions can be used to obtain expressions for the maximum likelihood estimators of these parameters. Often it is not possible to find explicit analytic formulae for the different parameters. In such a case a numerical procedure needs to be implemented such as the Newton-Raphson iterative algorithm. A general updating equation for the Newton-Raphson iterative process [111] is given by

$$x^* = x - \frac{f(x)}{f'(x)}$$

where x^* represents the root of the function $f(X)$. If the parametric distribution has only one parameter, this form can be utilised to find the value of its parameter β that maximises the log-likelihood function iteratively as follows:

$$\beta^{(m+1)} = \beta^{(m)} - \left.\frac{l(\beta)}{l'(\beta)}\right|_{\beta=\beta^{(m)}} \text{ for } m = 1, 2, 3, ... \tag{6.4}$$

where $\beta^{(m)}$ is the updated estimate of β that is obtained in the m^{th} step of the iterative procedure.

If the parametric distribution has more than one parameter, one can use a Taylor series expansion [111] to write $l(\underline{\beta}^{(m)})$ as follows:

$$l(\underline{\beta}^{(m)}) \cong l(\underline{\beta}) + \left(\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}}\right)^T (\underline{\beta}^{(m)} - \underline{\beta})$$
$$+ \frac{1}{2}(\underline{\beta}^{(m)} - \underline{\beta})^T \left(\frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T}\right) (\underline{\beta}^{(m)} - \underline{\beta}),$$

where T denotes the transpose of the vector.

The aim is to find a value $\underline{\beta}^{(m)}$ that maximises the log-likelihood. This can be found by taking the partial derivative of the log-likelihood function (as expanded using a Taylor series) and setting the derivative equal to $\underline{0}$:

$$\left(\frac{\partial l(\underline{\beta}^{(m)})}{\partial \underline{\beta}^{(m)}}\right) \cong \left(\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}}\right) + \left(\frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T}\right) \left(\underline{\beta}^{(m)} - \underline{\beta}\right).$$

Now let

$$\left(\frac{\partial l(\underline{\beta}^{(m)})}{\partial \underline{\beta}^{(m)}}\right) = 0$$

from which follows that

$$\underline{\beta}^{(m)} = \underline{\beta} - \left(\frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T}\right)^{-1} \left(\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}}\right)$$

which can be used to set up an updating formulae for an iterative estimation procedure as follows:

$$\underline{\beta}^{(m+1)} = \underline{\beta}^{(m)} - \left(\frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T}\right)^{-1} \Bigg|_{\underline{\beta}=\underline{\beta}^{(m)}} \left(\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}}\right) \Bigg|_{\underline{\beta}=\underline{\beta}^{(m)}}. \tag{6.5}$$

6.3 Simulating Parametric Loss Distributions

For the distributions discussed in Chapter 4 algorithms that can be used to fit these distributions to real-life data will be derived in section 6.5. In order to assess the reasonability of these algorithms we will simulate data from each of these distributions (with specified parameters) after which the derived algorithm will be applied in order to assess the performance of the fitting algorithm. In simulating these distributions we make use of either one of three main techniques:

1. Using SAS software's built-in functions to generate data points on a random basis from specific parametric distributions. These built-in functions are limited though. It is important to also consider what parameterization of the distribution is assumed in the the built-in function. The simulated data points can be easily assessed in terms of its empirical distributions and empirical moments and compared to their theoretical counterparts. It is generally seen that as the number of simulated data points is increased, the simulated distribution and its associated moments tends to be closer to the theoretical counterparts.

2. Using the probability integral transformation. This is a well-known technique whereby the cumulative distribution function of a parametric distribution can be set equal to a uniform distributed random variable with domain $(0, 1)$. This means that we can simulate uniform random variables and find random values under the proposed parametric distributions by applying the inverse of the cumulative distribution function to the random uniform variable [11]. This technique is dependent on whether it is possible to algebraically find the inverse of the cumulative distribution function.

3. In many instances neither one of the two approaches described above can be used, since neither a built-in function exists nor does the inverse of the cumulative distribution function exists. In some cases the cumulative distribution function isn't even known.

   The following pragmatic approach is suggests in this case:

   - Consider a sufficiently large domain on the real line. For the distributions considered in this study a sufficiently large portion on the positive half of the real line should be considered.
   - Partition the chosen portion of the real line into arbitrarily small intervals.
   - On each interval calculate the probability density at either the smallest, largest or middle value of the interval.
   - Use a numerical integration procedure to calculate a cumulative distribution function for the chosen partitioning.
   - Test that the chosen domain together with the partitioning is sufficient such that the maximum value of the resulting cumulative distribution function is sufficiently close to 1.

   The approach outlined above will give one a close representation of the true cumulative distribution function. Having this representation available, the probability integral transformation can be used as follows:

- Simulate a random uniform variable u, where $0 \leq u \leq 1$.
- Use the constructed cumulative distribution function, $F(x)$, to obtain a simulated value from the distribution by finding the value of X as $\inf\{x|u \leq F(x)\}$.

Slight variations on the approaches given above are followed and in some cases a combination of these techniques are followed. The technique used in each instance is briefly described in each distribution's respective subsection in section 6.5.

6.4 Fitting Analysis

In Section 6.5 the following are included for each parametric distribution:

- Derivation of the parameter estimation algorithm.
- A brief description of the simulation technique used to obtain observations from the parametric distribution being discussed.
- Discussion on the performance of the estimation algorithm when applied to the simulated data.

The simulation and the application of these estimation algorithms are facilitated with illustrative SAS programs. These SAS programs are included in Appendix A. A separate program is provided for each parametric distribution and is structured as follows:

1. The program starts off with the simulation component. The simulation is done using one of the techniques as discussed in Section 6.3.
2. A basic univariate analysis is performed using the built-in procedures which gives insight to how the underlying distribution looks.
3. For the distributions where method-of-moments estimates can be found, this estimation will follow using the results obtained in the univariate analysis.
4. The last part of the SAS code contains the portion necessary to perform maximum likelihood estimation. This will be done using the derived estimation algorithms. In some instances this estimation can be performed directly, but in most cases this involves numerical estimation. The iterative numerical estimation procedures are built into the PROC IML environment in SAS which involves applying the updating equation as given in (6.5) until the procedure converges.

6.5 Methods of Fitting Parametric Loss Distributions

6.5.1 Gamma Distribution

Deriving method-of-moments estimators and maximum likelihood estimation algorithm

Using the mean and variance as given in (B.64) and (B.65) the method-of-moments estimates follows directly. The method-of-moments estimators are given for θ and κ in (B.68) and (B.68), respectively. It is not clear how to evaluate these estimates' bias. Fisher [53], [30] indicate that the method-of-moments estimates may be insignificant and that maximum likelihood estimates should instead be considered.

The likelihood function in terms of κ and θ follows from (B.61):

$$L(\kappa,\theta) = \prod_{j=1}^{n} \frac{1}{\theta^k \Gamma(\kappa)} x_j^{\kappa-1} e^{-\frac{x_j}{\theta}}$$

$$l(\kappa,\theta) = n\left(\kappa \ln\left(\frac{1}{\theta}\right) - \ln(\Gamma(\kappa))\right) + (\kappa-1)\sum_{j=1}^{n}\ln(x_j) - \frac{1}{\theta}\sum_{j=1}^{n} x_j \qquad (6.6)$$

The estimation of κ and θ requires numerical techniques. The Newton-Raphson algorithm [111], [15] can be used to find values of κ and θ that maximizes (6.6). The iterative procedure will have an updating formula of the form [30]:

$$\hat{\kappa}_{k+1} = \hat{\kappa_k} - \frac{\ln(\hat{\kappa}_k) - \Psi(\hat{\kappa}_k) - \ln(\overline{x}) - \overline{w}}{\frac{1}{\hat{\kappa}_k} - \Psi'(\hat{\kappa}_k)} \text{ for } k \in \mathbb{N} \qquad (6.7)$$

where

$$\overline{w} = \frac{\sum_{i=1}^{n}}{\ln(x_i)},$$

from which the estimate of θ can be obtained as:

$$\hat{\theta} = \frac{\overline{x}}{\hat{\kappa}}$$

The digamma and trigamma functions are defined and used by several authors; see Kotz and Nadarajah [76] for example. Choi and Wette [30] gave exact formulae for the digamma and trigamma functions as follows:

$$\Psi(\kappa) = -\gamma - \frac{1}{\kappa} + \kappa \sum_{i=1}^{\infty} \frac{1}{i(i+\kappa)} \tag{6.8}$$

$$\text{where } \gamma \equiv \text{ Euler-Mascheroni constant [76]} \tag{6.9}$$

$$\Psi^{'}(\kappa) = \sum_{i=1}^{\infty} \left(\frac{1}{i+\kappa} \right)^2 \tag{6.10}$$

Exact values for $\Psi(\kappa)$ and $\Psi^{'}(\kappa)$ can be found using tables given by Pairman [96]. Software packages, such as SAS and MATLAB, have built-in functions to give values for the digamma and trigamma function for any valid value of κ. Approximations to the digamma and trigamma are given by Jordan [72], [30]. These approximations are obtained by expanding the expressions of the digamma and trigamma functions as Bernoulli series.

The iterative scheme as given in (6.7) is convergent for any positive initial value $\hat{\kappa}_0$ [30]. Furthermore, Choi and Wette considered a numerical study of possible bias of the maximum likelihood estimates (for which no analytic formula is known). It is shown that the bias decreases as the sample size increases. Empirical results showed that estimates for κ is generally higher than the true value of κ; that is $\mathrm{E}(\hat{\kappa}) \geq \kappa$. Estimates for θ appears to be very close to the true value of θ, especially for larger sample sizes and as such the maximum likelihood estimator for θ is considered to have no bias for larger samples.

Simulation technique

SAS code is given in Appendix A.1 which can be used to simulate from a specified Gamma distribution. This approach utilizes the built-in SAS function to generate observations from a Gamma distribution.

Performance of the estimation algorithm

The SAS code in Appendix A.1 also includes the coded iterative algorithm that can be used to fit a Gamma distribution to observed data. Generally the larger the sample size, the closer the maximum likelihood estimates are to the parameters that were specified for the purpose of the simulation. The method-of-moments estimates of the parameters are used as initial values in the iterative procedure and appears to be working well in ensuring convergence of the procedure.

6.5.2 Exponential Distribution

Deriving method-of-moments and maximum likelihood estimators

The method-of-moments estimator, as given in (B.41), can be found by letting $\mathrm{E}(X) = \overline{X}$.

One can also use maximum likelihood estimation to obtain an estimate for θ as follows:

$$l(\theta) = \ln(L(\theta)) = \ln\left(\prod_{j=1}^{n} \frac{1}{\theta} e^{-\frac{x_j}{\theta}}\right) = -n\ln(\theta) + \frac{n\overline{x}}{\theta} \text{ from (B.34)}$$

By taking the first order derivative of the log-likelihood function with respect to θ and setting it equal to 0 yields the maximum likelihood estimate as given in (B.42). This is the same as the method-of-moments estimate which is an unbiased estimate of θ in that

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}(\overline{X}) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}(X_j) = \theta.$$

The Exponential distribution is a special case of the Gamma distribution, which suggests that in practice it would be more worthwhile to first fit the Gamma distribution. If the true underlying distribution is an Exponential distribution, then this will be indicated by the maximum likelihood estimate of the κ parameter being very close to 1.

Simulation technique

SAS code is given in Appendix A.2 which can be used to simulate from a specified Exponential distribution. A built-in function in SAS is used to simulate random observations from an Exponential distribution.

Accuracy of estimates

For the Exponential distribution the maximum likelihood estimator for paramater θ is simply the sample mean which is an unbiased estimator for θ. Generally the larger the sample size, the closer the maximum likelihood estimate is to the parameter that was specified for the purpose of the simulation.

6.5.3 Chi-square Distribution

Deriving method-of-moments estimators and maximum likelihood estimation algorithm

The method-of-moments estimate of ν given in (B.20) can be found by letting the $\mathrm{E}(X)$ equal to the sample mean. This is an unbiased estimator of ν, since

$$\mathrm{E}(\hat{\nu}) = \mathrm{E}(\overline{X}) = \nu.$$

Since the Chi-square distribution is a special case of the Gamma distribution with $\theta = 2$ and $\kappa = \frac{\nu}{2}$, one can use the log-likelihood function as derived for the Gamma distribution in (6.6):

$$l(\nu) = -\frac{n\nu}{2}\ln(2) - n\ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \sum_{j=1}^{n}\ln\left(x_j^{\frac{\nu}{2}-1}\right) - \frac{1}{2}\sum_{j=1}^{n}x_j$$

To maximize the log-likelihood function, one needs to find the solution to $\frac{\partial l}{\partial \nu} = 0$. This can be obtained using the Newton-Raphson iteration algorithm which takes on the form:

$$\hat{\nu}_{k+1} = \hat{\nu}_k - \frac{\frac{\partial l}{\partial \nu}}{\frac{\partial l^2}{\partial \nu^2}}$$

One can leverage of the methodology given by Choi and Wette [30] for the Gamma distribution in (6.7) by letting $\frac{\nu}{2} = \kappa$ in which case the log-likelihood function reduces to the log-likelihood function as derived for the Gamma distribution with $\theta = 2$:

$$\hat{\kappa}_{k+1} = \hat{\kappa}_k - \frac{\Psi(\kappa) - \frac{1}{n}\sum_{j=1}^{n}\ln(x_j) + \ln(2)}{\Psi'(\kappa)} \tag{6.11}$$

Expression (6.11) can then be used as an iterative algorithm to find the value for κ that maximizes the log-likelihood function from which one can find the maximum likelihood estimate of ν as $\hat{\nu}_{MLE} = 2\hat{\kappa}_{MLE}$.

The bias of this estimate can be studied by means of numerical techniques as discussed by Choi and Wette [30], since no explicit expression is available for the bias of this estimate.

Since the Chi-square distribution is a special case of the Gamma distribution, it would be more worthwhile to instead fit the Gamma distribution. If the true underlying distribution is a Chi-square distribution, then this will be indicated by the maximum likelihood estimate of the θ parameter being very close to 2.

Simulation technique

SAS code is given in Appendix A.3 which can be used to simulate from a specified Chi-square distribution. The relationship that exists between the Chi-square and Gamma distributions are used to simulate from the Chi-square distribution by means of using the built-in function in SAS to simulate from a Gamma distribution.

Performance of the estimation algorithm

The SAS code given in Appendix A.3 also includes code for estimating the maximum likelihood estimates using the procedure as described above. The method-of-moments estimate of ν can be used as initial value in the iterative procedure. Generally the larger the sample size, the closer the maximum likelihood estimate is to the parameter that was specified for the purpose of the simulation.

6.5.4 Two-parameter Exponential Distribution

Deriving the maximum likelihood estimators

The likelihood function can be obtained using the probability density function given in (B.171):

$$L(\theta, \eta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\left(\sum_{j=1}^{n} x_j - n\eta\right)}$$

hence

$$l(\theta, \eta) = -n\ln(\theta) - \frac{\overline{x}}{\theta} + \frac{n\eta}{\theta}.$$

Taking the first order derivatives of the log-likelihood function with respect to θ and η, respectively, and setting these derivative equal to 0 yield:

$$\theta = \overline{x} - \eta$$

and

$$\frac{n}{\theta} = 0$$

These equations do not give any means of finding an estimate of η. Bain and Engelhardt [11] give reasoning on how the maximum likelihood estimate can be found for η. It is given that the likelihood function is only valid for values of $x \geq \eta$. Also it can be noted that

$$\sum_{j=1}^{n} (x_j - \eta) \tag{6.12}$$

will be minimized if $\eta = x_{(1)}$ where $x_{(1)}$ is the smallest observed value of X. Consequently the likelihood function will be maximized of $\eta = x_{(1)}$. Hence the maximum likelihood estimators (as also given in (B.177) and (B.178)) are $\hat{\eta} = X_{(1)}$ and $\hat{\theta} = \overline{X} - X_{(1)}$ for η and θ, respectively.

To study the bias the of these estimates, firstly consider the distribution of the first order statistic, $X_{(1)}$, of a set of independent, identically distributed random variables $X_1, X_2, ..., X_n$.

$$\begin{aligned} \mathrm{P}(X_{(1)} \leq x) &= 1 - (1 - F_X(x))^n \text{ where } X \sim EXP(\theta, \eta) \\ &= 1 - e^{-\frac{x-\eta}{\frac{\theta}{n}}}. \end{aligned}$$

It can then be concluded that $X_{(1)} \sim EXP\left(\frac{\theta}{n}, \eta\right)$. Using (B.37) and (B.38), it can be concluded that $\mathrm{E}(\hat{\eta}) = \dfrac{\theta}{n} + \eta$ which means that the maximum likelihood estimator of η is biased. It should be noted that $\mathrm{E}(\hat{\eta})$ tends to η as n tends to ∞. Also

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}(\overline{X}) - \mathrm{E}(Y) = \left(\frac{n-1}{n}\right)\theta$$

which means that this estimator is biased, but $\mathrm{E}(\hat{\theta})$ tends to θ as n tends to ∞. Since $\mathrm{E}\left(\dfrac{n}{n-1}\hat{\theta}\right) = \theta$, it follows that $\left(\frac{n-1}{n}\right)\left(\overline{X} - X_{(1)}\right)$ is an unbiased estimator of θ.

Simulation technique

The SAS code that can be used to simulate from the Two-parameter Exponential distribution is given in Appendix A.4. It follows from (4.3) that if a random variable X is Two-parameter Exponentially distributed with parameters θ and η, then $Y = X - \eta$ is Exponentially distributed with parameter θ. This relationship is used to simulate from the Two-parameter Exponential distribution by first simulating Y from an Exponential distribution and then calculating $X = Y + \eta$.

Accuracy of estimates

The maximum likelihood estimates can be determined directly from the sample mean and first (smallest) order statistic. The accuracy of the maximum likelihood estimates improve with an increasing sample size.

6.5.5 Erlang Distribution

Using the expression for the mean given in (B.28), the method-of-moments estimate for λ as given by (B.32) can be found.

The maximum likelihood estimator can be obtained from taking the first order derivative of the log-likelihood function with respect to λ. The likelihood function can be obtained using (B.25) from which the log-likelihood function can be found:

$$l(\lambda) = nN\ln(\lambda) - N\ln\left((n-1)!\right) + \sum_{j=1}^{N} \ln\left(x_j^{n-1}\right) - \lambda \sum_{j=1}^{N} x_j$$

where N is the number of samples considered and n is the size of each of the N samples. Setting the first order derivative of the log-likelihood function with respect to λ equal to 0 yields the maximum likelihood estimator given in (B.33), which is the same as the method-of-moments estimator.

Simulation technique

Consider the probability density function of the Gamma distribution function as given in (B.61). If we let parameters $\theta = \lambda^{-1}$ and $\kappa = n$, the probability density function of the Erlang distribution, as given in (B.25), is obtained. One can therefore simulate observations from the Erlang distributions using built-in function in SAS to simulate from a Gamma distribution. The SAS code that can be used is given in Appendix A.5.

Accuracy of estimates

As with the previous distributions discussed, the accuracy of this estimate improves with increasing sample sizes.

6.5.6 Generalized Extreme Value Distributions

In this section we introduce the likelihood function and log-likelihood function for the Generalized Extreme Value distribution which incorporates the Frechet, Gumbel and Weibull distributions. In Sections 6.5.7, 6.5.10 and

6.5.8 each distribution's specific estimation algorithms, simulation techniques and performance of the estimation algorithms are discussed.

Coles [31] expresses the generalized extreme value as follows:

$$\begin{aligned} F(x) &= e^{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}} \\ &= \begin{cases} e^{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}} & \text{if } \xi \neq 0 \\ e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}} & \text{if } \xi = 0. \end{cases} \end{aligned} \tag{6.13}$$

which is purely a reparameterization of the form given in (4.4) with $\xi = -k$, $\mu = \xi$ and $\sigma = \alpha$. Coles gives the log-likelihood functions for the cases where (1) $\xi \neq 0$ and (2) $\xi = 0$. Recall that the non-zero case relates to the Frechet (where $\xi > 0$) and Weibull (where $\xi < 0$) distributions and the zero case relates to the Gumbel distribution.

If $\xi \neq 0$:

$$\begin{aligned} l(\mu, \sigma, \xi) = &-m \ln(\sigma) - \sum_{i=1}^{m} \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}} \\ &- \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \ln \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right) \end{aligned} \tag{6.14}$$

and if $\xi = 0$:

$$l(\mu, \sigma) = -m \ln(\sigma) - \sum_{i=1}^{m} e^{-\frac{x_i - \mu}{\sigma}} - \sum_{i=1}^{m} \frac{x_i - \mu}{\sigma}. \tag{6.15}$$

The maximisation should be done using numerical techniques, since explicit formulae are not available for the maximum likelihood estimators for these parameters. The general expression for an iterative estimation process is given by (6.5).

Coles gives conditions in terms of the true underlying value of the parameter ξ to describe the nature of the estimates and likely these are to be found. We can also translate these in terms of parameter k as given in (4.4):

- When $\xi > -0.5$ (that is if $k < 0.5$):
  The maximum likelihood estimates are obtainable and exhibit the usual asymptotic results. This will therefore apply to the Frechet and Gumbel distributions and for the Weibull (only if $0 < k < 0.5$).

- When $-1 < \xi < -0.5$ (that is if $0.5 < k < 1$):
  The maximum likelihood estimates are obtainable, but don't have the standard asymptotic properties. Based on the range of values to which this applies, it can only be applicable to the Weibull distribution.

- When $\xi < -1$ (that is if $k > 1$):
  The maximum likelihood estimates are unlikely to be obtained. Again this can only be applicable to the Weibull distribution.

- When $\xi \leq -0.5$ (that is if $k \geq 0.5$):
  This corresponds to a distribution with a very short bounded upper tail which is rarely encountered. It should be noted the case for $k > 0$ refers to the Weibull distribution and in particular this is a form for minima. We are instead interested in the form for maxima as is discussed in Section 4.1.8.

The estimation of the parameters of the Frechet, Weibull and the Gumbel distributions are discussed in the following three sections using the parameterizations as discussed in Sections 4.1.7, 4.1.8 and 4.1.10. The links with the form as given by Coles are shown while the log-likelihood expressions given by Coles (see (6.14) and (6.15)) are utilised in deriving iterative estimation processes.

6.5.7 Frechet Distribution

Deriving the maximum likelihood estimation algorithm

Considering the form of the Frechet distribution as given by (4.14) and using the form as given by Coles in (6.13), it can be seen that (4.14) is a reparameterization of (6.13) with $\xi = \frac{1}{\beta}$, $\sigma = \frac{\delta}{\beta}$ and $\mu = \lambda + \delta$.

One can therefore obtain the log-likelihood function in terms of the λ, δ and β parameters from (6.14) as follows:

$$
\begin{aligned}
l(\lambda, \delta, \beta) &= l(\mu, \sigma, \xi)\big|_{\mu=\lambda+\delta, \sigma=\frac{\delta}{\beta}, \xi=\beta^{-1}} \\
&= m\ln(\beta) - \sum_{i=1}^{m}\left(\frac{\delta}{x_i - \lambda}\right)^{\beta} - (1+\beta)\sum_{i=1}^{m}\ln(x_i - \lambda) + \beta m\ln(\delta).
\end{aligned}
$$

Taking partial derivatives with respect to λ, δ and β and by then setting these expressions equal to 0 do not yield any explicit formulae for the maximum likelihood estimators of these parameters. As such the numerical procedure as given in (6.5) can be used. To implement this procedure we need the following vector of first order partial derivates and the matrix of second order partial derivatives (known as the Hessian matrix):

$$
\left(\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}}\right) = \begin{pmatrix} \frac{\partial l(\underline{\beta})}{\partial \lambda} \\ \frac{\partial l(\underline{\beta})}{\partial \delta} \\ \frac{\partial l(\underline{\beta})}{\partial \beta} \end{pmatrix}
$$

and

$$\left(\frac{\partial^2 l(\underline{\beta})}{\partial\underline{\beta}\partial\underline{\beta}^T}\right) = \begin{pmatrix} \frac{\partial^2 l(\underline{\beta})}{\partial\lambda^2} & \frac{\partial^2 l(\underline{\beta})}{\partial\delta\partial\lambda} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\lambda} \\ \frac{\partial^2 l(\underline{\beta})}{\partial\lambda\partial\delta} & \frac{\partial^2 l(\underline{\beta})}{\partial\delta^2} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\delta} \\ \frac{\partial^2 l(\underline{\beta})}{\partial\lambda\partial\beta} & \frac{\partial^2 l(\underline{\beta})}{\partial\delta\partial\beta} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta^2} \end{pmatrix}$$

where

$$\frac{\partial l(\underline{\beta})}{\partial\lambda} = -\frac{\beta}{\delta}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta+1} + (1+\beta)\sum_{i=1}^{m}\frac{1}{x_i-\lambda}$$

$$\frac{\partial l(\underline{\beta})}{\partial\delta} = -\beta\delta^{\beta-1}\sum_{i=1}^{m}(x_i-\lambda)^{-\beta} + \frac{\beta m}{\delta}$$

$$\frac{\partial l(\underline{\beta})}{\partial\beta} = \frac{m}{\beta} - \sum_{i=1}^{m}\ln\left(\frac{\delta}{x_i-\lambda}\right)\left(\left(\frac{\delta}{x_i-\lambda}\right)^{\beta} - 1\right)$$

and

$$\frac{\partial^2 l(\underline{\beta})}{\partial\lambda^2} = -\beta(\beta+1)\delta^{-2}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta+2} + (1+\beta)\sum_{i=1}^{m}\frac{1}{(x_i-\lambda)^2}$$

$$\begin{aligned}\frac{\partial^2 l(\underline{\beta})}{\partial\delta\partial\lambda} = \; & -\delta^{-1}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta+1} - \beta\delta^{-1}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta+1}\ln\left(\frac{\delta}{x_i-\lambda}\right) \\ & + \sum_{i=1}^{m}\frac{1}{x_i-\lambda}\end{aligned}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\lambda} = -\beta^2\delta^{-2}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta+1}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\delta^2} = -\beta(\beta-1)\delta^{-2}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta} - \frac{m\beta}{\delta^2}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\delta} = \frac{m}{\delta} - \delta^{-1}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta} - \beta\delta^{-1}\sum_{i=1}^{m}\left(\frac{\delta}{x_i-\lambda}\right)^{\beta}\ln\left(\frac{\delta}{x_i-\lambda}\right)$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial \beta^2} = \frac{-m}{\beta^2} - \sum_{i=1}^{m} \left(\frac{\delta}{x_i - \lambda} \right)^{\beta} \left(\ln \left(\frac{\delta}{x_i - \lambda} \right) \right)^2$$

These expressions can also be obtained using first order and second order partial derivatives given in terms of an alternative parameterization given by Otten and Van Montfort [94].

Simulation technique

This distribution can be simulated using the probability integral transformation where the random from the Frechet distribution X can be obtained from a simulated random uniform distributed variable U as follows:

$$X = \delta \left(-\ln(U) \right)^{-\frac{1}{\beta}} + \lambda$$

The SAS code to perform this simulation is given in Appendix A.6.

Performance of estimation algorithm

The SAS code to perform the maximum likelihood estimation is included in Appendix A.6. The iterative procedure reaches convergence, but is very sensitive to the choice of initial values. If initial values are not well set, convergence is often not obtained whilst the procedure breaks in some instances due to the arguments of the natural log functions becoming negative. If the initial value of λ is not very close to the true value, convergence is difficult to obtain. In order to choose an initial value for λ that is more likely to ensure convergence, we considered the fact that theoretically the there exists a relationship between the expected value of X and the parameter values as given in (B.58). Based on this relationship, a natural choice for the initial value of λ is as follows:

$$\lambda_0 = \overline{x} - \delta_0 \Gamma \left(1 - \frac{1}{\beta_0} \right)$$

where λ_0, δ_0 and β_0 denote the initial values of λ, δ and β. Using this approach to set the initial value for λ appeared to improve the likelihood of convergence of the iterative procedure.

By varying the values of the parameters of the distribution from which the observations were simulated, it was seen that the greater the value of λ relative to the values of β and δ, the more accurate the maximum likelihood estimate of λ whilst the maximum likelihood estimates of β and δ are often less accurate.

Due to the fact that convergence of the iterative procedure may be difficult to obtain together with the fact that the degree of accuracy of the maximum likelihood estimates may not always be very high, it is very important to consider measures such as quantile plots to assess the goodness-of-fit of the fitted distribution. The use of quantile plots to assess goodness-of-fit is discussed in Section 7.2.2.

6.5.8 Weibull Distribution

Deriving the maximum likelihood estimation algorithm

Considering the form of the Weibull distribution as given by (4.16) and using the form as given by Coles in (6.13), it can be seen that (4.16) is a reparameterization of (6.13) with $\xi = \frac{-1}{\beta}$, $\sigma = \frac{\delta}{\beta}$ and $\mu = \lambda - \delta$.

From Klugman et al [75] follow that this form of the Weibull distribution only has support for values of $x < \lambda$. This distribution is therefore suitable for modeling minima of distributions.

If $X_{(1)}, X_{(2)}, X_{(3)}, ..., X_{(n)}$ are the order statistics of an observed sample $(X_1, X_2, X_3, ..., X_n)$ where $X_1, X_2, X_3, ..., X_n$ are identical and independently distributed. If we let $Y_i = -X_i$, then we have that $Y_{(1)} = -X_{(n)}$ is the smallest order statistic of $Y_1, Y_2, Y_3, ..., Y_n$. Hence

$$\mathrm{P}(X_{(n)} \leq x) = \mathrm{P}(-X_{(n)} > x) = 1 - \mathrm{P}(-X_{(n)} \leq x) = 1 - \mathrm{P}(Y_{(1)} \leq x)$$

This relationship can be used in conjunction with the log-likelihood function as given by Coles for minima [31] in order to get the log-likelihood function for maxima.

Denote $F(x) = \mathrm{P}(X_{(1)} \leq x)$ and $H(x) = \mathrm{P}(X_{(n)} \leq x)$.. Furthermore, if we can find the relationship between the likelihood function for maxima and the likelihood function for minima, we can use the relationship to get the log-likelihood function for maxima using the function as given by Coles in (6.14).

$$L(\lambda, \delta, \beta) = \prod_{i=1}^{m} h(x_i) = \prod_{i=1}^{m} \frac{d}{dx_i} H(x_i) = \prod_{i=1}^{m} \frac{d}{dx_i}(1 - F(-x_i)) = \prod_{i=1}^{m} f(-x_i)$$

but $f(-x_i)$ is associated with $F(-x_i)$ where $F(x)$ is a special case of the Generalized Extreme Value distribution as given by (6.13) with $x = -x$, $\xi = -\beta^{-1}$, $\sigma = \delta\beta^{-1}$ and $\mu = \lambda - \delta$. This enables us to use the likelihood function given by Coles in (6.14) as:

$$\begin{aligned} l(\lambda,\delta,\beta) &= l(\mu,\delta,\xi;\underline{x})|_{x_i=-x_i,\xi=-\beta^{-1},\sigma=\delta\beta^{-1},\mu=\lambda-\delta} \\ &= -m\beta\ln(\delta)+m\ln(\beta)-\sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^{\beta}+(\beta-1)\sum_{i=1}^{m}\ln(x_i+\lambda). \end{aligned} \tag{6.16}$$

Taking partial derivatives with respect to λ, δ and β and by then setting these expressions equal to 0 do not yield any explicit formulae for the maximum likelihood estimators of these parameters. Thus the numerical procedure as given in (6.5) can be used. To implement this procedure we need the following vector of first order partial derivates and the Hessian matrix:

$$\left(\frac{\partial l(\underline{\beta})}{\partial\underline{\beta}}\right)=\begin{pmatrix}\frac{\partial l(\underline{\beta})}{\partial\lambda}\\ \frac{\partial l(\underline{\beta})}{\partial\delta}\\ \frac{\partial l(\underline{\beta})}{\partial\beta}\end{pmatrix}$$

and

$$\left(\frac{\partial^2 l(\underline{\beta})}{\partial\underline{\beta}\partial\underline{\beta}^T}\right)=\begin{pmatrix}\frac{\partial^2 l(\underline{\beta})}{\partial\lambda^2} & \frac{\partial^2 l(\underline{\beta})}{\partial\delta\partial\lambda} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\lambda}\\ \frac{\partial^2 l(\underline{\beta})}{\partial\lambda\partial\delta} & \frac{\partial^2 l(\underline{\beta})}{\partial\delta^2} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\delta}\\ \frac{\partial^2 l(\underline{\beta})}{\partial\lambda\partial\beta} & \frac{\partial^2 l(\underline{\beta})}{\partial\delta\partial\beta} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta^2}\end{pmatrix}$$

where

$$\frac{\partial l(\underline{\beta})}{\partial\lambda}=-\frac{\beta}{\delta}\sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^{\beta-1}+(\beta-1)\sum_{i=1}^{m}\frac{1}{x_i+\lambda}$$

$$\frac{\partial l(\underline{\beta})}{\partial\delta}=\frac{\beta}{\delta}\sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^{\beta}-\frac{m\beta}{\delta}$$

$$\frac{\partial l(\underline{\beta})}{\partial\beta}=-\sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^{\beta}\ln\left(\frac{x_i+\lambda}{\delta}\right)+\sum_{i=1}^{m}\ln(x_i+\lambda)-m\ln(\delta)+\frac{m}{\beta}$$

and

$$\frac{\partial^2 l(\underline{\beta})}{\partial\lambda^2}=-\frac{\beta(\beta-1)}{\delta^2}\sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^{\beta-2}-(\beta-1)\sum_{i=1}^{m}(x_i+\lambda)^{-1}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\delta\partial\lambda} = \left(\frac{\beta}{\delta}\right)^2 \sum_{i=1}^{m} \left(\frac{x_i+\lambda}{\delta}\right)^{\beta-1}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\lambda} = -\frac{1}{\delta}\sum_{i=1}^{m}\left(1+\beta\ln(\frac{x_i+\lambda}{\delta})\right)\left(\frac{x_i+\lambda}{\delta}\right)^{\beta-1} + \sum_{i=1}^{m}\frac{1}{x_i+\lambda}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\delta^2} = \frac{\beta}{\delta^2}\left(m - \frac{\beta+1}{\delta^\beta}\sum_{i=1}^{m}(x_i+\lambda)^\beta\right)$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\delta} = -\frac{m}{\delta} + \frac{1}{\delta}\sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^\beta\left(1+\beta\ln\left(\frac{x_i+\lambda}{\delta}\right)\right)$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial\beta^2} = -\frac{m}{\beta^2} - \sum_{i=1}^{m}\left(\frac{x_i+\lambda}{\delta}\right)^\beta\left(\ln\left(\frac{x_i+\lambda}{\delta}\right)\right)^2$$

If we let $\tilde{x} = x + \lambda$ and $\delta = \theta$, we have the cumulative distribution function for the two-parameter version of the Weibull distribution as given in (4.16). In this case the log-likelihood function given in (6.16) simplifies to

$$l(\theta,\beta) = -m\beta\ln(\theta) + m\ln(\beta) - \sum_{i=1}^{m}\left(\frac{\tilde{x}_i}{\theta}\right)^\beta + (\beta-1)\sum_{i=1}^{m}\ln(\tilde{x}_i). \quad (6.17)$$

Similar to the three parameter case of the Weibull distribution, maximization of this likelihood function requires numerical procedures. In order to apply the multivariate iterative procedure given in (6.5), the following is required:

$$\frac{\partial l(\underline{\beta})}{\partial\underline{\beta}} = \begin{pmatrix} \frac{\partial l(\underline{\beta})}{\partial\theta} \\ \frac{\partial l(\underline{\beta})}{\partial\beta} \end{pmatrix}$$

and

$$\frac{\partial l(\underline{\beta})}{\partial\underline{\beta}} = \begin{pmatrix} \frac{\partial^2 l(\underline{\beta})}{\partial\theta^2} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta\partial\theta} \\ \frac{\partial^2 l(\underline{\beta})}{\partial\theta\partial\beta} & \frac{\partial^2 l(\underline{\beta})}{\partial\beta^2} \end{pmatrix}$$

where

$$\frac{\partial l(\underline{\beta})}{\partial \theta} = -\frac{m\beta}{\theta} + \frac{\beta}{\theta}\sum_{i=1}^{m}\left(\frac{x_i}{\theta}\right)^{\beta}$$

$$\frac{\partial l(\underline{\beta})}{\partial \beta} = \frac{m}{\beta} + (\beta - 1)\sum_{i=1}^{m}\ln(x_i) - m\ln(\theta) - \sum_{i=1}^{m}\left(\frac{x_i}{\theta}\right)^{\beta}\ln\left(\frac{x_i}{\theta}\right)$$

and

$$\frac{\partial^2 l(\underline{\beta})}{\partial \theta^2} = \frac{m\beta}{\theta^2} - \beta(\beta+1)\theta^{-(\beta+2)}\sum_{i=1}^{m}x_i^{\beta}$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial \theta \partial \beta} = -\frac{m}{\theta} + \frac{1}{\theta}\sum_{i=1}^{m}\left(\frac{x_i}{\theta}\right)^{\beta}\left(1 + \beta\ln\left(\frac{x_i}{\theta}\right)\right)$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial \beta^2} = -\frac{m}{\beta^2} + \sum_{i=1}^{m}\ln(x_i) - \sum_{i=1}^{m}\left(\frac{x_i}{\theta}\right)^{\beta}\left(\ln\left(\frac{x_i}{\theta}\right)\right)^2$$

Simulation technique

The probability integral transformation is used to simulate random variables X from this distribution as follows:

$$X = \delta\left(-\ln(1-U)\right)^{\frac{1}{\beta}} - \lambda$$

where U is a random uniform distributed variable. The SAS code given in Appendix A.7.

Performance of estimation algorithm

The iterative procedure described above that can be used to find the maximum likelihood estimates for the three-parameter version of the Weibull distribution appears to have poor convergence properties. This may be attributed to the structural form of the Weibull distribution where we have that $x \geq \lambda$ which suggests that the true value of λ is some value close to the first (smallest) order statistic. Initial values of λ above its true values is likely to be too far away from the true value in which case the iterative procedure moves even further away from this true value and consequently of the true values of β and δ as well.

Since λ is some value close to the smallest observed order statistic, it may be reasonable to have that $\hat{\lambda} = X_{(1)}$ in which case the data can then be adjusted for this threshold value as follows:

$$\tilde{X} = X - \hat{\lambda} = X - X_{(1)}$$

in which the iterative procedure as given for the two parameter version of the Weibull distribution may be used. This approach is followed in the SAS code given in Appendix A.7. An estimate for λ is obtained by setting it equal to the first order statistic and then adjusting the observed values by subtracting the first order statistic from all observed values.

The SAS code in Appendix A.7 also includes a section where a regression approach is followed in fitting the Weibull distribution. If the true underlying distribution of a set of observations is the Weibull distribution, then the empirical cumulative distribution function would be a reasonable representation of the theoretical distribution. After adjusting the data based on the observed first order statistic, the problem of fitting a two-parameter Weibull distribution can be translated to a regression problem as follows [105]:

$$\ln(\tilde{X}) = a + b\ln\Big(-\ln(1 - \hat{F}_X(x))\Big) \tag{6.18}$$

where

$$\hat{F}_X(x) = \frac{1}{n}\{\text{Number of observations} \leq x\}$$

This is the result of letting $F_X(x) = \hat{F}_X(x)$, then:

$$\hat{F}_X(x) = 1 - e^{-\left(\frac{x}{\theta}\right)^{\beta}} \text{ from (4.16)}$$

hence

$$\ln(X) = \ln(\theta) + \left(\frac{1}{\beta}\right)\ln(1 - \hat{F}_X(x))$$

This means that once we have obtained the estimates of a and b in (6.18), we can obtain the estimates of θ (or δ for the three-parameter case) as well as β as follows:

$$\hat{\theta} = \hat{\delta} = e^{a} \text{ and } \hat{\beta} = \frac{1}{b}$$

The answers coming from the two approaches are different. In some instances the regression approach answers are more accurate. It may therefore be useful to consider the regression approach prior to the maximum likelihood estimation in order to use these estimates as initial values in the iterative procedure.

6.5.9 Rayleigh Distribution

Deriving method-of-moments estimators and maximum likelihood estimation algorithm

The Rayleigh distribution is a special case of the two parameter Weibull distribtion with $\beta = 2$. This means that to fit this distribution requires only the estimation of a single parameter. The method-of-moments estimator, given in (B.151), can be obtained using the expression for the mean in (B.147) and setting is equal to $\overline{x}$. This is an unbiased estimator of θ, since $\mathrm{E}(\hat{\theta}) = \theta$.

Maximum likelihood can also be used in which case we can leverage off the log-likelihood function for the two parameter version of the Weibull distribution, given in (6.17), with $\beta = 2$. Setting the first derivative equal to 0 yields:

$$\theta^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

We have that $\mathrm{E}(\hat{\theta^2}) = \left(\frac{4+\pi}{4}\right)\theta^2$, hence

$$\tau(\theta) = \frac{4}{4+\pi}\hat{\theta^2}$$

is an unbiased estimator of θ^2. Using the invariance property [1] one can find the unbiased estimator of θ as given in (B.152).

Since the Rayleigh distribution is a special case of the Weibull distribution, it makes sense to rather consider fitting the Weibull distribution first. If the true underlying distribution is in fact Rayleigh (suggested by an estimate for β close to 2), one can proceed by calculating the maximum likelihood estimate of θ specifically for the Rayleigh distribution as given above.

Simulation technique

Simulation from the Rayleigh distribution can be done using the probability integral transformation as follows:

$$X = \theta\sqrt{-\ln(1-U)}$$

where U is a random uniform distributed variable.

[1] If $\hat{\theta}$ is a maximum likelihood estimator of θ and $u(\theta)$ is a function of θ, then $u(\hat{\theta})$ is a maximum likelihood estimator of $u(\theta)$ [11].

Accuracy of estimators

SAS code that can be used to perform the calculation of the method-of-moments and maximum likelihood estimates is given in Appendix A.8. The accuracy of these estimates increases with increasing sample sizes.

6.5.10 Gumbel Distribution

Deriving the maximum likelihood estimation algorithm

The log-likelihood function in (6.15) for the Gumbel distribution is in terms of the parameterization of the distribution function as given by Coles [31]. By recognising that Coles's form is purely a reparameterization of the form given in (4.4) where $k = 0$ with $\mu = \xi$ and $\sigma = \alpha$, the log-likelihood function can be written as:

$$l(\xi, \alpha) = l(\mu, \sigma)|_{\mu=\xi,\sigma=\alpha} = -m\ln(\alpha) - \sum_{i=1}^{m} e^{-\frac{x_i-\xi}{\alpha}} - \frac{1}{\alpha}\sum_{i=1}^{m}(x_i - \xi).$$

By taking the first derivatives with respect to α and ξ and setting these equal to 0 no explicit expressions in terms of α or ξ can be obtained. For this reason a numerical procedure such as the iterative process given by (6.5) can be implemented in which case the vector of first derivatives

$$\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}} = \begin{pmatrix} \frac{\partial l(\xi,\alpha)}{\partial \xi} \\ \frac{\partial l(\xi,\alpha)}{\partial \alpha} \end{pmatrix}$$

and the Hessian matrix

$$\frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T} = \begin{pmatrix} \frac{\partial^2 l(\xi,\alpha)}{\partial^2 \xi^2} & \frac{\partial^2 l(\xi,\alpha)}{\partial \alpha \partial \xi} \\ \frac{\partial^2 l(\xi,\alpha)}{\partial \xi \partial \alpha} & \frac{\partial^2 l(\xi,\alpha)}{\partial^2 \alpha^2} \end{pmatrix}$$

need to be calculated, where

$$\frac{\partial l(\xi, \alpha)}{\partial \xi} = \frac{m}{\alpha} - \frac{1}{\alpha}\sum_{i=1}^{m} e^{-\frac{x_i-\xi}{\alpha}}$$

and

$$\frac{\partial l(\xi, \alpha)}{\partial \alpha} = -\frac{m}{\alpha} + \frac{1}{\alpha^2}\sum_{i=1}^{m}(x_i - \xi)\left(1 - e^{-\frac{x_i-\xi}{\alpha}}\right)$$

and

$$\frac{\partial^2 l(\xi,\alpha)}{\partial \xi^2} = -\frac{1}{\alpha^2}\sum_{i=1}^{m} e^{-\frac{x_i-\xi}{\alpha}}$$

$$\frac{\partial^2 l(\xi,\alpha)}{\partial \alpha \partial \xi} = \frac{1}{\alpha^2}\sum_{i=1}^{m}\left(e^{-\frac{x_i-\xi}{\alpha}} - \frac{1}{\alpha}(x_i-\xi)e^{-\frac{x_i-\xi}{\alpha}} - 1\right)$$

and

$$\frac{\partial^2 l(\xi,\alpha)}{\partial \alpha^2} = \frac{m}{\alpha^2} - \frac{2}{\alpha^3}\sum_{i=1}^{m}(x_i-\xi)\left(1-e^{-\frac{x_i-\xi}{\alpha}}\right) + \frac{1}{\alpha^2}\sum_{i=1}^{m}(x_i-\xi)\left(-e^{-\frac{x_i-\xi}{\alpha}}\right)\left(\frac{x_i-\xi}{\alpha^2}\right)$$

Simulation technique

To simulate from the Gumbel distribution one can use the probability integral transformation [11] as follows:

$$X = -\alpha \ln(-\ln(U)) + \xi$$

where U is the simulated random uniform variable. The SAS code for this simulation is shown in Appendix A.9.

Performance of estimation algorithm

Generally it is seen that with increasing sample sizes the algorithm converges faster with improved accuracy. What can also be seen is that the larger the value of the α parameter relative to the ξ parameter, the less accurate and less stable the estimate of the ξ parameter.

Convergence of the maximum likelihood estimation algorithm tends to be highly affected by the choice of the initial values of α and ξ. For this reason we determine initial values using a regression approach where we set the theoretical expression for the cumulative distribution function equal to the empirical distribution function $\hat{F}_X(x)$, rewrite the expression with X as linear a function of $\hat{F}_X(x)$ as follows:

$$X = \xi + \alpha\left(-\ln\left(-\ln\left(\hat{F}_X(x)\right)\right)\right)$$

from which the intercept and coefficient can then be solved and serve as initial values for ξ and α, respectively. This approach of finding initial values is also shown and incorporated in the SAS code given in Appendix A.9.

6.5.11 Pareto Distribution

Deriving the method-of-moments estimators and maximum likelihood estimation algorithm

Method-of-moments estimation can be performed using the expressions for the mean and variance for the Pareto Type II distribution as given in (B.141) and (B.142). Setting $\mathrm{E}(X) = \overline{x}$ and $\mathrm{var}(X) = s_x^2$ yields the estimators for κ and θ that are given in (B.144) and (B.145), respectively.

Estimates for κ and θ can also be obtained using maximum likelihood estimation. The likelihood function cen be obtained using the probability density function, given in (B.139), as:

$$L(\kappa, \theta) = \kappa^n \theta^{n\kappa} \prod_{j=1}^{n} (x_j + \theta)^{-(\kappa+1)}$$

from which the log-likelihood function follows:

$$n \ln(\kappa) + n\kappa \ln(\theta) - (\kappa + 1) \sum_{j=1}^{n} \ln(x_j + \theta).$$

Setting the log-likelihood function's first order partial derivatives with respect to κ and θ does not yield explixit expressions in terms of κ and θ. The maximum likelihood estimates therefore need to be obtained using the Netwon-Raphson numerical optimization technique as used in previous sections for which the vector of first order derivatives:

$$\frac{\partial l}{\partial \kappa} = \frac{n}{\kappa} + n \ln(\theta) - \sum_{j=1}^{n} \ln(x_j + \theta) \text{ and } \frac{\partial l}{\partial \theta} = \frac{n\kappa}{\theta} - (\kappa + 1) \sum_{j=1}^{n} (x_j + \theta)^{-1}$$

is required as well as the Hessian matrix for which the entries of second order partial derivatives are given by:

$$\begin{aligned}
\frac{\partial^2 l}{\partial \kappa^2} &= -\frac{n}{\kappa^2}, \\
\frac{\partial^2 l}{\partial \theta^2} &= -\frac{n\kappa}{\theta^2} + (\kappa + 1) \sum_{j=1}^{n} (x_j + \theta)^{-2}, \text{ and} \\
\frac{\partial^2 l}{\partial \kappa \partial \theta} &= \frac{n}{\theta} - \sum_{j=1}^{n} (x_j + \theta)
\end{aligned}$$

Simulation technique

To simulate from the Pareto distribution the probability integral transformation can be used [11] by letting the cumulative density function of the Pareto distribution (as given in (4.25) with $a = \theta$ and $k = \kappa$) equal to the random uniform random variable and then solving for X. This means that we obtain a simulated value for X as follows:

$$X = \theta \left(1 - U\right)^{-\frac{1}{\kappa}} - \theta$$

where U is the simulated random uniform variable. The SAS code for this simulation is shown in Appendix A.10.

Performance of the estimation algorithm

The method-of-moments estimates for θ and κ can be used as initial values for the iterative alogorithm. The algorithm appears to have good convergence properties if the method-of-moments estimates are used as initial values whilst increasing sample size improves the accuracy of these estimates.

6.5.12 Generalized Pareto Distribution

Deriving the maximum likelihood estimation algorithm

The likelihood function can then obtained from (B.74) from which the log-likelihood follows:

$$\begin{aligned} l(\kappa, \tau, \theta) &= n \ln(\Gamma(\kappa + \tau)) - n \ln(\Gamma(\kappa)) - n \ln(\Gamma(\tau)) + n\kappa \ln(\theta) \\ &\quad + (\tau - 1) \sum_{j=1}^{n} \ln(x_j) - (\kappa + \tau) \sum_{j=1}^{n} \ln(x_j + \theta). \end{aligned}$$

To maximize the log-likelihood function requires the use of the numerical technical technique with updating iterative updating formula given in (6.5). Hence we need to obtain expressions for the first and second order derivatives with respect to κ, τ and θ.

$$\begin{aligned} \frac{\partial l}{\partial \kappa} &= n\Psi(\kappa + \tau) - n\Psi(\kappa) + \sum_{j=1}^{n} \ln\left(\frac{\theta}{x_j + \theta}\right) \\ \frac{\partial l}{\partial \tau} &= n\Psi(\kappa + \tau) - n\Psi(\tau) + \sum_{j=1}^{n} \ln\left(\frac{x_j}{x_j + \theta}\right) \\ \frac{\partial l}{\partial \theta} &= \frac{n\kappa}{\theta} - (\kappa + \tau) \sum_{j=1}^{n} (x_j + \theta)^{-1} \end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \kappa^2} &= n\Psi^{'}(\kappa+\tau) - n\Psi^{'}(\kappa) \\
\frac{\partial^2 l}{\partial \tau \partial \kappa} &= n\Psi^{'}(\kappa+\tau) \\
\frac{\partial^2 l}{\partial \theta \partial \kappa} &= \frac{n}{\theta} - \sum_{j=1}^{n}(x_j+\theta)^{-1} \\
\frac{\partial^2 l}{\partial \tau^2} &= n\Psi^{'}(\kappa+\tau) - n\Psi^{'}(\tau) \\
\frac{\partial^2 l}{\partial \theta \partial \tau} &= -\sum_{j=1}^{n}(x_j+\theta)^{-1} \\
\frac{\partial^2 l}{\partial \theta^2} &= -\frac{n\kappa}{\theta^2} + (\kappa+\tau)\sum_{j=1}^{n}(x_j+\theta)^{-2}.
\end{aligned}
$$

Simulation technique

In order to simulate from this distribution is not straight forward, since random number generation functions for this particular distribution generally don't exist in software packages. Furthermore, a closed form expression for the cumulative distribution function is not available which means that we can't use the probability integral transformation. A solution is proposed here which is based on the relationship that exists between the Generalized Pareto distribution and the Compound Gamma distribution [45].

Consider a conditional random variable $X|\beta$ that is Gamma distributed with parameters α and β^{-1}. The probability density function is given by (B.61) where $\theta = \beta^{-1}$ and $\kappa = \alpha$. Furthermore, suppose β is again Gamma distributed with parameters γ and δ. The probability density function is given by (B.61) where $\theta = \delta^{-1}$ and $\kappa = \gamma$. The unconditional probability density funtion of X is then given by:

$$f_X(x) = \frac{1}{\left(\frac{1}{\delta}\right)^{\gamma} B(\alpha,\gamma)} \left(\frac{x}{x+\delta}\right)^{\alpha-1} \left(\frac{1}{x+\delta}\right)^{\gamma+1} \tag{6.19}$$

which is the probability density function of a Compound Gamma distribution [45]. If we now let $\delta = \theta$, $\gamma = \kappa$ and $\alpha = \tau$, then the probability density function given in (6.19) simplifies to the probability density function for the Generalized Pareto distribution as given in (B.74). It can therefore be seen that the Generalized Paretro distribution can be viewed as a reparameterization of the Compound Gamma distribution. Hence, to simulate from a Generalized Pareto distribution, the following steps can be followed:

1. Simulate a random β from a $\text{GAM}\left(\kappa, \frac{1}{\theta}\right)$ distribution.
2. Simulate a random X from a $\text{GAM}\left(\tau, \frac{1}{\beta}\right)$ distribution, where the β is the value as simulated in the previous step.
3. Repeat the previous two steps n times to have a simulated sample of n observations from a Generalized Pareto distribution with parameters θ, κ and τ.

The SAS code that can be used to perform the simulation as described above is given in Appendix A.11.

Performance of estimation algorithm

Generally the algorithm yields convergence given that suitable initial values are chosen. The accuracy of these estimates improve with increasing sample sizes.

6.5.13 Lognormal Distribution

Deriving the maximum likelihood estimators

Using the probability density function as given in (B.126) for the Lognormal distribution, the likelihood function can be obtained from which the log-likelihood function follows as given by:

$$l(\mu, \sigma) = -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \sum_{j=1}^{n} \ln(x_j) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} \left(\ln(x_j) - \mu\right)^2$$

After taking the first order partial derivatives with respect to μ and σ and setting these derivatives equal to 0, we get the maximum likelihood estimators that are given in (B.131) and (B.132).

It follows from Result 2 in Section 4.1.13 that if $X \sim LN(\mu, \sigma^2)$, then $Y = \ln(X) \sim N(\mu, \sigma)$ and using this relationship enables one to determine the expected values of the maximum likelihood estimators of the Lognormal distribution:

$$\text{E}(\hat{\mu}_{MLE}) = \frac{1}{n} \sum_{j=1}^{n} \text{E}(\ln(X_j)) = \frac{1}{n} \sum_{j=1}^{n} \text{E}(Y_j) = \frac{1}{n} \sum_{j=1}^{n} \mu = \mu$$

and

$$\begin{aligned} \text{E}(\hat{\sigma}^2_{MLE}) &= \frac{1}{n} \sum_{j=1}^{n} \text{E}\left(\left(\ln(X_j) - \frac{1}{n} \sum_{j=1}^{n} \ln(x_j)\right)^2\right) = \frac{1}{n} \sum_{j=1}^{n} \text{E}\left((Y_j - \overline{Y})^2\right) \\ &= \sigma^2 \end{aligned}$$

This gives that $\hat{\mu}_{MLE}$ and $\hat{\sigma}_{MLE}$ are unbiased estimators for μ and σ.

Simulation technique

The fact that there exists a relationship between the Lognormal and the Normal distribution makes it possible to simulate Lognormal random variables from simulated Normal random variables. This is shown in the SAS code given in Appendix A.12.

Accuracy of estimates

The maximum likelihood estimates of the parameters for the Lognormal distribution can be calculated directly from sample data with the expressions given in (B.131) and (B.132). The accuracy of these estimates improve with increasing sample size, but with sample sizes as small as 500 these estimates are already very close to the true values when observations are simulated from the distribution.

6.5.14 Beta-prime Distribution

Deriving the maximum likelihood estimation algorithm

The likelihood can be obtained using (B.1) from which the log-likelihood function follows as given here:

$$\begin{aligned} l(\delta_1, \delta_2) &= n\ln(\Gamma(\delta_1+\delta_2)) - n\ln(\Gamma(\delta_1)) - n\ln(\Gamma(\delta_2)) \\ &\quad + (\delta_1 - 1)\sum_{j=1}^{n}\ln(x_j) - (\delta_1+\delta_2)\sum_{j=1}^{n}\ln(x_j+1) \end{aligned}$$

respectively. To optimize this log-likelihood using numerical methods, requires the first order and second order partial derivatives:

$$\begin{aligned} \frac{\partial l}{\partial \delta_1} &= n\Psi(\delta_1+\delta_2) - n\Psi(\delta_1) + \sum_{j=1}^{n}\ln\left(\frac{x_j}{x_j+1}\right) \\ \frac{\partial l}{\partial \delta_2} &= n\Psi(\delta_1+\delta_2) - n\Psi(\delta_2) - \sum_{j=2}^{n}\ln(x_j+1). \end{aligned}$$

The second order derivatives are given below:

$$\frac{\partial^2 l}{\partial \delta_1^2} = n\Psi^{'}(\delta_1+\delta_2) - n\Psi^{'}(\delta_1)$$
$$\frac{\partial^2 l}{\partial \delta_1 \partial \delta_2} = n\Psi^{'}(\delta_1+\delta_2)$$
$$\frac{\partial^2 l}{\partial \delta_2^2} = n\Psi^{'}(\delta_1+\delta_2) - n\Psi^{'}(\delta_2).$$

Simulation technique

Similar to Generalized Pareto distribution, the simulation from the Beta-prime distribution is more complex, since built-in functions generally don't exist in software packages whilst the probability integral transformation cannot be used if a closed form expression for the cumulative distribution function is not available.

If we consider the probability density function of the Compound Gamma distribution [45] as given in (6.19) and let $\delta = 1$, $\gamma = \delta_2$ and $\alpha = \delta_1$, then (6.19) reduces to the probability density function for the Beta-prime distribution as given in (B.1). Hence we can use the same simulation technique as derived for the Generalized Pareto distribution. In terms of the Beta-prime distribution this will entail the following:

1. Simulate β from a $GAM(\delta_2, 1)$ distribution.
2. Simulate an X from a $GAM\left(\delta_1, \frac{1}{\beta}\right)$ distribution.
3. Repeat n times to obtain a simulated sample with n observations.

The SAS code to simulate observations from a Beta-prime distribution using this algorithm is given in Appendix A.13.

Performance of estimation algorithm

In some instances this algorithm does not converge, especially for samples with observed values that are larger in magnitude. Since this distribution is a special case of the Generalized Pareto distribution, one can start off by first considering to fit the Generalized Pareto after which parameter estimates' values can be assessed to determine if a Beta-prime distribution is in fact implied.

6.5.15 Birnbaum-Saunders Distribution

Deriving the maximum likelihood estimation algorithm

The probability density function, given in (B.6), can be used to derive the likelihood and log-likelihood functions. These functions are given by Lemonte et al [78]. The log-likelihood function proposed is given by:

$$l(\alpha,\beta) \propto -n\ln(\alpha) - \frac{1}{2}\sum_{j=1}^{n}\alpha^{-2}\left(\sqrt{\frac{t_j}{\beta}} - \sqrt{\frac{\beta}{t_j}}\right)^2 + \sum_{j=1}^{n}\ln(t_j^{-\frac{1}{2}}\beta^{-\frac{1}{2}} + \beta^{\frac{1}{2}}t_j^{-\frac{3}{2}}).$$

The optimization of this function needs to be done numerically. We will consider using the iterative approach as given in (6.5) for which the first and second order partial derivatives are required. The first order partial derviatives are given by:

$$\begin{aligned}
\frac{\partial l(\alpha,\beta)}{\partial\alpha} &= -\frac{n}{\alpha} + \frac{1}{\alpha^3}\sum_{j=1}^{n}\left(\sqrt{\frac{t_j}{\beta}} - \sqrt{\frac{\beta}{t_j}}\right)^2 \\
\frac{\partial l(\alpha,\beta)}{\partial\beta} &= -\frac{1}{2\alpha^2}\sum_{j=1}^{n}\left(\frac{1}{t_j} - \frac{t_j}{\beta^2}\right) + \frac{1}{2\beta}\sum_{j=1}^{n}\frac{\beta - t_j}{\beta + t_j}.
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 l(\alpha,\beta)}{\partial\alpha^2} &= \frac{n}{\alpha^2} - \frac{3}{\alpha^4}\sum_{j=1}^{n}\left(\sqrt{\frac{t_j}{\beta}} - \sqrt{\frac{\beta}{t_j}}\right)^2 \\
\frac{\partial^2 l(\alpha,\beta)}{\partial\beta^2} &= -\frac{1}{\alpha^2}\sum_{j=1}^{n}\frac{t_j}{\beta^3} - \frac{1}{2\beta^2}\sum_{j=1}^{n}\left(\frac{\beta - t_j}{\beta + t_j}\right) + \frac{1}{\beta}\sum_{j=1}^{n}\frac{t_j}{(\beta + t_j)^2} \\
&\text{and} \\
\frac{\partial^2 l}{\partial\alpha\partial\beta} &= \frac{1}{\alpha^3}\sum_{j=1}^{n}\left(\frac{1}{t_j} - \frac{t_j}{\beta^2}\right).
\end{aligned}$$

Simulation technique

The relationship between the Normal distribution and the Birnbaum-Saunders distribution as given in (4.30) can be used to simulate random variables from the Birnbaum-Saunders distribution. The SAS code utilising this simulation technique is given in Appendix A.14.

Performance of estimation algorithm

The iterative algorithm appears to easily yield convergence and not to be too sensitive to the choice of the initial values of α and β.

6.5.16 Burr and related distributions

In this section the maximum likelihood estimation of the parameters and the simulation of the Burr and Singh-Maddala dsitributions are discussed.

Deriving the maximum likelihood estimation algorithm

Using the probability density function as given in (4.32) the likelihood function can be derived in order to estimate the distribution parameters using maximum likelihood estimation. From this function the log-likelihood function follows:

$$\begin{aligned} l(k,c,\alpha) = n\ln(k) + n\ln(c) - n\ln(\alpha) - (k+1)\sum_{j=1}^{n}\ln\left(1+\left(\frac{x_j}{\alpha}\right)^c\right) \\ + (c-1)\sum_{j=1}^{n}\ln\left(\frac{x_j}{\alpha}\right). \end{aligned}$$

The log-likelihood function can be optimized numerically using (6.5). The first order partial derivatives with respect to k, c and α are given by:

$$\begin{aligned} \frac{\partial l}{\partial k} &= \frac{n}{k} - \sum_{j=1}^{n}\ln\left(1+\left(\frac{x_j}{\alpha}\right)^c\right) \\ \frac{\partial l}{\partial c} &= \frac{n}{c} + \sum_{j=1}^{n}\ln\left(\frac{x_j}{\alpha}\right) - (k+1)\sum_{j=1}^{n}\left(\frac{\left(\frac{x_j}{\alpha}\right)^c\ln\left(\frac{x_j}{\alpha}\right)}{1+\left(\frac{x_j}{\alpha}\right)^c}\right) \\ \frac{\partial l}{\partial \alpha} &= -\frac{n}{\alpha} + \frac{c(k+1)}{\alpha}\sum_{j=1}^{n}\left(\frac{x_j^c}{\alpha^c+x_j^c}\right) - \frac{n(c-1)}{\alpha}. \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial k^2} &= -\frac{n}{k^2} \\ \frac{\partial^2}{\partial c\partial k} &= -\sum_{j=1}^{n}\frac{x_j^c\ln\left(\frac{x_j}{\alpha}\right)}{\alpha^c+x_j^c} \\ \frac{\partial^2}{\partial \alpha\partial k} &= \frac{c}{\alpha}\sum_{j=1}^{n}\frac{x_j^c}{\alpha^c+x_j^c} \end{aligned}$$

$$\frac{\partial^2 l}{\partial c^2} = -\frac{n}{c^2} - (k+1)\sum_{j=1}^{n}\left(\frac{\left(\ln\left(\frac{x_j}{\alpha}\right)\right)^2\left(\frac{x_j}{\alpha}\right)^c}{\left(1+\left(\frac{x_j}{\alpha}\right)^c\right)^2}\right)$$

$$\frac{\partial^2 l}{\partial \alpha \partial c} = -\frac{n}{\alpha} + \left(\frac{k+1}{\alpha}\right)\sum_{j=1}^{n}\left(\frac{x_j^c}{x_j^c+\alpha^c}\right)\left(1+\frac{c\alpha^c\ln\left(\frac{x_j}{\alpha}\right)}{x_j^c+\alpha^c}\right)$$

and

$$\frac{\partial^2 l}{\partial \alpha^2} = \frac{nc}{\alpha^2} - \frac{c(k+1)}{\alpha^2}\sum_{j=1}^{n}\left(\frac{x_j^c}{\alpha^c+x_j^c}\right)\left(1+\frac{c\alpha^c}{\alpha^c+x_j^c}\right).$$

In Section 4.1.19 it was shown that the Singh-Maddala distribution is a special case of the Burr Type XII distribution. Therefore the same procedure for fitting the Burr distribution can be followed without having to make an assumption upfront of whether the suitable distribution is the Burr Type XII or in fact one of its special cases.

Simulation technique

The probability integral transformation can be used to simulate from the Burr or any of its related distributions. The use of this technique is illustrated in the SAS code given in appendix A.15.

Performance of estimation algorithm

Various simulations with varying values of the parameters were considered. It was seen that for large values of c the iterative algorithm lacks convergence whilst in some cases it moves towards combinations of values for c, k and α that results in arguments to the log functions that are not permissable. For larger values of k the accuracy of the estimation decreases whilst it appears less likely that the algorithm will converge.

6.5.17 Dagum Distribution

Deriving the maximum likelihood estimation algorithm

The probability density function for a random variable X from the Dagum distribution is given by (B.21)from which the likelihood function can be obtained from which the log-likelihood function follows. This likelihood is proposed by Domma [44] and is given here in terms of the parameterization as introduced in Section 4.1.17:

$$l(a,b,p) = n\ln(a) + na\ln(b) + n\ln(p) - (a+1)\sum_{j=1}^{n}\ln(x_j) - \\ (p+1)\sum_{j=1}^{n} ln\left(1 + b^a x_j^{-a}\right)$$

This function needs to be maximized using numerical techniques; therefore the first and second order partial derivatives with respect to a, b and p need to be obtained [44]. These partial derivatives are as given by Domma:

$$\frac{\partial l}{\partial a} = \frac{n}{a} + n\ln(b) - \sum_{j=1}^{n}\ln(x_j) - (p+1)\sum_{j=1}^{n}\left(\frac{b^a}{x_j^a + b^a}\right)\ln\left(\frac{b}{x_j}\right)$$
$$\frac{\partial l}{\partial b} = \frac{na}{b} - \frac{a(p+1)}{b}\sum_{j=1}^{n}\left(\frac{b^a}{x_j^a + b^a}\right)$$
$$\frac{\partial l}{\partial p} = \frac{n}{p} - \sum_{j=1}^{n}\ln\left(1 + \left(\frac{b}{x_j}\right)^a\right)$$

and

$$\frac{\partial^2 l}{\partial a^2} = -\frac{n}{a^2} - (p+1)\sum_{j=1}^{n}\frac{b^a x_j^a\left(\ln\left(\frac{b}{x_j}\right)\right)^2}{\left(x_j^a + b^a\right)^2}$$
$$\frac{\partial^2 l}{\partial b \partial a} = \frac{n}{b} - \left(\frac{p+1}{b}\right)\sum_{j=1}^{n}\left(\frac{b^a}{x_j^a + b^a}\right)\left(1 + \frac{a x_j^a \ln\left(\frac{b}{x_j}\right)}{x_j^a + b^a}\right)$$
$$\frac{\partial^2 l}{\partial p \partial a} = -\sum_{j=1}^{n}\left(\frac{b^a}{x_j^a + b^a}\right)\ln\left(\frac{b}{x_j}\right)$$

$$\frac{\partial^2 l}{\partial b^2} = -\frac{na}{b^2} + \frac{a(p+1)}{b^2}\sum_{j=1}^{n}\left(\frac{b^a}{x_j^a + b^a}\right)\left(1 - \frac{a x_j^a}{x_j^a + b^a}\right)$$
$$\frac{\partial^2 l}{\partial p \partial b} = -\frac{a}{b}\sum_{j=1}^{n}\frac{b^a}{x_j^a + b^a}$$
$$\frac{\partial^2 l}{\partial p^2} = -\frac{n}{p^2}.$$

Simulation technique

One can simulate from the Dagum distribution using the probability integral transformation as follows:

$$X = b\left(U^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}}$$

where U is a Uniform random variable defined on the domain $(0, 1)$.The SAS code that can be used to simulate observations from a Dagum distribution using the probability integral transformation is given in Appendix A.16.

Performance of estimation alogorithm

The iterative algorithm does not possess strong convergence properties. By considering multiple combinations of parameters and simulating from these distributions and then trying to fit these distributions yielded convergence very rarely. The algorithm is also very sensitive to the choice of initial values for these parameters. For the cases where convergence is obtained, accuracy improves only with very large sample sizes. These properties potentially make this distribution less likely to be successful in practical applications.

6.5.18 Generalized Beta Distribution of the Second Kind

Deriving the maximum likelihood estimation algorithm

The likelihood function can be obtained using the probability density function as given by (B.70) from which the log-likelihood function, as given here, follows:

$$\begin{aligned} l(a, b, p, q) = {} & n\ln(a) - nap\ln(b) + n\ln(\Gamma(p+q)) - n\ln(\Gamma(p)) - n\ln(\Gamma(q)) \\ & + (ap-1)\sum_{j=1}^{n}\ln(x_j) - (p+q)\sum_{j=1}^{n}\ln\left(1 + \left(\frac{x_j}{b}\right)^a\right) \end{aligned}$$

This log-likelihood function needs to be maximized numerically using (6.5). The vector of first order partial derivatives is required which contains the following partial derivatives:

$$\frac{\partial l}{\partial a} = \frac{n}{a} - np\ln(b) + p\sum_{j=1}^{n}\ln(x_j) - (p+q)\sum_{j=1}^{n}\frac{x_j^a}{b^a + x_j^a}\ln\left(\frac{x_j}{b}\right)$$

$$\frac{\partial l}{\partial b} = -\frac{nap}{b} + (p+q)\sum_{j=1}^{n}\left(\frac{x_j^a}{b^a + x_j^a}\right)\left(\frac{a}{b}\right)$$

$$\frac{\partial l}{\partial p} = -na\ln(b) + n\Psi(p+q) - n\Psi(p) + a\sum_{j=1}^{n}\ln(x_j) - \sum_{j=1}^{n}\ln\left(1+\left(\frac{x_j}{b}\right)^a\right)$$
$$\frac{\partial l}{\partial q} = n\Psi(p+q) - n\Psi(q) - \sum_{j=1}^{n}\ln\left(1+\left(\frac{x_j}{b}\right)^a\right).$$

We also need the Hessian matrix with the following entries:

$$\frac{\partial^2 l}{\partial a^2} = -\frac{n}{a^2} - (p+q)\sum_{j=1}^{n}\frac{x_j^a b^a \ln(\frac{x_j}{b})}{\left(b^a + x_j^a\right)^2}\ln\left(\frac{x_j}{b}\right)$$
$$\frac{\partial^2 l}{\partial b\partial a} = -\frac{np}{b} + (p+q)\sum_{j=1}^{n}ab^{a-1}\ln\left(\frac{x_j}{b}\right)\frac{x_j^a}{\left(b^a+x_j^a\right)^2} + \left(\frac{p+q}{b}\right)\sum_{j=1}^{n}\frac{x_j^a}{b^a+x_j^a}$$
$$\frac{\partial^2 l}{\partial p\partial a} = -n\ln(b) + \sum_{j=1}^{n}\ln(x_j) - \sum_{j=1}^{n}\frac{x_j^a}{b^a+x_j^a}\ln\left(\frac{x_j}{b}\right)$$
$$\frac{\partial^2 l}{\partial q\partial a} = -\sum_{j=1}^{n}\frac{x_j^a}{b^a+x_j^a}\ln\left(\frac{x_j}{b}\right)$$

$$\frac{\partial^2 l}{\partial b^2} = \frac{nap}{b^2} - a^2 b^{a-2}(p+q)\sum_{j=1}^{n}\frac{x_j^a}{\left(b^a+x_j^a\right)^2} - \frac{a(p+q)}{b^2}\sum_{j=1}^{n}\frac{x_j^a}{b^a+x_j^a}$$
$$\frac{\partial^2 l}{\partial p\partial b} = -\frac{na}{b} + \frac{a}{b}\sum_{j=1}^{n}\frac{x_j^a}{b^a+x_j^a}$$
$$\frac{\partial^2 l}{\partial q\partial b} = \frac{a}{b}\sum_{j=1}^{n}\frac{x_j^a}{b^a+x_j^a}$$

$$\frac{\partial^2 l}{\partial p^2} = n\left(\Psi^{'}(p+q) - \Psi^{'}(p)\right)$$
$$\frac{\partial^2 l}{\partial q\partial p} = n\Psi^{'}(p+q)$$
$$\frac{\partial^2 l}{\partial q^2} = n\left(\Psi^{'}(p+q) - \Psi^{'}(q)\right)$$

The algorithm as given in (6.5) can now be used.

Simulation technique

To simulate from this distribution, a similar technique can be used to what was used to simulate from the Generalized Pareto and Beta-prime distributions. If the probability density function of the Compound Gamma distribution [45] as given in (6.19) is considered and we let $X = Y^a$, $\delta = b^a$, $\gamma = q$ and $\alpha = p$, then the resulting probability density function of Y is the same as the probability density function for the Generalized Beta distribution as given in (B.70). Hence to simulate from this distribution, the following steps can be followed:

1. Simulate a β from a $\text{GAM}\left(q, \frac{1}{b^a}\right)$ distribution.
2. Then, by using the simulated value of β in the previous step, simulate and X from a $\text{GAM}\left(p, \frac{1}{\beta}\right)$ distribution.
3. Obtain $Y = X^{\frac{1}{a}}$
4. Repeat steps (1) to (3) n times to obtain a simulated sample of size n from a Generalized Beta distribution of the Second Kind.

The SAS code to perform the simulation as described above is given in Appendix A.17.

Performance of estimation algorithm

Various parameter values to simulate from together with various combinations of initial values for the iterative procedure were considered. Generally it was seen that it is difficult to obtain convergence, even when initial values are chosen close to the true values and despite having the sample size very large. It is therefore expected that it might be difficult to fit this distribution to real-life data.

6.5.19 Kappa Family of Distributions

Deriving the maximum likelihood estimation algorithm

Consider the cumulative distribution function of the three-parameter case, as given in (4.42), and let $\theta = \alpha$ and $\alpha = \alpha^{-1}$, then the cumulative distribution function simplifies to the cumulative distribution for the two-parameter case (which is given in (Kappa2CDF)).

This means that we need a distribution fitting algorithm for the three-parameter case only, from which we can assess whether a two-parameter parameterization is implied based in the maximum likelihood estimates of α and θ.

The likelihood function follows directly from (B.109) while the associated log-likelihood is given by:

$$l(\alpha,\beta,\theta) = n\ln(\alpha) + n\ln(\theta) - n\alpha\ln(\beta) + (\alpha\theta - 1)\sum_{j=1}^{n}\ln(x_j) - (\alpha+1)\sum_{j=1}^{n}\ln\left(1+\left(\frac{x_j}{\beta}\right)^{\theta}\right).$$

To find the values of α, β and θ that maximises this log-likelihood function, we need to first consider the first order partial derivatives with respect to α, β and θ, respectively.

$$\begin{aligned}
\frac{\partial l}{\partial \alpha} &= \frac{n}{\alpha} - n\theta\ln(\beta) + \theta\sum_{j=1}^{n}\ln(x_j) - \sum_{j=1}^{n}\ln\left(1+\left(\frac{x_j}{\beta}\right)^{\theta}\right)\\
\frac{\partial l}{\partial \beta} &= -\frac{n\alpha\theta}{\beta} + \frac{\theta(\alpha+1)}{\beta}\sum_{j=1}^{n}\frac{x_j^{\theta}}{\beta^{\theta}+x_j^{\theta}}\\
\frac{\partial l}{\partial \theta} &= \frac{n}{\theta} - n\alpha\ln(\beta) + \alpha\sum_{j=1}^{n}\ln(x_j) - (\alpha+1)\sum_{j=1}^{n}\frac{x_j^{\theta}}{x_j^{\theta}+\beta^{\theta}}\ln\left(\frac{x_j}{\beta}\right).
\end{aligned}$$

By setting these partial derivatives equal to 0, explicit expressions in terms of α, β and θ cannot be obtained from which the maximum likelihood estimates can directly be calculated from. We therefore also need to derive the second order partial derivatives which are required to estimate these parameters numerically using (6.5). The second order partial derivatives are given below:

$$\begin{aligned}
\frac{\partial^2 l}{\partial \alpha^2} &= -\frac{n}{\alpha^2}\\
\frac{\partial^2 l}{\partial \beta\partial\alpha} &= -\frac{n\theta}{\beta} + \frac{\theta}{\beta}\sum_{j=1}^{n}\left(\frac{x_j^{\theta}}{\beta^{\theta}+x_j^{\theta}}\right)\\
\frac{\partial^2 l}{\partial \theta\partial\alpha} &= -n\ln(\beta) + \sum_{j=1}^{n}\ln(x_j) - \sum_{j=1}^{n}\left(\frac{x_j^{\theta}}{\beta^{\theta}+x_j^{\theta}}\right)\ln\left(\frac{x_j}{\beta}\right)
\end{aligned}$$

and

$$\frac{\partial^2 l}{\partial \beta^2} = \frac{n\alpha\theta}{\beta^2} - \frac{\theta(\alpha+1)}{\beta^2}\sum_{j=1}^{n}\left(\frac{x_j^\theta}{\beta^\theta + x_j^\theta}\right)\left(1 + \frac{\theta\beta^\theta}{\beta^\theta + x_j^\theta}\right)$$

$$\frac{\partial^2 l}{\partial \theta \partial \beta} = -\frac{n\alpha}{\beta} + \left(\frac{\alpha+1}{\beta}\right)\sum_{j=1}^{n}\left(\frac{x_j^\theta}{\beta^\theta + x_j^\theta}\right)\left(1 + \frac{\theta\beta^\theta \ln\left(\frac{x_j}{\beta}\right)}{\beta^\theta + x_j^\theta}\right)$$

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{n}{\theta^2} - (\alpha+1)\sum_{j=1}^{n}\frac{x_j^\theta\beta^\theta\left(\ln\left(\frac{x_j}{\beta}\right)\right)^2}{\left(\beta^\theta + x_j^\theta\right)^2}.$$

Simulation technique

The probability integral transformation is used to simulate from this distribution using the following expression:

$$X = \beta\left(U^{-\frac{1}{\alpha}} - 1\right)^{-\frac{1}{\theta}}$$

where U is a uniform random variable on the domain $(0,1)$.

Performance of estimation algorithm

It was seen that convergence is not always guaranteed, even when initial values are chosen close to the true values and despite having the sample size very large. Thus it may be difficult to fit this distribution in practice.

6.5.20 Log-logistic Distribution

Deriving the maximum likelihood estimation algorithm

The probability density function for a random variable X from a Log-logistic distribution is given by (B.122). Using this function, the log-likelihood in terms of its parameters, α and β, can be expressed as follows:

$$l(\alpha,\beta) = n\ln(\beta) + \beta\sum_{j=1}^{n}\ln\left(\frac{x_j}{\alpha}\right) - \sum_{j=1}^{n}\ln\left(x_j\right) - 2\sum_{j=1}^{n}\ln\left(1 + \left(\frac{x_j}{\alpha}\right)^\beta\right)$$
$$- 2\sum_{j=1}^{n}\ln\left(1 + \left(\frac{x_j}{\alpha}\right)^\beta\right).$$

The maximum likelihood estimates are obtained numerically, hence the first and second order partial derivatives of the likelihood with respect to α and β are required in order to make use of (6.5). The first order partial derivatives are given by the following expressions:

$$\frac{\partial l}{\partial \alpha} = \frac{\beta n}{\alpha} + 2\frac{\beta}{\alpha}\sum_{j=1}^{n} \frac{x_j^\beta}{\alpha^\beta + x_j^\beta}$$

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} + \sum_{j=1}^{n} \ln(x_j) - n\ln(\alpha) - 2\sum_{j=1}^{n} \frac{x_j^\beta}{\alpha^\beta + x_j^\beta} \ln\left(\frac{x_j}{\alpha}\right).$$

The second order partial derivatives are given below:

$$\frac{\partial^2 l}{\partial \alpha^2} = \frac{\beta n}{\alpha^2} - 2\frac{\beta}{\alpha^2}\sum_{j=1}^{n} \frac{x_j^\beta}{\alpha^\beta + x_j^\beta} - 2\frac{\beta}{\alpha^2}\sum_{j=1}^{n} \frac{\beta x_j^\beta \alpha^\beta}{\left(\alpha^\beta + x_j^\beta\right)^2}$$

$$\frac{\partial^2 l}{\partial \beta \partial \alpha} = -\frac{n}{\alpha} + \frac{2}{\alpha}\sum_{j=1}^{n} \frac{x_j^\beta}{\alpha^\beta + x_j^\beta} + \frac{2\beta}{\alpha}\sum_{j=1}^{n} \frac{\alpha^\beta x_j^\beta \ln\left(\frac{x_j}{\alpha}\right)}{\left(\alpha^\beta + x_j^\beta\right)^2}$$

and

$$\frac{\partial^2 l}{\partial \beta^2} = -\frac{n}{\beta^2} - 2\sum_{j=1}^{n} \frac{x_j^\beta \alpha^\beta \left(\ln\left(\frac{x_j}{\alpha}\right)\right)^2}{\left(x_j^\beta + \alpha^\beta\right)^2}.$$

Simulation technique

One can simulate from the Log-logistic distribution easily by using the probability integral transformation. The SAS code to perform this simulation is included in Appendix A.19.

Performance of estimation algorithm

Convergence of the estimation algorithm is often not found whilse in some cases the algorithm converges quickly and with reasonable accuracy. Generally the accuracy increases with increasing sample sizes.

Generally iterative numerical algorithms' performance can be influenced by the choice of initial values. In this instance a regression approach can be used to find initial values for α and β. Consider the empirical distribution function $\hat{F}_X(x)$ and set it equal to the expression for the cumulative distribution function in (4.43). The equation can then be rewritten as follows:

$$\ln(X) = \ln(\alpha) - \frac{1}{\beta}\ln\left(\frac{1}{\hat{F}_X(x)} - 1\right)$$

which is in the form of a linear regression

$$Y = a + bX \text{ with } Y = \ln(X), a = \ln(\alpha), b = -\beta^{-1} \text{ and } X = \ln\left(\left(\hat{F}_X(x)\right)^{-1}\right)$$

The estimates from this expression can then be used to find initial values for α and β. This is also included in Appendix A.19.

6.5.21 Folded Normal Distribution

Deriving the maximum likelihood estimation algorithm

The likelihood function for the Folded Normal distribution is derived directly from the probability density function as given in (B.43) from which an expression for the log-likelihood function follows to be:

$$l(\mu,\sigma) = -n\ln(\sigma) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma}\right)^2 + \sum_{j=1}^{n}\ln\left(1 + e^{-\frac{2x_j\mu}{\sigma^2}}\right).$$

To find the values of μ and σ that will maximise the log-likelihood function, the partial derivatives with respect to μ and σ are required, which is given in the expressions below:

$$\frac{\partial l}{\partial \mu} = \sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma^2}\right) - 2\sum_{j=1}^{n}\left(\frac{x_j}{\sigma^2}\right)\frac{e^{-\frac{2x_j\mu}{\sigma^2}}}{1+e^{-\frac{2x_j\mu}{\sigma^2}}}$$
$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{j=1}^{n}\left(\frac{(x_j-\mu)^2}{\sigma^3}\right) + 4\sum_{j=1}^{n}\left(\frac{x_j\mu}{\sigma^3}\right)\frac{e^{-\frac{2x_j\mu}{\sigma^2}}}{1+e^{-\frac{2x_j\mu}{\sigma^2}}}.$$

Setting these partial derivatives equal to 0 yields the following relationship between the maximul likelihood estimators for μ and σ [79], [46]:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}x_j^2 - \hat{\mu}^2.$$

This still doesn't provide explicit expression in terms of μ and σ. Therefore the Newton-Raphson algorithm needs to be implemented in order to derive maximimum likelihood estimators numerically. This will require the evaluation of second order partial derivatives, which are given below:

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{\sigma^2} + \frac{4}{\sigma^4}\sum_{j=1}^{n} x_j^2 \frac{e^{2\frac{x_j\mu}{\sigma^2}}}{\left(1+e^{2\frac{x_j\mu}{\sigma^2}}\right)^2}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \sigma \partial \mu} &= -2\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma^3}\right) + 4\sum_{j=1}^{n}\left(\frac{x_j}{\sigma^3}\right)\left(1+e^{\frac{2x_j\mu}{\sigma^2}}\right)^{-1} \\ &\quad - 8\sum_{j=1}^{n}\left(\frac{x_j^2\mu}{\sigma^5}\right)e^{\frac{2x_j\mu}{\sigma^2}}\left(1+e^{\frac{2x_j\mu}{\sigma^2}}\right)^{-2}\end{aligned}$$

$$\frac{\partial^2 l}{\partial \sigma^2} = -\frac{2}{\sigma^3}\sum_{j=1}^{n}(x_j-\mu) + \frac{4}{\sigma^3}\sum_{j=1}^{n}\frac{x_j}{e^{2\frac{x_j\mu}{\sigma^2}}+1} - \frac{8\mu}{\sigma^5}\sum_{j=1}^{n}\frac{e^{2\frac{x_j\mu}{\sigma^2}}}{\left(1+e^{2\frac{x_j\mu}{\sigma^2}}\right)^2}.$$

Simulation technique

The simulation follows directly from how the Folded Normal distribution was defined in Section 4.1.24 where a random observation W is first simulated from a Normal distribution with parameters μ and σ. Obtain a random observation X from the Folded Normal distribution by taking $X = |W|$. The SAS code which can be used to perform both the simulation is given in Appendix A.20.

Performance of estimation algorithm

The estimation algorithm appears to perform well with quick convergence and reasonable accuracy for various combinations of parameter values and initial values. For larger values of σ and values for μ close to 0 the estimation becomes less accurate.

6.5.22 Inverse Gamma and related distributions

Deriving the method-of-moments estimators and maximum likelihood estimation algorithm

- For the Inverse Gamma distribution

Expressions for the mean and variance are given in (B.91) and (B.92). Setting $\mathrm{E}(X) = \overline{x}$ and $\mathrm{var}(X) = s_x^2$ yields the method-of-moments estimators for α and θ that are given in (B.94) and (B.95).

The expression for the probability density function for a random variable that is Inverse Gamma distributed with parameters α and θ is given in (B.89) from which the likelihood function can be obtained from which the log-likelihood function follows to be:

$$l(\alpha,\theta) = n\alpha\ln(\theta) - n\ln(\Gamma(\alpha)) - (\alpha+1)\sum_{j=1}^{n}\ln(x_j) - \theta\sum_{j=1}^{n}x_j^{-1}.$$

The maximum likelihood estimators of μ and σ can be obtained numerically using (6.5). This algorithm requires us to derive the first and second order partial derivatives with respect to μ and σ. These partial derivatives are given below:

$$\frac{\partial l}{\partial\alpha} = n\ln(\theta) - n\Psi(\alpha) - \sum_{j=1}^{n}\ln(x_j)$$
$$\frac{\partial l}{\partial\theta} = \frac{n\alpha}{\theta} - \sum_{j=1}^{n}x_j^{-1}$$

and

$$\frac{\partial^2 l}{\partial\alpha^2} = -n\Psi^{'}(\alpha)$$
$$\frac{\partial^2 l}{\partial\theta\partial\alpha} = \frac{n}{\theta}$$
$$\frac{\partial^2 l}{\partial\theta^2} = -\frac{n\alpha}{\theta^2}$$

- For the Inverse Chi-square distribution

Since the Inverse Chi-square distribution is a special case of the Inverse Gamma distribution with $\alpha = \frac{\nu}{2}$ and $\theta = 2$ (as shown in Section 4.1.26). Therefore the log-likelihood function can be derived from the log-likelihood function of the Inverse Gamma distribution:

$$l(\nu) = -\frac{n\nu}{2}\ln(2) - n\ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \sum_{j=1}^{n}(2x_j)^{-1} - \left(\frac{\nu}{2}+1\right)\sum_{j=1}^{n}\ln(x_j).$$

Setting the first order partial derivative equal to 0 doesn't yield an explicit expression in terms of ν giving its maximum likelihood estimator; hence numerical estimation should be performed using the Newton-Raphson algorithm as given in (6.4). For this particular problem this updating equation will take on the following form:

$$\nu^{(New)} = \nu^{(Old)} - \frac{\frac{n}{2}\ln(2) + \frac{n}{2}\Psi\left(\frac{\nu^{(Old)}}{2}\right) + \frac{1}{2}\sum_{j=1}^{n}\ln(x_j)}{\frac{n}{4}\Psi'\left(\frac{\nu^{(Old)}}{2}\right)}.$$

Simulation technique

- For the Inverse Gamma distribution

Using Result 3, it follows that we can simulate from the Inverse Gamma distribution using the following steps:

1. Simulate Y from GAM$\left(\frac{1}{\theta}, \alpha\right)$ distribution.
2. Calculate $X = Y^{-1}$, which gives a simulated value from an Inverse Gamma distribution with parameters θ and α.

The SAS code that can be used to simulate from an Inverse Gamma distribution is included in Appendix A.21.

- For the Inverse Chi-square distribution

Given the relationship between the Inverse Chi-square and Inverse Gamma, the same steps used to simulate from the Inverse Gamma distribution with $\theta = 1$ and $\alpha = \frac{\nu}{2}$. The SAS code to perform this simulation is given in Appendix A.22.

Performance of estimation algorithm

- For the Inverse Gamma distribution

The algorithm performs reasonable in terms of convergence and giving estimates that are accurate. It is sensitive to the choice of initial values. In the SAS code provided in Appendix A.21 the observed mean is used as intial value for θ in the iterative estimation algorithm as the mean value provides an indication of the scale of the observed values.

- For the Inverse Chi-square distribution

The algorithm easily converges while it generally provides reasonably accurate estimates of the parameter ν.

6.5.23 Loggamma Distribution

Deriving the maximum likelihood estimation algorithm

The likelihood function for the Loggamma distribution can be derived from the probability density function as given in (B.113). The log-likelihood function is then given by:

$$l(a,\lambda) = na\ln(\lambda) - n\ln(\Gamma(a)) + (a-1)\sum_{j=1}^{n}\ln(\ln(x_j)) - (\lambda+1)\sum_{j=1}^{n}\ln(x_j).$$

The maximum likelihood estimators for a and λ need to be found which requires the derivation of the first order partial derivatives of the log-likelihood function in terms of a and λ.

$$\begin{aligned}\frac{\partial l}{\partial a} &= n\ln(\lambda) - n\Psi(a) + \sum_{j=1}^{n}\ln(\ln(x_j)) \\ \frac{\partial l}{\partial \lambda} &= \frac{na}{\lambda} - \sum_{j=1}^{n}\ln(x_j).\end{aligned}$$

When these first order partial derivatives are set to 0, it is not possible to derive explicit expressions for the maximum likelihood estimators of a and λ. It can, however, be derived that

$$\hat{\lambda}_{MLE} = \left(\sum_{j=1}^{n}\ln(x_j)\right)^{-1} n\hat{a}_{MLE}.$$

The maximum likelihood estimates will therefore have to be determined numerically using (6.5). The second order partial derivatives are therefore also required and are given below:

$$\begin{aligned}\frac{\partial^2 l}{\partial a^2} &= -n\Psi^{'}(a) \\ \frac{\partial^2 l}{\partial \lambda \partial a} &= \frac{n}{\lambda} \\ \frac{\partial^2 l}{\partial \lambda^2} &= -\frac{na}{\lambda^2}.\end{aligned}$$

Simulation technique

One can simulate random variables from a Loggamma distribution by first simulating an Y from a Gamma distribution and then calculate $X = e^Y$ which is then a simulated value from a Loggamma distribution. The SAS code that can be used to perform this simulation is given in Appendix A.23.

Performance of estimation algorithm

It was found that the algorithm is sensitive to the choice of initial values, in particular the combination of values used. In order to choose initial values for a and λ that are in a relation to each other that makes sense in terms of the distribution form, it was decided to consider the expression for the expected value as given in (B.115). If we set this expression equal to the observed sample mean, $\overline{x}$, we can rewrite the equation to have λ expressed as a function of a and $\overline{x}$:

$$\lambda = \left(1 - \overline{x}^{-\frac{1}{a}}\right)^{-1}$$

This expression was utilised to set-up initial values for a and λ by choosing some value for a and to then calculate an associated initial value for λ. This significantly improved the performance of the iterative estimation procedure in terms of its likelihood to reach convergence. This approach is incorporated in the SAS code in Appendix A.23.

6.5.24 Snedecor's F Distribution

Deriving the method-of-moments estimators and maximum likelihood estimation algorithm

Setting the expressions for the mean and variance (as given in (B.166) and (B.167)) equal to their observed sample counterparts, $\overline{x}$ and s_x^2 yields method-of-moments estimates for ν_1 and ν_2 as given in (B.169) and (B.170), respectively.

An iterative estimation algorithm to obtain the maximum likelihood estimates will be derived below in which case the method-of-moments estimates can be used as starting values.
The likelihood function follows from the product of the probability density function as given by Bain and Engelhardt [11]. The associated log-likelihood function is given by:

$$\begin{aligned} l(\nu_1, \nu_2) = {} & n \ln\left(\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\right) - n\ln\left(\Gamma\left(\frac{\nu_1}{2}\right)\right) - n\ln\left(\Gamma\left(\frac{\nu_2}{2}\right)\right) \\ & + \frac{n\nu_1}{2}\ln\left(\frac{\nu_1}{\nu_2}\right) + \left(\frac{\nu_2}{2}-1\right)\sum_{j=1}^{n}\ln(x_j) - \\ & \left(\frac{\nu_1+\nu_2}{2}\right)\sum_{j=1}^{n}\ln\left(1+\frac{\nu_1}{\nu_2}x_j\right). \end{aligned}$$

This function cannot be maximized by means of finding the first order partial derivatives and solving for maximum likelihood estimators after setting the first order partial derivatives equal to 0. For this reason both the first and second order partial derivatives are required as set out below:

$$\frac{\partial l}{\partial \nu_1} = \frac{n}{2}\Psi\left(\frac{\nu_1+\nu_2}{2}\right) - \frac{n}{2}\Psi\left(\frac{\nu_1}{2}\right) + \frac{n}{2}\ln\left(\frac{\nu_1}{\nu_2}\right) + \frac{n}{2} + \frac{1}{2}\sum_{j=1}^{n}\ln(x_j)$$
$$- \frac{1}{2}\sum_{j=1}^{n}\ln\left(1+\frac{\nu_1}{\nu_2}x_j\right) - \left(\frac{\nu_1+\nu_2}{2}\right)\sum_{j=1}^{n}\left(\frac{x_j}{\nu_2+\nu_1 x_j}\right)$$

$$\frac{\partial l}{\partial \nu_2} = \frac{n}{2}\Psi\left(\frac{\nu_1+\nu_2}{2}\right) - \frac{n}{2}\Psi\left(\frac{\nu_2}{2}\right) - \frac{n\nu_1}{2\nu_2} - \frac{1}{2}\sum_{j=1}^{n}\ln\left(1+\frac{\nu_1}{\nu_2}x_j\right)$$
$$+ \left(\frac{\nu_1+\nu_2}{2}\right)\sum_{j=1}^{n}\frac{\nu_1 x_j}{\nu_2(\nu_2+\nu_1 x_j)}.$$

$$\frac{\partial^2 l}{\partial \nu_1^2} = \frac{n}{4}\Psi'\left(\frac{\nu_1+\nu_2}{2}\right) - \frac{n}{4}\Psi'\left(\frac{\nu_1}{2}\right) + \frac{n}{2\nu_1} - \sum_{j=1}^{n}\frac{x_j}{\nu_2+\nu_1 x_j}$$
$$+ \left(\frac{\nu_1+\nu_2}{2}\right)\sum_{j=1}^{n}\left(\frac{x_j}{\nu_2+\nu_1 x_j}\right)^2$$

$$\frac{\partial^2 l}{\partial \nu_2 \partial \nu_1} = \frac{n}{4}\Psi'\left(\frac{\nu_1+\nu_2}{2}\right) - \frac{n}{2\nu_2} - \frac{1}{2}\sum_{j=1}^{n}\frac{(\nu_2-\nu_1)x_j}{(\nu_2+\nu_1 x_j)\nu_2}$$
$$+ \left(\frac{\nu_1+\nu_2}{2}\right)\sum_{j=1}^{n}\frac{x_j}{(\nu_2+\nu_1 x_j)^2}$$

and

$$\frac{\partial^2 l}{\partial \nu_2^2} = \frac{n}{4}\Psi'\left(\frac{\nu_1+\nu_2}{2}\right) - \frac{n}{4}\Psi'\left(\frac{\nu_2}{2}\right) + \frac{n\nu_1}{2\nu_2^2} + \sum_{j=1}^{n}\frac{\nu_1 x_j}{\nu_2(\nu_2+\nu_1 x_j)}$$
$$- \left(\frac{\nu_1+\nu_2}{2}\right)\sum_{j=1}^{n}\frac{\nu_1 x_j(2\nu_2+\nu_1 x_j)}{\nu_2^2(\nu_2+\nu_1 x_j)^2}$$

These partial derivatives can be used in (6.5) to estimate ν_1 and ν_2 numerically.

Simulation technique

The simulation from this distribution can be done using built-in SAS functions. This is illustrated in the SAS code given in Appendix A.24.

Performance of estimation algorithm

The algorithm generally converges, but doesn't always have a high level of accuracy.

6.5.25 Inverse Gaussian Distribution

Deriving the method-of-moments estimators and maximum likelihood estimation algorithm

Using the expressions for the mean and variance as given in (B.99) and (B.100) and setting these expressions equal to the sample mean and the sample variance, respectively, yields the method-of-moments estimates for μ and λ as given in (B.102) and (B.103).

These estimates can be used as starting values in the iterative process used to obtain the maximum likelihood estimates, since the maximum likelihood estimation requires the use of the iterative algorithm as given in (6.5).

The log-likelihood function is given by:

$$l(\mu, \lambda) = \frac{n}{2}\ln(\lambda) - \frac{n}{2}\ln(2\pi) - \frac{3}{2}\sum_{j=1}^{n}\ln(x_j) - \frac{\lambda}{2\mu^2}\sum_{j=1}^{n}\frac{(\mu - x_j)^2}{x_j}.$$

The first order partial derivatives with respect to μ and λ are given below:

$$\begin{aligned}
\frac{\partial l}{\partial \mu} &= \frac{\lambda}{\mu^3}\sum_{j=1}^{n}\frac{(\mu - x_j)^2}{x_j} - \frac{\lambda}{\mu^2}\sum_{j=1}^{n}\frac{\mu - x_j}{x_j} \\
\frac{\partial l}{\partial \lambda} &= \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_{j=1}^{n}\frac{(\mu - x_j)^2}{x_j}.
\end{aligned}$$

The second order partial derivatives are given below:

$$\begin{aligned}
\frac{\partial^2 l}{\partial \mu^2} &= -\frac{3\lambda}{\mu^4}\sum_{j=1}^{n}\frac{(\mu - x_j)^2}{x_j} + \frac{4\lambda}{\mu^3}\sum_{j=1}^{n}\frac{(\mu - x_j)}{x_j} - \frac{\lambda}{\mu^2}\sum_{j=1}^{n}x_j^{-1} \\
\frac{\partial^2 l}{\partial \lambda \partial \mu} &= \frac{1}{\mu^3}\sum_{j=1}^{n}\frac{(\mu - x_j)^2}{x_j} - \frac{1}{\mu^2}\sum_{j=1}^{n}\frac{(\mu - x_j)}{x_j} \\
\frac{\partial^2 l}{\partial \lambda^2} &= -\frac{n}{2\lambda^2}.
\end{aligned}$$

Simulation technique

The simulation cannot be done using built-in functions in SAS; neither can the probability integral transformation be applied, since a closed-form expression for the cumulative distribution function is not available. The following approach can be followed to simulate from this distribution:

1. Consider a sufficiently large subset of the positive real line, say $[0, \zeta]$. Various iterations may have to be run to determine whether the selected subset is in fact large enough.

2. Partition the subset into intervals of size ϵ. This will result into having intervals $[0, \epsilon]$, $(\epsilon, 2\epsilon]$, $(2\epsilon, 3\epsilon]$, $\ldots$. Decreasing the interval size will refine the partitioning from which more accurate numerical estimation results are expected, whilst it will lead to an increase in processing time.

3. Calculate the cumulative distribution function's value at each interval's upper bound using the cumulative distribution function as given in 4.46. This means that for the i^{th} interval one will calculate the following:

$$F_X(i\epsilon) = \Phi\left(\sqrt{\frac{\lambda}{i\epsilon}}\left(\frac{i\epsilon}{\mu} - 1\right)\right) + e^{2\frac{\lambda}{\mu}}\Phi\left(-\sqrt{\frac{\lambda}{i\epsilon}}\left(\frac{i\epsilon}{\mu} + 1\right)\right)$$

   One may also consider the mid-point or lower bound of each interval to calculate the cumulative distribution at. It is important to assess at this point whether the calculated value of the cumulative distribution function associated with the upper bound of the last interval is sufficiently close to 1. This will be an indication of whether the subset of the positive real line considered is sufficiently large enough.

4. One can now simulate random uniform random variables on the domain $[0, 1]$, say U. This can be done using a built-in function in SAS.

5. Now one can apply an approach similar to the probability integral transformation. The largest value $X = i\epsilon$ such that $F_X(x) \leq U$ is then considered a simulated value from the Inverse Gaussian distribution.

This simulation approach is used and illustrated in the SAS code given in Appendix A.25.

Performance of estimation algorithm

When the method-of-moment estimates for the parameters μ and λ were used as initial values in the iterative maxmimum likelihood estimation procedure, convergence was easily obtained with the estimates being reasonably accurate.

6.5.26 Skew-normal Distribution

Deriving the method-of-moments estimators and maximum likelihood estimation algorithm

In comparison to the estimation of the other distributions considered in this study, the estimation of the parameters of the Skew-normal distribution using maximum likelihood is more complex.

In this section the method-of-moments estimation is discussed for the direct parameterization and both the method-of-moments estimation and the maximum likelihood estimation are discussed for the centered parameterization as given by Pewsey [100].

- **Direct Parameterization**

  - **Method-of-Moments Estimation**:

    Suppose we observe a sample of size n of observations $(y_1, y_2, \ldots, y_n)$ associated with a random variable $Y = \xi + \eta X$ where $X \sim SN(\lambda)$. Let $\overline{y}$ and s_y denote the sample mean and standard deviation, respectively. The observed sample can now be standardized to obtain $(y_{1S}, y_{2S}, \ldots, y_{nS})$ which is now associated with a random variable $Y_S = \xi_S + \eta_S X$, where

    $$\hat{\xi}_S = \frac{\hat{\xi} - \overline{y}}{s_y} \text{ and } \hat{\eta}_S = \frac{\hat{\eta}}{s_y}$$

    If one now equates the theoretical expresssions for the moments of the standardized random variable with the observed sample central moments (m_1, m_2 and m_3), one can find the method-of-moments estimates [100]:

$$\hat{\xi}_S = -\left(\frac{2}{4-\pi}\right)^{\frac{1}{3}}\left(\frac{m_3}{s_y^3}\right)^{\frac{1}{3}}$$

$$\hat{\eta}_S = \sqrt{1+\hat{\xi}_S^2}$$

and

$$\hat{\delta} = -\frac{\hat{\xi_S}}{b\hat{\eta_S}}$$

from which the estimate of λ follows to be:

$$\hat{\lambda} = \frac{\hat{\delta}}{\sqrt{1-\hat{\delta}^2}} \tag{6.20}$$

Sampling distributions of $\hat{\xi}$ and $\hat{\delta}$ are bimodal for values of λ that are close to 0 which means that if the true distribution is close to normal, these estimates may be inaccurate.

From the arguments above for the method-of-moments estimation it follows that the direct paramterization may not be ideal for the purpose of fitting the Skew-normal distribution.

- **Centered Parameterization**

  - **Method-of-Moments Estimation**:
    In the context of this parameterization, the expressions for the mean, variance and skewness coefficient (as given in (B.158) to (B.160)) can be set equal to the sample mean, variance and skewness to obtain the method-of-moments estimates for μ and σ as given in (B.161) and (B.162).
    This consequently means that the method-of-moments estimate for δ is as given in (B.164) from which the method-of-moments estimate of λ can be calculated exactly as for the direct parameterization (given in (6.20)).

  - **Maximum Likelihood Estimation**:
    Consider an observed sample of size n, $(y_1, y_2, \ldots, y_n)$, where $Y \sim SN_C(\mu, \sigma, \gamma_1)$ (similar to the parameterization and notation as given in Section 4.1.28). One can then standardize observed outcomes using the sample mean and sample variance as follows:

$$y_i^{(S)} = \frac{y_i - \overline{y}}{s_y} \text{ for } i = 1, 2, 3, ..., n$$

then one effectively considers a standardized random variable

$$Y^{(S)} \sim SN_C(\mu_S, \sigma_S, \gamma_1).$$

A constraint in terms of the direct parameterization is given which can equivalently be introduced in the centered parameterization, which requires the following relationship to hold, [8], [100]:

$$\hat{\sigma}_S = \left((\hat{\mu}_S)^2 (1+\hat{\tau}^2)+1\right)^{\frac{1}{2}} - \hat{\mu}_S\hat{\tau}$$

where $\hat{\tau} = \left(\frac{2}{4-\pi}\right)^{\frac{1}{3}} \hat{\gamma}_1^{\frac{1}{3}}$.

The values of $\hat{\mu}_S$ and $\hat{\gamma}_1$ are maximizing the log-likelihood given by:

$$l(\mu_S,\tau) = -n\ln\left(\sqrt{(\mu_S)^2(1+\tau^2)+1} - \mu_S\tau\right) - \frac{n}{2}\ln(1+\tau^2) + \sum_{j=1}^{n}\ln\left(\Phi\left(\frac{\frac{(y_{jS}-\mu_S)\tau}{\sqrt{\mu_S^2(1+\tau^2)+1}-\mu_S\tau}+\tau^2}{\sqrt{(b^2+\tau^2(b^2-1))(1+\tau^2)}}\right)\right) \quad (6.21)$$

From (6.21) follows that τ can take on values in the range $\left(-\sqrt{\frac{2}{\pi-2}}, \sqrt{\frac{2}{\pi-2}}\right)$.

The final estimates of μ, σ and γ can be found from $\hat{\mu_S}$ and $\hat{\tau}$ as follows:

$$\begin{aligned}
\hat{\mu} &= \overline{y} + s_y\hat{\mu}_S \\
\hat{\sigma} &= s_y\hat{\sigma}_S \\
&= s_y\left(\sqrt{\hat{\mu}_S(1+\hat{\tau}^2)+1} - \hat{\mu}_S\hat{\tau}\right) \\
&\text{and} \\
\hat{\gamma}_1 &= \left(\frac{4-\pi}{2}\right)\hat{\tau}^3
\end{aligned}$$

If $\hat{\mu} = \mathrm{E}(X) = b\delta$ and $\hat{\sigma} = \sqrt{\mathrm{var}(X)} = \sqrt{1-b^2\delta^2}$ the centered parameterization reduces to the standard parameterization; i.e. $Y = X$. Therefore one will be able to assess from the estimates whether the centered or the standard paramterization is implied.

Simulation technique

Since a closed form expression for the cumulative distribution function is not available, the probability integral transformation can't be used to simulate from this distribution. A technique similar to what was introduced in Section 6.5.25 for the Inverse Gaussian is suggested here. The following steps can be followed for the simulation:

1. Consider a sufficient subset of the real line. Since the standard Skew-normal parameterization considers a transformation of a standard Normal random variable, it will be sufficient to consider the interval $[-10, 10]$ as 99.73% of observations under a standard Normal distribution will lie between -3 and 3.

2. Partition the subset into intervals of size ϵ. This will result into having intervals $[-10, -10+\epsilon]$, $(-10+\epsilon, -10+2\epsilon]$, $(-10+2\epsilon, -10+3\epsilon]$, $\ldots$, $(10-\epsilon, 10]$. Decreasing the interval size will refine the partitioning from which more accurate numerical estimation results are expected, whilst it will lead to an increase in processing time.

3. Calculate the probability density function's value at each interval's upper bound using the probability density function as given in (4.47). This means that for the i^{th} interval one will calculate the following:

$$h(-10+i\epsilon) = 2\Phi(\lambda(-10+i\epsilon))\phi(-10+i\epsilon)$$

   One may also consider the mid-point or lower bound of each interval to calculate the density at.

4. The cumulative distribution function can now be constructed by using a numerical integration procedure to calculate the area under the constructed probability density function.

5. One can now simulate random uniform random variables on the domain $[0, 1]$, say U.

6. Now one can apply an approach similar to the probability integral transformation. The largest value $X = -10 + i\epsilon$ such that $F_X(x) \leq U$ is then considered a simulated value from the Skew-normal distribution.

The SAS code illustrating this simulation approach is included in Appendix A.26.

Accuracy of estimation

- The method-of-moments estimates on all three parameterizations appeared to be fairly accurate. Instability on the standard and direct parameterizations started to appear with values of λ very close to 0, whilst it remained stable for the centered parameterization.

- It follows that the centred parameterization is the most stable and due to its relationship with the standard parameterization it suggests that it may work well as a general approach to fitting a Skew-normal disribution to consider the centered parameterization.

- Assess the maximum likelihood estimation by testing whether the values for μ_S and τ that maximizes the log- likelihood function given in (6.21) are in fact leading to accurate maximum likelihood estimates for μ, σ and λ for both the standard and centered parameteizations. The SAS code for this assessment is included in appendix A.26.

  It was found that the values for μ_s and τ that maximizes the log-likelihood function are associated with values for μ, σ and λ that are very close to the values used when simulating for both the standard and centered parameterizations.

Chapter 7

Practical Application

7.1 Introduction

The properties and techniques that have been studied in order to fit and classify distributions can now be applied to real-life data of general insurance claims sizes. In Chapter 2 the key components of a general insurance risk model were introduced. The techniques gathered through Chapters 3 to 6 is in particular useful to model and understand the risks associated with claims severity. This may in turn enable the practitioner to have an improved view of the portfolio of insured risks and how risks coming from the upper tails of distributions may influence factors such as reserving, negotiating reinsurance agreements and qunatifying the likelihood of ruin.

Furthermore it is of importance to interpret results once these techniques are applied to gain an understanding of which distributions perform well and how the goodness-of-fit compares amongst the various fitted distributions. These aspects will be discussed in the remainder of this chapter.

7.2 Goodness-of-fit Assessment

Once a distribution is fitted, it is of importance to assess how well it fits and in particular to assess the ability of the fitted distribution to capture the behaviour of the observed data across the entire range of observed values including the upper tails.

For the purpose of evaluating the goodness-of-fit in this practical application, we will consider the use of the following criteria:

- The likelihood ratio test to assess the significance of the fitted parameters.

- Quantile plots to assess the goodness-of-fit over the range of observed values together with the correlation between the observed and theoretical quantiles as a single measure to quantify the overall goodness-of-fit.

7.2.1 Likelihood ratio test

Consider a general hypothesis for the fitted parameters of a distribution, where $\underline{\theta}$ is the vector of distribution parameters. Let $\underline{\theta}_0$ be the vector of parameters that miximizes the likelihood function under the null hypothesis and $\underline{\theta}_1$ be the vector of parameters that can vary over all possible values of the alternative hypothesis. Let the likelihood functions under the null and alternative hypotheses be denoted as follows:

$$L_0 = L(\underline{\theta}_0) \text{ and } L_1 = L(\underline{\theta}_1).$$

Define a test statistic $T = 2\ln\left(\frac{L_1}{L_0}\right) = 2\left(\ln(L_1 - L_0)\right) = 2(l_1 - l_0)$. Reject H_0 if $T > c$ with critical value c a value such that $P(T > c) = \alpha$ with α being the chosen level of significance [75].

Coles [31] highlights the use of this test for the purpose of model selection and explains it as follows. Suppose M_1 represents the model with parameter vector $\underline{\theta}$ and M_0 represents a subset of the model M_1 obtained by constraining k of the parameters in $\underline{\theta}$.

Let $l(M_1)$ be the maximized log-likelihood for model M_1 and let $l(M_0)$ be the maximized log-likelihood for model M_0. Let $D = 2(l(M_1) - l(M_0))$, which is the deviance test statistic. This statistic will have a Chi-square distribution with k degrees of freedom. The fitted model parameters for model M_1 will be considered to be significant if $D > c$ with the critical value c such that $P(D > c) = \alpha$. This will be the case where the constraint on model M_0 is for all parameters to be 0.

To conclude, in order to test a global hypothesis for the fitted model of the form:

$$H_0 : \hat{\underline{\theta}} = \underline{0}$$
$$H_1 : \hat{\underline{\theta}} \neq \underline{0},$$

let $L(\underline{\theta})\Big|_{\underline{\theta}=\underline{0}} = L_0$ and $L(\underline{\theta})\Big|_{\hat{\underline{\theta}}=\underline{0}} = L_{max}$.

The test statistic is $T = -2\ln\left(\frac{L_0}{L_{max}}\right) \sim \chi^2(r)$ where r denotes the number of estimated parameters. Reject the null hypothesis if $T > c$ with the critical value c a value such that $P(T > c) = \alpha$ with α be the chosen level of significance. This will test if the model parameters are jointly significant.

7.2.2 Quantile plots

Rolski et al [105] discuss the use of quantile plots to evaluate goodness-of-fit of fitted distribtions as well as a technique providing visual insight to the success of the fitted distribution to capture the tail behaviour seen in the observed data.

The quantile function is defined in Definition 26 in terms of the generalized inverse function which is defined in Definition 25. One can also construct the empirical equivalent of the quantile function, which is given by $Q_n(y) = Q_{F_n}(y)$. For an ordered sample of size n the following two events are equivalent:

$$\{Q_n(y) = U_{(k)}\} = \left\{\frac{k-1}{n} < y \leq \frac{k}{n}\right\}$$

where $U_{(k)}$ denotes the empirical cumulative distribution function associated with the k^{th} order statistic from the ordered sample.

Due to the fact that most of the distributions we consider are defined on the complete positive real line, we will have that for the largest order statistic the empirical cumulative distribution function will have a value of 1. This will relate to a theoretical quantile function value of $Q_F(1) = \infty$. For this reason it makes sense to apply the continuity correction in which case we'll have the following two events to be equivalent:

$$\{Q_n(y) = U_{(k)}\} = \left\{\frac{k-1}{n+1} < y \leq \frac{k}{n+1}\right\}$$

For increasing sample sizes the impact of this correction will become almost negligible, but will still ensure that a finite value for the theoretical quantile function will be obtained.

One can derive theoretical quantile functions for various theoretical distributions dependeing on whether the cumulative distribution functions exist and whether closed-form expressions for these cumulative distribution functions can be found algebraically. Beirlant et al [16] also indicate the use of, amongst others, the Pareto quantile function as a basis for evaluating the goodness-of-fit for fitted Pareto distributions and to also assess the accurate capturing of the observed tail behaviour.

The correlation between the empirical quantile function and the theoretical quantile function can give a quantification of the overall goodness-of-fit in a single value. In the context of Extreme Value distributions Kotz and Nadarajah state that a correlation coefficient test can be done based on the correlation between the sample order statistics and their expected values.

These expected values are essentially the quantiles implied by the proposed parametric distribution [76].

Quantile plots provide quantitative value, in addition the visual interpretation. For the Exponential distribution it follows from (4.1) that $Q_F(y) = -\frac{1}{\lambda}\ln(1-y)$. If the plot suggests that the relationship between $Q_F(y)$ and $Q_n(y)$ is close to linear then (i) the underlying claim size distribution is close to an Exponential distribution and (ii) hence a straight line can be fitted through the scatter plot using simple linear regression.

The simple linear regression model can be represented as a linear model as follows [65]:

$$y = \beta_1 X + \varepsilon.$$

If we have a sample of n observations, the coefficient β_1 can be estimated using ordinary least square estimation. From Steyn et al [112] follows that a least square estimator for β_1 is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}.$$

In the context of our quantile function based on the observed claim sizes we have that

$$y_i = \Theta_{(i)} \text{ and } x_i = Q_F\left(\frac{i}{n+1}\right).$$

If the underlying distribution is the same as the proposed theoretical distribution then the slope of the line passing through the coordinates given by $\left(Q_F\left(\frac{i}{n+1}\right), U_{(i)}\right)$ should theoretically be 1 and in practical applications close to 1. Hence if we consider the following:

$$U_{(i)} = Q_F\left(\frac{i}{n+1}\right) = -\frac{1}{\lambda}\ln(1-y),$$

which is of the form $y = \beta_1 X$. Hence we can estimate λ as

$$\hat{\lambda} = -\frac{\sum_{i=1}^{n}\left(Q_F\left(\frac{i}{n+1}\right)\right)^2}{\sum_{i=1}^{n} Q_F\left(\frac{i}{n+1}\right)U_{(i)}}.$$

The reasonability of fitted regression is usually assessed by means of the associated R^2 (coefficient of determination) value. Given the relationship between R^2 and the correlation between the dependent and independent variables in a simple linear regression model, one can evaluate the correlation from which the R^2 value automatically follows [112]. Consider the correlation between the observed quantiles and quantiles implied by the proposed distribution:

$$corr(u, Q_F(\cdot)) = \frac{\sum_{i=1}^{n}\left(u_{(i)} - \overline{u}\right)\left(Q_F\left(\frac{i}{n+1}\right) - \overline{Q}_F\right)}{\sqrt{\sum_{i=1}^{n}\left(u_{(i)} - \overline{u}\right)^2 \sum_{i=1}^{n}\left(Q_F\left(\frac{i}{n+1}\right) - \overline{Q}_F\right)^2}}$$

where

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{n} u_{(i)} \text{ and } \overline{Q}_F = \frac{1}{n}\sum_{i=1}^{n} Q_F\left(\frac{i}{n+1}\right).$$

7.3 Claims Size Data

7.3.1 Background

Data from a South African provider of non-life insurance was obtained. This included claims sizes arising from various types of insured risks. The insurer provides cover of special risks generally not covered by most insurance providers. Claims may be submitted directly by individuals or by a direct insurer with whom an individual has an insurance contract with.

As mentioned in Chapter 2 a suitable segmentation needs to be applied to consider subgroups of insured risks that are homogeneous. For this purpose we considered the main lines of insurance provided as a criteria for applying a suitable segmentation and decided to use the line of insured risks with the largest number of claims.

Based on the insurance provider's claims procedure that needs to be followed when submitting a claim, it follows that policies in respect of the line of insurance we considered do not necessarily have an excess included, unless it has been voluntarily elected by the insured party. With the data made available for the purpose of this application, such excess amounts were not provided and as such all claims in the data were treated as if no excess amounts were in force.

In practice most insurance providers introduce excess amounts to certain insured risks in order to eliminate high frequency of very small claims and to reduce premiums payable by insured parties in respect of their policies. This means that observed claims are truncated or censored in terms of both the likelihood of a loss event and in terms of the size of the claim, since the loss suffered by the insurance provider excludes the amount of the excess payable by the insured party. This means that modifications need to be made to the observed frequency of claims as well as the observed claims sizes [105]. One may use theory developed in for truncated distributions - see for example Jawitz [69] - when excess amounts are present. Alternatively the excess amounts, if kept on a database, can be added to the observed claim in order to calculate the true full loss amount. One may then proceed in modelling these loss amounts. Such an approach may work well, especially in cases where excess amounts may vary from one insured party to the next.

7.3.2 Description of data and segmentation

Some lines of insured risks had very few claims historically which makes it to difficult to apply the techniques developed in Chapters 3 to 6. The purpose here is to illustrate how these techniques can be applied in practice. For this reason it was decided to focus only on one of the segments for which a sufficient number of claims were available.

The historical claims sizes, expressed in ZAR (South African Rand) spanned over a few calendar years. As such it was deemed prudent to adjust the historical values for the effect of inflation. The inflation as implied by the consumer price index (CPI), as publised by Statistics South Africa (StatsSA) was used to perform this adjustment.

The final dataset considered consisted of 2966 historical claims from a single segment of insured risks adjusted for inflation. Tests for outliers were not performed, since we specifically wanted to keep the more extreme observations in the data in order to assess the different distributions' capability of capturing the upper tail behaviour accurately.

Table 7.1 shows the key summary statistics of the set of observed claims. The empirical distribution is shown in Figure 7.1. From this figure it can be seen that the distribution of the observed claims is extremely skew. This is supported by a very large value for the skewness given in Table 7.1 together with the fact that median is so much smaller than the mean - 85% of the observed values are smaller than the mean.

Number of observations	2966	Mean	57275.61
Median	9755.52	Mode	6239.52
Standard Deviation	178040.50	Skewness	5.70
Kurtosis	39.59		
Maximum	2345099.55	99^{th} percentile	995336.68
Minimum	127.95	1^{st} percentile	257.38

Table 7.1: Descriptive Statistics of the Observed Claims

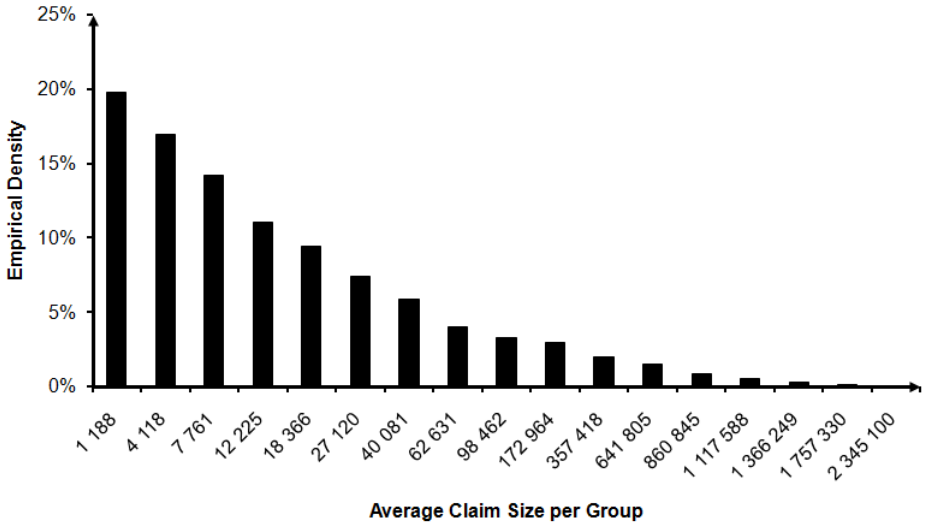

Figure 7.1: Empirical Distribution of the Observed Claims

The descriptive statistics in Table 7.1 and Figure 7.1 indicate that there is a considerable weighth in the upper tail of the observed claims distribution. The ratio of the sample variance to the square of the sample mean is 9.66, which case it follows from Section 3.5.4 that based on this ratio there is no evidence against the hypothesis of the distribution having a heavy tail.

7.4 Fitting of Distributions

In the literature it is often seen that claim sizes are being modelled by positively skewed distributions such as the Lognormal, Loggamma, Gamma, Pareto and Burr distributions [66], [25], [41]. In the context of reserving and estimating IBNR claims the Lognormal appears to be considered most often as an assumed distribution for the incremental claims sizes [102], [83], whilst the Normal distribution is also assumed in some instances [108], [84].

With the summary statistics in Table 7.1 clearly indicating that the data under consideration is highly skewed it makes sense in investigating the skew distributions discussed in Chapters 4, 5 and 6. Furthermore, due to the fact that the Lognormal is used so often, this practical application is also aimed to assess the goodness-of-fit of this distribution relative to alternative distributions and to assess its robustness in being able to fit the distribution easily to various datasets with claims sizes.

7.4.1 Approach

The approach followed was to consider each of the distributions discussed in Chapter 4 and to then apply the estimation techniques as derived in Chapter 6 to fit the specific parametric distributions to the observed claims data.

After fitting each of the distributions an assessment was conducted by means of a quantile-quantile plot to determine whether the fitted distribution provides an adequate fit to the observed data across the full range of observed values including the upper tail. The findings related to the quality of each of the parametric distributions' goodness-of-fit are discussed in the next section.

7.4.2 Results

In this section we briefly discuss the goodness-of-fit as observed from the quantile-quantile plot for each of the fitted parametric distributions.

1. **Gamma Distribution**

   The estimation algorithm reached convergence quite quickly. From the The quantile-quantile plot in Figure 7.2 it can be seen that the fit was generally poor with the density too high on the lower range values whilst the tail of the observed data is heavier than the fitted distribution's tail.

2. **Birnbaum-Saunders Distribution**

   The estimation algorithm reached convergence quite quickly. Similar to the Gamma distribution, as shown in Figure 7.3, the fit was generally poor with the density too high on the lower range values whilst

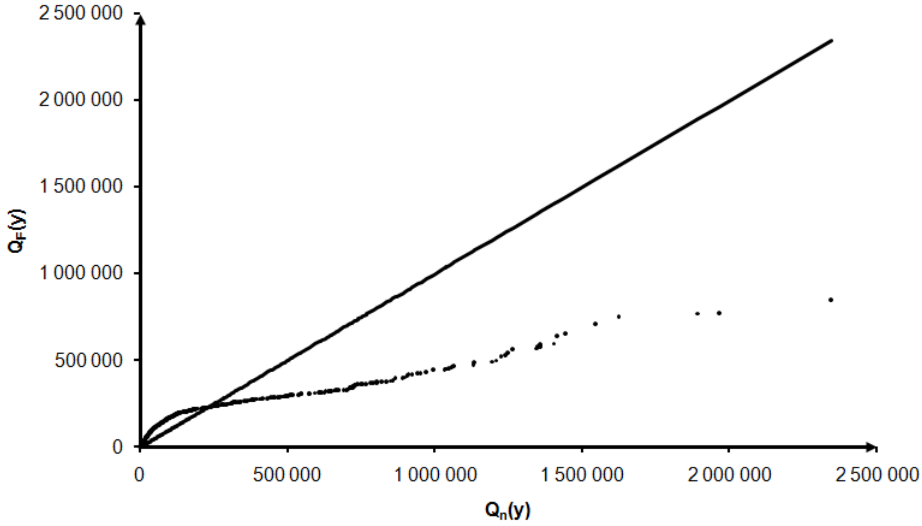

Figure 7.2: Quantile-quantile plot: Fitted Gamma Distribution

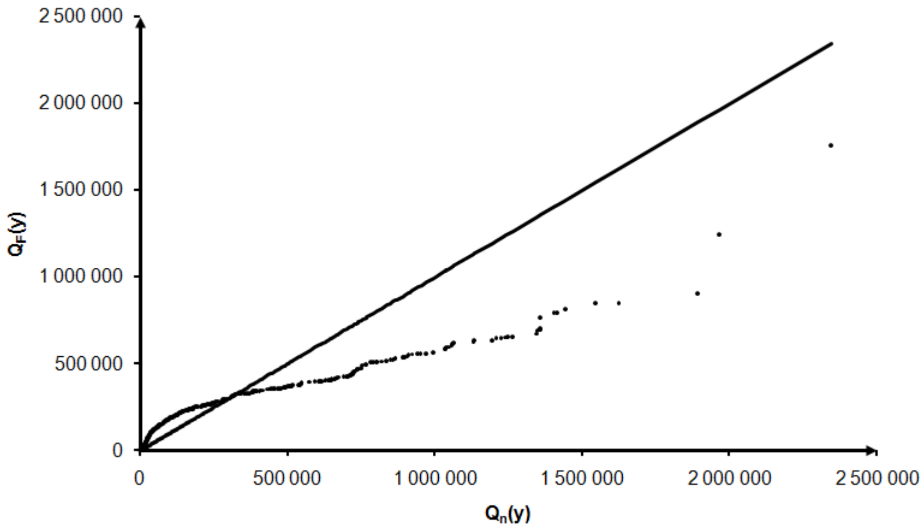

Figure 7.3: Quantile-quantile plot: Fitted Birnbaum-Saunders Distribution

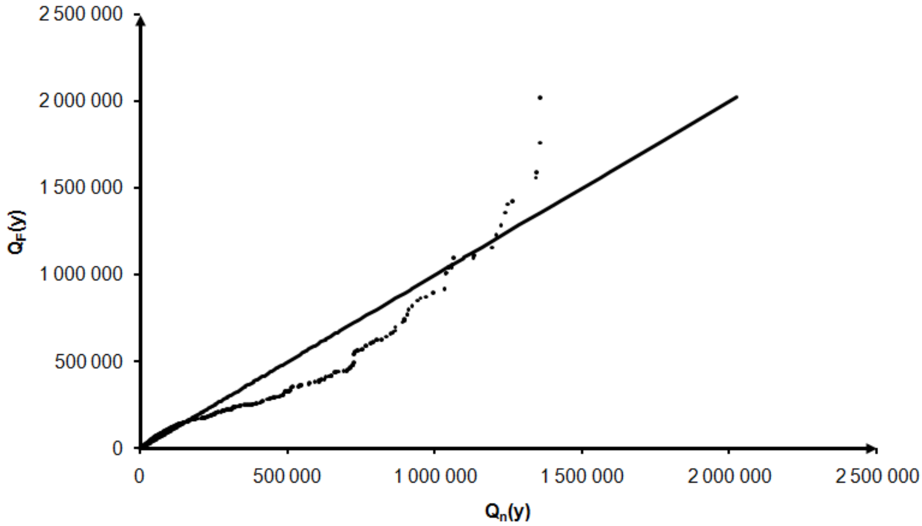

Figure 7.4: Quantile-quantile plot: Fitted Loggamma Distribution

the tail of the observed data was heavier than the fitted distribution's tail.

3. **Loggamma Distribution**

   The estimation algorithm reached convergence. In Figure 7.4 it can be seen that the tail of the fitted distribution was too heavy. On the lower 90% of the values the fitted distribution provided a fit to the data that was reasonable, but with densities being too low in the middle range of the observed values.

4. **Inverse Gaussian Distribution**

   The estimation algorithm converged. Similar to the Loggamma distribution, the tail of the fitted distribution is too heavy. On the lower 90% of the values the fitted distribution provides a fit to the data that is reasonable , but with densities being too low in the middle range of the observed values. Figure 7.5 gives a comparison of the theoretical and observed quantiles.

5. **Generalized Pareto Distribution**

   The estimation algorithm is very sensitive to the choice of initial values. The final fitted distribution had a tail that was too heavy in comparison with the empirical distribution of the observed values. The

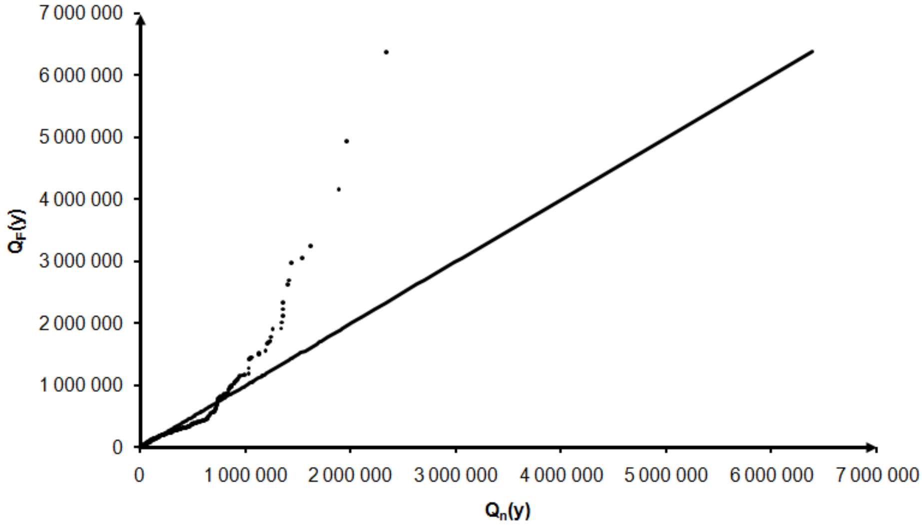

Figure 7.5: Quantile-quantile plot: Fitted Inverse Gaussian Distribution

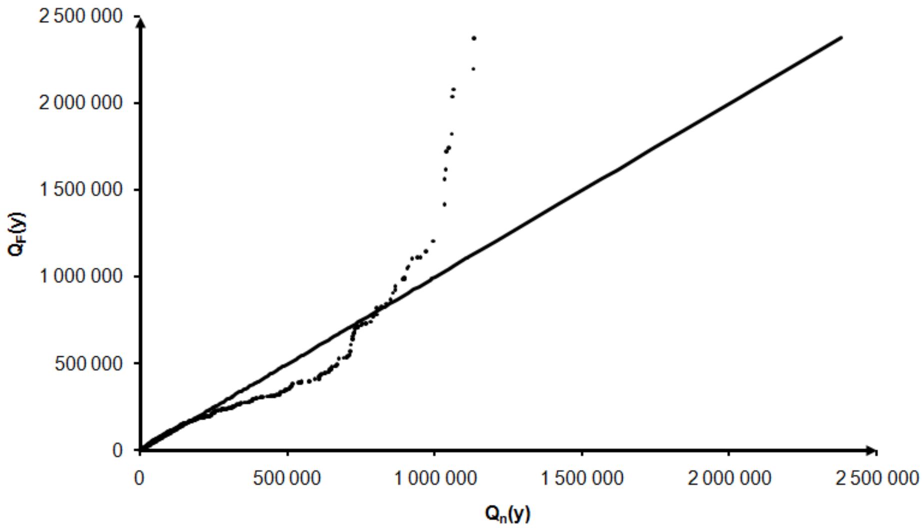

Figure 7.6: Quantile-quantile plot: Fitted Generalized Pareto Distribution

goodness-of-fit on the middle range of values was also not good. A comparison of the fitted distribution's theoretical quantiles with the observed quantiles is given in Figure 7.6.

6. **Pareto Distribution**

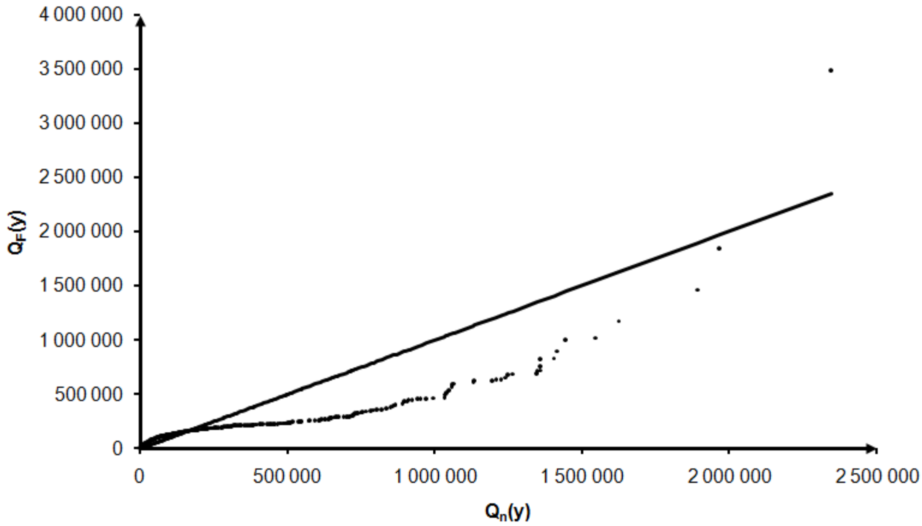

Figure 7.7: Quantile-quantile plot: Fitted Pareto Distribution

In Figure 7.7 it can be seen that distribution does not provide an adequate fit to the observed data.

7. **Lognormal Distribution**

   As graphically presented in the quantile-quantile plot given Figure 7.8, the fitted distribution appeared to be reasonable in that it provided an adequate fit over a very large range of values. In the upper tail of the fitted distribution the tail was too heavy, but the heaviness was to a lesser extent to what was observed for the Loggamma, Inverse Gaussian and Generalized Pareto distributions.

8. **Folded Normal Distribution**

   As suggested by the quantile-quantile plot given in Figure 7.9, the fitted distribution completely lacked any fit to the observed values with the upper tail weight being completely too light.

9. **Rayleigh Distribution**

   This fitted distribution complete lacked any fit to the observed values with the upper tail weight being completely too light - see Figure 7.10.

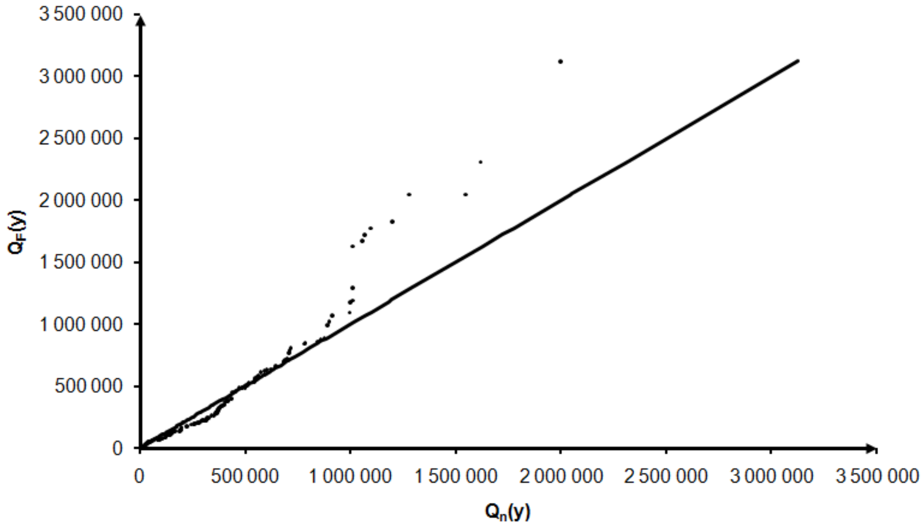

Figure 7.8: Quantile-quantile plot: Fitted Lognormal Distribution

10. **Log-logistic Distribution**

    The fitted distribution had properties that were similar to that of the Loggamma and Inverse Gaussian distributions , as shown in Figure 7.11, in that it provided a fit that was reasonable to the lower range of observed values with the middle range's probability densities being underestimated and the upper range having a tail that was too heavy.

11. **Burr Distribution**

    The estimation algorithm is sensitive to the choice of initial values in order to obtain convergence. Convergence was obtained when applied to the observed values. The heaviness of the tail of the fitted distribution was too extreme. When excluding the 20 largest values the quantile-quantile plot as given in Figure 7.12 indicates that the fitted distribution provided a good fit on the lower range of observed values only.

    It was shown in Section 4.1.19 that the Singh-Maddala distribution is a special case of the Burr distribution with parameters $c = k = q$ and $\alpha = b$. For the distribution that we fitted to the observed claims, the estimated values of c and k were 1.19 and 0.71, respectively. This indicated that the fitted Burr distribution was not the special case.

12. **Weibull Distribution**

    The fitted distribution provided a very poor fit to the observed values with a tail weight that was extremely heavy as shown in Figure 7.13.

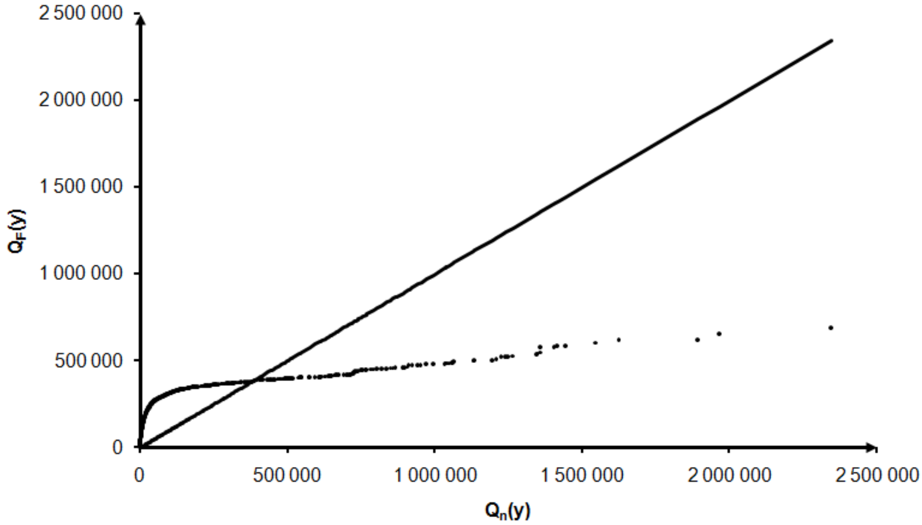

Figure 7.9: Quantile-quantile plot: Fitted Folded Normal Distribution

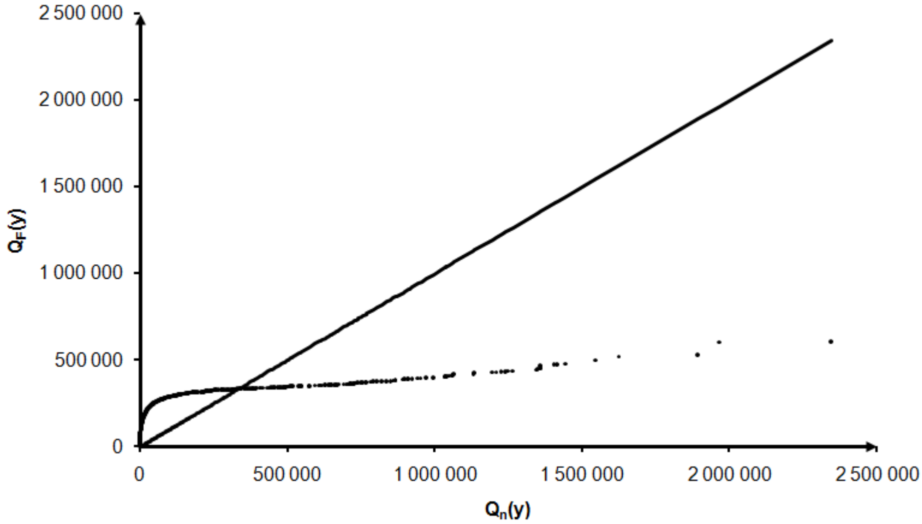

Figure 7.10: Quantile-quantile plot: Fitted Rayleigh Distribution

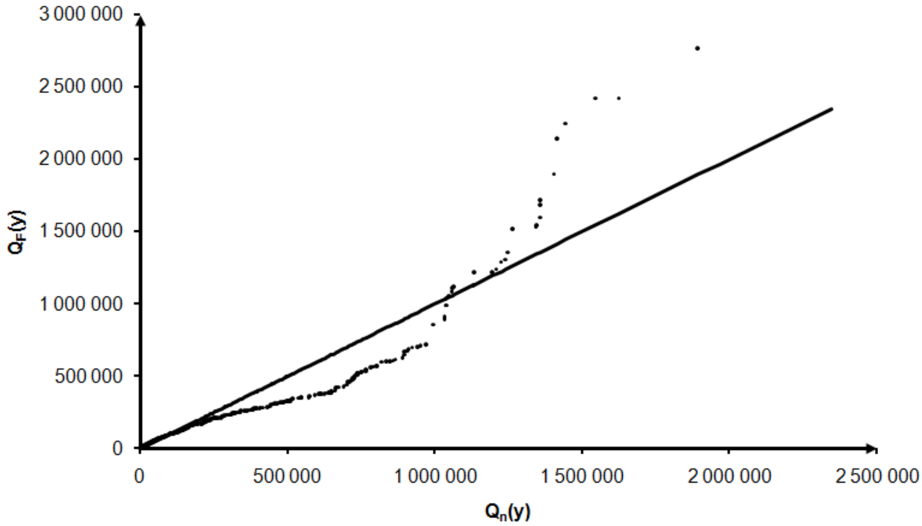

Figure 7.11: Quantile-quantile plot: Fitted Log-logistic Distribution

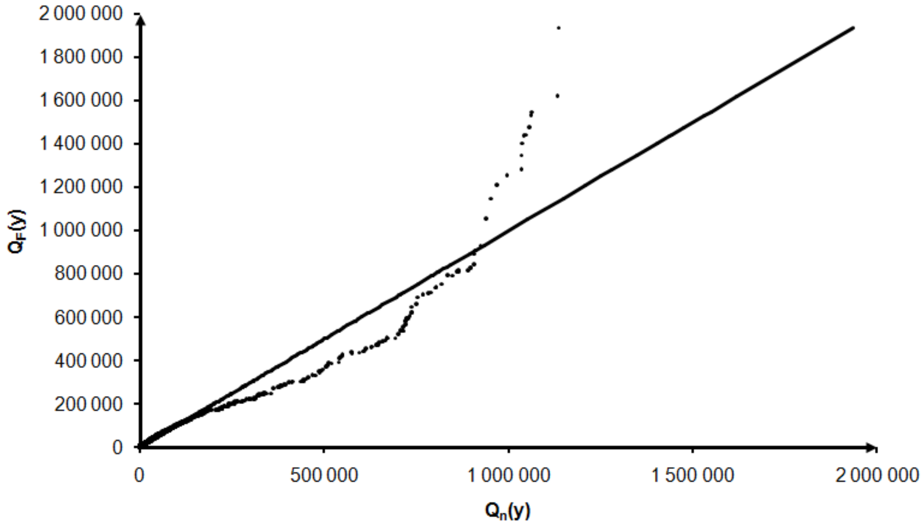

Figure 7.12: Quantile-quantile plot: Fitted Burr Distribution

13. **Gumbel Distribution**

    The maximum likelihood estimation algorithm did not converge. For this reason it was decided to use the regression approach as described in Section 6.5.10 to obtain parameter estimates for the Gumbel distribution based on the observed values. This fitted distribution did not provide a good fit overall or on any smaller range of values - see Figure 7.14.

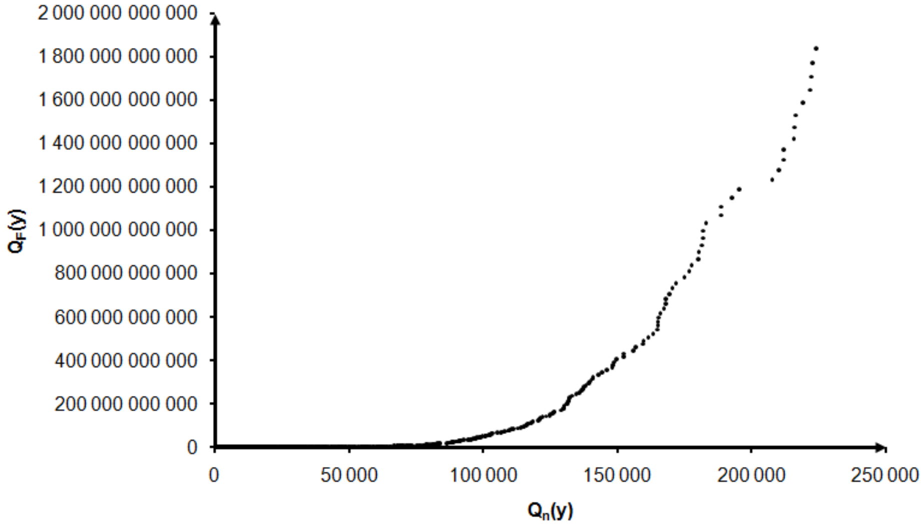

Figure 7.13: Quantile-quantile plot: Fitted Weibull Distribution

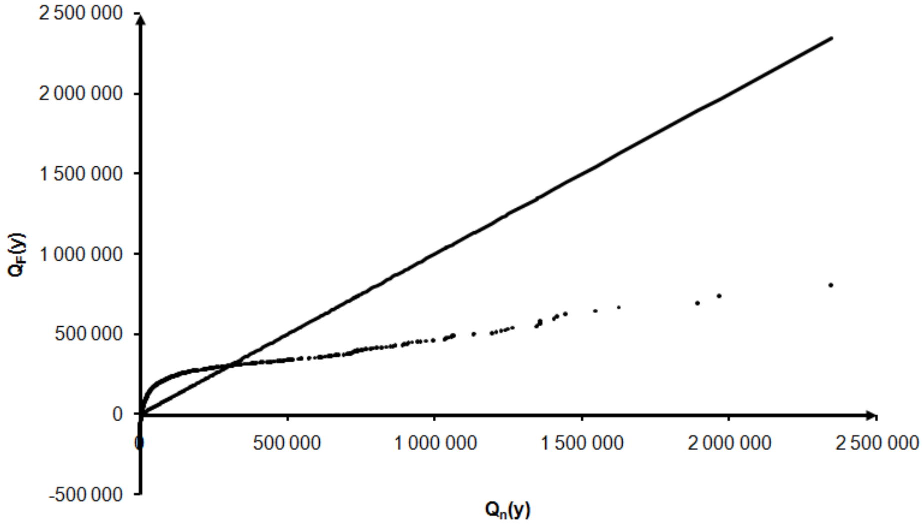

Figure 7.14: Quantile-quantile plot: Fitted Gumbel Distribution

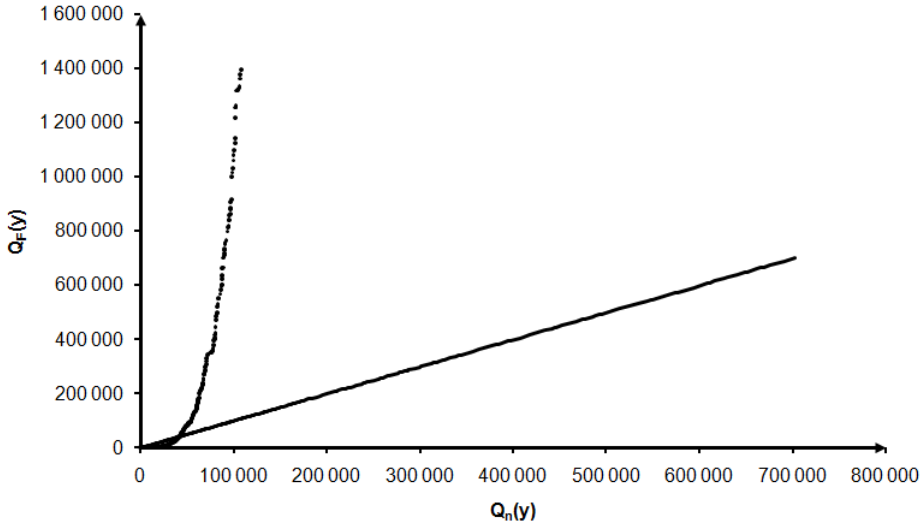

Figure 7.15: Quantile-quantile plot: Fitted Frechet Distribution

14. **Frechet Distribution**

    It was difficult to obtain good convergence as it was occurring at a slow rate. Figure 7.15 shows that the final fitted distribution had an extremely heavy tail and therefore did not provide an adequate fit to the observed values. This then led to probability densities on the lower values as implied by the fitted distribution that were too low.

15. **Exponential and Two-parameter Exponential Distribution**

    For both the Exponential distribution and the Two-parameter Exponential distribution the fitted distribution lacked adequate fits to the observed data with the upper tails of these fitted distributions being too light. This is clearly illustrated by the quantile-quantile plot given in Figure 7.16.

16. **Chi-square Distribution**

    The shape of this distribution implies a relationship between the mean and the variance in that both depends on the parameter ν which led to the distribution not providing an adequate fit to the observed values.

17. **Beta-prime Distribution**

    Figure 7.17 shows that the fitted distribution did not provide a good fit to the observed values at all while the fitted tail was too heavy.

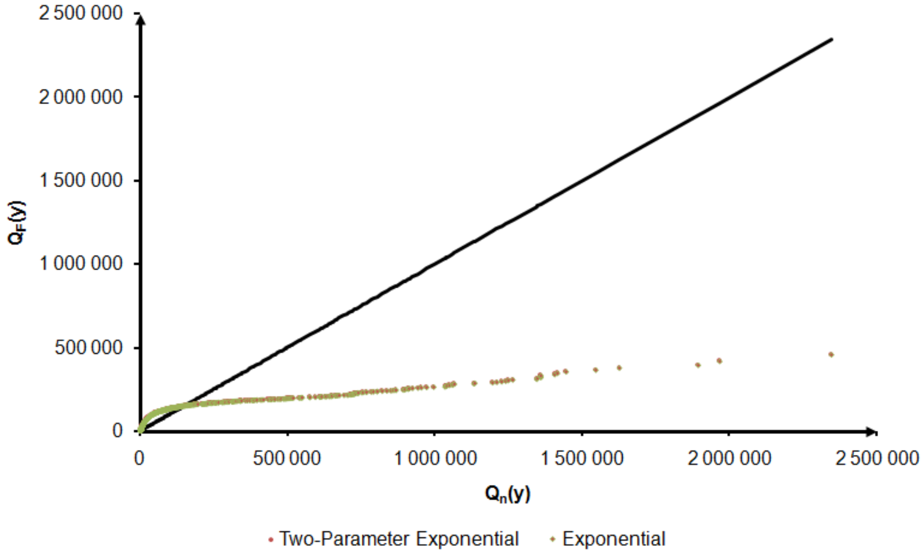

Figure 7.16: Quantile-quantile plot: Fitted Exponential and Two-parameter Exponential Distributions

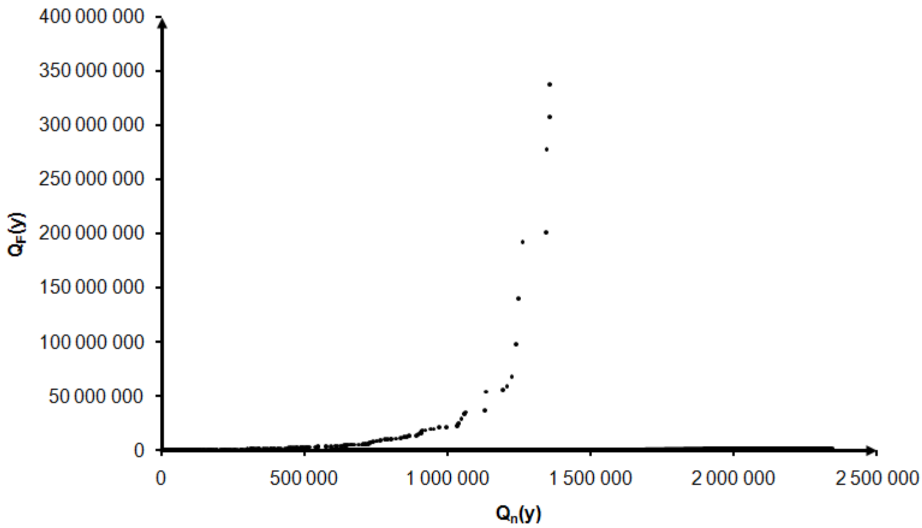

Figure 7.17: Quantile-quantile plot: Fitted Beta-prime Distribution

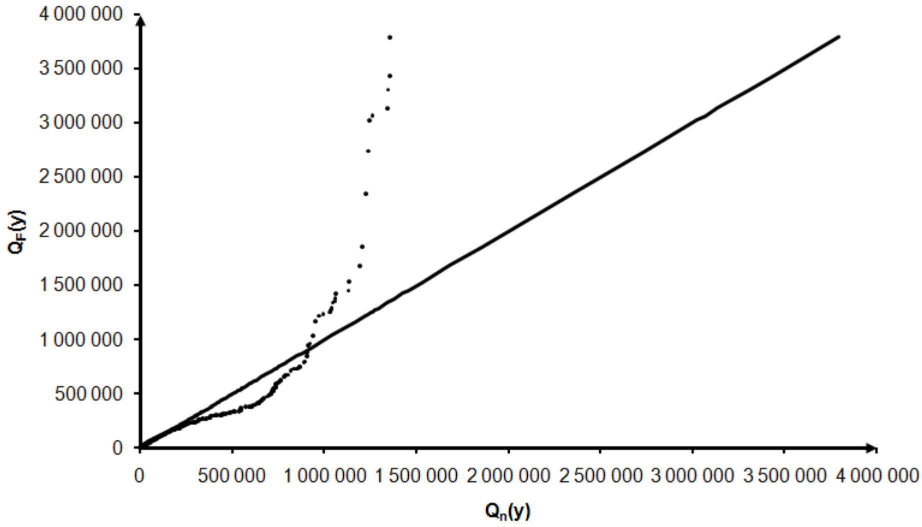

Figure 7.18: Quantile-quantile plot: Fitted Dagum Distribution

18. **Dagum Distribution**

    The estimation algorithm initially moved into a state of oscillation, where this oscillation was between two sets of parameter values being calculated in the iterative algorithm. In an attempt to break this oscillation cycle smaller step sizes were introduced in the iterative process. This helped to obtain convergence of the estimation algorithm.

    The fitted distribution had properties that were comparable with those of the Loggamma and Inverse Gaussian distribution in that the distribution gave a reasonable fit for the lower range of observed values. Figure 7.18 shows that the probability densities of the fitted distribution on the middle range were underestimated and that the upper tail weight was too heavy.

19. **Kappa Family of Distribution**

    The estimation algorithm appeared to be sensitive to the choice of initial values. The fitted distribution three- parameter version gave a reasonable fit for the lower range of observed values only. It can be seen in Figure 7.19 that the upper tail weight of the fitted distribution was too heavy.

20. **Inverse Gamma Distribution**

    Method-of-moments estimates of the parameters yielded a distribution with an upper tail that was too light. The maximum likelihood estimation algorithm is very sensitive to the choice of initial values, but

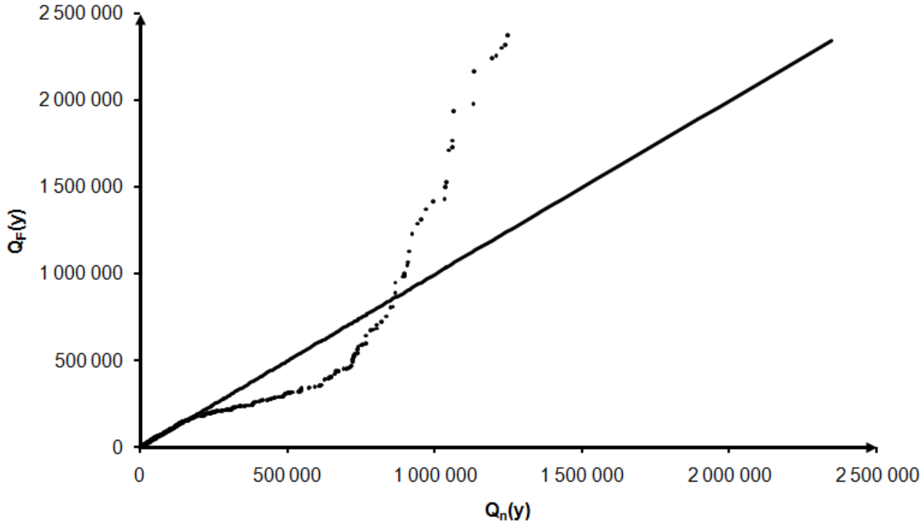

Figure 7.19: Quantile-quantile plot: Fitted Kappa Distribution

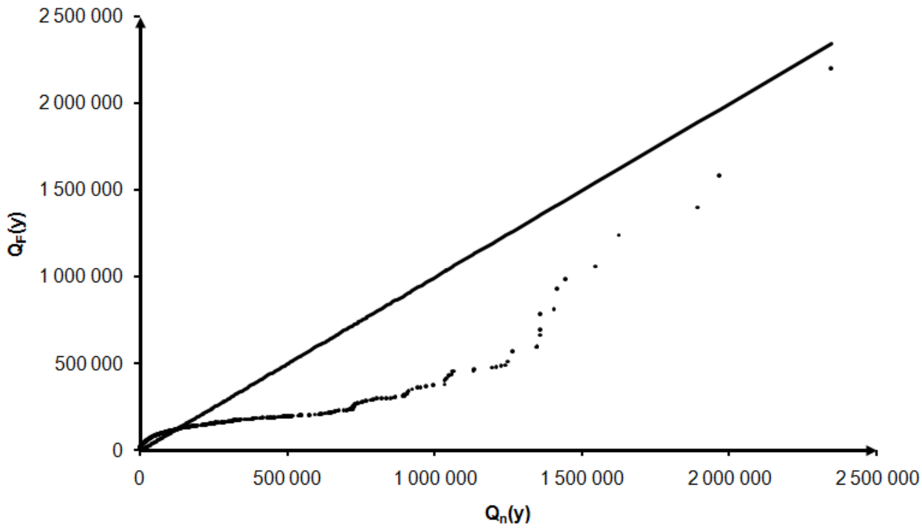

Figure 7.20: Quantile-quantile plot: Fitted Inverse Gamma Distribution

convergence was obtained. Figure 7.20 shows that this yielded a fitted distribution that also had a tail that was too light with the overall being poor.

21. **Inverse Chi-square Distribution**

    This fitted distribution did not provide an adequate fit.

22. **Snedecor's F Distribution**

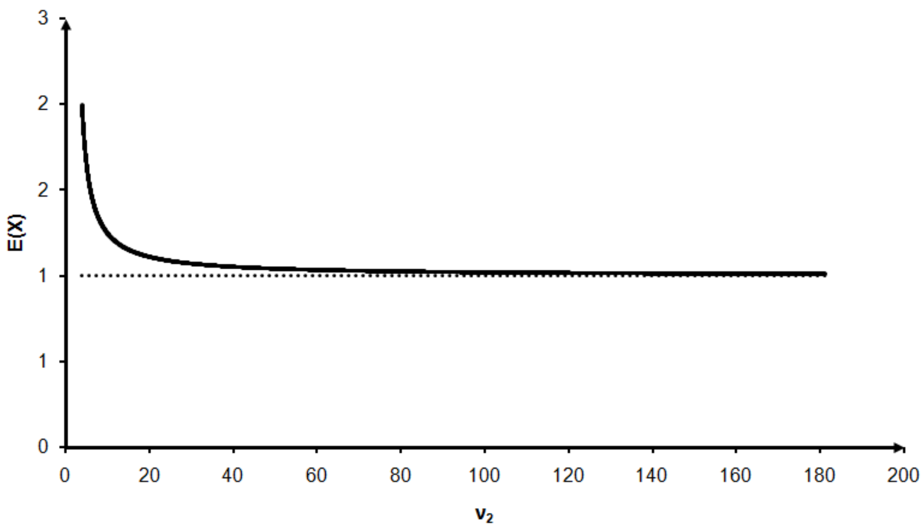

Figure 7.21: Expected value of Snedecor F distributed variable for varying values of ν_2

Algebraically it can be argued that this distribution is not adequate to model the claims sizes as is. For a random variable X that is F distributed with parameters ν_1 and ν_2, we have from Section 4.1.22

$$\mathrm{E}(X^2) = \frac{\left(\frac{\nu_2^2}{\nu_1}\right)(\nu_1 + 2)}{(\nu_2 - 2)(\nu_2 - 4)}.$$

This value can only be positive which implies that $\nu_1 > 0$ and $\nu_2 > 4$. Also for values of $\nu_2 > 4$ the expected value of X is a decreasing function of ν_2 - see Figure 7.21. This means that we will not be able to find an estimate for ν_2 that will yield an expected value that corresponds to the observed mean.

The F distribution will therefore not be suitable to model claims sizes as is. Instead one should consider the use of the F distribution when some transformation of the claims sizes is considered for which the range of possible values from the F distribution makes sense in explaining the behaviour of the transformed random variable.

23. **Skew-normal Distribution**

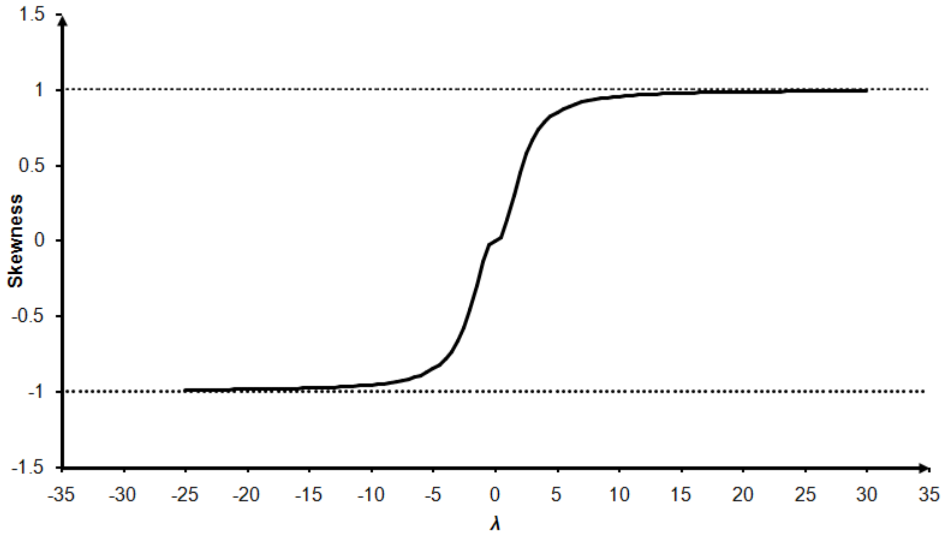

Figure 7.22: Skewness of Skew-normal distribution for varying values of λ

In Figure 7.22 the skewness for the Skew-normal for varying values of λ is represented graphically. From this graph it is evident that as $|\lambda|$ tends to infinity, the skewness tends in absolute value to 1.

In Section 4.1.28 it was given that as λ tends to infinity, the Skew-normal distribution tends to the half-normal distribution as discussed in Section 4.1.24, [8]. The skewness of the half-normal distribution is 1.

The skewness of the observed claims is 5.70 which exceeds the skewness of the Half Normal distribution completely and therefore also exceeds the skewness of the Skew-normal distribution. This therefore suggests that the Skew-normal distribution cannot be fitted to the observed data. An attempt to obtain the method-of-moments estimates supported this argument in that the estimation of the value for λ could not be performed. It is a function of the observed skewness coefficient and the value of the skewness coefficient yielded an argument to the square root being negative.

24. **Generalized Beta Distribution of the Second Kind**

Convergence to the estimation algorithm could not be found. A grid search approach was followed to find a set of parameter values that

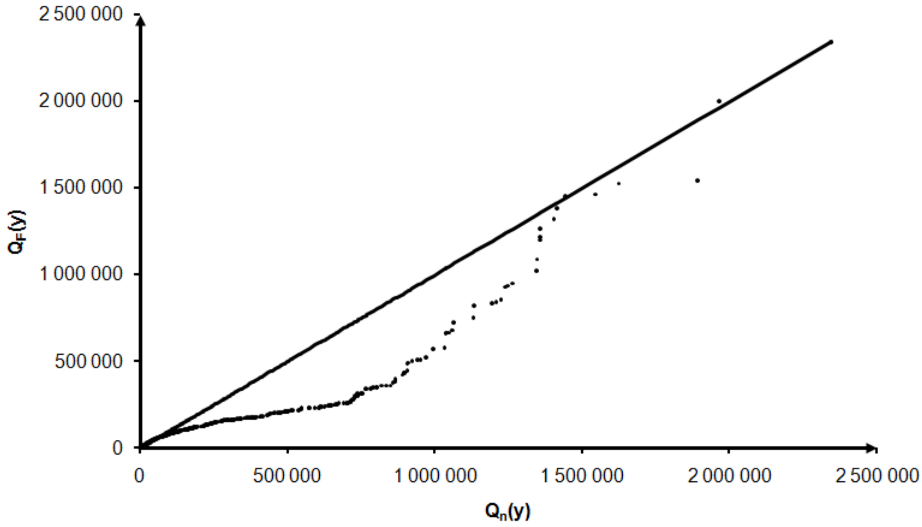

Figure 7.23: Quantile-quantile plot: Fitted Generalized Beta Distribution of the Second Kind

maximized the likelihood function. In Figure 7.23 it can be seen that the goodness-of-fit was poor with the upper tail being too light.

In the items above the goodness-of-fit of the distributions were discussed individually. It is evident that out of the distributions we considered there is no specific distribution that provided an overall good fit to the observed data. From the visual inspection of the goodness-of-fit by using the quantile-quantile plots the Loggamma, Inverse Gaussian and Lognormal distributions provided the best fit the observed data.

Figure 7.24 shows a comparison of the quantile plots for the fitted Lognormal, Loggamma and Inverse Gaussian distributions. The probability density functions of the three fitted distributions are shown in Figure 7.25. This indicates that the Lognormal captures the overall distribution the best with the tail behaviour reasonably well captured, but with a significant underestimation of probability densities in the middle range of observed values. On the middle range of values the Loggamma and Inverse Gaussian also shows an underestimation of densities but to some extent capture it better than the Lognormal.

In Table 7.2 statistics are given that can be used to evaluate the goodness-of-fit of these three distributions as described in Sections 7.2.2 and 7.2.1 based on:

- The correlation between the observed values and the values predicted by the fitted distribution.

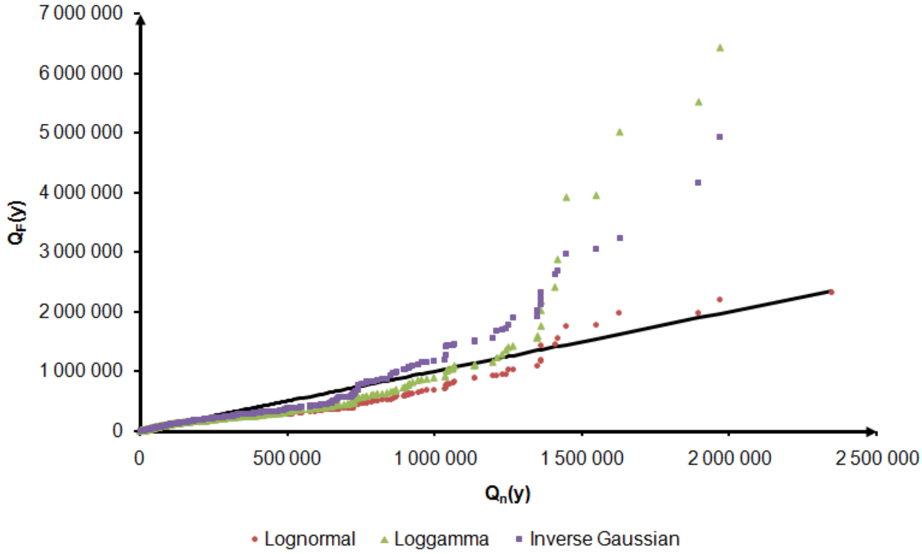

Figure 7.24: Quantile-quantile plots comparison for fitted Loggamma, Lognormal and Inverse Gaussian distributions

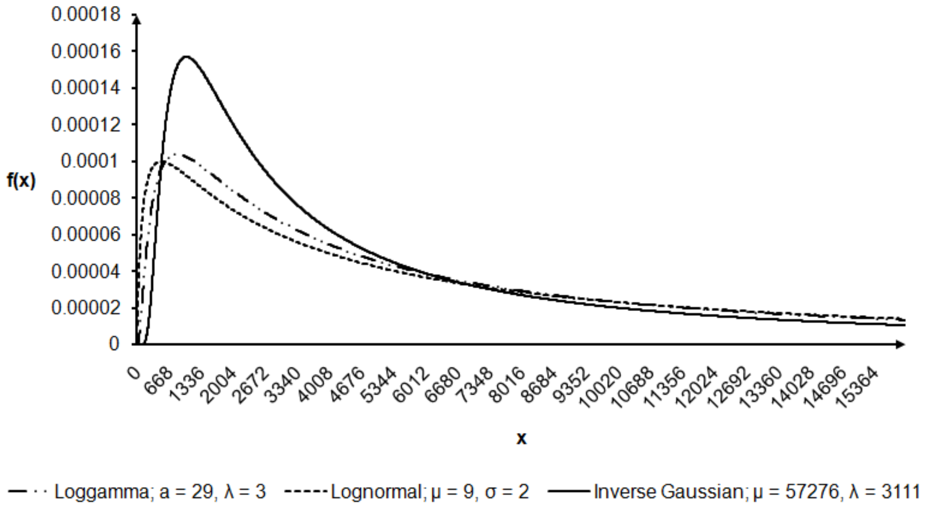

Figure 7.25: Density functions for fitted Loggamma, Lognormal and Inverse Gaussian distributions

- The R^2 and slope of a linear regression fitted to assess on an overall level how much of the variation in the observed claims sizes is explained by the fitted distribution. This is done by fitting the linear regression through the origin and assessing how close the slope is to 1.

Distribution	**Loggamma**	**Lognormal**	**Inverse Gaussian**
Correlation	0.86	0.97	0.95
Slope	1.16	0.77	1.26
R^2	0.73	0.93	0.89
LR Test Statistic	42541	$> 2.62 \times 10^{11}$	$> 3.13 \times 10^{12}$
α	0.05	0.05	0.05
Degrees of Freedom	2	2	2
Critical Value	5.99	5.99	5.99
Parameters Significant	Yes	Yes	Yes

Table 7.2: Goodness-of-fit Statistics of the fitted distributions

- The likelihood ratio (abreviated as LR) test statistic.

From the statistics in Table 7.2 it is evident that all three fitted distributions' parameters are significant.

In some instances industry knowledge and experience can be used to support the use of the one distribution rather than the other. Even though claims have not been experienced as extreme as being suggested by the Loggamma and Inverse Gaussian distributions, the business expectation may be that such large claims are likely to occur in the near future in which case some reasoning exists in order to choose the Loggamma or Inverse Gaussian distribution to model the claims rather than the Lognormal distribution.

The one aspect that is not satisfactory about the goodness-of-fit of any one of these three fitted distributions is the lack of fit in the middle range. The lack of fit gives an underestimation of the likelihood of fairly large claims. The underestimation is therefore likely to result in understatement of expected losses and may therefore severely impact aspects such as reinsurance treaties being modelled, pricing of risks with larger risk exposure in monetary terms, capital reserve being held and potentially misstatement of ruin probabilities.

Klugman et al [75] discuss the use of a splicing method to create new distributions. This technique is somewhat similar to the technique of mixing which is applied to cases where it appears that two or more separate random processes are responsible for generating claims. With the technique of splicing it is assumed that one model is adequate to capture the claims behaviour over one interval while other models are adequate on other intervals. A two-component spliced distribution's density function is defined as follows:

$$f(x) = \begin{cases} a_1 f_1(x) & c_0 < x < c_1, \\ a_2 f_2(x) & c_1 \leq x < c_2 \end{cases}$$

where a_1 and a_2 are constants with values such that $f(x)$ is a legitimate density function and $a_1 + a_2 = 1$.

The tail behaviour of the observed claims may be different from the behaviour of the smaller claims in which case a single parametric distribution may not suffice to capture the behaviour of both the small claims and the upper extreme claims sizes [66]. This serves as a motivation for using a technique such as splicing [75]. We, however, focus in this study on the fitting of the individual distributions.

In recent literature the use of techniques similar to splicing are considered. Cooray and Ananda [32] as well as Teodorescu [117] use a Lognormal-Pareto composite model with the Lognormal distribution for the smaller losses up to some threshold and a Pareto distribution for the values above the threshold. Nadarajah and Bakar [91] discuss methods of finding the weights a_1 and a_2 to find a composite density function that is legitimate, i.e. to ensure that the total area under the probability density function sum to 1. The Lognormal-Pareto and Lognormal-Burr composite distributions are applied by them to model fire insurance claims. Also see Preda and Ciumara [101] for a comparative study of a Weibull-Pareto composite and a Lognormal-Pareto composite distribution. Composites of the Lognormal, Loggamma and Gamma distributions are also discussed by Hewitt and Lefkowitz [66].

This notion of composite distributions is generally described in the literature in a situation where the upper right tail is better described by one distribution and the majority of the data left of the upper right tail better described by another. So generally it means that the domains of these distributions largely overlap with conditions specified to indicate which to use for values above the threshold and which to use for values below the threshold.

In the data that we are currently considering, the scenario is somewhat different in that it appears as if there is one random process underlying the lower part of the range of observed values and another random process underlying the upper part of the range of observed values. This is very likely to occur where different events leading to losses on the same insured risks can be completely different in terms of the likely size of the losses resulting from these events. For property insurance, for example, the loss resulting from a burglary might be very small in comparison to the loss resulting from a fire. This means that one can fit distributions related specific to the less severe and more severe classes of losses, respectively. Note that this is somewhat different from splicing, perhaps closer to the idea of segmentation.

If we inspect the observed claims we have a bit closer and consider splitting the dataset at various threshold levels it does appear as if there may potentially be two random process present generating these claims. More specifically it appears that if we split the data at a value of 350000 the two sets of claims exhibit reasonable distribution forms as shown in Figures 7.26 and 7.27.

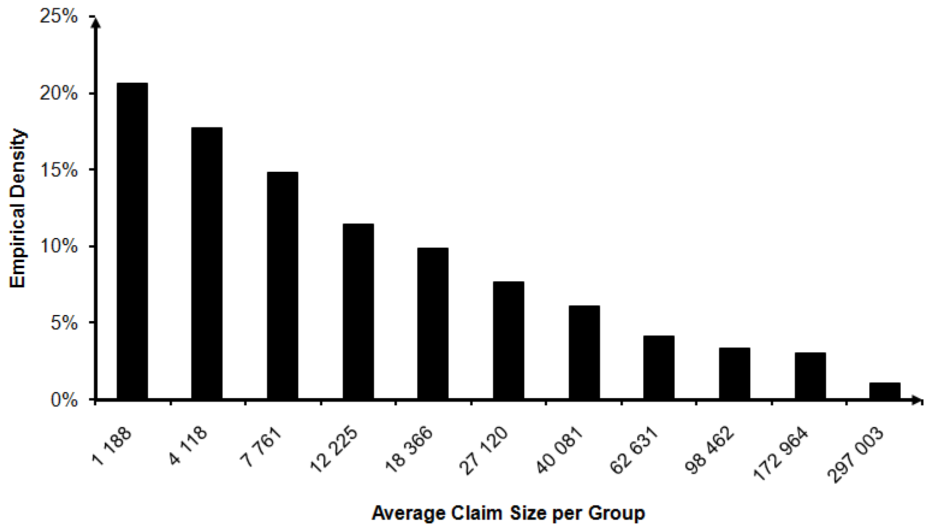

Figure 7.26: Distribution of observed claims smaller than or equal to 350000

In light of the fact that the use of composite distributions are referred to in the literature, it may give an improvement to the fits we obtained with the Lognormal, Loggamma and Inverse Gaussian distributions. For this reason we fitted the parametric distributions to two sets of claims and first visually assessed the goodness-of-fit using the quantile-quantile plots. We present the results of the distributions for which reasonable fits were obtained below.

1. **Claims smaller than the threshold of** 350000
   The distribution of observed claims with values less than 350000 is still extremely skew with a skewness coefficient with a value of 3.87. For this reason the tails of the fitted distributions are generally too heavy.

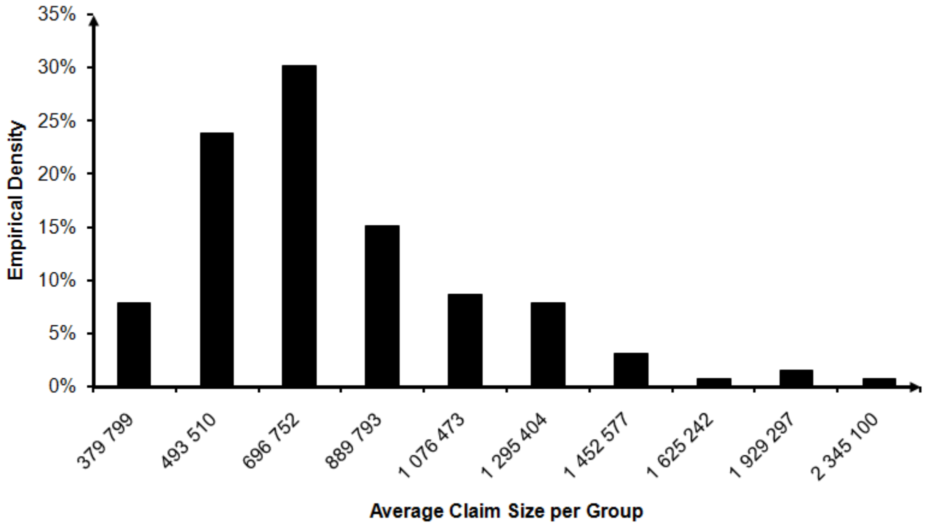

Figure 7.27: Distribution of observed claims larger than 350000

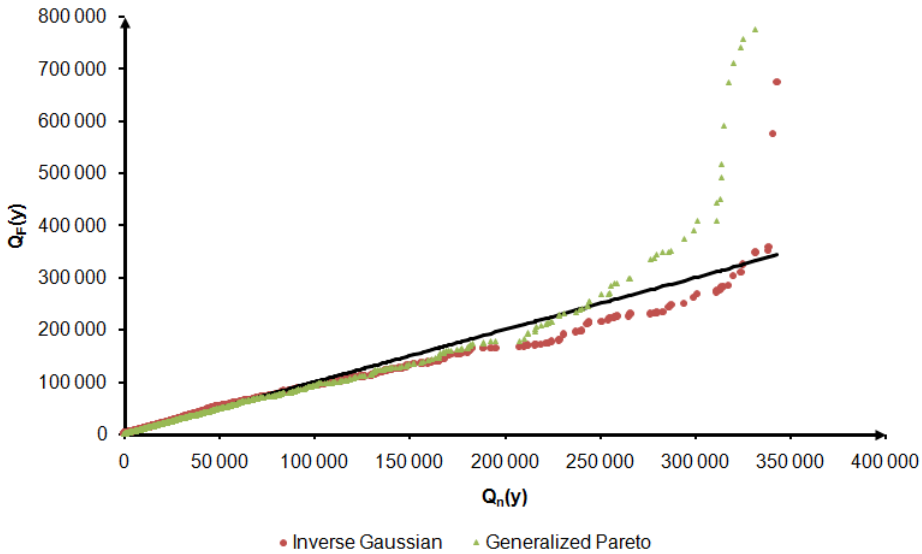

Figure 7.28: Comparison of quantile-quantile plots for distributions fitted to the small claims

The metrics shown in table 7.3 also support the fact that the observed data is very skew. As can be expected, the skewness is lower once the larger claims are excluded to be considered separately. The ratio of the variance to the the square of the expected value gives a value of 3.48 which does not give any prove against the hypothesis of the distribution having a heavy tail.

Number of observations	2840	Mean	24457.41
Median	8981.98	Mode	6239.52
Standard Deviation	45652.18	Skewness	3.87
Kurtosis	17.48		

Table 7.3: Descriptive Statistics of the Observed Claims smaller than 350000

The following distributions yielded the best fits:

- ***Inverse Gaussian Distribution***
  This distribution provides a good fit over the lower half of observed claims with some degree of underestimation of the upper tail weight. The extreme upper tail weight is too heavy, but this is present for less than 0.07% of the observations.
- ***Generalized Pareto Distribution***
  The distribution provides a good fit for the lower and middle ranges of observed values, but has a tail that is too heavy.

The quantile-quantile plots of these two fitted distributions are shown in Figure 7.28. From these plots it can be seen that both distribution show some lack of fit with an underestimation in the middle range while the upper tails are too heavy. The fitted Inverse Gaussian has a tail that is lighter than the tail of the Generalized Pareto distribution and for this reason appears to provide the better. The goodness-of-fit statistics as discussed in Sections 7.2.2 and 7.2.1 are given in Table 7.4. These statistics support the visual interpretation from the quantile-quantile plots and also indicate the parameters of the fitted distributions are significant. The probability density functions of the fitted distributions are graphically compared in Figure 7.29. In general it is evident that there is no specific distribution that provides a very good fit to the observed claims with values less than 350000.

2. **Claims larger than the threshold of** 350000
   There are 126 observed claims larger than the threshold value. The distribution of these claims is still positively skewed with a skewness coefficient of 1.45. As can be seen in Figure 7.27 and from the metrics given in table 7.5, the upper tail of the observed claims' distribution is not that heavy and it can therefore be expected that it will be easier to fit a distribution to these observed claims. The ratio of the variance to the square of the mean yields a value of 0.64 which indicates that the distribution does not have a heavy tail.

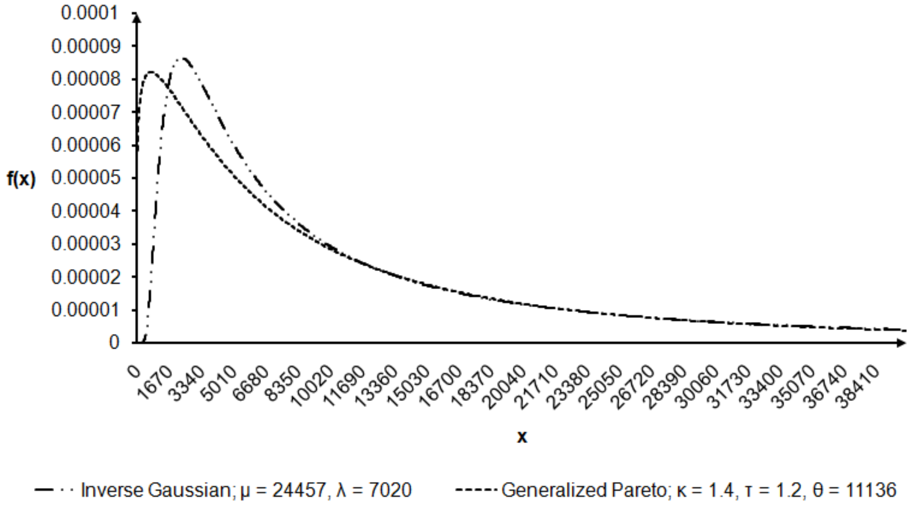

Figure 7.29: Density functions of the fitted Inverse Gaussian and Generalized Pareto distributions

Since most of the distributions considered have domains on $\mathbb{R}^+$, it was decided to subtract the threshold value from all of the claims under consideration and to then fit distributions to these transformed claims sizes. The following distributions yielded the best fits:

- ***Gamma Distribution***
  The overall fit of the distribution is very good across the whole range of observed values. There is a slight indication that the tail might be too light.

- ***Inverse Gaussian Distribution***
  We considered the method-of-moments estimates, since convergence of the maximum likelihood estimation algorithm could not be obtained. This yields a reasonable fit, but with some overestimation of the probability densities on the lower values.

- ***Lognormal Distribution***
  The fitted distribution provides a reasonable fit across a large portion of the range of transformed claim values. The tail of the fitted distribution is too heavy, though.

- ***Folded Normal Distribution***
  Despite the rate of convergence of the iterative estimation being quite slow, the distribution provides a very good fit to the trans-

Distribution	**Inverse Gaussian**	**Generalized Pareto**
Correlation	0.98	0.83
Slope	0.93	1.31
R^2	0.95	0.69
LR Test Statistic	$> 5.31 \times 10^{11}$	32822
α	0.05	0.05
Degrees of Freedom	2	3
Critical Value	5.99	7.81
Parameters Significant	Yes	Yes

Table 7.4: Goodness-of-fit Statistics of the fitted distributions to claims smaller than 350000

Number of observations	126	Mean	446987.37
Median	373045.90	Mode	373045.90
Standard Deviation	358511.94	Skewness	1.45
Kurtosis	2.85		

Table 7.5: Descriptive Statistics of the Observed Claims larger than 350000

formed claims.

- ***Gumbel Distribution***
  The fitted distribution provides a reasonable fit to a large portion of the range of transformed claim values, but in the context of modeling claims sizes the distribution has a shortcoming in that it is defined on the whole real line and therefore has non-zero probability density on the negative half of the real line. The fitted distributions upper tail is also slightly too light.

- ***Two-parameter Exponential Distribution***
  The fitted distribution provides a reasonable fit to the transformed claim values. The upper tail of the fitted distribution is too heavy with the probability densities on the lower range of values being underestimated.

- ***Three-parameter Kappa Family of Distributions***
  The fitted three-parameter Kappa distribution provides a reasonable fit. There is an overestimation of the probability densities on the extreme lower values which results in a slight underestimation of probability densities across the whole range of values. Furthermore the upper tail of the fitted distribution is slightly too light.

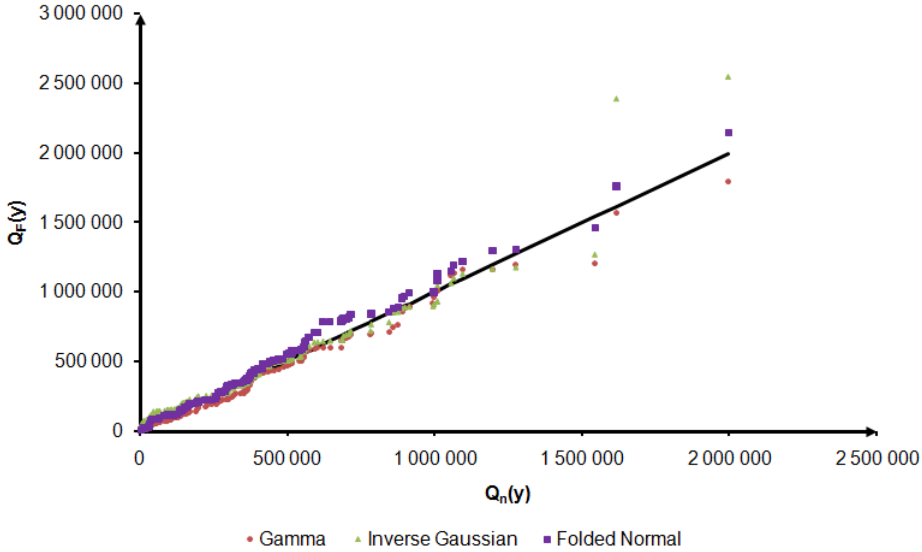

Figure 7.30: Comparison of quantile-quantile plots for distributions fitted to the large claims

Generally the distributions fitted to these larger values provide very good fits to the transformed claim values. The ability of the distributions to capture the tail behaviour is also better compared the ability of the distributions fitted to the smaller claims. From the observations made above for the 7 distributions it is evident that the Gamma, Inverse Gaussian and Folded Normal distributions provide the best fits across the whole range of transformed claim values. A visual comparison of the these three distributions are given in the form of quantile-quantile plots in Figure 7.30. The goodness-of-fit statistics as discussed in Sections 7.2.2 and 7.2.1 are given in Table 7.6.

It is seen in Figure 7.30 that all three fitted distributions provide reasonable fits to the observed claims, but with the Inverse Gaussian having slightly too much weight in the upper tail. The density functions of these three fitted distributions are graphically compared in

Distribution	**Gamma**	**Inverse Gaussian**	**Folded Normal**
Correlation	0.99	0.97	0.99
Slope	0.94	1.04	1.07
R^2	0.98	0.94	0.99
LR Test Statistic	$> 1.13 \times 10^{11}$	$> 1.22 \times 10^{11}$	4.12×10^{19}
α	0.05	0.05	0.05
Degrees of Freedom	2	2	2
Critical Value	5.99	5.99	5.99
Parameters Significant	Yes	Yes	Yes

Table 7.6: Goodness-of-fit Statistics of the fitted distributions to claims larger than 350000

Figure 7.31.

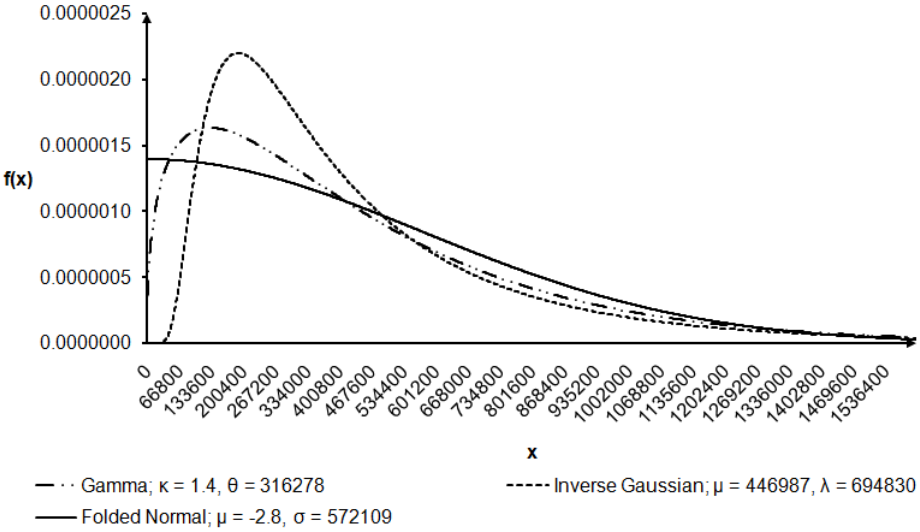

Figure 7.31: Density functions of the fitted Gamma, Inverse Gaussian and Folded Normal distributions

7.4.3 Conclusion

The application of the estimation algorithms as derived in Chapter 6 to the real claims size data illustrated how these techniques can be used in practice. The application highlighted the importance of having suitable initial values chosen for the iterative estimation algorithms. Furthermore it echoed the arguments often found in the literature that it is often difficult to obtain a single distribution that provides a good fit to both the smaller and extremely large observed values [75], [32], [91].

The concept of composite or spliced distributions was considered. The aim of looking at such an approach was not focussed at finding the final composite distribution, but instead used as an argument as to why it would make sense to fit different distributions on different ranges of the observed values. In addition to fitting distributions to all claims together, we used a threshold value to form two sets of claims on two different ranges and then fitted distributions to both the sets, respectively.

With our choice of a threshold value of 350000 we have seen that the distribution shape for the smaller claims is completely different from the shape for the larger claims. This supported the approach to model the small and larger claims separately. Due to the small claims' distribution still being highly skewed, it was difficult to obtain good fitted distributions. For the large claims a few distributions provided very good fits to the observed data. In such a context where it is difficult to obtain a good fitted distribution on a portion of the range, Klugman et al specifically mention the use of composite or spliced distributions where one distribution is parametric and the other empirical [75].

Generally it was seen that the Inverse Gaussian and Lognormal distributions provide good fits to the observed claims irrespective of whether all, only the small or only the large claims are considered. Both of these distributions are related to the Normal distribution which adds some appeal to using these two distributions, since the Normal distribution has well-known properties, is widely known and it is easy to use in analysis, modeling and simulations.

Our choice of threshold value used here of 350000 to distinguish between small and large claims have been decided on after considering a range of possible thresholds. One may, however, consider fitting composite distributions on smaller and larger claims based on various thresholds in order to evaluate if using a smaller or larger threshold value yields better fits to either one or both of the smaller and larger claims. This suggests that the exercise of finding a threshold and fitting multiple distributions may very likely become an iterative process in which the threshold is changed from iteration to another and paramteric distributions re-fitted and the goodness-of-fit re-assessed until satisfactory fitted distributions are obtained.

Finally, more than one potential distribution or more than one set of composite distributions may be obtained that can be fitted to the data. Like we have seen in our real-life application in some instances it might be very difficult to capture behaviour of the underlying risk process on certain ranges of value in which case expert judgement and industry knowledge should play a role to ascertain whether, for instance, a heavier or lighter tail is required. Modeling of claim sizes will ultimately affect estimated aggregate losses which in turn affects pricing strategies, reserving and the ceedance of risks (if reinsurance treaties are present or are considered). Consequently it ultimately has an influence on the probability of ruin.

Chapter 8

Conclusion

The aim of this chapter is to summarize the different components this study consisted of, the results obtained in the application of the derived techniques and to suggest further research topics that may supplement the study conducted.

8.1 Overview of study conducted

The key focus of this study was to consider claims sizes as a stochastic component forming part of a typical general insurance risk model.

The study was started with a detailed literature review which indicated frequent use of and reference to such a risk model. It further revealed that research has been focussed on all components within this model with a strong focus on studying techniques in modeling ruin probabilities in the presence of reinsurance treaties and uncertainty associated with the returns on the investment of premiums collected. Research has also been done on modeling claim frequencies and claims sizes with the Lognormal, Pareto, Weibull and Burr distributions being used most frequently.

The general insurance risk model together with its individual stochastic elements were introduced in Chapter 2. Going forward the focus of this study was to develop techniques that can be used to understand, describe and model claims sizes and in particular the behaviour associated with claims size distributions that are heavy-tailed.

It was argued in Chapter 1 that the distributions of claims sizes often exhibit skewness where the skewness is associated with events leading to very large to extreme losses. From a reserving perspective as well as from a ruin probability perspective it is important to have an understanding of the likelihood of such large losses to be able to predict large claim occurrence with a reasonable level of accuracy. This means that one will have to know how to model the distribution of these losses. Furthermore, if there is a fairly large probability of experiencing very large losses, an understanding of distributions' tail heaviness is required and it should be allowed for in the modeling process in which case distributions with heavy tails can be considered.

A variety of univariate parametric distributions were introduced in Chapter 4 that can be used for modeling claims sizes. These distributions are all either positively skewed or can take on positively skewed forms depending on their parameter values. The key properties of these distributions were introduced together with special parameterizations. Links between parametric distributions have been discussed.
In Chapter 6 techniques were studied that can be used for fitting the parametric distributions that were introduced in Chapter 4. Generally either or both the maximum likelihood and method-of-moments estimation techniques were considered. Since most of the parametric distributions have more than one parameter, the maximum likelihood estimation resulted in becoming a multivariate numerical problem. Consequently, iterative multivariate estimation algorithms were derived and SAS PROC IML programs developed to perform these parameter estimations numerically. For each distribution the estimation algorithm was tested on a sample simulated from the distribution itself after which the efficiency, likelihood of the convergence of the algorithm and the accuracy of the estimates were evaluated and discussed.

Since it was indicated that it is important to have an idea of the degree of upper tail heaviness, an in-depth study was conducted in Chapter 3 to understand what is meant by tail heaviness. The concept of a heavy tail was formally defined. Measures to evaluate tail behaviour such as hazard rates, hazard rate functions, mean residual hazard rate functions and exponential bounds were introduced together with illustrative examples. These measures can be used to describe the upper tail weight of a parametric distribution while some measures can also be used to compare the upper tail weight of two distributions. Based on these measures a set of techniques was derived that were used in Chapter 5 to evaluate the heaviness of the tails of the parametric distributions that were introduced in Chapter 4. It was seen that it is generally not vey easy to confirm that a distribution has a heavy tail or not.

The statistical techniques discussed in Chapters 3 to 6 were then applied to a real-life general insurance claims dataset in order to ascertain how well these techniques may work in practice. This application was discussed in detail in Chapter 7.

8.2 Results and Observations

Based on the literature review it is clear that a great amount of research is available on all components of the general insurance risk model, even on the use of statistical distributions to model claims sizes.

The research conducted revealed that there exists many parametric distributions suitable for modeling claims sizes based on their domain and shape. There are also many variations on parametric distributions, such as the composite and spliced distributions [91], [101] that are considered.

Maximum likelihood is a well-known technique for fitting distributions. It appeared to not always be as easy to use on real-life data with difficulties experienced in order to obtain convergence of the estimation algorithms.

Techniques and metrics to describe and detect tail heaviness proved to often be algebraically complex when applied to parametric distributions. Furthermore it was found that whilst it is often possible to get a counter-argument against the hypothesis of heavy-tailedness in order to confirm that a distribution does not have a heavy tail, it is generally more difficult to confirm that a distribution does have a heavy tail.

The application of the techniques gathered in this study to real-life claims data revealed that it may be difficult to find a single parametric distribution to desribe the distribution of the observed claims. Whilst recognising that one may segment data and perform modeling on the segmented portions, it may still be difficult to find a distribution that fits the data well, especially across the whole range of observed values. In this light it was acknowledged that judgment cannot be applied solely from a theoretical perspective, but also requires business expert judgment on whether fitted distributions make sense.
In the practical application parametric distributions were fitted with very high levels of goodness-of-fit. The general shortcoming was on the upper tails where the fitted distributions' upper tails were generally heavier than the observed distributions' upper tail weights. Such fits may still be adequate if such tail heaviness is considered by business practitioners to be likely, despite not having observed such extreme events historically.

8.3 Future Research

Based on the literature review and study conducted together with the results from the application, the following items were identified to consider for further research:

- **Composite distributions**
  It was realised only during the application stage that it can become difficult to find a single distribution providing an adequate fit to the whole range of observed claims. In the application a segmentation based on the claim size was performed and separate models fitted to these segments with reasonably good results. In the literature the use of composite distributions received considerable attention by authors such as Klugman et al [75], Nadarajah and Bakar [91], Cooray and Ananda [32], Teodorescu [117] and Preda and Ciumara [101]. This technique performs very well if there is a specific distribution that can provide a good fit on the tail or upper range of observed values while another distribution fits well on the lower range of observed values. An overlay of these distributions can then be considered in the form of a branching function where the final fitted model is a single distribution, but with two branches coming from two different parametric distributions. As a result, one will have to incorporate these two parametric forms and the cut-off point between the two branches into the model fitting algorithm.

- **Censoring and treatment of excess amounts**
  The data used in the application did not include any indication of whether excess agreements were in place, what the excess values were and whether the claims sizes in the dataset were inclusive or exclusive of excess amounts. The presence of an excess generally affects both the likelihood of a policyholder to claim as well as the actual claim size [22]. This means that if excesses were in place for the insured events of which we modelled the claims, a degree of censoring is present and should ideally be allowed for. Fot this reason it would make sense to consider left censored distributions - see for example [75], [11]. Furthermore, the values modelled should consistently be inclusive or exclusive of the excess amounts, depending on whether total loss values or net claims paid amounts are required to be modelled.

- **Dependent risks**
  Whilst it is often assumed that insured risks are independent, certain classes of risk are exposed to a degree of dependence. An example would be where a single natural disaster such as a hail storm can affect multiple policyholders at the same time in which case they are not

completely independent in terms of the insured event. Cossette et al [33] considered copulas whereby allowance is made for the dependence between risks. In other studies copulas are used to describe influences from macro economic factors on the dependencies that exist among processes.- see [60].

- **Generalized distributions**
  Parametric distributions are popular to use because their properties are often well-known and it enables the user to easily perform analysis on a specific problem if the occurence of a specific problem can be described by a parametric distribution. In our application it was seen that it may be difficult to always get a parametric distribution providing a good fit to the underlying data, especially on the full range of observed values. Generalized skew distributions have been considered in the literature - see Goerg [58] and Azzalini [8]. In these generalized distributions a flexible approach is introduced to model skewed data by means of generalizing a symmetric parametric distribution such as the Normal or t distribution to allow for the asymmetry present in the data.

- **Non-parameteric methods**
  As was seen in the application of the techniques to fit the parametric distributions to real-life data, it is not easy to find distributions that provide a good fit to the whole range of observed values. Similar observations may be made even when considering alternatives such as the composite distributions. In such cases it may be worthwhile to consider non-parametric techniques such as kernel density estimation which is an unsupervised learning procedure [65]. Essentially a refinement of a histogram is constructed based on observed values. This technique is useful to summarise the distribution of data in a non-parametric way and may be a useful step towards constructing a parametric model [57], [81].

Appendix A

SAS Code

The follwing sections show the SAS code for the simulation and maxmimum likelihood estimation of the parameters for the distributions as discussed in chapter 4.

A.1 Gamma Distribution

```
/* ================================================== */
/* FITTING GAMMA DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                     ESTIMATION                     */
/* ================================================== */

/* Simulate n observations from a Gamma(alpha,kappa)
   distribution */
%let theta = 5;
%let kappa = 3;
%let n = 100000;

data Gamma;
do i = 1 to &n;
X = &theta*rangam(0,&kappa);
output;
end;
drop i;
run;

proc univariate data = Gamma;
var X;
histogram / gamma;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
```

```
output out=Moments mean=_MEAN_ std=_STDDEV_
 skewness=_SKEWNESS_
 kurtosis=_KURTOSIS_;
run;

data Moments;
set Moments;
kappa_hat = (_MEAN_**2)/(_STDDEV_**2);
theta_hat = (_STDDEV_**2)/_MEAN_;

call symput("initial_kappa",kappa_hat);
call symput("initial_theta",theta_hat);

run;

/* Fit Gamma distribution using maximum likelihood
   estimation */

proc iml;

use Gamma;
read all into X_Obs;

use Moments;
read all into Moments;
Moments = Moments‘;

theta_old = &initial_theta;
kappa_old = &initial_kappa;
iteration = J(1000,1,-1);
kappa_old = J(1,1,&initial_kappa);
kappa_new = J(1,1,1);

xbar = (1/nrow(X_Obs))*sum(X_Obs);
wbar = (1/nrow(X_Obs))*sum(log(X_Obs));

do i=1 to 1000;
kappa_new[1,1] = kappa_old - (log(kappa_old)-digamma(kappa_old)
-log(xbar)+wbar)/((1/kappa_old)-trigamma(kappa_old));
if i=1 then difference = kappa_new-kappa_old;
iteration[i,1] = kappa_new;
difference = kappa_new-kappa_old;
kappa_old[1,1] = kappa_new[1,1];
```

```
if abs(difference) < 0.0001 then i=1000;

end;

kappa_hat = kappa_new;
theta_hat = xbar/kappa_hat;

print kappa_new theta_hat;

quit;
```

A.2 Exponential Distribution

```
/* ========================================================= */
/* FITTING EXPONENTIAL DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                        ESTIMATION                         */
/* ========================================================= */

/* Simulate n observations from a Exponential(theta)
   distribution */
%let theta = 5;
%let n = 100000;

data Exponential;
do i = 1 to &n;
X = &theta*rangam(0,1);
output;
end;
drop i;
run;

proc univariate data = Exponential;
var X;
histogram / gamma;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_
                   skewness=_SKEWNESS_
 kurtosis=_KURTOSIS_;
run;

/* Note that the maximum likelihood estimate is equal to the
   method-of-moments estimate */
data Moments;
```

```
set Moments;
theta_hat = _MEAN_;

call symput("initial_theta",theta_hat);

run;
```

A.3 Chi-square Distribution

```
/* ======================================================= */
/* FITTING CHI-SQUARE DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                        ESTIMATION                         */
/* ======================================================= */

/* Simulate n observations from Chi-square(nu) distribution */
%let nu = 5;
%let n = 100000;

data ChiSquare;
do i = 1 to &n;
X = 2*rangam(0,&nu/2);
output;
end;
drop i;
run;

proc univariate data = ChiSquare;
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_
                   skewness=_SKEWNESS_
 kurtosis=_KURTOSIS_;
run;

data Moments;
set Moments;
nu_hat = _MEAN_;

call symput("nu_hat",nu_hat);

run;
```

```
/* Fit the Chi-square distribution using maximum likelihood
 estimation */

proc iml;

use ChiSquare;
read all into X_Obs;

kappa_old = &nu_hat;

do i = 1 to 1000;

kappa_new = kappa_old - (digamma(kappa_old)-(1/nrow(X_Obs))
*sum(log(X_Obs))+log(2))/(trigamma(kappa_old));
difference = kappa_new - kappa_old;
kappa_old = kappa_new;

if abs(difference) < 0.0001 then i=1000;

end;

nu_hat = 2*kappa_new;

print nu_hat;

quit;
```

A.4 Two-parameter Exponential Distribution

```
/* ====================================================== */
/* FITTING TWO-PARAMETER EXPONENTIAL DISTRIBUTION USING */
/*              MAXIMUM LIKELIHOOD ESTIMATION           */
/* ====================================================== */

/* Simulate n observations from a Two-parameter
   Exponential(theta,eta) distribution */
%let theta = 5;
%let eta = 3;
%let n = 100000;

data TwoParmaterExponential;
do i = 1 to &n;
X = &theta*rangam(0,1)+&eta;
```

```
output;
end;
drop i;
run;

proc univariate data = TwoParmaterExponential;
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ min=_MIN_
                   skewness=_SKEWNESS_
         kurtosis=_KURTOSIS_;
run;

/* Note that the maximum likelihood estimate is equal to the
   method-of-moments estimate */
data Moments;
set Moments;
eta_hat = _MIN_;
theta_hat = _MEAN_-_MIN_;
run;
```

A.5 Erlang Distribution

```
/* ==================================================== */
/* FITTING ERLANG DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                      ESTIMATION                      */
/* ==================================================== */

/* Simulate n observations from a Erlang(n,lambda)
   distribution */
%let lambda = 5;
%let n = 10;
%let k = 100000;

data Erlang;
do i = 1 to &k;
X = (1/&lambda)*rangam(0,&n);
output;
end;
drop i;
run;
```

```
proc univariate data = Erlang;
var X;
histogram / gamma;
/* Note that 1 divided by the scale parameter estimate is the
   estimate for lambda */
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ n=_COUNT_
           skewness=_SKEWNESS_
 kurtosis=_KURTOSIS_;
run;

/* Note that the maximum likelihood and method-of-moments
   estimates of lambda are the same */
data Moments;
set Moments;
lambda_hat = &n / _MEAN_;
run;
```

A.6 Frechet Distribution

```
/* ===================================================== */
/* FITTING FRECHET DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                        ESTIMATION                     */
/* ===================================================== */

/* Simulate n observations from a Frechet(beta,delta,lambda)
 distribution */
%let beta = 8;
%let delta = 2;
%let lambda = 100;
%let n = 100000;

data Frechet;
do i = 1 to &n;
U = ranuni(0); /* Random uniform(0,1) variable */
X = &delta*((-log(U))**(-1/&beta))+&lambda;
/* Using the probability integral transformation */
output;
end;
drop i U;
run;

proc univariate data = Frechet;
```

```
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_
                   skewness=_SKEWNESS_
 kurtosis=_KURTOSIS_;
run;
quit;

data _NULL_;
set Moments;
call symput("Mean",_MEAN_);
run;

/* Fit the Frechet distribution using maximum likelihood
 estimation */

proc iml;

beta_old = 10;
delta_old = 1;
lambda_old= &Mean-delta_old*Gamma(1-1/beta_old);

use Frechet;
read all into X_Obs;

N_obs = nrow(X_Obs);

Partial_1 = J(3,1,0);
Partial_2 = J(3,3,0);

do i=1 to 2500;

B_Old = beta_old // delta_old // lambda_old;

/* First Order Partial Derivatives */

dl_dbeta =  N_Obs/beta_old -sum(log(J(N_obs,1,delta_old)
/(X_Obs-J(N_Obs,1,lambda_old)))#((J(N_obs,1,delta_old)
/(X_Obs-J(N_Obs,1,lambda_old)))##beta_old-J(N_Obs,1,1)));

dl_ddelta = -beta_old*(delta_old**(beta_old-1))
  *sum((X_Obs-J(n_obs,1,lambda_old))##(-beta_old))
+beta_old*N_obs/delta_old;
```

```
dl_dlambda =
     -((beta_old)/(delta_old))*sum((J(N_obs,1,delta_old)
 /(X_Obs-J(N_Obs,1,lambda_old))##(beta_old+1))
 +(1+beta_old)*sum((x_Obs-J(n_obs,1,lambda_old))##(-1));

/* Second Order Partial Derivatives */

dl2_dbeta_dbeta = -N_Obs/(beta_old**2)
- sum(((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##beta_old)
#((log(((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old))))##(2)));

dl2_ddelta_dbeta = N_Obs/delta_old-(1/delta_old)
*sum((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##(beta_old))
-beta_old/delta_old
*sum(((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##(beta_old))
#(log((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old))))));

dl2_dlambda_dbeta = -(((beta_old)/(delta_old))**(2))
*sum((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##(beta_old+1));

dl2_ddelta_ddelta = -((beta_old*(beta_old-1))/(delta_old**2))
  *sum((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##(beta_old))
-(N_Obs*beta_old)/(delta_old**2);

dl2_dlambda_ddelta = -(1/delta_old)
 *sum((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##(beta_old+1))
-(beta_old/delta_old)
*sum(((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old)))##(beta_old+1))
#(log((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
#(J(N_Obs,1,delta_old))))))
+sum((X_Obs-J(N_Obs,1,lambda_old))##(-1));

dl2_dlambda_dlambda = -(beta_old*(beta_old+1))/(delta_old**2)
*sum((((X_Obs-J(N_Obs,1,lambda_old))##(-1))
```

```
#(J(N_Obs,1,delta_old))##(beta_old+2))
+(beta_old+1)
*sum(((X_Obs-J(N_Obs,1,lambda_old))##(-2)));

Partial_1 = dl_dbeta // dl_ddelta // dl_dlambda;
Partial_2 =
(dl2_dbeta_dbeta   || dl2_ddelta_dbeta
|| dl2_dlambda_dbeta)  //
(dl2_ddelta_dbeta  || dl2_ddelta_ddelta
|| dl2_dlambda_ddelta) //
(dl2_dlambda_dbeta || dl2_dlambda_ddelta
|| dl2_dlambda_dlambda);

B_New = B_Old - ((Partial_2)**(-1))*(Partial_1);
Difference = B_New - B_Old;

B_Old = B_New;
beta_old = B_Old[1];
delta_old = B_Old[2];
lambda_old = B_Old[3];
diff = sum(abs(Difference));

if diff < 0.00001 then i = 2500;

end;

print beta_old delta_old lambda_old diff;

quit;
```

A.7 Weibull Distribution

```
/* ===================================================== */
/* FITTING WEIBULL DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                      ESTIMATION                       */
/* ===================================================== */

/* Simulate n observations from Weibull(beta,delta,lambda)
   distribution */
%let beta = 2.5;
%let delta = 10;
%let lambda = 5;
%let n = 100000;
```

```
data Weibull;
do i = 1 to &n;
U = ranuni(0); /* Random uniform(0,1) variable */
/* FOR MAXIMA */
X = ( &delta*((-log(1-U))**(1/&beta))-&lambda);
/* Using the probability integral transformation */
output;
end;
drop i U;
run;

proc means data = Weibull noprint;
var X;
output out=Lambda_Estimate min=Lambda_Hat;
run;

data _NULL_;
set Lambda_Estimate;
call symput("Lambda_Hat",-Lambda_Hat);
run;

data Weibull_Adj(keep = X_Adj);
set Weibull;
X_Adj = X+&Lambda_Hat+0.0001;
run;

/* Fit the Weibull distribution using maximum likelihood
   estimation */

proc iml;

beta_old = 5;
theta_old = 8;

use Weibull_Adj;
read all into X_Obs;

/*print X_Obs;*/

N_obs = nrow(X_Obs);

Partial_1 = J(2,1,0);
Partial_2 = J(2,2,0);
```

```
B_Old = beta_old // theta_old;

do i=1 to 1000;

Ratio = X_Obs#J(N_Obs,1,1/theta_old);

/* First Order Partial Derivatives */

dl_dbeta = N_Obs/beta_old+(beta_old-1)*sum(log(X_Obs))-N_Obs
           *log(theta_old)-sum((Ratio##beta_old)#log(Ratio)) ;
dl_theta = -N_Obs*beta_old/theta_old+(beta_old/theta_old)
           *sum(Ratio##beta_old);

/* Second Order Partial Derivatives */

dl2_dbeta_dbeta = -N_Obs/(beta_old**2)+sum(log(X_Obs))
                  -sum((Ratio##beta_old)#((log(Ratio))##2));
dl2_dtheta_dbeta = -N_Obs/theta_old+(1/theta_old)
                   *sum((Ratio##beta_old)#(J(N_Obs,1,1)
 +(J(N_Obs,1,beta_old))#(log(Ratio))));
dl2_dtheta_dtheta = N_Obs*beta_old/(theta_old**2)
 -beta_old*(beta_old+1)/(theta_old**2)
 *sum(Ratio##beta_old);

partial_1 = dl_dbeta // dl_theta;
partial_2 = (dl2_dbeta_dbeta  || dl2_dtheta_dbeta ) //
(dl2_dtheta_dbeta  || dl2_dtheta_dtheta) ;

B_New = B_Old - (inv(partial_2))*partial_1;

beta_old = B_New[1];
theta_old = B_New[2];
difference=sum(abs(B_New-B_Old));
B_Old = B_New;

if difference < 0.00001 then i=1000;

end;

print difference B_New;

quit;
```

```
/* Regression Approach */

proc sort data = Weibull_Adj;
by X_Adj;
run;

%let co = %sysfunc(open(Weibull_Adj));
%let cn = %sysfunc(attrn(&co,nobs));
%let cc = %sysfunc(close(&co));

data Weibull_Adj;
set Weibull_Adj;
by X_Adj;
F_X = _N_ / (&cn+1);
log_X = log(X_Adj);
log_Q = log(-log(1-F_X));
run;

proc reg data=Weibull_Adj outest=test;
model log_X = log_Q;
output out=Residuals r=Res;
run;
quit;

proc gplot data = Residuals;
plot Res*X_Adj;
run;
```

A.8 Rayleigh Distribution

```
/* ====================================================== */
/* FITTING RAYLEIGH DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                      ESTIMATION                        */
/* ====================================================== */

/* Simulate n observations from a Rayleigh(theta)
   distribution */
%let theta = 5;
%let n = 100000;

data Rayleigh;
do i = 1 to &n;
U = ranuni(0); /* Random uniform(0,1) variable */
```

```
X = &theta*sqrt(-log(1-U)); /* Using the probability integral
                               transformation */
output;
end;
drop i U;
run;

proc univariate data = Rayleigh;
var X;
histogram / Rayleigh;
/* The estimate of Theta is given by sqrt(2)*Sigma */
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ skewness=_SKEWNESS_
kurtosis=_KURTOSIS_ N = _COUNT_;
run;

data Moments;
set Moments;
Theta_Hat_MME = 2*_MEAN_/sqrt(CONSTANT('PI'));
Theta_Hat_MLE = sqrt((_STDDEV_**2+_MEAN_**2));
run;
```

A.9 Gumbel Distribution

```
/* ==================================================== */
/* FITTING GUMBEL DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                     ESTIMATION                       */
/* ==================================================== */

/* Simulate n observations from a Gumbel(xi,alpha)
   distribution */
%let xi = 0.8;
%let alpha =5;
%let n = 10000;

data Gumbel;
do i = 1 to &n;
U = ranuni(0); /* Random uniform(0,1) variable */
X = -&alpha*log(-log(U))+&xi;
/* Using the probability integral transformation */
output;
end;
drop i U;
```

```
run;

/* Regression Approach */
%let co = %sysfunc(open(Gumbel));
%let cn = %sysfunc(attrn(&co,nobs));
%let cc = %sysfunc(close(&co));

proc sort data = Gumbel;
by X;
run;

data Gumbel;
set Gumbel;
by X;
F_X = _N_/(&cn+1);
Q = -log(-log(F_X));
run;

proc reg data = Gumbel outest=Parameter_Estimates rsquare;
model X = Q;
run;
quit;

data _NULL_;
set Parameter_Estimates;
call symput("Intercept",Intercept);
call symput("Q",Q);
run;

data Gumbel;
set Gumbel (keep = X);
run;

/* Fit the Gumbel distribution using maximum likelihood
   estimation */

proc iml;

xi_old =&Intercept;
alpha_old =&Q;

use Gumbel;
read all into X_Obs;
```

```
N_obs = nrow(X_Obs);

B_Old = xi_old // alpha_old;

do i=1 to 10;

/* First Order Partial Derivatives */

dl_dxi = N_Obs/alpha_old - (1/alpha_old)
*sum(exp(-((X_Obs - J(N_Obs,1,xi_old))
#((J(N_obs,1,alpha_old))##(-1)))));
dl_dalpha = -N_Obs/alpha_old+(1/(alpha_old**2))
*sum((X_Obs - J(N_Obs,1,xi_old))#(J(N_Obs,1,1)
-exp(-((X_Obs - J(N_Obs,1,xi_old))
#((J(N_obs,1,alpha_old))##(-1))))));

Partial_1 =  dl_dxi // dl_dalpha;

/* Second Order Partial Derivatives */

dl2_dxi_dxi = -(1/(alpha_old**2))
*sum(exp(-((X_Obs - J(N_Obs,1,xi_old))
#((J(N_obs,1,alpha_old))##(-1)))));
dl2_dalpha_dalpha = N_Obs/alpha_old**2-(2/(alpha_old**3))
*sum((X_Obs - J(N_Obs,1,xi_old))
#(J(N_Obs,1,1)-exp(-((X_Obs - J(N_Obs,1,xi_old))
#((J(N_obs,1,alpha_old))##(-1))))))
-(1/(alpha_old**4))*sum((X_Obs - J(N_Obs,1,xi_old))
#(exp(-((X_Obs - J(N_Obs,1,xi_old))
#((J(N_obs,1,alpha_old))##(-1)))))
#(X_Obs - J(N_Obs,1,xi_old)));

dl2_dalpha_dxi = -N_Obs/(alpha_old**2)+(1/alpha_old**2)
 *sum(exp((-1/alpha_old)#(X_Obs-J(N_Obs,1,xi_old))))
 -(1/alpha_old)*sum(((1/alpha_old**2)
 #(X_Obs-J(N_Obs,1,xi_old)))#(exp((-1/alpha_old)
 #(X_Obs-J(N_Obs,1,xi_old)))));

Partial_2 = (dl2_dxi_dxi  || dl2_dalpha_dxi) //
(dl2_dalpha_dxi || dl2_dalpha_dalpha);

B_New = B_Old-(inv(Partial_2))*(Partial_1);
difference = sum(abs(B_New-B_Old));
xi_old = B_New[1];
```

```
alpha_old = B_New[2];

if difference < 0.0001 then i=10000;
B_Old = B_New;

end;

xi_hat = xi_old;
alpha_hat = alpha_old;

print xi_hat alpha_hat difference;

quit;
```

A.10 Pareto Distribution

```
/* ===================================================== */
/* FITTING PARETO DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                       ESTIMATION                      */
/* ===================================================== */

/* Simulate n observations from a Pareto(theta,kappa)
   distribution */
%let theta = 9;
%let kappa = 5;
%let n = 100000;

data Pareto;
do i = 1 to &n;
U = ranuni(1234); /* Random uniform(0,1) variable */
X = &theta*((1-U)**(-1/&kappa)-1);
/* Using the probability integral transformation */
output;
end;
drop i U;
run;

proc univariate data = Pareto;
var X;
histogram;* / Pareto;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_
                   skewness=_SKEWNESS_
```

```
         kurtosis=_KURTOSIS_ N = _COUNT_;
run;

data Moments;
set Moments;
Theta_Hat_MME =_MEAN_*(_STDDEV_**2+_MEAN_**2)/
               (_STDDEV_**2-_MEAN_**2);
Kappa_Hat_MME =2*(_STDDEV_**2)/(_STDDEV_**2 - _MEAN_**2);
run;

data _NULL_;
set Moments;
call symput("theta_hat_MME",Theta_Hat_MME);
call symput("Kappa_Hat_MME",Kappa_Hat_MME);
run;

proc iml;

use Pareto;
read all into X_Obs;

N_Obs  = nrow(X_Obs);

kappa_old = &Kappa_Hat_MME;
theta_old = &Theta_Hat_MME;

B_Old = kappa_old // theta_old;

do i = 1 to 10000;

/* First Order Partial Derivatives */
dl_dkappa = N_obs/kappa_old+N_Obs*log(theta_old)
-sum(log(X_Obs+J(N_Obs,1,theta_old)));
dl_dtheta = N_Obs*kappa_old/theta_old-(kappa_old+1)
*sum((X_Obs+J(N_Obs,1,theta_old))##(-1));

/* Second Order Partial Derivatives */

dl2_dkappa_dkappa = -N_Obs/(kappa_old**2);
dl2_dtheta_dkappa = N_Obs/theta_old
                    -sum(X_Obs+J(N_Obs,1,theta_old)) ;
dl2_dtheta_dtheta = -N_Obs*kappa_old/(theta_old**2)
                    +(kappa_old+1)
*sum((X_Obs+J(N_Obs,1,theta_old))##(-2)) ;
```

```
Partial_1 = dl_dkappa // dl_dtheta;
Partial_2 = (dl2_dkappa_dkappa || dl2_dtheta_dkappa) //
(dl2_dtheta_dkappa || dl2_dtheta_dtheta);

B_New = B_Old-(inv(Partial_2))*Partial_1;

difference = sum(abs(B_New-B_Old));
kappa_old = B_New[1];
theta_old = B_Old[2];
B_Old = B_New;

if difference < 0.0001 then i = 10000;

end;

print difference B_New;

quit;
```

A.11 Generalized Pareto Distribution

```
/* ===================================================== */
/* FITTING GENERALIZED PARETO DISTRIBUTION USING MAXIMUM */
/*                LIKELIHOOD ESTIMATION                  */
/* ===================================================== */

/* Simulate n observations from a Generalized
   Pareto(theta,kappa,tau) distribution */
%let theta = 18;
%let kappa = 2;
%let tau = 4 ;
%let n = 100000;

data Two_Stage_Simulation;

do j=1 to &n;
B = (1/&theta)*rangam(0,&kappa);
X = (1/B)*rangam(0,&tau);
output;
end;
keep X;
```

```
run;

proc univariate data = Two_Stage_Simulation;
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ skewness=_SKEWNESS_
 kurtosis=_KURTOSIS_ N = _COUNT_;
run;

data _NULL_;
set Moments;
call symput("mean",_MEAN_);
run;

proc iml;

use Two_Stage_Simulation;
read all into X_Obs;

N_Obs = nrow(X_Obs);

theta_old = 2;
kappa_old = 1.5;
tau_old = (kappa_old-1)*&mean/theta_old;

B_Old = theta_old // kappa_old // tau_old;

do i=1 to 100000;

/* First Order Partial Derivatives */

dl_dtheta = N_Obs*kappa_old/theta_old-(kappa_old+tau_old)
*sum((X_Obs+J(N_Obs,1,theta_old))##(-1)) ;
dl_dkappa = N_Obs*Digamma(kappa_old+tau_old)
            -N_Obs*Digamma(kappa_old)
  +sum(log(J(N_Obs,1,theta_old)
#((X_Obs+J(N_Obs,1,theta_old))##(-1))));
dl_dtau   = N_Obs*Digamma(kappa_old+tau_old)
            -N_Obs*Digamma(tau_old)
+sum(log(X_Obs#((X_Obs+J(N_Obs,1,theta_old))##(-1))));

/* Second Order Partial Derivatives */
```

```
dl2_dtheta_dtheta = -N_Obs*kappa_old/(theta_old**2)
+(kappa_old+tau_old)
*sum((X_Obs+J(N_Obs,1,theta_old))##(-2));
dl2_dkappa_dtheta = N_Obs/theta_old
-sum((X_Obs+J(N_Obs,1,theta_old))##(-1));
dl2_dtau_dtheta = -sum((X_Obs+J(N_Obs,1,theta_old))##(-1));

dl2_dkappa_dkappa = N_Obs*trigamma(kappa_old+tau_old)
-N_Obs*trigamma(kappa_old);
dl2_dtau_dkappa = N_Obs*trigamma(kappa_old+tau_old);

dl2_dtau_dtau = N_Obs*trigamma(kappa_old+tau_old)
-N_Obs*trigamma(tau_old);

Partial_1 =  dl_dtheta // dl_dkappa // dl_dtau;
Partial_2 =
 (dl2_dtheta_dtheta||dl2_dkappa_dtheta||dl2_dtau_dtheta)//
 (dl2_dkappa_dtheta||dl2_dkappa_dkappa||dl2_dtau_dkappa)//
 (dl2_dtau_dtheta  ||dl2_dtau_dkappa  ||dl2_dtau_dtau);

B_New = B_Old - (inv(Partial_2))*Partial_1;

difference = sum(abs(B_New - B_Old));
theta_old = B_New[1];
kappa_old = B_New[2];
tau_old = B_New[3];

if difference<0.00000001 then i=100000;

B_Old = B_New;
print B_Old difference;

end;

theta_hat = theta_old;
kappa_hat = kappa_old;
tau_hat = tau_old;

print theta_hat kappa_hat tau_hat difference;

quit;
```

A.12 Lognormal Distribution

```
/* ======================================================== */
/* FITTING LOGNORMAL DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                        ESTIMATION                        */
/* ======================================================== */

/* Simulate n observations from a Lognormal(mu,sigma)
   distribution */
%let mu = 0.8;
%let sigma = 0.5;
%let n = 10000;

data LogNormal;
do i = 1 to &n;
U = &mu+&sigma*rannor(0);
X = exp(U);
output;
end;
drop i U;
run;

/* Fit Lognormal distribution using MLE*/
proc iml;

use LogNormal;
read all into X_Obs;

N_Obs = nrow(X_Obs);

mu_hat = (1/N_Obs)*sum(log(X_Obs));
sigma_hat = sqrt((1/N_Obs)*sum((log(X_Obs)-
            J(N_Obs,1,mu_hat))##(2)));

print mu_hat sigma_hat;

quit;
```

A.13 Beta-prime Distribution

```
/* ========================================================== */
/* FITTING BETA-PRIME DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                        ESTIMATION                          */
/* ========================================================== */

/* Simulate n observations from a Beta-prime(delta_1,delta_2)
   distribution */

%let delta_1 = 4;
%let delta_2 = 20;
%let n = 100000;
%let increment = 0.005;

data Two_Stage_Simulation;

do j=1 to &n;
B = 1*rangam(0,&delta_2);
X = (1/B)*rangam(0,&delta_1);
output;
end;
keep X;
run;

proc univariate data = Two_Stage_Simulation;
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out = Moments mean=_MEAN_ std=_STDDEV_
                     skewness=_SKEWNESS_
                     kurtosis=_KURTOSIS_ N = _COUNT_;
run;

data _NULL_;
set Moments;
call symput("Mean",_MEAN_);
run;

proc iml;

use Two_Stage_Simulation;
read all into X_Obs;
```

```
N_Obs = nrow(X_Obs);

delta_1_old = 2;
delta_2_old = (delta_1_old/&mean)+1;

B_Old = delta_1_old // delta_2_old;

do i=1 to 1000;

/* First Order Partial Derivatives */

dl_ddelta_1 = N_Obs*Digamma(delta_1_old+delta_2_old)
-N_Obs*Digamma(delta_1_old)
+sum(log(X_Obs#((X_Obs+J(N_Obs,1,1))##(-1))));
dl_ddelta_2 = N_Obs*Digamma(delta_1_old+delta_2_old)
-N_Obs*Digamma(delta_2_old)-sum(log(X_Obs+J(N_Obs,1,1)));

/* Second Order Partial Derivatives */

dl2_ddelta_1_ddelta_1=N_Obs*Trigamma(delta_1_old+delta_2_old)
-N_Obs*Trigamma(delta_1_old);
dl2_ddelta_2_ddelta_1=N_Obs*Trigamma(delta_1_old+delta_2_old);
dl2_ddelta_2_ddelta_2=N_Obs*Trigamma(delta_1_old+delta_2_old)
-N_Obs*Trigamma(delta_2_old);

/*print dl2_ddelta_1_ddelta_1  dl2_ddelta_2_ddelta_1;*/
/*print  dl2_ddelta_2_ddelta_2;*/

Partial_1 = dl_ddelta_1 // dl_ddelta_2;
Partial_2 = (dl2_ddelta_1_ddelta_1||dl2_ddelta_2_ddelta_1) //
(dl2_ddelta_2_ddelta_1||dl2_ddelta_2_ddelta_2);

B_New = B_Old - (inv(Partial_2))*Partial_1;
delta_1_old = B_New[1];
delta_2_old = B_New[2];
difference = sum(abs(B_New-B_Old));
B_Old = B_New;

if difference < 0.00001 then i =1000;

end;

delta_1_hat = delta_1_old;
```

```
delta_2_hat = delta_2_old;

print delta_1_hat delta_2_hat difference;

quit;
```

A.14 Birnbaum-Saunders Distribution

```
/* ==================================================== */
/* FITTING BIRNBAUM-SAUNDERS DISTRIBUTION USING MAXIMUM */
/*                 LIKELIHOOD ESTIMATION                */
/* ==================================================== */
%let alpha = 0.3;
%let beta = 2000;

data Birnbaum_Saunders;
do i=1 to 100000;
X = ((&alpha)/2)*rannor(0);
T = &beta*(1+2*X**2+2*X*sqrt(X**2+1));
output;
end;
drop X i;
run;

proc iml;

use Birnbaum_Saunders;
read all into T_Obs;

N_Obs = nrow(T_Obs);

alpha_old = 1;
beta_old = 1;

B_Old = alpha_old // beta_old;

do i = 1 to 1000;

/* First Order Partial Derivatives */
dl_dalpha = -N_Obs/alpha_old+(1/alpha_old**3)
            *sum((sqrt(T_Obs#J(N_Obs,1,1/beta_old))
-sqrt(J(N_Obs,1,beta_old)#(T_Obs##(-1))))##(2));
dl_dbeta = -(1/(2*alpha_old**2))*sum((T_Obs##(-1)
```

```
 - T_Obs#J(N_Obs,1,beta_old**(-2))))
 +(1/(2*beta_old))*sum((J(N_Obs,1,beta_old)-T_Obs)
 #((J(N_Obs,1,beta_old)+T_Obs)##(-1)));

/* Second Order Partial Derivatives */

dl2_dalpha_dalpha = N_Obs/alpha_old**2-(3/alpha_old**4)
*sum((sqrt(T_Obs#J(N_Obs,1,1/beta_old))
-sqrt(J(N_Obs,1,beta_old)#(T_Obs##(-1))))##(2));
dl2_dbeta_dalpha =  (1/(alpha_old**3))*sum(T_Obs##(-1)
-T_Obs#J(N_Obs,1,beta_old**(-2)));

dl2_dbeta_dbeta = -(1/alpha_old**2)
                  *sum(T_Obs#J(N_Obs,1,beta_old**(-3)))
-(1/(2*beta_old**(2)))*sum((J(N_Obs,1,beta_old)-T_Obs)
#((J(N_Obs,1,beta_old)+T_Obs)##(-1)))
+(1/beta_old)*sum(T_Obs#((beta_old+T_Obs)##(-2)));


Partial_1 = dl_dalpha // dl_dbeta;
Partial_2 = (dl2_dalpha_dalpha || dl2_dbeta_dalpha) //
(dl2_dbeta_dalpha  || dl2_dbeta_dbeta);


B_New = B_Old-(inv(Partial_2))*Partial_1;
alpha_old = B_New[1];
beta_old = B_New[2];
difference = sum(abs(B_New - B_Old));
B_Old = B_New;

if difference < 0.00001 then i=1000;

end;

alpha_hat = alpha_old;
beta_hat = beta_old;

print alpha_hat beta_hat difference;

quit;
```

A.15 Burr Distribution

```
/* ================================================== */
/* FITTING BURR DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                    ESTIMATION                      */
/* ================================================== */

%let alpha = 50000;
%let c = 2;
%let k = 10;

data Burr;
do i = 1 to 100000;
X = &alpha*((1-ranuni(0))**(-1/&k)-1)**(1/&c);
output;
end;
drop i;
run;

proc iml;

use Burr;
read all into X_Obs;

N_Obs = nrow(X_Obs);

alpha_old = 1;
c_old = 1;
k_old = 1;

B_Old = alpha_old // c_old // k_old;

do i=1 to 1000;

/* First Order Partial Derivatives */

dl_dalpha = -N_Obs*c_old/alpha_old+(c_old*(k_old+1)/alpha_old)
*sum((X_Obs##c_old)#((J(N_Obs,1,alpha_old)##c_old
+X_Obs##c_old)##(-1)));
dl_dc = N_Obs/c_old
        +sum(log(X_Obs#((J(N_Obs,1,alpha_old))##(-1))))
-(k_old+1)*sum((((1/alpha_old)#X_Obs)##c_old)
#(log(((1/alpha_old)#X_Obs))
```

```
#((J(N_Obs,1,1)+((1/alpha_old)#X_Obs)##c_old)##(-1)));
dl_dk = N_Obs/k_old-sum(log(J(N_Obs,1,1)
        +((1/alpha_old)#X_Obs)##c_old));‘‘‘‘

/* Second Order Partial Derivatives */

dl2_dalpha_dalpha = N_Obs*c_old/(alpha_old**2)
-(c_old*(k_old+1)/(alpha_old**2))
  *sum(((X_Obs##c_old)
#((X_Obs##c_old+J(N_Obs,1,alpha_old)##c_old)##(-1)))
#(J(N_Obs,1,1)+(J(N_Obs,1,c_old*alpha_old**(c_old)))
#((J(N_Obs,1,alpha_old**c_old)+X_Obs##c_old)##(-1))));

dl2_dc_dalpha = -N_Obs/alpha_old+((k_old+1)/alpha_old)
*sum(((X_Obs##c_old)
#((X_Obs##c_old+J(N_Obs,1,alpha_old**c_old))##(-1)))
#(J(N_Obs,1,1)+((c_old*alpha_old**c_old)
#log((1/alpha_old)#X_Obs))#((X_Obs##c_old
+J(N_Obs,1,alpha_old**c_old))##(-1))));

dl2_dk_dalpha = (c_old/alpha_old)*sum((X_Obs##c_old)
#((X_Obs##c_old+J(N_Obs,1,alpha_old**c_old))##(-1)));

dl2_dc_dc = -N_Obs/(c_old**2)-(k_old+1)*sum(((log((1/alpha_old)
#X_Obs))##(2))#(((1/alpha_old)#X_Obs)##c_old)
#(((J(N_Obs,1,1)+((1/alpha_old)#X_Obs)##c_old)##2)##(-1)));
dl2_dk_dc = -sum((X_Obs##c_old)#(log((1/alpha_old)#X_Obs))
#((X_Obs##c_old+J(N_obs,1,alpha_old**c_old))##(-1)));

dl2_dk_dk = -N_Obs/(k_old**2);

Partial_1=dl_dalpha // dl_dc // dl_dk;
Partial_2=(dl2_dalpha_dalpha||dl2_dc_dalpha||dl2_dk_dalpha)//
(dl2_dc_dalpha  ||dl2_dc_dc    ||dl2_dk_dc)//
(dl2_dk_dalpha  ||dl2_dk_dc    ||dl2_dk_dk);


B_New = B_Old - (inv(Partial_2))*Partial_1;
difference = sum(abs(B_New-B_Old));
alpha_old = B_New[1];
c_old = B_New[2];
k_old = B_New[3];
B_Old = B_New;
```

```
if difference < 0.000001 then i = 1000;

end;

alpha_hat = alpha_old;
c_hat = c_old;
k_hat = k_old;

print alpha_hat c_hat k_hat difference;

quit;
```

A.16 Dagum Distribution

```
/* =================================================== */
/* FITTING DAGUM DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                     ESTIMATION                      */
/* =================================================== */

/* Simulate n observations from a Dagum(a,b,p) distribution */
%let a = 46;
%let b = 372;
%let p = 110;
%let n = 100000;

data Dagum;
do j=1 to &n;
X = &b*(ranuni(0)**(-1/&p)-1)**(-1/&a);
output;
end;
keep X;
run;

proc iml;

use Dagum;
read all into X_Obs;

N_Obs = nrow(X_Obs);

a_old = 3;
b_old = 5;
p_old = 6;
```

```
Beta_Old = a_old // b_old // p_old;

do i=1 to 10000;

/* First Order Partial Derivatives */

dl_da = N_Obs/a_old+N_Obs*log(b_old)-sum(log(X_Obs))
    -(p_old+1)*sum((J(N_Obs,1,b_old**a_old)
#((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1)))
#log(b_old#(X_Obs##(-1))));

dl_db = N_Obs*a_old/b_old-(a_old*(p_old+1)/b_old)
        *sum((b_old**a_old)#((X_Obs##a_old
+J(N_Obs,1,b_old**a_old))##(-1)));

dl_dp = N_Obs/p_old-sum(log(J(N_Obs,1,1)
        +(b_old**a_old)#(X_Obs##(-a_old))));

/* Second Order Partial Derivatives */

dl2_da_da = -N_Obs/(a_old**2)
      -(p_old)*sum((b_old**a_old)#(X_Obs##a_old)
      #((log(b_old#X_Obs##(-1)))##2)
      #((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-2)));

dl2_db_da = N_Obs/b_old-((p_old+1)/b_old)*sum(((b_old**a_old)
            #(X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1))
#(J(N_Obs,1,1)+(a_old#(X_Obs##a_old)
#(log(b_old#(X_Obs##(-1)))))
#((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1))));

dl2_dp_da = -sum((b_old**a_old)
            #((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1))
#(log(b_old#(X_Obs##(-1)))));

dl2_db_db = -N_Obs*a_old/(b_old**2)
            +(a_old*(p_old+1)/(b_old**2))*sum(((b_old**a_old)
      #((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1)))
      #(J(N_Obs,1,1)-a_old#(X_Obs##a_old)
#((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1))));

dl2_dp_db = -(a_old/b_old)*sum((b_old**a_old)
            #((X_Obs##a_old+J(N_Obs,1,b_old**a_old))##(-1)));
```

```
dl2_dp_dp = -N_Obs/(p_old**2);

Partial_1 =  dl_da // dl_db // dl_dp;
Partial_2 =  (dl2_da_da || dl2_db_da || dl2_dp_da) //
 (dl2_db_da || dl2_db_db || dl2_dp_db) //
 (dl2_dp_da || dl2_dp_db || dl2_dp_dp);

Beta_New = Beta_Old -(inv(Partial_2))*Partial_1;

difference = sum(abs(Beta_New - Beta_Old));

a_old = Beta_New[1];
b_old = Beta_New[2];
p_old = Beta_New[3];

if i>990 then print a_old b_old p_old difference;

if difference<0.00000001 then i=10000;

Beta_Old = Beta_New;
end;

a_hat = a_old;
b_hat = b_old;
p_hat = p_old;

print a_hat b_hat p_hat difference;

quit;
```

A.17 Generalized Beta Distribution of the Second Kind

```
/* ======================================================== */
/* FITTING GENERALIZED BETA DISTRIBUTION OF THE SECOND KIND */
/*          USING MAXIMUM LIKELIHOOD ESTIMATION             */
/* ======================================================== */

/* Simulate n observations from a GB2(a,b,q,p) distribution */
%let a = 0.8;
%let b = 10;
```

```
%let q = 8;
%let p = 0.5;
%let n = 1000000;

data Two_Stage_Simulation;

do j=1 to &n;
B = (1/(&b**&a))*rangam(0,&q);
X = ((1/B)*rangam(0,&p))**(1/&a);
output;
end;
keep X;
run;

proc iml;

use Two_Stage_Simulation;
read all into X_Obs;

N_Obs = nrow(X_Obs);

a_old = 1;
b_old = 2;
q_old = 2;
p_old = 1;

Beta_Old = a_old // b_old // q_old // p_old;

do i=1 to 10000;

Term_1 = log((1/b_old)#X_Obs);
Term_2 = X_Obs##a_old+J(N_Obs,1,b_old**a_old);

/* First Order Partial Derivatives */

dl_da = N_Obs/a_old-q_old*sum(Term_1)+(p_old+q_old)
        *sum((b_old**a_old)#(Term_1)#((Term_2)##(-1)));
dl_db = (N_obs*a_old*q_old)/b_old-(p_old+q_old)
        *(a_old*b_old**(a_old-1))*sum(Term_2##(-1));
dl_dq = N_Obs*a_old*log(b_old)+N_Obs*Digamma(p_old+q_old)
        -N_Obs*Digamma(q_old)-sum(log(Term_2));
dl_dp = N_Obs*Digamma(p_old+q_old)-N_Obs*Digamma(p_old)
        +a_old*sum(log(X_Obs))-sum(log(Term_2));
```

```
/* Second Order Partial Derivatives */

dl2_da_da = -N_Obs/a_old**2-(p_old+q_old)*(b_old**a_old)
            *sum((X_Obs##a_old)#(Term_1##2)#((Term_2)##(-2)));
dl2_db_da = N_Obs*q_old/b_old-(p_old+q_old)
*sum((J(N_Obs,1,b_old**(2*a_old-1))+(b_old**(a_old-1))
#X_Obs##a_old-(a_old*(b_old**(a_old-1)))
#(X_Obs##a_old)#Term_1)#(Term_2##(-2)));
dl2_dq_da = N_Obs*log(b_old)
      -sum((J(N_Obs,1,(b_old**a_old)*log(b_old))+(X_Obs##a_old)
      #log(X_Obs))#(Term_2##(-1)));
dl2_dp_da = sum(log(X_Obs))
            -sum((J(N_Obs,1,(b_old**a_old)*log(b_old))
            +(X_Obs##a_old)#log(X_Obs))#(Term_2##(-1)));

dl2_db_db = -N_Obs*a_old*q_old/(b_old**2)
            -(a_old**2)*(b_old**(a_old-2))
            *(p_old+q_old)*sum((X_Obs##a_old)#(Term_2##(-2)))
      +a_old*(b_old**(a_old-2))
*(p_old+q_old)*sum(Term_2##(-1));
dl2_dq_db = N_Obs*a_old/b_old-(a_old*(b_old**(a_old-1)))
            *sum(Term_2##(-1));
dl2_dp_db = -(a_old*(b_old**(a_old-1)))*sum(Term_2##(-1));

dl2_dq_dq = N_Obs*(Trigamma(p_old+q_old)-Trigamma(q_old));
dl2_dp_dq = N_Obs*Trigamma(p_old+q_old);

dl2_dp_dp = N_Obs*(Trigamma(p_old+q_old)-Trigamma(p_old));

Partial_1=dl_da // dl_db // dl_dq // dl_dp;
Partial_2=(dl2_da_da||dl2_db_da||dl2_dq_da||dl2_dp_da)//
 (dl2_db_da||dl2_db_db||dl2_dq_db||dl2_dp_db)//
 (dl2_dq_da||dl2_dq_db||dl2_dq_dq||dl2_dp_dq)//
 (dl2_dp_da||dl2_dp_db||dl2_dp_dq||dl2_dp_dp);

Beta_New = Beta_Old - (inv(Partial_2))*Partial_1;
difference = sum(abs(Beta_New - Beta_Old));
a_old = Beta_New[1];
b_old = Beta_New[2];
q_old = Beta_New[3];
p_old = Beta_New[4];

if difference<0.00000001 then i=10000;
```

```
Beta_Old = Beta_New;

end;

a_hat = a_old;
b_hat = b_old;
q_hat = q_old;
p_hat = p_old;

print a_hat b_hat q_hat p_hat difference;

quit;
```

A.18 Kappa Family of Distributions - Three-parameter case

```
/* ======================================================== */
/* FITTING THREE-PARAMETER KAPPA DISTRIBUTION USING MAXIMUM */
/*                 LIKELIHOOD ESTIMATION                    */
/* ======================================================== */

/* Simulate n observations from a Three-parameter
   Kappa(alpha,beta,theta) distribution */
%let alpha  = 4;
%let beta   = 15;
%let theta = 3;
%let n = 10000;

data Kappa_Three_Par;
do j=1 to &n;
X = &beta*((ranuni(0))**(-1/&alpha)-1)**(-1/&theta);
output;
end;
keep X;
run;

proc iml;

use Kappa_Three_Par;
read all into X_Obs;

N_Obs = nrow(X_Obs);
```

```
alpha_old = 2;
beta_old = 5;
theta_old = 2;

B_Old = alpha_old // beta_old // theta_old;

do i=1 to 10000;

Term_1 = (X_Obs##theta_old)#((X_Obs##theta_old
         +J(N_Obs,1,beta_old**theta_old))##(-1));
Term_2 = (1/beta_old)#X_Obs;

/* First Order Partial Derivatives */

dl_dalpha = N_Obs/alpha_old-N_Obs*theta_old*log(beta_old)
            +theta_old*sum(log(X_Obs))-sum(log(J(N_Obs,1,1)
+Term_2##theta_old));
dl_dbeta  = -N_Obs*alpha_old*theta_old/beta_old
            +(theta_old*(alpha_old+1)/beta_old)*sum(Term_1);
dl_dtheta = N_Obs/theta_old-N_Obs*alpha_old*log(beta_old)
            +alpha_old*sum(log(X_Obs))-(alpha_old+1)
*sum(Term_1#log(Term_2));

/* Second Order Partial Derivatives */

dl2_dalpha_dalpha = -N_Obs/alpha_old**2;
dl2_dbeta_dalpha = -N_Obs*theta_old/beta_old
                   +(theta_old/beta_old)*sum(Term_1);
dl2_dtheta_dalpha = -N_Obs*log(beta_old)+sum(log(X_Obs))
                    -sum(Term_1#log(Term_2));

dl2_dbeta_dbeta = (N_Obs*alpha_old*theta_old)/(beta_old**2)
  -((theta_old*(alpha_old+1))/(beta_old**2))*sum((Term_1)
#((J(N_Obs,1,1)+(theta_old*beta_old**theta_old)
  #((J(N_Obs,1,beta_old**theta_old)
+X_Obs##theta_old)##(-1)))));

dl2_dtheta_dbeta = -N_Obs*alpha_old/beta_old
                   +((alpha_old+1)/beta_old)
 *sum(Term_1#(J(N_Obs,1,1)
 +(theta_old*beta_old**theta_old)
         #log((1/beta_old)#X_Obs)#((X_Obs##theta_old
 +J(N_Obs,1,beta_old**theta_old))##(-1))));
```

```
dl2_dtheta_dtheta = -N_Obs/theta_old**2-(alpha_old+1)
                    *sum(Term_1#((beta_old**theta_old)
#((X_Obs##theta_old
+J(N_Obs,1,beta_old**theta_old))##(-1)))
#((log(Term_2))##2));

Partial_1 =  dl_dalpha // dl_dbeta // dl_dtheta;
Partial_2 =
 (dl2_dalpha_dalpha||dl2_dbeta_dalpha||dl2_dtheta_dalpha)//
 (dl2_dbeta_dalpha ||dl2_dbeta_dbeta  ||dl2_dtheta_dbeta )//
 (dl2_dtheta_dalpha||dl2_dtheta_dbeta||dl2_dtheta_dtheta);

B_New = B_Old - (inv(Partial_2))*Partial_1;
difference = sum(abs(B_New - B_Old));
alpha_old = B_New[1];
beta_old  = B_New[2];
theta_old = B_New[3];

if difference<0.000001 then i=10000;

B_Old = B_New;

end;

alpha_hat = alpha_old;
beta_hat  = beta_old;
theta_hat = theta_old;

print alpha_hat beta_hat theta_hat difference;

quit;
```

A.19 Log-logistic Distribution

```
/* =============================================== */
/* FITTING LOG_LOGISTIC DISTRIBUTION USING MAXIMUM */
/*            LIKELIHOOD ESTIMATION                */
/* =============================================== */

/* Simulate n observations from a Log-logistic(alpha,beta)
   distribution */
%let alpha  = 9697;
```

```
%let beta  = 18;
%let n = 100000;

data Log_Logistic;
do j=1 to &n;
X = &alpha*((ranuni(0))**(-1)-1)**(-1/&beta);
output;
end;
keep X;
run;

/* Find initial values for alpha and beta using a regression
   approach */

proc sort data = Log_Logistic;
by X;
run;

%let co = %sysfunc(open(Log_Logistic));
%let cn = %sysfunc(attrn(&co,nobs));
%let cc = %sysfunc(close(&co));

%put &cn;

data Log_Logistic_Regression;
set Log_Logistic;
by X;
F_X = _N_/(&cn+1);
Y = log(X);
X = -log(F_X**(-1)-1);
run;

proc reg data = Log_Logistic_Regression noprint
        outest=Parameter_Estimates rsquare;
model Y = X;
output out=Test;
run;
quit;

data _NULL_;
set Parameter_Estimates;
call symput("alpha_initial",exp(Intercept));
call symput("beta_initial",1/X);
run;
```

```
/* Maximum Likelihood Estimation */

proc iml;

use Log_Logistic;
read all into X_Obs;

N_Obs = nrow(X_Obs);

alpha_old = &alpha_initial;
beta_old = &beta_initial;

B_Old = alpha_old // beta_old;

do i=1 to 1000;

Term_1 = (X_Obs##beta_old)#((X_Obs##beta_old
         +J(N_Obs,1,alpha_old**beta_old))##(-1));
Term_2 = (J(N_Obs,1,alpha_old**beta_old))
         #((X_Obs##beta_old
 +J(N_Obs,1,alpha_old**beta_old))##(-1));
Term_3 = log((1/alpha_old)#X_Obs);

/* First Order Partial Derivatives */

dl_dalpha = -(N_Obs*beta_old/alpha_old)+(2*beta_old/alpha_old)
*sum(Term_1);
dl_dbeta  = (N_Obs/beta_old)+sum(Term_3)-2*sum(Term_1#Term_3);

/* Second Order Partial Derivatives */

dl2_dalpha_dalpha = (N_Obs*beta_old/(alpha_old**2))
                    -(2*beta_old/(alpha_old**2))*sum(Term_1)
-(2*(beta_old**2)/(alpha_old**2))
*sum(Term_1#Term_2);

dl2_dbeta_dalpha = -N_Obs/alpha_old+(2/alpha_old)*sum(Term_1)
                   +(2*beta_old/alpha_old)
 *sum(Term_1#Term_2#Term_3);

dl2_dbeta_dbeta = -N_Obs/(beta_old**2)
                  -2*sum(Term_3#Term_3#Term_1#Term_2);
```

```
Partial_1 =  dl_dalpha // dl_dbeta;
Partial_2 =  (dl2_dalpha_dalpha || dl2_dbeta_dalpha) //
       (dl2_dbeta_dalpha || dl2_dbeta_dbeta ) ;

B_New = B_Old - (inv(Partial_2))*Partial_1;

difference = sum(abs(B_New - B_Old));
alpha_old = B_New[1];
beta_old  = B_New[2];

if difference<0.00000001 then i=1000;

B_Old = B_New;

end;

alpha_hat = alpha_old;
beta_hat  = beta_old;

print alpha_hat beta_hat difference;

quit;
```

A.20 Folded Normal Distribution

```
/* ================================================ */
/* FITTING FOLDED NORMAL DISTRIBUTION USING MAXIMUM */
/*              LIKELIHOOD ESTIMATION               */
/* ================================================ */

%let mu = 3354;
%let sigma = 7678;

data Folded_Normal;
do i = 1 to 100000;
X = abs(&mu+&sigma*rannor(0));
output;
end;
drop i;
run;

proc univariate data = Folded_Normal;
var X;
```

```
histogram;
output out=Moments Mean = _MEAN_ std = _STDEV_
                   skewness=_SKEW_
                   kurtosis = _KURTOSIS_;
run;

data _NULL_;
set Moments;
sigma_MME = _MEAN_*sqrt(constant('PI')/2);
call symput ("MEAN",_MEAN_);
call symput ("STDEV",_STDEV_);
run;

proc iml;

use Folded_Normal;
read all into X_Obs;

N_Obs = nrow(X_Obs);

mu_old = &MEAN;
sigma_old = &STDEV;

B_Old = mu_old // sigma_old;

do i = 1 to 1000;

Term_1 = 2#(X_Obs#J(N_Obs,1,mu_old))
         #((J(N_Obs,1,sigma_old**2))##(-1));
Term_2 = (X_Obs-J(N_Obs,1,mu_old))
         #(J(N_Obs,1,sigma_old**(-2)));
Term_3 = ((X_Obs-J(N_Obs,1,mu_old))##2)
         #(J(N_Obs,1,sigma_old**(-3)));
Term_4 = (X_Obs#J(N_Obs,1,mu_old))
         #(J(N_Obs,1,sigma_old**(-3)));

/* First Order Partial Derivatives */

dl_dmu = sum(Term_2)-2*sum((X_Obs#J(N_Obs,1,sigma_old**(-2)))
         #((exp(-Term_1))#((J(N_Obs,1,1)
 +exp(-Term_1))##(-1))));

dl_dsigma = -N_Obs/sigma_old+sum(Term_3)
            +4*sum(Term_4#((exp(-Term_1))
```

```
            #((J(N_Obs,1,1)+exp(-Term_1))##(-1))));

/* Second Order Partial Derivatives */

dl2_dmu_dmu = -N_Obs/sigma_old**2
        +4*sum(((X_Obs#J(N_Obs,1,sigma_old**(-2)))##(2))
#((exp(Term_1))#((J(N_Obs,1,1)+exp(Term_1))##(-2))));

dl2_dsigma_dmu = -2*sum((X_Obs-J(N_Obs,1,mu_old))
 #(J(N_Obs,1,sigma_old**(-3))))
 +4*sum(X_Obs#J(N_Obs,1,sigma_old**(-3))
 #((J(N_Obs,1,1)+exp(Term_1))##(-1)))
 -8*sum((X_Obs##2)#J(N_Obs,1,mu_old)
 #(J(N_Obs,1,sigma_old**(-5)))#((exp(Term_1))
 #((J(N_Obs,1,1)+exp(Term_1))##(-2))));

dl2_dsigma_dsigma = N_Obs/sigma_old**2-3*sum(Term_2##2)
-12*sum((1/sigma_old)#Term_4#((exp(Term_1)
+J(N_Obs,1,1))##(-1)))+(16*mu_old**2/sigma_old**6)
*sum(X_Obs#X_Obs#exp(Term_1)#((exp(Term_1)
+J(N_Obs,1,1))##(-2)));

Partial_1 =  dl_dmu // dl_dsigma;
Partial_2 =  (dl2_dmu_dmu || dl2_dsigma_dmu) //
             (dl2_dsigma_dmu || dl2_dsigma_dsigma);

B_New = B_Old - (inv(Partial_2))*Partial_1;
mu_old = B_New[1];
sigma_old = B_New[2];
difference = sum(abs(B_New-B_Old));

if difference < 0.00001 then i = 1000;

B_Old = B_New;

end;

mu_hat = mu_old;
sigma_hat = sigma_old;

print mu_hat sigma_hat difference;

quit;
```

A.21 Inverse Gamma Distribution

```
/* ================================================ */
/* FITTING INVERSE GAMMA DISTRIBUTION USING MAXIMUM */
/*              LIKELIHOOD ESTIMATION               */
/* ================================================ */

/* Simulate n observations from a Inverse Gamma(thet,kappa)
   distribution */

%let theta = 6750;
%let alpha = 4;
%let n = 100000;

data Inverse_Gamma;
do i = 1 to &n;
X = ((1/&theta)*rangam(0,&alpha))**(-1);
output;
end;
drop i;
run;

proc univariate data = Inverse_Gamma;
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ skewness=_SKEWNESS_
                   kurtosis=_KURTOSIS_;
run;

data _NULL_;
set Moments;
call symput("Mean",_MEAN_);
run;

/* Fit the Inverse Gamma distribution using maximum likelihood
   estimation */

proc iml;

use Inverse_Gamma;
read all into X_Obs;
```

```
N_Obs = nrow(X_Obs);

theta_old = &Mean;
alpha_old = 1;

B_Old = alpha_old // theta_old;

do i=1 to 10000;

/* First Order Partial Derivatives */

dl_dalpha = N_Obs*log(theta_old)-N_Obs*Digamma(alpha_old)
            -sum(log(X_Obs));
dl_dtheta = N_Obs*alpha_old/theta_old-sum(X_Obs##(-1));

/* Second Order Partial Derivatives */

dl2_dalpha_dalpha = -N_Obs*Trigamma(alpha_old);
dl2_dtheta_dalpha = N_Obs/theta_old;

dl2_dtheta_dtheta = -N_Obs*alpha_old/(theta_old**2);


/*print dl2_dalpha_dalpha dl2_dtheta_dalpha;*/
/*print dl2_dtheta_dtheta;*/

Partial_1 = dl_dalpha // dl_dtheta;
Partial_2 = (dl2_dalpha_dalpha || dl2_dtheta_dalpha) //
      (dl2_dtheta_dalpha || dl2_dtheta_dtheta);

B_New = B_Old - (inv(Partial_2))*Partial_1;
difference = sum(abs(B_New - B_Old));
alpha_old = B_New[1];
theta_old = B_New[2];

if difference<0.000001 then i=10000;

B_Old = B_New;

end;

alpha_hat = alpha_old;
theta_hat = theta_old;
```

```
print alpha_hat theta_hat difference;

quit;
```

A.22 Inverse Chi-square Distribution

```
/* ===================================================== */
/* FITTING INVERSE CHI-SQUARE DISTRIBUTION USING MAXIMUM */
/*                 LIKELIHOOD ESTIMATION                 */
/* ===================================================== */

/* Simulate n observations from a Inverse Chi-square(nu)
   distribution */
%let nu = 30;
%let n = 100000;

data Inverse_Chi_Square;
do i = 1 to &n;
X = ((rangam(0,&nu/2))**(-1))/2;
output;
end;
drop i;
run;

/* Fit the Inverse Gamma distribution using maximum likelihood
   estimation */

proc iml;

use Inverse_Chi_Square;
read all into X_Obs;

N_Obs = nrow(X_Obs);

nu_old = 1;

do i=1 to 1000;

nu_new = nu_old-((N_Obs/2)*log(2)+(N_Obs/2)*Digamma(nu_old/2)
         +(1/2)*sum(log(X_Obs)))/((N_Obs/4)
 *Trigamma(nu_old/2));

difference = sum(abs(nu_new - nu_old));
```

```
if difference<0.000001 then i=1000;

nu_old = nu_new;

end;

nu_hat = nu_old;

print nu_hat difference;

quit;
```

A.23 Loggamma Distribution

```
/* ====================================================== */
/* FITTING LOGGAMMA DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                      ESTIMATION                        */
/* ====================================================== */

/* Simulate n observations from a Loggamma(lambda,a)
   distribution */
%let lambda = 67;
%let a = 465;
%let n = 100000;

data Loggamma;
do i = 1 to &n;
Y = (1/&lambda)*rangam(0,&a);
X=exp(Y);
output;
end;
drop i Y;
run;

proc univariate data = Loggamma;
var X;
histogram;
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ skewness=_SKEWNESS_
                   kurtosis=_KURTOSIS_;
run;
```

```
data _NULL_;
set Moments;
call symput("Mean",_MEAN_);
run;

/* Fit the Gamma distribution using maximum likelihood
   estimation */

proc iml;

use Loggamma;
read all into X_Obs;
N_Obs = nrow(X_Obs);

a_old = 100;
lambda_old = (1-&mean**(-1/a_old))**(-1);

B_Old = a_old // lambda_old;

do i=1 to 10000;

/* First Order Partial Derivatives */

dl_da = N_Obs*log(lambda_old)-N_Obs*Digamma(a_old)
        +sum(log(log(X_Obs)));

dl_dlambda = N_Obs*a_old/lambda_old-sum(log(X_Obs));

/* Second Order Partial Derivatives */

dl2_da_da = -N_Obs*Trigamma(a_old);
dl2_dlambda_da = N_Obs/lambda_old;

dl2_dlambda_dlambda = -N_Obs*a_old/(lambda_old**2);

Partial_1 = dl_da // dl_dlambda;
Partial_2 = (dl2_da_da  || dl2_dlambda_da) //
(dl2_dlambda_da || dl2_dlambda_dlambda);

B_New = B_Old - (inv(Partial_2))*Partial_1;
difference = sum(abs(B_New-B_Old));
a_old = B_New[1];
lambda_old = B_New[2];
```

```
if difference < 0.000001 then i=10000;

B_Old = B_New;

end;

a_hat = a_old;
lambda_hat = lambda_old;

print a_hat lambda_hat difference;

quit;
```

A.24 Snedecor's F Distribution

```
/* ========================================================= */
/* FITTING SNEDECOR F DISTRIBUTION USING MAXIMUM LIKELIHOOD */
/*                         ESTIMATION                        */
/* ========================================================= */

/* Simulate n observations from a F(nu1,nu2) distribution */
%let nu1 = 8;
%let nu2 = 10;
%let n = 100000;

data Snedecor_F;
do i = 1 to &n;
X = FINV(ranuni(0),&nu1,&nu2);
output;
end;
drop i;
run;

proc univariate data = Snedecor_F;
var X;
histogram; /*  */
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ skewness=_SKEWNESS_
                   kurtosis=_KURTOSIS_ N = _COUNT_;
run;

data Moments;
set Moments;
```

```
nu_1_MME = (2*_MEAN_**2)/((_STDDEV_**2+_MEAN_**2)
           *(2-_MEAN_)-_MEAN_**2);
nu_2_MME = 2*_MEAN_/(_MEAN_-1);
call symput("nu_1_MME",nu_1_MME);
call symput("nu_2_MME",nu_2_MME);
run;

proc iml;

use Snedecor_F;
read all into X_Obs;

N_Obs = nrow(X_Obs);

nu_1_old = &nu_1_MME;
nu_2_old = &nu_2_MME;

B_Old = nu_1_old // nu_2_old;

do i=1 to 100;

Term_1 = J(N_Obs,1,1)+(nu_1_old/nu_2_old)#X_Obs;
Term_2 = X_Obs#((J(N_Obs,1,nu_2_old)+nu_1_old#X_Obs)##(-1));
Term_3 = X_Obs#((J(N_Obs,1,nu_2_old)+nu_1_old#X_Obs)##(-2));

/* First Order Partial Derivatives */

dl_dnu1 = (N_Obs/2)*Digamma((nu_1_old+nu_2_old)/2)-(N_Obs/2)
          *Digamma(nu_1_old/2)+(N_Obs/2)*log(nu_1_old/nu_2_old)
      +(N_Obs/2)+0.5*sum(log(X_Obs))-0.5*sum(log(Term_1))
-((nu_1_old+nu_2_old)/2)*sum(Term_2);

dl_dnu2 = (N_Obs/2)*Digamma((nu_1_old+nu_2_old)/2)
          -(N_Obs/2)*Digamma(nu_2_old/2)
-((N_Obs*nu_1_old)/(2*nu_2_old))-0.5*sum(log(Term_1))
+((nu_1_old+nu_2_old)/2)*sum((nu_1_old/nu_2_old)#Term_2);

/* Second Order Partial Derivatives */

dl2_dnu1_dnu1 = (N_Obs/4)*Trigamma((nu_1_old+nu_2_old)/2)
                -(N_Obs/4)*Trigamma(nu_1_old/2)
+(N_Obs/(2*nu_1_old))
        -sum(Term_2)+((nu_1_old+nu_2_old)/2)*sum(Term_2##2);
```

```
dl2_dnu2_dnu1 = (N_Obs/4)*Trigamma((nu_1_old+nu_2_old)/2)
               -(N_Obs/(2*nu_2_old))
 -0.5*((nu_2_old-nu_1_old)/nu_2_old)
 *sum(Term_2)+((nu_1_old+nu_2_old)/2)*sum(Term_3);

dl2_dnu2_dnu2 = (N_Obs/4)*Trigamma((nu_1_old+nu_2_old)/2)
               -(N_Obs/4)*Trigamma(nu_2_old/2)
 +((N_Obs*nu_1_old)/(2*nu_2_old**2))
       +(nu_1_old/nu_2_old)*sum(Term_2)
 -((nu_1_old+nu_2_old)/2)*(nu_1_old/(nu_2_old**2))
       *sum(Term_3#(J(N_Obs,1,2*nu_2_old)+nu_1_old#X_Obs));

Partial_1 = dl_dnu1 // dl_dnu2;
Partial_2 = (dl2_dnu1_dnu1 || dl2_dnu2_dnu1) //
            (dl2_dnu2_dnu1 || dl2_dnu2_dnu2);

B_New = B_Old - 0.5*(inv(Partial_2))*Partial_1;
difference = sum(abs(B_New-B_Old));
nu_1_old = B_New[1];
nu_2_old = B_New[2];

if difference < 0.000001 then i = 1000;

B_Old = B_New;

end;

nu_1_hat = nu_1_old;
nu_2_hat = nu_2_old;

print nu_1_hat nu_2_hat difference;

quit;
```

A.25 Inverse Gaussian Distribution

```
/* ================================================== */
/* FITTING INVERSE GAUSSIAN DISTRIBUTION USING MAXIMUM */
/*               LIKELIHOOD ESTIMATION                */
/* ================================================== */

/* Simulate n observations from a Inverse Gaussian(mu,lambda)
   distribution */
```

```
%let mu = 4;
%let lambda = 10;
%let n = 10000;
%let increment = 0.0001;

data Real_Line_Partition;
do i = 1 to 1000001;
X = &increment*(i-1);
output;
end;
run;

data Inverse_Gaussian_CDF;
set Real_Line_Partition;

if X=0 then F_X=0;
else F_X = CDF('NORMAL',sqrt(&lambda/X)*((X/&mu)-1))
           +exp(2*(&lambda/&mu))
 *CDF('NORMAL',-sqrt(&lambda/X)*((X/&mu)+1));
run;

data Random_Numbers;
do i=1 to &n;
u=ranuni(0);
output;
end;
run;

proc sql;
create table Random_Inverse_Gaussian
as select a.U, max(b.X) as X
from Random_Numbers as a left join Inverse_Gaussian_CDF as b
on a.U>=b.F_X
group by U;
quit;

proc univariate data = Random_Inverse_Gaussian;
var X;
histogram; /*  */
inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3);
output out=Moments mean=_MEAN_ std=_STDDEV_ skewness=_SKEWNESS_
                   kurtosis=_KURTOSIS_ N = _COUNT_;
run;
```

```
data Inverse_Gaussian;
set Random_Inverse_Gaussian (keep = X);
run;

data Moments;
set Moments;
mu_MME = _MEAN_;
lambda_MME = (_MEAN_**3)/(_STDDEV_**2);
call symput("mu_MME",mu_MME);
call symput("lambda_MME",lambda_MME);
run;

proc iml;

use Inverse_Gaussian;
read all into X_Obs;
N_Obs = nrow(X_Obs);

mu_old = &mu_MME;
lambda_old = &lambda_MME;

B_Old = mu_old // lambda_old;

do i=1 to 1000;

Term_1 = ((J(N_Obs,1,mu_old)-X_Obs)##2)#((X_Obs)##(-1));
Term_2 = ((J(N_Obs,1,mu_old)-X_Obs))#((X_Obs)##(-1));

dl_dmu = (lambda_old/mu_old**3)*sum(Term_1)
         -(lambda_old/mu_old**2)*sum(Term_2);

dl_dlambda = N_Obs/(2*lambda_old)
             -(1/(2*mu_old**2))*sum(Term_1);

dl2_dmu_dmu = -(3*lambda_old/mu_old**4)*sum(Term_1)
              +(4*lambda_old/mu_old**3)*sum(Term_2)
-(lambda_old/mu_old**2)*sum(X_Obs##(-1));

dl2_dlambda_dmu = (mu_old**(-3))*sum(Term_1)
                  -(mu_old**(-2))*sum(Term_2);

dl2_dlambda_dlambda = -N_Obs/(2*lambda_old**2);
```

```
Partial_1 = dl_dmu // dl_dlambda;
Partial_2 = (dl2_dmu_dmu    || dl2_dlambda_dmu) //
            (dl2_dlambda_dmu  || dl2_dlambda_dlambda);

B_New = B_Old - (inv(Partial_2))*Partial_1;
difference = sum(abs(B_New - B_Old));
mu_old = B_New[1];
lambda_old = B_New[2];

if difference < 0.00001 then i = 1000;

B_Old = B_New;

end;

mu_hat = mu_old;
lambda_hat = lambda_old;

print mu_hat lambda_hat difference;

quit;
```

A.26 Skew-normal Distribution

```
/* ======================================================= */
/* FITTING AZZALINI SKEW_NORMAL DISTRIBUTION USING MAXIMUM */
/*                 LIKELIHOOD ESTIMATION                   */
/* ======================================================= */
%let lambda = 2;

%let xi = 7;
%let eta = 5;

%let mu = 45;
%let sigma = 234;

%let min = -10;
%let max = 10;
%let increment = 0.0005;

%let n = 3000;

/* Simulate observations from a Skew-normal distribution and
```

```
   apply parameterizations */
data Skew_Normal;
retain cap_h_X 0;
retain pre_h_X 0;
do i = 1 to (&max-&min)/&increment;
X = &min+(i)*&increment;
h_X = 2*(CDF('NORMAL',&lambda*X,0,1))*(PDF('NORMAL',X,0,1));
cap_h_X = cap_h_X + 0.5*(pre_h_X+h_X)*&increment;
output;
pre_h_X = h_X;
end;
drop i;
run;

data Random_Numbers;
do i = 1 to &n;
U = ranuni(0);
output;
end;
run;

proc sql;
create table Azzalini_Skew_Normal
as select unique(a.U) as U, max(b.X) as X
from Random_Numbers as a left join Skew_Normal as b
on a.U >= b.cap_h_X
group by U;
quit;

data Azzalini_Skew_Normal;
set Azzalini_Skew_Normal;
b = sqrt(2/constant('PI'));
delta = &lambda/sqrt(1+&lambda**2);
Y_Direct   = &xi+&eta*X;
Y_Centered = &mu+&sigma*((X-b*delta)
             /(sqrt(1-(b**2)*(delta**2))));
drop b delta;
run;

/* Assess the simulated distributions */
proc univariate data = Azzalini_Skew_Normal;
var X;
histogram;
output out=Moments_X Mean = _MEAN_ std = _STDEV_
```

```
                     skewness=_SKEW_
                     kurtosis = _KURTOSIS_;
run;

proc univariate data = Azzalini_Skew_Normal;
var Y_Direct;
histogram;
output out=Moments_Y_Direct Mean = _MEAN_ std = _STDEV_
                            skewness=_SKEW_
                            kurtosis = _KURTOSIS_;
run;

proc univariate data = Azzalini_Skew_Normal;
var Y_Centered;
histogram;
output out=Moments_Y_Centered Mean = _MEAN_ std = _STDEV_
                              skewness=_SKEW_
                              kurtosis = _KURTOSIS_;
run;

/* METHOD-OF-MOMENTS ESTIMATION */
/* Standard Parameterization */
data Moments_X;
set Moments_X;
b = sqrt(2/CONSTANT('PI'));
delta_hat = sqrt((_SKEW_**(2/3))/((b**2)*(_SKEW_**(2/3))
            +(b**(2/3))*(2*b**2-1)**(2/3)));
lambda_hat = delta_hat/sqrt(1-delta_hat**2);
run;

/* Direct Parameterization */
data Moments_Y_Direct;
set Moments_Y_Direct;
xi_s = -((2/(4-CONSTANT('PI')))**(1/3))*(_SKEW_**(1/3));
eta_s = sqrt(1+xi_s**2);

b = sqrt(2/CONSTANT('PI'));

delta_hat = -xi_s/(b*eta_s);
xi_hat = _STDEV_*xi_s+_MEAN_;
eta_hat = eta_s*_STDEV_;
lambda_hat = delta_hat/sqrt(1-delta_hat**2);
run;
```

```
/* Centered Parameterization */
data Moments_Y_Centered;
set Moments_Y_Centered;

mu_hat = _MEAN_;
sigma_hat = _STDEV_;
gamma_hat = _SKEW_;

b = sqrt(2/CONSTANT('PI'));
delta_hat = sqrt((gamma_hat**(2/3))/((b**2)*(gamma_hat**(2/3))
            +(b**(2/3))*(2*b**2-1)**(2/3)));
lambda_hat = delta_hat/sqrt(1-delta_hat**2);
run;

/* Assess whether the centered parameterization is valid for
   the standard case */
data Moments_X_with_Centered_Approach;
set Moments_X;

mu_hat = _MEAN_;
sigma_hat = _STDEV_;
gamma_hat = _SKEW_;

b = sqrt(2/CONSTANT('PI'));
delta_hat = sqrt((gamma_hat**(2/3))/((b**2)*(gamma_hat**(2/3))
            +(b**(2/3))*(2*b**2-1)**(2/3)));
lambda_hat = delta_hat/sqrt(1-delta_hat**2);

test_E_X = b*delta_hat;
test_S_X = sqrt(1-(b**2)*(delta_hat**2));
run;


/* Assess the accuracy of the maximum likelihood estimation
   using a grid-search approach */
%let Parameterization = X;

data Keep_X;
set Azzalini_Skew_Normal (keep = &Parameterization);
run;

data _NULL_;
set Moments_&Parameterization(obs=1);
call symput("STDEV",_STDEV_);
```

```
call symput("MEAN",_MEAN_);
run;

proc iml;
reset;
use Keep_X;
read all into X_Obs;

min_tau=-sqrt(2/(CONSTANT('PI')-2))+0.000001;
max_tau=-min_tau;

min_mu_s = -3;
max_mu_s = 3;

Loops = 100;

N_Obs = nrow(X_Obs);

Parameters = J(Loops*Loops,2,0);
log_likelihood = J(Loops*Loops,1,0);

do i = 1 to Loops;
do j = 1 to Loops;

b = sqrt(2/CONSTANT('PI'));

mu_s = min_mu_s+i*((max_mu_s-min_mu_s)/Loops);
tau = min_tau+j*((max_tau-min_tau)/Loops);

Parameters[(i-1)*Loops+j,1] = mu_s;
Parameters[(i-1)*Loops+j,2] = tau;

Numerator = (((1/&STDEV)#(X_Obs-J(N_Obs,1,&MEAN))
-J(N_Obs,1,mu_s))#J(N_Obs,1,tau))
#((J(N_Obs,1,sqrt((mu_s**2)*(1+tau**2)+1)
-mu_s*tau))##(-1))+tau**2;

Denominator = sqrt((b**2+(tau**2)*(b**2-1))*(1+tau**2));
VAL = (1/Denominator)*Numerator;

CDF = CDF('NORMAL',VAL,0,1);

do k=1 to N_Obs;
if CDF[k,1]=0 then CDF[k,1]=0.000000000001;
```

```
end;

log_likelihood[(i-1)*Loops+j,1] =
-N_Obs*log(sqrt((mu_s**2)*(1+tau**2)+1)-mu_s*tau)
-(N_Obs/2)*log(1+tau**2)+sum((log(CDF)));
end;
end;

Grid = Parameters || log_likelihood;
Column_Labels = {'mu_s' 'tau' 'log_likelihood' };
create Grid_Data from Grid[colname = Column_Labels];
append from Grid;

quit;

proc sort data = Grid_Data;
by descending log_likelihood;
run;

proc print data = Grid_Data (obs=10);
run;

data MLE;
set Grid_data (obs=1);
b = sqrt(2/CONSTANT('PI'));
gamma_hat = ((tau)**3)*((4-constant('PI'))/2);
mu_hat = &STDEV*mu_s+&MEAN;
sigma_hat=&STDEV*(sqrt(mu_s*(1+tau**2)+1)-mu_s*tau);
delta_hat = sqrt((gamma_hat**(2/3))/((b**2)
            *(gamma_hat**(2/3))
            +(b**(2/3))*(2*b**2-1)**(2/3)));
lambda_hat = delta_hat/sqrt(1-delta_hat**2);
run;
```

Appendix B

Summary of Continuous Distributions

B.1 Beta-prime Distribution - δ_1, δ_2

Probability Density Function:

$$f_{\delta_1,\delta_2}(x) = \frac{1}{B(\delta_1,\delta_2)}\left(\frac{x}{x+1}\right)^{\delta_1-1}\left(\frac{1}{x+1}\right)^{\delta_2+1} \text{ for } x \geq 0 \text{ with } \delta_1, \delta_2 > 0, \tag{B.1}$$

where $B(\delta_1, \delta_2)$ is the beta function as given in Definition 28.

Moments:

$$\mathrm{E}(X^r) = \frac{B(\delta_1+r,\delta_2-r)}{B(\delta_1,\delta_2)} \tag{B.2}$$

$$\mathrm{E}(X) = \frac{\delta_1}{\delta_2-1} \tag{B.3}$$

$$\mathrm{var}(X) = \frac{\delta_1(\delta_1+\delta_2-1)}{(\delta_2-1)^2(\delta_2-2)} \tag{B.4}$$

$$\begin{aligned}\text{skewness}(X) = &\frac{\delta_1(\delta_1+2)(\delta_1+1)(\delta_2-1)^2}{\delta_1(\delta_2-2)^{1/2}(\delta_2-3)(\delta_1+\delta_2-1)} \\ &- 3\frac{\delta_1^2(\delta_1+1)(\delta_2-1)(\delta_2-3)}{\delta_1(\delta_2-2)^{1/2}(\delta_2-3)(\delta_1+\delta_2-1)} \\ &+ 2\frac{\delta_1^3(\delta_2-2)(\delta_2-3)}{\delta_1(\delta_2-2)^{1/2}(\delta_2-3)(\delta_1+\delta_2-1)}\end{aligned} \tag{B.5}$$

Tail Weight:
Heavy tail cannot be confirmed while a contradiction to the condition for heavy-tailedness was not found.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. In some instances the iterative algorithm does not converges, especially for larger observed values.

B.2 Birnbaum-Saunders Distribution - α, β

Probability Density Function:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\tau^2} \frac{1}{2\alpha} \left(\frac{1}{\sqrt{x\beta}} + \frac{\sqrt{\beta}}{x^{\frac{3}{2}}} \right) \text{ for } x > 0 \text{ with } \alpha, \beta > 0 \tag{B.6}$$

Moments:

$$\mathrm{E}(X) = \beta + \left(\frac{\alpha^2 \beta}{2} \right) \tag{B.7}$$

$$\mathrm{var}(X) = \alpha^2 \beta^2 \left(1 + \frac{5}{4}\alpha^2 \right) \tag{B.8}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. The iterative algorithm converges reasonably easily.

B.3 Burr Type XII Distribution - c, k

Probability Density Function:

$$f_X(x) = kcx^{c-1} \left(1 + x^c\right)^{-(k+1)}, \text{ for } x > 0 \text{ with } c, k > 0 \tag{B.9}$$

Moments:

$$\mathrm{E}(X^r) = k \frac{\Gamma(\frac{r}{c} + 1)\Gamma(k - \frac{r}{c})}{\Gamma(k+1)} \tag{B.10}$$

$$\mathrm{E}(X) = k\frac{\Gamma(\frac{1}{c}+1)\Gamma(k-\frac{1}{c})}{\Gamma(k+1)} \tag{B.11}$$

$$\begin{aligned}\mathrm{var}(X) = &\frac{k}{\Gamma(k+1)^2}\left(\Gamma\left(\frac{2}{c}+1\right)\Gamma\left(k-\frac{2}{c}\right)\Gamma(k+1)\right)\\ &-\frac{k^2}{\Gamma(k+1)^2}\left(\Gamma\left(\frac{1}{c}+1\right)\Gamma\left(k-\frac{1}{c}\right)\right)^2\end{aligned} \tag{B.12}$$

Tail Weight:
Heavy tail suggested by decreasing hazard rate function.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. For large values of c the iterative algorithm lacks convergence. For larger values of k the accuracy of the estimation decreases while convergence of the iterative algorithm is less likely.

B.4 Chi-square Distribution - ν

Probability Density Function:

$$f_X(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}x^{\frac{\nu}{2}-1}e^{-\frac{1}{2}x} \text{ for } x \geq 0 \text{ with } \nu > 0 \tag{B.13}$$

Moment Generating Function:

$$M_X(t) = (1-2t)^{-\frac{\nu}{2}}. \tag{B.14}$$

Moments:

$$\mathrm{E}(X^r) = 2^r\prod_{j=0}^{r-1}\left(\frac{\nu}{2}+j\right). \tag{B.15}$$

$$\mathrm{E}(X) = \nu \tag{B.16}$$

$$\mathrm{var}(X) = 2\nu \tag{B.17}$$

$$\mathrm{skewness}(X) = \frac{2^{\frac{3}{2}}}{\sqrt{\nu}} \tag{B.18}$$

$$\mathrm{kurtosis}(X) = 3 + \frac{12}{\nu} \tag{B.19}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\nu} = \overline{x} \tag{B.20}$$

Maximum likelihood:
Numerical optimization of the log-likelihood function is used. The algorithm generally converges easily when using the method-of-moments estimate as initial value. Accuracy increases with larger sample sizes.

B.5 Dagum Distribution - a, b, p

Probability Density Function:

$$f_X(x) = ab^a p x^{-(a+1)} \left(1 + b^a x^{-a}\right)^{-(p+1)} \text{ for } x > 0 \text{ with } a > 0 \text{ and } b, p \geq 0 \tag{B.21}$$

Moments:

$$\mathrm{E}(X^r) = b^r p B\left(\frac{r}{a} + p, 1 - \frac{r}{a}\right) \tag{B.22}$$

$$\mathrm{E}(X) = pb\frac{\Gamma\left(\frac{1}{a} + p\right)\Gamma\left(1 - \frac{1}{a}\right)}{\Gamma(p+1)} \tag{B.23}$$

$$\mathrm{var}(X) = pb^2 B\left(\frac{2}{a} + p, 1 - \frac{2}{a}\right) - p^2 b^2 \left(B\left(\frac{1}{a} + p, 1 - \frac{1}{a}\right)\right)^2 \tag{B.24}$$

Tail Weight:
The hazard rate function is increasing, indicating that the distribution has a heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Convergence of the iterative estimation algorithm is rarely obtained. For cases where convergence is obtained the accuracy of estimates increase with increasing sample sizes.

B.6 Erlang Distribution - λ, n

Probability Density Function:

$$f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \text{ for } n \in \mathbb{N}, \lambda > 0 \text{ and } x \in \mathbb{R}^+ \tag{B.25}$$

Moment Generating Function:

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n \tag{B.26}$$

Moments:

$$\mathrm{E}(X^r) = \frac{\Gamma(n+r)}{\lambda^r \Gamma(n)} \tag{B.27}$$

$$\mathrm{E}(X) = \frac{n}{\lambda} \tag{B.28}$$

$$\mathrm{var}(X) = \frac{n}{\lambda^2} \tag{B.29}$$

$$\mathrm{skewness}(X) = \frac{2}{\sqrt{n}} \tag{B.30}$$

$$\mathrm{kurtosis}(X) = 3 + \frac{6}{n} \tag{B.31}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\lambda} = \frac{n}{\bar{x}} \tag{B.32}$$

Maximum likelihood:

$$\hat{\lambda} = \frac{n}{\bar{x}} \tag{B.33}$$

B.7 Exponential Distribution - θ

Probability Density Function:

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \text{ for } x \geq 0 \text{ with } \theta > 0 \tag{B.34}$$

Moment Generating Function:

$$M_X(t) = \left(\frac{1}{1-\theta t}\right). \tag{B.35}$$

Moments:

$$\mathrm{E}(X^r) = r!\theta^r \tag{B.36}$$

$$\mathrm{E}(X) = \theta \tag{B.37}$$

$$\mathrm{var}(X) = \theta^2 \tag{B.38}$$

$$\mathrm{skewness}(X) = 2 \tag{B.39}$$

$$\mathrm{kurtosis}(X) = 9 \tag{B.40}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\theta} = \overline{x} \tag{B.41}$$

Maximum likelihood:

$$\hat{\theta} = \overline{x} \tag{B.42}$$

B.8 Folded Normal Distribution - μ, σ

Probability Density Function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\left(e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{x+\mu}{\sigma}\right)^2}\right) \tag{B.43}$$

for $x \geq 0$ with $-\infty < \mu < \infty$ and $\sigma > 0$

Moments:

$$\mathrm{E}(X^r) = \sigma^r \sum_{j=0}^{r} \binom{r}{j} \theta^{r-j} \left(I_j(-\theta) + (-1)^{r-j} I_j(\theta)\right) \tag{B.44}$$

where

$$I_j(a) = \frac{1}{\sqrt{2\pi}} \int_a^\infty y^j e^{-\frac{1}{2}y^2} dy \text{ for } j = 1, 2, ...$$

$$I_0(a) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}y^2} dy$$

and

$$\theta = \frac{\mu}{\sigma}$$

For $r = 1$:

$$I_0(-\theta) - I_0(\theta) = -(1 - 2F(\theta)) \tag{B.45}$$

$$I_1(-\theta) + I_1(\theta) = \sqrt{\frac{2}{\pi}} e^{-\frac{\theta^2}{2}} \tag{B.46}$$

For $r = 2$:

$$I_0(-\theta) + I_0(\theta) = 1 \tag{B.47}$$

$$I_1(-\theta) - I_1(\theta) = 0 \tag{B.48}$$

$$I_2(-\theta) + I_2(\theta) = 1 \tag{B.49}$$

For $r = 3$:

$$I_2(-\theta) - I_2(\theta) = -\left(\sqrt{\frac{2}{\pi}}\theta e^{-\frac{1}{2}\theta^2} + (1 - 2I_0(-\theta))\right) \tag{B.50}$$

$$I_3(-\theta) + I_3(\theta) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\theta^2}(\theta^2 + 2) \tag{B.51}$$

For $r = 4$:

$$I_3(-\theta) - I_3(\theta) = 0 \tag{B.52}$$

$$I_4(-\theta) + I_4(\theta) = 3 \tag{B.53}$$

$$\mathrm{E}(X) = \sqrt{\frac{2}{\pi}}\sigma e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} - \mu\left(1 - 2F\left(\frac{\mu}{\sigma}\right)\right) \tag{B.54}$$

$$\mathrm{var}(X) = \mu^2 + \sigma^2 - \left(\sqrt{\frac{2}{\pi}}\sigma e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} - \mu\left(1 - 2F\left(\frac{\mu}{\sigma}\right)\right)\right)^2 \tag{B.55}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. The algorithm has good convergence properties. For larger values of σ and values for μ closer to 0 the estimation becomes less accurate.

B.9 Frechet Distribution - β, δ, λ

Probability Density Function:

$$f_X(x) = \begin{cases} exp\left(-\left(\frac{\delta}{x-\lambda}\right)^{\beta}\right)\beta\delta\left(\frac{\delta}{x-\lambda}\right)^{\beta-1}\left(\frac{1}{x-\lambda}\right)^{2} & \text{if } x \geq \lambda, \\ 0 & \text{otherwise} \end{cases} \tag{B.56}$$

for $x \geq \lambda$ with $\delta, \beta > 0$ and $\lambda \in \mathbb{R}$

Moments:

$$\mathrm{E}(X^r) = \int_0^{\infty} \left(\delta y^{-\frac{1}{\beta}} + \lambda\right)^r e^{-x} dx \tag{B.57}$$

$$\mathrm{E}(X) = \delta\Gamma\left(1 - \frac{1}{\beta}\right) + \lambda \tag{B.58}$$

$$\mathrm{var}(X) = \delta^2\left(\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma\left(1 - \frac{1}{\beta}\right)^2\right) \tag{B.59}$$

$$\mathrm{skewness}(X) = \frac{\Gamma\left(1-\frac{3}{\beta}\right) - 3\Gamma\left(1-\frac{1}{\beta}\right)\left(1-\frac{2}{\beta}\right) + 2\Gamma\left(1-\frac{1}{\beta}\right)^3}{\left(\delta^2\left(\Gamma\left(1-\frac{2}{\beta}\right) - \Gamma\left(1-\frac{1}{\beta}\right)^2\right)\right)^{\frac{3}{2}}} \tag{B.60}$$

Tail Weight:
The hazard rate function is decreasing which indicates a heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Convergence of the estimation algorithm is highly dependent on the choice of initial values. When the true value of λ is large in relation to the values of β and δ, estimates of β and δ becomes less accurate.

B.10 Gamma Distribution - θ, κ

Probability Density Function:

$$f_X(x) = \frac{1}{\theta^{\kappa}\Gamma(\kappa)} x^{\kappa-1} e^{-\frac{x}{\theta}} \text{ for } x \geq 0 \text{ with } \theta, \kappa > 0 \tag{B.61}$$

Moment Generating Function:

$$M_X(t) = (1-\theta t)^{-\kappa} \tag{B.62}$$

Moments:

$$\mathrm{E}(X^r) = \frac{\Gamma(\kappa+r)}{\Gamma(\kappa)}\theta^r = \left(\prod_{j=0}^{r-1}(\kappa+j)\right)\theta^r \tag{B.63}$$

$$\mathrm{E}(X) = \kappa\theta \tag{B.64}$$

$$\mathrm{var}(X) = \kappa\theta^2 \tag{B.65}$$

$$\mathrm{skewness}(X) = \frac{2}{\sqrt{\kappa}} \tag{B.66}$$

$$\mathrm{kurtosis}(X) = \frac{3(\kappa+2)}{\kappa} \tag{B.67}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\theta} = \frac{s_x^2}{\overline{x}}, \text{ and} \tag{B.68}$$

$$\hat{\kappa} = \frac{\overline{x}^2}{s_x^2} \tag{B.69}$$

Maximum likelihood:
Numerical optimization of the log-likelihood function is used. The algorithm generally converges easily when using the method-of-moments estimates as initial values. Accuracy increases with larger sample sizes.

B.11 Generalized Beta Distribution of the Second Kind - a, b, p, q

Probability Density Function:

$$f_X(x) = \frac{a x^{ap-1}}{b^{ap}B(p,q)\left(1+\left(\frac{x}{b}\right)^a\right)^{p+q}} \text{ for } x \geq 0 \text{ with } a,b,p,q > 0 \tag{B.70}$$

Moments:

$$\mathrm{E}(X^r) = b^r\frac{\Gamma\left(\frac{r}{a}+p\right)\Gamma\left(q-\frac{r}{a}\right)}{\Gamma(p)\Gamma(q)} \tag{B.71}$$

$$\mathrm{E}(X) = b\frac{\Gamma\left(\frac{1}{a}+p\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(p)\Gamma(q)} \tag{B.72}$$

$$\mathrm{var}(X) = b^2\frac{\Gamma\left(\frac{2}{a}+p\right)\Gamma\left(q-\frac{2}{a}\right)}{\Gamma(p)\Gamma(q)} - b^2\left(\frac{\Gamma\left(\frac{1}{a}+p\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(p)\Gamma(q)}\right)^2 \tag{B.73}$$

Tail Weight:
Heavy-tailedness cannot be confirmed algebraically. There exist combinations of values for a, b, p and q for which the distribution does not have a heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Convergence of the estimation algorithm is generally difficult to obtain.

B.12 Generalized Pareto Distribution - κ, θ, τ

Probability Density Function:

$$f_X(x) = \frac{\Gamma(\kappa+\tau)}{\Gamma(\kappa)\Gamma(\tau)}\frac{\theta^\kappa x^{\tau-1}}{(x+\theta)^{\kappa+\tau}} \text{ for } x > 0 \text{ with } \kappa,\tau,\theta > 0 \tag{B.74}$$

Moments:

$$\mathrm{E}(X^r) = \theta^r\frac{\Gamma(\tau+r)\Gamma(\kappa-r)}{\Gamma(\kappa)\Gamma(\tau)} \text{ for } -\tau < r < \kappa \tag{B.75}$$

$$\mathrm{E}(X) = \frac{\theta\tau}{\kappa-1} \tag{B.76}$$

$$\mathrm{var}(X) = \frac{\theta^2\tau(\kappa+\tau-1)}{(\kappa-1)^2(\kappa-2)} \tag{B.77}$$

$$\mathrm{skewness}(X) = \frac{2(2\tau+\kappa-1)}{(\kappa-3)\sqrt{\tau(\kappa-2)(\kappa+\tau-1)}} \tag{B.78}$$

Tail Weight:
The distribution has heavy tail, but this is dependent on the values of κ and τ.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. the algorithm yields convergence given that suitable initial values are chosen. The accuracy of these estimates improve with increasing sample sizes.

B.13 Gumbel Distribution - α, ξ

Probability Density Function:

$$f(x) = \frac{1}{\alpha} e^{-\left(e^{-\left(\frac{x-\xi}{\alpha}\right)} + \left(\frac{x-\xi}{\alpha}\right)\right)} \text{ with } \xi, \alpha > 0 \tag{B.79}$$

Moments:

$$\mathrm{E}(X) = \xi + 0.57721566\alpha \tag{B.80}$$

$$\mathrm{var}(X) = 1.64493407\alpha^2 \tag{B.81}$$

$$\mathrm{skewness}(X) = 2.4042\alpha^3 \tag{B.82}$$

$$\mathrm{kurtosis}(X) = \frac{\pi^4}{15}\alpha^3 \tag{B.83}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
<u>Maximum likelihood</u>:
Numerical optimization of the log-likelihood function is used. Convergence of the estimation algorithm is highly dependent on the choice of initial values for α and ξ. Accuracy of estimates increases with increased sample sizes. When value of α is large relative to the value of ξ, accuracy of estimate for ξ decreases.

B.14 Inverse Chi-square Distribution - ν

Probability Density Function:

$$f_X(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{x}\right)^{\frac{\nu}{2}+1} e^{-\frac{1}{2x}} \text{ for } x > 0 \text{ with } \nu > 0 \tag{B.84}$$

Moments:

$$\mathrm{E}(X^r) = \frac{\Gamma\left(\frac{\nu}{2} - r\right)}{2^r\Gamma\left(\frac{\nu}{2}\right)} \tag{B.85}$$

$$\mathrm{E}(X) = \frac{1}{\nu - 2} \tag{B.86}$$

$$\mathrm{var}(X) = \frac{2}{(\nu-2)^2(\nu-4)} \tag{B.87}$$

$$\mathrm{skewness}(X) = \frac{2^{\frac{5}{2}}\sqrt{\nu-4}}{\nu-6} \tag{B.88}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Generally the algorithm reaches convergence easily with reasonable accuracy of the estimate for ν.

B.15 Inverse Gamma Distribution - α, θ

Probability Density Function:

$$f_X(x) = \frac{\left(\frac{\theta}{x}\right)^{\alpha} e^{-\frac{\theta}{x}}}{x\Gamma(\alpha)} \text{ for } x > 0 \text{ with } \alpha, \beta > 0 \tag{B.89}$$

Moments:

$$\mathrm{E}(X^r) = \frac{\Gamma(\alpha - r)}{\Gamma(\alpha)}\theta^r \tag{B.90}$$

$$\mathrm{E}(X) = \frac{\theta}{\alpha - 1} \tag{B.91}$$

$$\mathrm{var}(X) = \frac{\theta^2}{(\alpha-2)(\alpha-1)^2} \tag{B.92}$$

$$\mathrm{skewness}(X) = \frac{4\sqrt{\alpha-2}}{\alpha-3} \tag{B.93}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-Moments:

$$\hat{\alpha} = \frac{\overline{x} + 2s_x^2}{s_x^2}, \text{ and} \tag{B.94}$$

$$\hat{\theta} = \frac{\overline{x}^2 + 2s_x^2\overline{x}^2}{s_x^2} \tag{B.95}$$

Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Generally the algorithm reaches convergence easily with reasonable accuracy of the parameter estimates, but is sensitive to the choice of initial values.

B.16 Inverse Gaussian Distribution - μ, λ

Probability Density Function:

$$f_T(t) = \sqrt{\frac{\lambda}{2\pi t^3}} e^{-\frac{\lambda(\mu-t)^2}{2\mu^2 t}} \text{ for } t > 0 \text{ with } \mu, \lambda > 0 \tag{B.96}$$

Moment Generating Function:

$$M_T(\tau) = \frac{e^{\left(\frac{\lambda}{\mu}\right)}}{e^{\sqrt{\left(\frac{\lambda}{\mu}\right)^2 - 2\lambda\tau}}} \tag{B.97}$$

Moments:

$$\mathrm{E}(T^{-r}) = \frac{\mathrm{E}(T^{r+1})}{\mu^{2r+1}} \tag{B.98}$$

$$\mathrm{E}(T) = \mu \tag{B.99}$$

$$\mathrm{var}(T) = \frac{\mu^3}{\lambda} \tag{B.100}$$

$$\mathrm{skewness}(X) = 3\sqrt{\frac{\mu}{\lambda}} \tag{B.101}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\mu} = \overline{x}, \text{ and} \tag{B.102}$$

$$\hat{\lambda} = \frac{\overline{x}^3}{s_x^2} \tag{B.103}$$

Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Generally the algorithm reaches convergence easily with reasonable accuracy of the parameter estimates. The method-of-moments estimates can be used as initial values to the iterative estimation algorithm.

B.17 Kappa Family (Two-parameter) Distribution - α, β

Probability Density Function:

$$f_X(x) = \frac{x^{-(\alpha+1)}}{\beta^{-\alpha}} \left(1 + \left(\frac{x}{\beta}\right)^{-\alpha}\right)^{-\left(\frac{1}{\alpha}+1\right)} \text{ for } x \geq 0 \text{ and } \alpha, \beta > 0 \tag{B.104}$$

Moments:

$$\mathrm{E}(X^r) = \frac{\beta^r}{\alpha} \frac{\Gamma\left(\frac{r+1}{\alpha}\right)\Gamma\left(\frac{\alpha-r}{\alpha}\right)}{\Gamma(\frac{\alpha+1}{\alpha})} \tag{B.105}$$

$$\mathrm{E}(X) = \frac{\beta}{\alpha} B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right) \tag{B.106}$$

$$\mathrm{var}(X) = \left(\frac{\beta}{\alpha}\right)^2 \left(\alpha B\left(\frac{3}{\alpha}, 1-\frac{2}{\alpha}\right) - \left(B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right)\right)^2\right) \tag{B.107}$$

$$\begin{aligned} \mathrm{skewness}(X) = &\frac{\alpha^2 B\left(\frac{4}{\alpha}, 1-\frac{3}{\alpha}\right)}{\left(\alpha B\left(\frac{3}{\alpha}, 1-\frac{2}{\alpha}\right) - \left(B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right)\right)^2\right)^{3/2}} \\ &- 3\frac{\alpha B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right) B\left(\frac{3}{\alpha}, 1-\frac{2}{\alpha}\right)}{\left(\alpha B\left(\frac{3}{\alpha}, 1-\frac{2}{\alpha}\right) - \left(B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right)\right)^2\right)^{3/2}} \\ &+ \frac{2\left(B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right)\right)^3}{\left(\alpha B\left(\frac{3}{\alpha}, 1-\frac{2}{\alpha}\right) - \left(B\left(\frac{2}{\alpha}, 1-\frac{1}{\alpha}\right)\right)^2\right)^{3/2}} \end{aligned} \tag{B.108}$$

Tail Weight:
Has a heavy tail, based on a decreasing hazard rate function.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function for the three-parameter version is used. If the underlying distribution is the two-parameter version, it will be suggested by the parameter estimates. In most cases it is difficult to obtain convergence for the estimation algorithm as it is very sensitive to the choice of initial values.

B.18 Kappa Family (Three-parameter) Distribution - α, β, θ

Probability Density Function:

$$f_X(x) = \frac{\alpha\theta\left(\frac{x^{\alpha\theta-1}}{\beta^{\alpha\theta}}\right)}{\left(1+\left(\frac{x}{\beta}\right)^\theta\right)^{\alpha+1}} \text{ for } x \geq 0 \text{ with } \theta, \beta, \alpha > 0 \tag{B.109}$$

Moments:

$$\mathrm{E}(X^r) = \beta^r \frac{\Gamma\left(\frac{r}{\theta}+\alpha\right)\Gamma\left(1-\frac{r}{\theta}\right)}{\Gamma(\alpha)} \tag{B.110}$$

$$\mathrm{E}(X) = \beta \frac{\Gamma\left(\frac{1}{\theta}+\alpha\right)\Gamma\left(1-\frac{1}{\theta}\right)}{\Gamma(\alpha)} \tag{B.111}$$

$$\begin{aligned} \mathrm{var}(X) = {} & \left(\frac{\beta}{\Gamma(\alpha)}\right)^2 \Gamma(\alpha)\Gamma\left(\frac{2}{\theta}+\alpha\right)\Gamma\left(1-\frac{2}{\theta}\right) - \\ & \left(\frac{\beta}{\Gamma(\alpha)}\right)^2 \left(\Gamma\left(\frac{1}{\theta}+\alpha\right)\Gamma\left(1-\frac{1}{\theta}\right)\right)^2 \end{aligned} \tag{B.112}$$

Tail Weight:
Has a heavy tail, based on a decreasing hazard rate function.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. In most cases it is difficult to obtain convergence for the estimation algorithm as it is very sensitive to the choice of initial values.

B.19 Loggamma Distribution - a, λ

Probability Density Function:

$$f_X(x) = \frac{\lambda^a}{\Gamma(a)} (\ln(x))^{a-1} x^{-(\lambda+1)} \text{ for } x > 1 \text{ and } a > 0 \tag{B.113}$$

Moments:

$$\mathrm{E}\left(X^r\right) = \left(\frac{\lambda}{\lambda - r}\right)^a \tag{B.114}$$

$$\mathrm{E}(X) = \left(\frac{\lambda}{\lambda-1}\right)^a \tag{B.115}$$

$$\mathrm{var}(X) = \left(\frac{\lambda}{\lambda-2}\right)^a - \left(\frac{\lambda}{\lambda-1}\right)^{2a} \tag{B.116}$$

$$\mathrm{skewness}(X) = \frac{\left(\frac{\lambda}{\lambda-3}\right)^a - 3\left(\frac{\lambda}{\lambda-1}\right)^a\left(\frac{\lambda}{\lambda-2}\right)^a + 2\left(\frac{\lambda}{\lambda-1}\right)^{3a}}{\sqrt[3/2]{\left(\frac{\lambda}{\lambda-2}\right)^a - \left(\frac{\lambda}{\lambda-1}\right)^{2a}}} \tag{B.117}$$

Tail Weight:
There are parameter values for which the distribution does not have a heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Depending on the choice of initial values of λ and a, the algorithm performs well in terms of obtaining convergence and with reasonable accuracy.

B.20 Logistic Distribution - θ, ξ

Probability Density Function:

$$f_X(x) = \frac{1}{\theta}\frac{e^{\frac{x-\xi}{\theta}}}{\left(1+e^{\frac{x-\xi}{\theta}}\right)^2} \text{ for } x \in \mathbb{R} \text{ with } -\infty < \xi < \infty \text{ and } \theta > 0 \quad (B.118)$$

Moments:

$$\mathrm{E}(X) = \xi \quad (B.119)$$

$$\mathrm{var}(X) = \frac{\pi^2\theta^2}{3} \quad (B.120)$$

$$\mathrm{skewness}(X) = 0 \quad (B.121)$$

B.21 Log-logistic Distribution - α, β

Probability Density Function:

$$f_X(x) = \frac{\beta\left(\frac{x}{\alpha}\right)^\beta}{x\left(1+\left(\frac{x}{\alpha}\right)^\beta\right)^2} \text{ for } \text{ with } \alpha,\beta > 0 \quad (B.122)$$

Moments:

$$\mathrm{E}(X^r) = \alpha^r\Gamma\left(1+\frac{r}{\beta}\right)\Gamma\left(1-\frac{r}{\beta}\right) \quad (B.123)$$

$$\mathrm{E}(X) = \alpha\Gamma\left(1+\frac{1}{\beta}\right)\Gamma\left(1-\frac{1}{\beta}\right) \quad (B.124)$$

$$\mathrm{var}(X) = \alpha^2\left(\Gamma\left(1+\frac{2}{\beta}\right)\Gamma\left(1-\frac{2}{\beta}\right)-\left(\Gamma\left(1+\frac{1}{\beta}\right)\Gamma\left(1-\frac{1}{\beta}\right)\right)^2\right) \quad (B.125)$$

Tail Weight:
Heavy-tail based on a decreasing hazard rate function.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. In some instances it is difficult to obtain convergence of the estimation algorithm as it is sensitive to the choice of initial values of α and β.

B.22 Lognormal Distribution - μ, σ

Probability Density Function:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(\ln(x)-\mu)^2}{\sigma^2}\right)} \text{ for } x > 0 \text{ with } -\infty < \mu < \infty \text{ and } \sigma > 0. \tag{B.126}$$

Moments:

$$\mathrm{E}(X^r) = e^{\mu r + \frac{1}{2}r^2\sigma^2} \tag{B.127}$$

$$\mathrm{E}(X) = e^{\mu + \frac{1}{2}\sigma^2} \tag{B.128}$$

$$\mathrm{var}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \tag{B.129}$$

$$\mathrm{skewness}(X) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1} \tag{B.130}$$

Tail Weight:
Heavy-tailed.

Parameter Estimation:
Maximum likelihood:

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} \ln(x_j) \tag{B.131}$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\ln(x_j) - \frac{1}{n}\sum_{j=1}^{n}\ln(x_j)\right)^2 \tag{B.132}$$

B.23 Normal Distribution - μ, σ

Probability Density Function:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{B.133}$$

Moment Generating Function:

$$M_X(t) = e^{\mu + \frac{t^2\sigma^2}{2}} \tag{B.134}$$

Moments:

$$\mathrm{E}(X) = \mu \tag{B.135}$$
$$\mathrm{var}(X) = \sigma^2 \tag{B.136}$$
$$\mathrm{skewness}(X) = 0 \tag{B.137}$$
$$\mathrm{kurtosis}(X) = 3 \tag{B.138}$$

B.24 Pareto Type II (Lomax) Distribution - κ, θ

Probability Density Function:

$$f_X(x) = \frac{\kappa}{\theta}\left(1 + \frac{x}{\theta}\right)^{-(\kappa+1)} \quad \text{for } x > 0 \text{ with } \theta, \kappa > 0 \tag{B.139}$$

Moments:

$$\mathrm{E}(X^r) = \theta^r \frac{\Gamma(r+1)\Gamma(\kappa - r)}{\Gamma(\kappa)} \tag{B.140}$$

$$\mathrm{E}(X) = \frac{\theta}{\kappa - 1} \tag{B.141}$$
$$\mathrm{var}(X) = \frac{\theta^2\kappa}{(\kappa-1)^2(\kappa-2)} \tag{B.142}$$
$$\mathrm{skewness}(X) = \frac{2(\kappa+1)}{(\kappa-3)\sqrt{(\kappa-2)(\kappa-1)}} \tag{B.143}$$

Tail Weight:
Heavy-tailed, but not for values of $\kappa < 2$.

Parameter Estimation:
Method-of-moments:

$$\hat{\kappa} = 2\frac{s_x^2}{s_x^2 - \overline{x}^2} \text{ and} \tag{B.144}$$
$$\hat{\theta} = \overline{x}\frac{s_x^2 + \overline{x}^2}{s_x^2 - \overline{x}^2} \tag{B.145}$$

Maximum likelihood:
Numerical estimation is used to optimize the log-likelihood. The estimation algorithm has good convergencies properties if the method-of-moments estimates are used as initial values. The accuracy of estimates improve with increasing sample size.

B.25 Rayleigh Distribution - θ

Probability Density Function:

$$f_X(x) = \frac{2}{\theta^2} x e^{-\left(\frac{x}{\theta}\right)^2}, x > 0 \text{ with } \theta > 0 \tag{B.146}$$

Moments:

$$\mathrm{E}(X) = \frac{\theta\sqrt{\pi}}{2} \tag{B.147}$$

$$\mathrm{var}(X) = \frac{\theta^2}{4}(4-\pi) \tag{B.148}$$

$$\mathrm{skewness}(X) = \frac{2\pi^{3/2} - 6\sqrt{\pi}}{(4-\pi)^{3/2}} \tag{B.149}$$

$$\begin{aligned} \mathrm{kurtosis}(X) &= \mathrm{E}\left((X - \mathrm{E}(X))^4\right) \\ &= \frac{32 - 12\pi^2}{(4-\pi)^2} \end{aligned} \tag{B.150}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\theta} = \frac{2\overline{X}}{\sqrt{\pi}} \tag{B.151}$$

Maximum likelihood:

$$\hat{\theta} = 2\sqrt{\frac{\sum_{i=1}^{m} X_i^2}{n(4+\pi)}} \tag{B.152}$$

B.26 Singh-Maddala Distribution - a, b, q

Probability Density Function:

$$f_X(x) = \frac{qa}{x}\left(1 + \left(\frac{x}{b}\right)^a\right)^{-(q+1)} \left(\frac{x}{b}\right)^a \text{ for } x \geq 0 \text{ with } a, b, q > 0. \tag{B.153}$$

Moments:

$$\mathrm{E}(X^r) = \frac{b^r}{(q-1)!}\Gamma\left(\frac{r}{a}+1\right)\Gamma\left(1 - \frac{r}{a}\right) \tag{B.154}$$

$$\mathrm{E}(X) = \frac{b}{(q-1)!}\Gamma\left(\frac{1}{a}+1\right)\Gamma\left(1-\frac{1}{a}\right) \tag{B.155}$$

$$\mathrm{var}(X) = \frac{b^2}{(q-1)!}\left(\Gamma\left(\frac{2}{a}+1\right)\Gamma\left(q-\frac{2}{a}\right) - \frac{\left(\Gamma\left(\frac{1}{a}+1\right)\Gamma\left(q-\frac{1}{a}\right)\right)^2}{(q-1)!}\right) \tag{B.156}$$

Tail Weight:
A deacreasing hazard rate function indicates a heavy tail.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. It is difficult to obtain convergence for the estimation algorithm.

B.27 Skew-normal Distribution - μ, σ, λ (Centered Paramaterization)

Probability Density Function:

$$f(x) = \frac{\sqrt{1-b^2\delta^2}}{\sigma}\Phi\left(\lambda\left((x-\mu)\frac{\sqrt{1-b^2\delta^2}}{\sigma}+b\delta\right)\right)$$
$$\phi\left((x-\mu)\frac{\sqrt{1-b^2\delta^2}}{\sigma}+b\delta\right) \text{ for } x\in\mathbb{R} \text{ with } \mu,\lambda\in\mathbb{R} \tag{B.157}$$

with $\Phi(\cdot)$ and $\phi(\cdot)$ the cumulative distribution and probability density functions of the standard Normal distribution where $b = \sqrt{\frac{2}{\pi}}$ and $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Moments:

$$\mathrm{E}(X) = \mu \tag{B.158}$$
$$\mathrm{var}(X) = \sigma^2 \tag{B.159}$$
$$\gamma_1 = \frac{b\delta^3(2b^2-1)}{(1-b^2\delta^2)^{\frac{3}{2}}} \tag{B.160}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\mu} = \overline{x} \tag{B.161}$$

$$\hat{\sigma} = s_x, \text{ and} \tag{B.162}$$

$$\hat{\lambda} = \frac{\hat{\delta}}{\sqrt{1-\hat{\delta}^2}}, \tag{B.163}$$

where s_x is the sample standard deviation, and

$$\hat{\delta} = \sqrt{\frac{\hat{\gamma_1}^{\frac{2}{3}}}{b^2\hat{\gamma_1}^{\frac{2}{3}} + b^{\frac{2}{3}}(2b^2-1)^{\frac{2}{3}}}}$$

with

$$\hat{\gamma_1} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})}{s_x^3} \tag{B.164}$$

Maximum likelihood:
Numerical optimization of the log-likelihood function is used.

B.28 Snedecor's F Distribution - ν_1, ν_2

Valid for values of $x > 0$ and $\nu_1, \nu_2 > 0$.

Moments:

$$\mathrm{E}(X^r) = \left(\frac{\nu_2}{\nu_1}\right)^r \frac{\Gamma\left(\frac{\nu_1}{2}+r\right)\Gamma\left(\frac{\nu_2}{2}-r\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \tag{B.165}$$

$$\mathrm{E}(X) = \frac{\nu_2}{\nu_2-2} \tag{B.166}$$

$$\mathrm{var}(X) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \tag{B.167}$$

$$\mathrm{skewness}(X) = \frac{\frac{\nu_2^3\nu_1(\nu_1+4)(\nu_1+2)}{\nu_1^3(\nu_2-2)(\nu_2-4)(\nu_2-6)} - 3\frac{\nu_2^3(\nu_1+2)}{\nu_1(\nu-2)^2(\nu_2-4)} + 2\frac{\nu_2^3}{(\nu_2-2)^3}}{\left(\frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}\right)^{3/2}} \tag{B.168}$$

Tail Weight:
A deacreasing hazard rate function indicates a heavy tail.

Parameter Estimation:
Method-of-moments:

$$\hat{\nu_1} = \frac{2\overline{x}^2}{(s_x^2 + \overline{x}^2)(2-\overline{x}) - \overline{x}^2} \tag{B.169}$$

and

$$\hat{\nu_2} = \frac{2\overline{x}}{\overline{x} - 1} \tag{B.170}$$

Maximum likelihood:
Numerical optimization of the log-likelihood function is used. Convergence of the estimation algorithm is reached reasonably easily when using the method-of-moments estimates as initial values. These estimates are, however not always very accurate.

B.29 Two-parameter Exponential Distribution - η, θ

Probability Density Function:

$$f_X(x) = \frac{1}{\theta} e^{\left(-\frac{x-\eta}{\theta}\right)} \text{ for } x > \eta \text{ with } -\infty < \eta < \infty \tag{B.171}$$

Moment Generating Function:

$$M_X(t) = \frac{e^{t\eta}}{1 - t\theta} \tag{B.172}$$

Moments:

$$\mathrm{E}(X) = \mathrm{E}((Y + \eta)) = \theta + \eta \tag{B.173}$$

$$\mathrm{var}(X) = \theta^2 \tag{B.174}$$

$$\mathrm{skewness}(X) = 2 \tag{B.175}$$

$$\mathrm{kurtosis}(X) = 9 \tag{B.176}$$

Tail Weight:
No heavy tail.

Parameter Estimation:
Maximum likelihood:

$$\hat{\eta} = X_{(1)}, \text{ and} \tag{B.177}$$

$$\hat{\theta} = \overline{X} - X_{(1)} \tag{B.178}$$

B.30 Weibull Distribution - β, θ

Probability Density Function:

$$f_X(x) = \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\theta}\right)^\beta} \text{ for } x > 0 \text{ with } \beta, \theta > 0 \tag{B.179}$$

Moments:

$$\mathrm{E}(X^r) = \theta^r \Gamma\left(\frac{r}{\beta} + 1\right) \text{ with } r > -\beta \tag{B.180}$$

$$\mathrm{E}(X) = \theta\Gamma\left(1 + \frac{1}{\beta}\right) \tag{B.181}$$

$$\mathrm{var}(X) = \theta^2 \left(\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2\right) \tag{B.182}$$

$$\mathrm{skewness}(X) = \frac{\Gamma\left(1+\frac{3}{\beta}\right) - 3\Gamma\left(1+\frac{1}{\beta}\right)\Gamma\left(1+\frac{2}{\beta}\right) + 2\Gamma\left(1+\frac{1}{\beta}\right)^3}{\sqrt[3/2]{\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma\left(1+\frac{1}{\beta}\right)^2}} \tag{B.183}$$

Tail Weight:
Heavy tail if $\beta \in (0, 1)$.

Parameter Estimation:
Maximum likelihood:
Numerical optimization of the log-likelihood function is used. The algorithm generally converges, given that suitable initial values for the parameters are chosen.

Bibliography

[1] A. Abtahi, J. Behboodian, and M. Sharafi. A general class of univariate skew distributions considering Stein's lemma and infinite divisibility. *Metrika*, 75(2):193–206, 2012.

[2] S. Adeyemi and M. Ojo. A genralization of the Gumbel distribution. *Kragujevac Journal of Mathematics*, 25:19–29, 2003.

[3] D. Akdemir and A. K. Gupta. A matrix variate skew distribution. *European Journal of Pure and Applied Mathematics*, 3(2):128–140, 2010.

[4] R. Ambagaspitiya. Ultimate ruin probability in the Sparre-Andersen model with dependent claim sizes and claim occurrence times. *Insurance: Mathematics and Economics*, 44(3):464–472, 2009.

[5] R. Arellano-Valle, H. Gómez, and F. Quintana. Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128(2):427–443, 2005.

[6] E. Arjas. The claims reserving problem in non-life insurance: Some structural ideas. *Astin Bulletin*, 19(2):139–152, 1988.

[7] S. Asmussen, F. Avrin, and M. Usabel. Erlangian approximations for finite-horizon ruin probabilities. *Astin Bulletin*, 32(2):267–281, 2002.

[8] A. Azzalini. A class of distributions which includeds the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985.

[9] A. Azzalini and A. Capitanio. *The Skew-Normal and Related Families*. Cambridge University Press, 2014.

[10] A. Azzalini and M. Genton. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76(1):106–129, 2008.

[11] L. Bain and M. Engelhardt. *Introduction to Probability and Mathematical Statistics*. Wiley, 1999.

[12] T. Bali and P. Theodossiou. Risk measurement performance of alternative distribution functions. *Journal of Risk and Insurance*, 75(2):411–437, 2008.

[13] R. Barlow, A. Marshall, and F. Proschan. Properties of probability distributions with monotone hazard rates. *The Annals of Mathematical Statistics*, 34(2):375–389, 1963.

[14] N. Barrera, M. Galea, S. Torres, and M. Villalón. Class of skew-distributions: Theory and applications in biology. *Statistics*, 40(4):365–375, 2006.

[15] R. Bartle. *The Elements of Real Analysis*, chapter 33: Uniform Convergence and Infinite Integrals, page 283. John Wiley and Sons Inc., 1964.

[16] J. Beirlant, P. Vynckier, and J. Teugels. Tail index estimation, pareto quantile plots and regression diagnostics. *Journal of the American Statistical Association*, 91(436):1659–1667, 1996.

[17] J. Bernardo and A. Smith. *Bayesian Theory*. Wiley, 1993.

[18] A. Birnbaum. On the analysis of factorial experiments without replication. *Technometrics*, I(4):9999, 1959.

[19] J. Birnbaum and J. Saunders. A new family of life distributions. *Technometrics*, 33(1):51–60, 1991.

[20] P. J. Boland. *Statistical and probabilistic methods in actuarial science.* CRC Press, 2007.

[21] E. Bradlow, B. Hardie, and P. Fader. Bayesian inference for the negative binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics*, 11(1):189–201, 2002.

[22] M. Braun, P. Fader, E. Bradlow, and H. Kunreuther. Modeling the pseudodeductible in insurance claims decisions. *Management Science*, 52(8):1258–1272, 2006.

[23] M. Brown. Further monotonicity properties for specialized renewal processes. *The Annals of Probability*, 9(5):891–895, 1981.

[24] M. Bryson. Heavy-tailed distributions: Properties and tests. *Technometrics*, 16(1):61–68, 1974.

[25] K. Burnecki, G. Kukla, and R. Weron. Property insurance loss distributions. *Physica*, A(287):269–278, 2000.

[26] I. Burr. Cumulative frequency functions. *Annals of Mathematical Statistics*, 13:215–232, 1942.

[27] F. Chen, J. Zhu, and Z. Li. Upper bounds for the ruin probabilities of the entrance-based risk model. *Communications in Statistics - Theory and Methods*, 37(16):2634–2652, 2008.

[28] Y. Chen and C. Su. Finite time ruin probability with heavy-tailed insurance and financial risks. *Statistics and Probability Letters*, 76(16):1812–1820, 2006.

[29] C. Chiang. Stochastic processes. *Encyclopedia of Biostatistics*, 1998.

[30] S. Choi and R. Wette. Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics*, 11(4):683–690, 1969.

[31] S. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics, 2001.

[32] K. Cooray and M. Ananda. Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal*, 5:321–334, 2005.

[33] H. Cossette, P. Gaillardetz, E. Marceau, and J. Rioux. On two dependent individual risk models. *Insurance: Mathematics and Economics*, 30:153–166, 2002.

[34] C. Dagum. A new model of personal income distribution: specification and estimation. *Economie Appliquée*, 30:413–437, 1977.

[35] C. Dagum. The generation and distribution of income, the Lorenz curve and the Gini ratio. *Economie Appliquée*, 33:327–367, 1980.

[36] C. Daniel. Use of half-normal plots in interpreting two-level experiments. *Technometrics*, 1(4):311–341, 1959.

[37] D. Davis. An analysis of some failure data. *Journal of the American Statistical Association*, 47(258):113–150, 1952.

[38] R. Davis and T. Mikosch. Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Processes and their Applications*, 118(4):560–584, 2008.

[39] M. Denuit, X. Maréchal, S. Pitrebois, and J. Walhin. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus System.* John Wiley and Sons, Ltd, Chichester, UK, 2007.

[40] M. Diasparra and R. Romera. Inequalities for the ruin probability in a controlled discrete-time risk process. *European Journal of Operational Research*, 204:496–504, 2010.

[41] O. Dickerson, S. Katti, and A. Hofflander. Loss distributions in non-life insurance. *The Journal of Insurance*, 28(3):45–54, 1961.

[42] D. Dickson and H. Waters. The distribution of the time to ruin in the classical risk model. *Astin Bulletin*, 32(2):299–313, 2002.

[43] D. C. M. Dickson. *Insurance risk and ruin.* International Series on Actuarial Science. Cambridge: Cambridge University Press., 2005.

[44] F. Domma. Asymptotic distribution of the maximum likelihood estimators of the parameters of the right-truncated dagum distribution. *Communication in Statistics - Simulation and Computation*, 36(6):1187–1199, 2007.

[45] S. Dubey. Compound gamma, beta and F distributions. *Metrika*, 16(1):28–31, 1970.

[46] R. Elandt. The folded normal distribution: Two methods of estimating parameters from moments. *Technometrics*, 3(4):551–562, 1961.

[47] P. Embrechts, R. Frey, and H. Furrer. *Stochastic Processes in Insurance and Finance.* To find, 1999.

[48] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, 1996.

[49] P. Embrechts and H. Schmidli. Ruin estimation for a general insurance risk model. *Advanced Applied Probability*, 26:404–422, 1994.

[50] P. England and R. Verral. Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(III):443–544, 2002.

[51] Z. Fabián. New measures of central tendency and variability of continuous distributions. *Communications in Statistics - Theory and Methods*, 37(2):159–174, 2008.

[52] T. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

[53] R. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, Series A(222):309–368, 1922.

[54] R. Fisher and H. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.

[55] J. Folks and R. Chhikara. The inverse Gaussian distribution and its statistical applications - a review. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):263–289, 1978.

[56] S. Foss, D. Korshunov, and S. Zachary. *An Introduction to Heavy-Tailed and Subexponential Distributions.* Springer, 2011.

[57] P. Giudici and S. Figini. *Applied Data Mining for Business and Industry, Second Edition.* John Wiley and Sons Ltd, 2009.

[58] G. Goerg. Lambert W random variables - a new family of generalized skewed distributions with applications to risk estimation. *The Annals of Applied Statistics*, 5(3):2197–2230, 2011.

[59] J. Greenwood, J. Landwehr, N. Matalas, and J. Wallis. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse forms. *Water Resources Research*, 15:1049–1054, 1979.

[60] D. Guégan and J. Zhang. Change analysis of a dynamic copula for measuring dependence in multivariate financial data. *Quantitative Finance*, 10(4):421–430, 2010.

[61] R. Gupta and D. Kundu. Generalized logistic distributions. *Journal of Applied Statistical Science*, 18(1):51–66, 2010.

[62] S. Gupta, A. Qureishi, and B. Shah. Best linear unbiased estimators of the parameters of the logistic distribution using order statistics. *Technometrics*, 9(1):43–56, 1967.

[63] E. Hashorva. On the asymptotic distribution of certain bivariate reinsurance treaties. *Insurance Mathematics and Economics*, 40(2):200–208, 2007.

[64] E. Hashorva and J. Hüsler. Estimation of tails and related quantities using the number of near-extremes. *Communications in Statistics - Theory and Methods*, 34(2):337–349, 2005.

[65] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning.* Wiley, 1999.

[66] C. Hewitt and B. Lefkowitz. Methods for fitting distributions to insurance loss data. *Proceedings of the Casualty Actuarial Society*, 66(125):139–160, 1979.

[67] L. Horvath and E. Willekens. Estimates for the probability of ruin starting with a large initial reserve. *Insurance: Mathematics and Economics*, 5:285–293, 1986.

[68] J. Hosking, J. Wallis, and E. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985.

[69] J. Jawitz. Moments of truncated continuous univariate distributions. *Advances in Water Resources*, 27:269–281, 2004.

[70] A. Jenkinson. The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171, 1969.

[71] L. Jingzhi, W. Kaiyong, and W. Yuebao. Finite-time ruin probability with NQD dominated varying-tailed claims and NLOD inter-arrival times. *Journal of Systems Science and Complexity*, 22(3):407–414, 2009.

[72] C. Jordan. *Calculus of finite differences*. Chelsea Publishing Company, New York, 1960.

[73] B. Kim and H. Kim. Moments of claims in a Markovian environment. *Insurance: Mathematics and Economics*, 40(3):485497, 2007.

[74] C. Kleiber. Dagum vs. Sing-Maddala income distributions. *Economics Letters*, 53:265–268, 1996.

[75] S. Klugman, H. Panjer, and G. Willmot. *Loss Models, Third edition*, pages 34–40,333. John Wiley and Sons Inc., 2008.

[76] S. Kotz and S. Nadarajah. *Extreme Value Distributions: Theory and Applications*. Imperial College Press, 2000.

[77] D. Kundu, N. Kannan, and N. Balakrishnan. On the hazard function of Birnbaum-Saunders distribution and associated inference. *Computational Statistics and Data Analysis*, 52:2692–2702, 2008.

[78] A. Lemonte, F. Cribari-Netor, and K. Vasconcellos. Improved statistical inference for the two-parameter Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis*, 51:4656–4681, 2007.

[79] F. Leone, L. Nelson, and R. Nottingham. The folded normal distribution. *Technometrics*, 3(4):543–550, 1961.

[80] Q. Li and J. Racine. *Nonparametric Econometrics Theory and Practice*. Princeton University Press, 2007.

[81] F. Lotti and E. Santarelli. Industry dynamics and the distribution of firm sizes: A nonparametric approach. *Southern Economic Journal*, 70(3):443–466, 2004.

[82] D. Luenberger. *Investment Science*, chapter 11: Models of Asset Dynamics, pages 308–309. Oxford University Press Inc., 1998.

[83] T. Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, 23(2):213–225, 1993.

[84] T. Mack. Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics and Economics*, 15:133–138, 1994.

[85] W. Mielke. Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology*, 12:275–280, 1973.

[86] W. Mielke and E. Johnson. Three parameter kappa distribution: Maximum likelihood estimates and likelihood ratio tests. *Monthly Weather Review*, 101:701–707, 1973.

[87] D. Montgomery. *Statistical Quality Control: A Modern Introduction, 6th edition, International Student Version.* John Wiley and Sons, Inc, 2009.

[88] A. Müller. Ordering of risks: A comparative study via stop-loss transforms. *Insurance: Mathematics and Economics*, 17(1):215–222, 1996.

[89] S. Nadaraja and S. Kotz. Skew models I. *Acta Applicandae Mathematicae*, 98(1):1–28, 2007.

[90] S. Nadaraja and S. Kotz. Skew models II. *Acta Applicandae Mathematicae*, 98(1):29–46, 2007.

[91] S. Nadarajah and S. Bakar. New composite models for the danish fire insurance data. *Scandinavian Actuarial Journal*, 1:1–8, 2012.

[92] S. Nadarajah and S. Kotz. Skew distributions generated from different families. *Acta Applicandae Mathematicae*, 91(1):1–37, 2006.

[93] O. Önalan. Financial risk management with normal inverse Gaussian distributions. *International Research Journal of Finance and Economics*, 38:104–115, 2010.

[94] A. Otten and M. Van Montfort. Maximum likelihood estimation of the general extreme-value distribution parameters. *Journal of Hydrology*, 47:187–192, 1980.

[95] V. Pacakova and B. Linda. Simulations of the extreme losses in non-life insurance. *E and M Ekonomie A Management*, 12(4):97–103, 2009.

[96] E. Pairman. *Tables of the Digamma and Trigamma Functions.* Cambridge University Press, 1954.

[97] P. Palmitesta and C. Provasi. Aggregation of dependent risks using the Koehler-Symanowski copula function. *Computational Economics*, 25(1–2):189–205, 2005.

[98] H. Park and H. Sohn. Parameter estimation of the generalized extreme value distribution for structural health monitoring. *Probabilistic Engineering Mechanics*, 21:366–376, 2006.

[99] M. Parmenter. *Theory of Interest and Life Contingencies with Pension Applications: A Problem-Solving Approach, Third Edition*. ACTEX Publications, Winsted, Connecticut, 1999.

[100] A. Pewsey. Problems of inference for Azzalini's skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870, 2000.

[101] V. Preda and R. Ciumara. On composite models Weibull-Pareto and lognormal-Pareto. A comparative study. *Romanian Journal of Economic Forecasting*, 2:32–46, 2006.

[102] A. Renshaw and R. Verral. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4(IV):903–923, 1998.

[103] J. Rieck and J. Nedelman. A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33(1):51–60, 1991.

[104] T. Rolski. Point processes. *Encyclopedia of Actuarial Science*, 2006.

[105] T. Rolski, H. Schmidli, V. Schmidt, and T. Teugels. *Stochastic Processes for Insurance and Finance*. Wiley, 1999.

[106] M. Schäl. Stochastic optimization for the ruin probability. *PAMM*, 3(1):17–19, 2003.

[107] H. Schmidli. On minimizing the ruin probability by investment and reinsurance. *The Annals of Applied Probability*, 12(3):890–907, 2002.

[108] K. Schmidt and M. Zocher. The bornheutter-ferguson principle. *Variance*, 2:85–110, 2008.

[109] J. Shuster. On the inverse Gaussian distribution function. *Journal of the American Statistical Association*, 63(324):1514–1516, 1968.

[110] D. Sondermann. Reinsurance in arbitrage-free markets. *Insurance: Mathematics and Economics*, 10(3):191–202, 1991.

[111] J. Stewart. *Calculus, Early Transcendentals, Fifth Edition*. Thomson Learning, 2003.

[112] A. Steyn, C. Smit, S. Du Toit, and C. Strasheim. *Modern Statistics for Practice*, chapter 5: Measures of Spread, Symmetry and kurtosis, page 136. J.L. van Schaik Publishers, 1996.

[113] P. Tadikamalla. A look at the Burr and related distributions. *International Statistical Review*, 48(3):337–344, 1980.

[114] M. Taksar. Optimal risk and dividend distribution control models for an insurance company. *Mathematical Methods of Operations Research*, 51:1–42, 2000.

[115] Q. Tang. Asymptotic ruin probabilities in finite horizon with subexponential losses and associated discount factors. *Probability in the Engineering and Informational Sciences*, 20(1):103–113, 2006.

[116] Q. Tang, G. Wang, and K. Yuen. Uniform tail asymptotics for the stochastic present value of aggregate claims in the renewal risk model. *Insurance Mathematics and Economics*, 46(2):362–370, 2010.

[117] S. Teodorescu. On the truncated composite lognormal-Pareto model. *Mathematical Reports*, 12(62):71–84, 2010.

[118] J. Teugels. The class of subexponential distributions. *The Annals of Probability*, 3(6):1000–1011, 1975.

[119] S. Wang. A framework for pricing financial and insurance risks. *Astin Bulletin*, 32(2):213–234, 2002.

[120] X. Wei and Y. Hu. Ruin probabilities for discrete time risk models with stochastic rates of interest. *Statistics and probabilities letters*, 78(6):707–715, 2008.