Comparison Studies of Estimators for the Generalized Gamma Distribution and New Findings

Jakka Chang
Student number: 5174376

August 23, 2022

Instructor:	Piao Chen
Thesis committee:	Piao Chen, Tina Nane
Faculty:	EEMCS, Delft

Delft University of Technology

Contents

1 Introduction

In industrial processes a lot of effort is put into reliability analysis of products. It is important to know how long one can expect a product or piece of equipment to last before failing. For example, this knowledge is important in the aviation industry to forecast when reparations should be scheduled for airplanes. In commercial industries this data can be used to price insurances to turn a profit for the seller. The main distributions for modelling these are the Weibull and the exponential distribution. However, it is common to run into reliability data that cannot be described accurately by either of these models. This prompted the development of a more general model for data analysis.

In 1962 the generalized gamma distribution was introduced to address this problem (Stacy, 1962). The generalized gamma distribution is, as its name suggests, a generalization of the gamma distribution and contains three parameters. The generalized gamma reduces to, the exponential, the Weibull and the gamma distribution for certain values of the parameters. This gives it great flexibility in data analysis as it can describe a wider range of data sets. In other cases, the generalized gamma can aid in discerning which model to use in an analysis, for example the choice of using the Weibull over the exponential distribution.

For all the advantages that the GGD brings with it, parameter estimation has remained a key barrier to the widespread adoption of the distribution. Often times, even for sample sizes of 200 to 300, maximum likelihood estimates fail to converge, meaning that there is still a large error between the estimated and the real value of the parameters. In this paper I will highlight the difficulties in parameter estimation for the GGD and compare multiple estimation methods that have been proposed.

In Chapter 2 I will give an introduction to the generalized gamma distribution by explaining its parameters and properties and concluding with examples of applications of the GGD to data analysis. In Chapter 3 I outline different estimation methods and give theoretical derivations before comparing the results of the methods in Chapter 4. I conclude the paper in Chapter 5 with new findings for parameter estimation of the generalized gamma distribution.

2 The Generalized Gamma Distribution

In this chapter I will give a brief introduction to the generalized gamma distribution. I will give an overview of what types of shapes the GGD can take on and some of the properties the distribution has. This will give an intuitive sense of what the distribution is. As mentioned in the introduction, the GGD is a very flexible model with lots of applications in statistics. I will highlight some of the aspects that give rise to the GGD's applicability.

The generalized gamma distribution (GGD) was introduced in (Stacy, 1962) as a unification of distributions such as the Weibull, gamma and lognormal distribution. Its density is given by

$$f(x|\alpha,\beta,k) = \frac{\beta x^{\beta k-1}}{\alpha^{\beta k}\Gamma(k)}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]$$

this is a reparametrization of the original GGD described in (Stacy, 1962),
with $\alpha = a, \beta = p, k = \frac{d}{p}$. The GGD encompasses distributions such as the Weibull distribution (for $k = 1$), the gamma distribution ($\beta = 1$), lognormal distribution ($k \to \infty$). A list of sub-distributions is given in Subsection 2.2.

The flexibility of the GGD lends itself to being used for data analysis in various fields. The GGD has been applied to: health cost, where the GGD was used in regression modelling (Manning, Basu, Mullahy, & Manning, 2002). Civil engineering, where it was used in a flood frequency model (Pham & Almhana, 1995). Economics, where it was used to model income distributions (Kleiber & Kotz, 2003).

Despite its flexibility, parameter estimation is a key issue in applying the GGD to data analysis. Estimators tend to peform poorly under low sample size. Additionally, parameter estimation through maximum likelihood result results in large mean square errors. Various estimators have been proposed to circumvent these problems. The most well-known of these being estimation through logarithmic cumulants (Gao, Ouyang, Luo, Liang, & Zhou, 2016) and (Zhang et al., 2020). As well as an estimator through bayesian inference (Ramos, Achcar, Moala, Ramos, & Louzada, 2017). We will also be analysing an improvement of the maximum likelihood estimation described in (Noufaily & Jones, 2013). In this paper we will provide a comparison of these estimators and analyse the strengths of one estimator over the other.

2.1 Parameters of the GGD

The GGD contains three parameters. One is a scale parameter, controlling for the concentration of the mass of the distribution. Two of the parameters are shape parameters, meaning simply that the general shape of the distribution is determined by these two values.

The parameter α is a scale parameter. As α increases, the distribution becomes more "spread out" while the distribution becomes more and more concentrated around a point as α decreases to zero. In Figure 1 we see the plot of the GGD with values of α between 1 and 4.

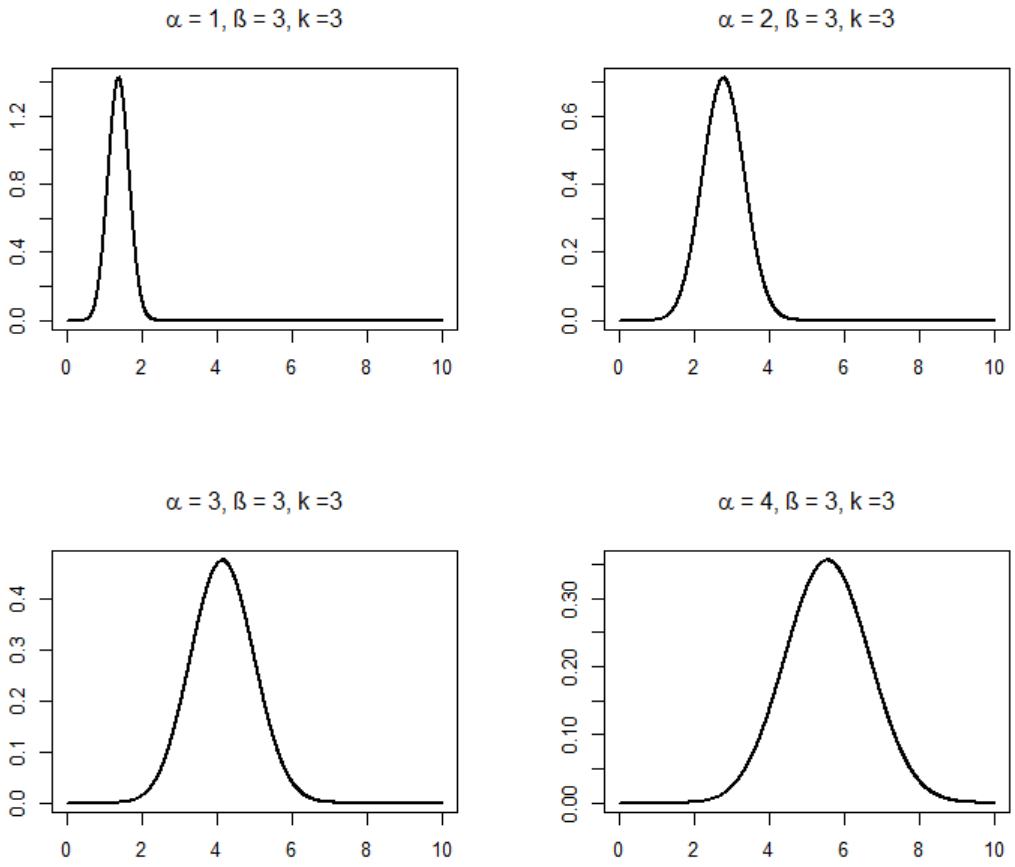

Figure 1: The GGD for $\beta = 3$, $k = 3$, $\alpha = 1, 2, 3, 4$

β and k are shape parameters. The graph of the GGD can vary wildly with different values for these two parameters. Nevertheless I try to paint a picture of how each parameter affects the GGD. In the following two figures, α is chosen equal to 10 as the graph will be very slim for small values, which gets in the way of showcasing the influence of the other parameters. In Figure 2 we see plots for varying values of k. We see that small values of k can turn the GGD into a decreasing function. For higher values of k the distribution tends to become more and more symmetrical and the peak moves asymptotically to the right. In fact, for $k \to \infty$, the GGD reduces to a lognormal distribution.

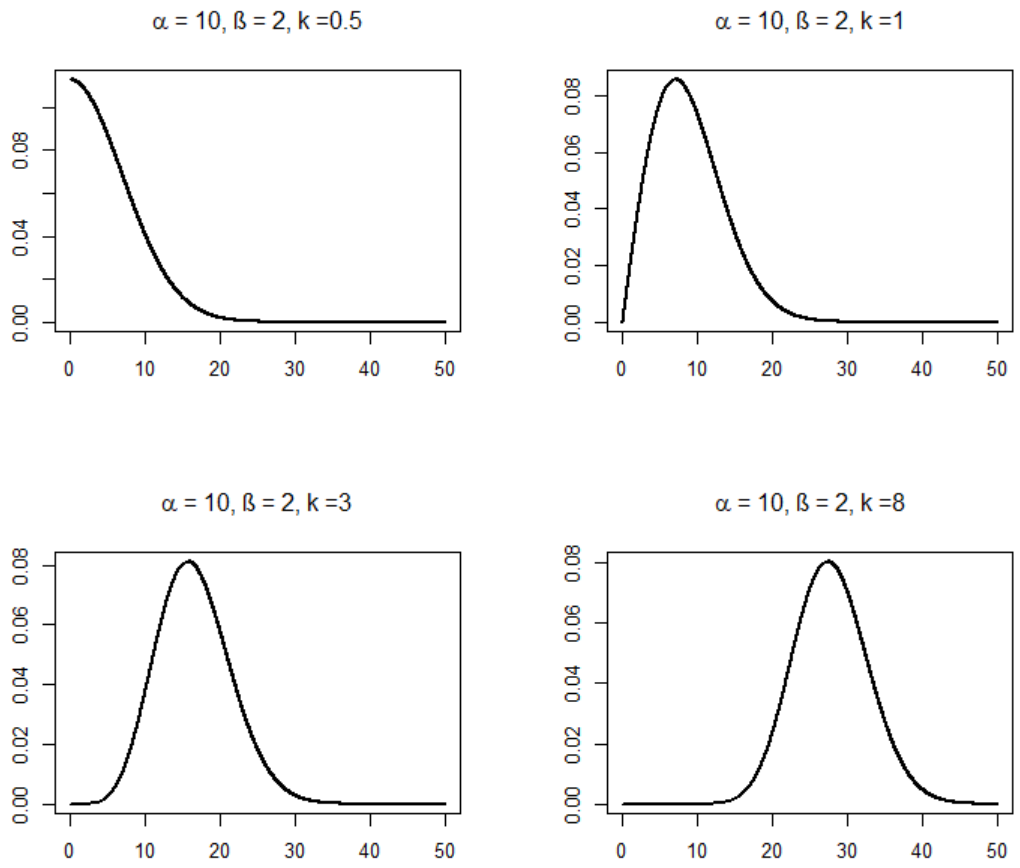

Figure 2: The GGD for $\alpha = 10$, $\beta = 1$, $k = 0.5, 1, 3, 8$

In Figure 3 we see plots for β. We see that for low values of the parameter, the graph turns into a decreasing function as well. As β increases, we see that the mass of the distribution gets more and more concentrated.

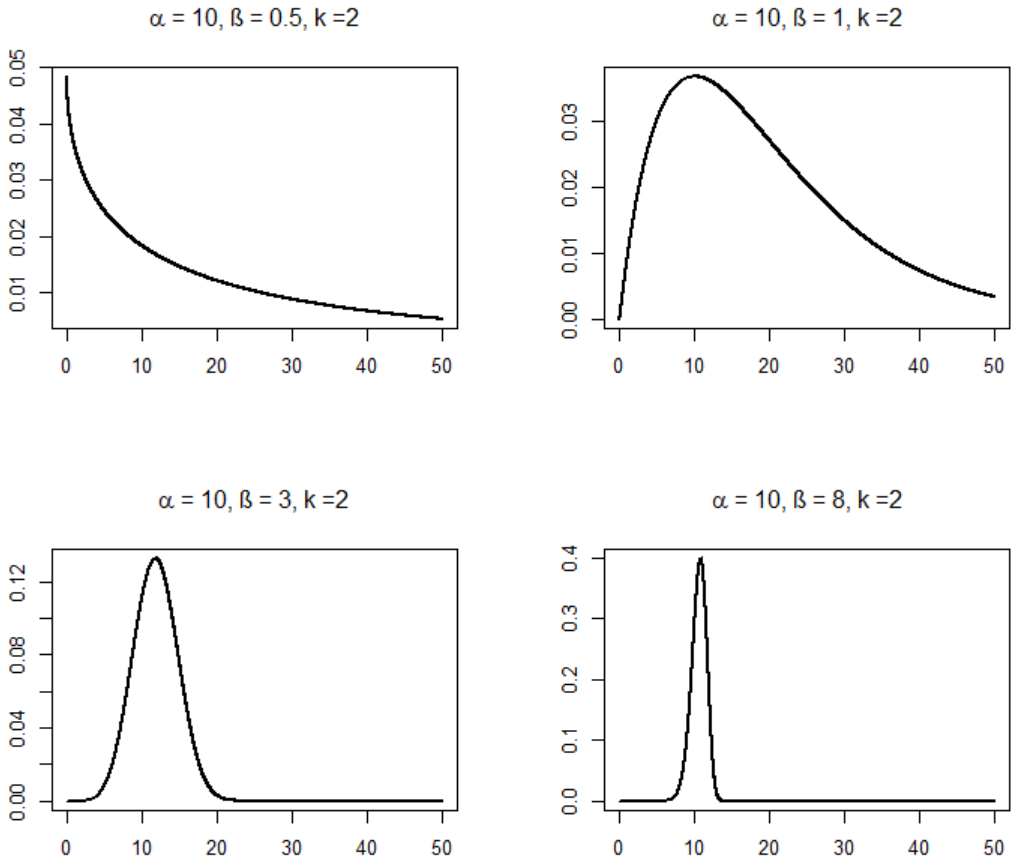

Figure 3: The GGD for $\alpha = 3$, $k = 2$, $\beta = 0.5, 1, 3, 8$

2.2 Properties of the GGD

As mentioned at the start of the chapter the GGD encompasses many other distributions used in statistics. In Table 1 we see a list of distributions that can be obtained for special parameters of the GGD.

Distribution	PDF	Parameter values
Weibull	$\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)$	k = 1
Gamma	$\frac{1}{\alpha^k\Gamma(k)}x^{k-1}\exp\left(-\frac{x}{\alpha}\right)$	$\beta = 1$
Chi-squared	$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}x^{n/2-1}\exp\left(-\frac{x}{2}\right)$	$\alpha = 2, \beta = 1, k = \frac{n}{2}$
Exponential	$\frac{1}{\alpha}\exp\left(-\frac{x}{\beta}\right)$	$\beta = 1, k = 1$
Half-normal	$\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$	$\alpha = \sqrt{2}, \beta = 2, k = \frac{1}{2}$
Rayleigh	$\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$	$\alpha = \sigma\sqrt{2}, \beta = 2, k = 1$

Table 1: Distributions encompassed by the GGD

Let X have pdf $f(x|\alpha,\beta,k)$, $X \sim f(x|\alpha,\beta,k)$, where f denotes the generalized gamma distribution. Then we have the following:

Property 1, a multiple of a rv from the GGD is also a rv from the GGD
If $X \sim f(x|\alpha,\beta,k)$, $cX \sim f(x|c\alpha,\beta,k)$, where $c > 0$

Property 2, powers of a rv from the GGD is also a rv from the GGD
If $X \sim f(x|\alpha,\beta,k)$, $X^m \sim f(x|a^m,\beta/m,k)$

Property 3, moments of the GGD
The rth moment of X is given by $E(X^r) = \begin{cases} a^r\Gamma\{(k\beta+r)/\beta\}/\Gamma(k), & \text{if } \frac{r}{p} > -k \\ \infty, & \text{otherwise} \end{cases}$.

2.3 Applications of the GGD as a lifetime distribution

A good way to show the applicability of the GGD is through the applications of its sub-distributions. Here I will explain the uses of three of the GGD's sub-distributions; the Weibull, the lognormal and the gamma distribution.

The Weibull distribution
A random variable X that has a Weibull distribution has PDF

$$f(x|\lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[\left(\frac{x}{\lambda}\right)^{k}\right]$$

The Weibull is used in lifetime analysis. Lifetime analysis is used in studying how long a piece of equipment can be expected to function without failing. This can be useful information when one wants to price their insurance policy for example. If an insurance company knows how long a product will last without failing or breaking (on average). And how much reparations will cost a company when insurance is exercised, then the price of the insurance can be chosen such that the company will make a profit in the long run. In vehicles such as airplanes this type of analysis can be used to schedule maintenance before the vehicle breaks down.

Once the parameters for a Weibull distribution have been fitted, one can obtain information such as:

- Reliability after a period of time: The probability that a product is still functional after, for example, 12 months of use.
- Average life span: How long a product will function on average before becoming unusable.
- Failure rate: How many units of a product can be expected to break after a certain period of time.

The gamma distribution A random variable X that has a gamma distribution has PDF

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)}$$

The gamma distribution can be used in reliability analysis. Reliability analysis is used in cases where it is possible for a product to fail multiple times, the gamma distribution can then describe the expected time in between two failures.

The lognormal distribution
A random variable X that has a lognormal distribution has PDF

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right]$$

Aside from being used in lifetime analysis like the Weibull distribution. The lognormal is often times used in engineering analysis where negative values for random variables are physically impossible. This makes the lognormal a preferred distribution over the normal distribution in some cases. This property is used in analyzing stock returns as well. Stock returns are often said to follow a lognormal distribution.

The GGD can contribute to research where one would like to use these distributions but is unsure of which exact distribution can be best used for a given data set. For example, let's say one wants to examine whether a given data set follows a Weibull distribution. As mentioned in Subsection 2.2, The GGD reduces to the Weibull for $k = 1$. Thus one can test hypothesis $k = 1$ (the data is described by a Weibull distribution) versus $k \neq 1$ (the data follows a different distribution). In the case that $k = 1$ seems likely, a researcher can then choose to model the data with a Weibull.

It can also happen that a given data set does not quite follow any of the previously listed sub-distributions of the GGD. In this case the GGD might be preferable for modelling the data, as it has similar properties to its sub-distributions but is more flexible.

3 Parameter estimation methods

This section will highlight why maximum likelihood estimation for the parameters of the GGD can be tricky. Section 3.2 until 3.5 will go into detail on 5 parameter estimation methods for the generalized gamma. Some details on my implementation of the methods will be given and each estimation method will be used to estimate samples from the GGD for various values of the parameters. The results will be given in tables with some comments provided alongside them. A side-to-side comparison of all the methods will be given in Section 4.

3.1 Problems with Maximum Likelihood Estimation

In this section we will further explore the problems one encounters with the MLEs for the GGD. Let $x_1, ..., x_n$ denote random samples from the GGD with parameters α, β and k. We then obtain the following system of equations.

$$-n\hat{k} + \sum_{i=1}^{n} \left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}} = 0 \tag{1}$$

$$\frac{n}{\hat{\beta}} + \hat{k}\sum_{i=1}^{n} \log\left(\frac{t_i}{\hat{\alpha}}\right) - \sum_{i=1}^{n} \left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}} \log\left(\frac{t_i}{\hat{\alpha}}\right) = 0 \tag{2}$$

$$-n\psi(\hat{k}) + \hat{\beta}\sum_{i=1}^{n} log\left(\frac{t_i}{\hat{\alpha}}\right) = 0 \tag{3}$$

Where $\psi(k) = \frac{d}{dk}\log(\Gamma(k))$. Also known as the digamma function. This system of equations are coupled and thus need to be solved simultaneously. Hager & Bain showed that this system can be reduced to a single non-linear equation in the unknown quantity $\hat{b}$ (Hager & Bain, 1970). We do this by rewriting Equation (1) to

$$\hat{a} = \left[\sum_{i=1}^{n} t_i^{\hat{b}}/(n\hat{k})\right]^{1/\hat{b}} \tag{4}$$

Substituting this expression for $\hat{a}$ into Equation (2), we obtain

$$-\hat{k} = \left\{\hat{b}\left[(\sum \log t_i)/n - (\sum t_i^{\hat{b}})/\sum t_i^{\hat{b}}\right]\right\}^{-1} \tag{5}$$

Equation (3) can then be rewritten to

$$-\psi(\hat{k}) + \hat{b}(\sum \log t_i)/n - \log(\sum t_i^{\hat{b}}) + \log(n\hat{k}) = 0 \tag{6}$$

with $\hat{k}$ as given above. With equation (6), we can find maximum likelihood estimates by finding the root $\hat{\beta}$ and then substituting into Equation (5). Finally, we obtain our estimate $\hat{\alpha}$ from Equation (4). Finding the root for Equation (6) can be done by methods such as Newton-Raphson and the bisection method. Unfortunately, Hager & Bain state that the Newton-Raphson method does not work well in solving this final expression. They add that the maximum likelihood estimators are not very well-behaved unless sample sizes are large, meaning maximum likelihood estimates tend to have a large margin of error for

lower sample sizes. Indeed, even for sample sizes of size three hundred, it can happen that the m.l.e.'s fail to converge to the true parameter values. The problem of finding estimates with maximum likelihood is further analysed in (Lawless, 1980). We reparametrise the GGD to a different form. This form will be equivalent to the parametrisation given in the introduction and will highlight why we run into trouble when estimating the parameters. Let $t_1, ..., t_n$ be samples from the generalized gamma distribution. Let $y_i = \log(t_i)$ for $i = 1, ..., n$. Then the samples y_i can be shown to have distribution function

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)} \exp\left\{\sqrt{k}\omega - ke^{\omega/\sqrt{k}}\right\}$$

where $\omega = \sigma^{-1}(y - \mu)$, $\mu = log(\alpha) + \beta^{-1}log(k)$, $\sigma = \frac{1}{\sqrt{k\beta^2}}$ and $k = \frac{\alpha}{\beta}$.

We will refer to this distribution as the log ggd. This parametrisation is equivalent to the GGD in the sense that the same distributions can still be obtained from the GGD as from this distribution, with the caveat that this distribution will produce the logarithm of the distributions that the GGD produces. For example, for values $k = 1$, the GGD reduces to the Weibull distribution. In the new distribution in y, the $k = 1$ produces the log of the Weibull distribution. Similarly $k = \infty$ produces the log-normal distribution in the GGD, and produces the normal (the log of the log-normal) distribution in the new distribution. Finally, where $\beta = 1$ produced the gamma distribution in the GGD, $\frac{\sigma}{\sqrt{k}} = 1$ now produces the log-gamma distribution.

In this distribution, it is clear why parameter estimation is difficult. As can be seen in the figure, different values of k can produce similar probability distribution functions. Going back to the (α, β, k) parametrisation, for similar values for σ and μ, but different values of k, will have very different values of α and β.

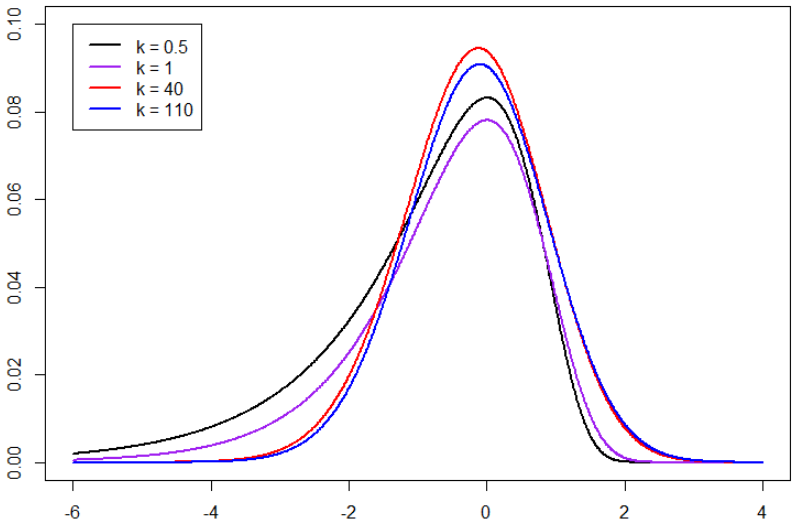

Figure 4: The GGD for various values of k

3.2 Lawless method

Lawless solves the issue with ML estimation by assuming k to be fixed in the log ggd and performing maximum likelihood estimation on the other two parameters, μ and σ. Lawless states two reasons for this approach:

1. At the time, it was hard to estimate all three parameters at the same time, due to the reasons stated in Section 3.1.
2. Often times, it is actually the case that the GGD is used with a fixed value for k. For example, take the Weibull distribution which is often used to analyse the lifespan of products. In this case k is chosen to be equal to 1.

After estimating values for μ and σ for a given k, we vary over k in order to maximize the likelihood over all three parameters. First we derive the maximum likelihood estimators for μ and σ. Again, let $T_1, ..., T_n$ be samples from the GGD, take $Y_i = \log(T_i)$, then a reparametrisation of the generalized gamma gives us probability density function.

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)} \exp\left\{\sqrt{k}\omega - ke^{\omega/\sqrt{k}}\right\}$$

where $\omega = \sigma^{-1}(y-\mu)$, $\mu = log(\alpha) + \beta^{-1}log(k)$, $\sigma = \frac{1}{\sqrt{k\beta^2}}$ and $k = \frac{\alpha}{\beta}$. It is mentioned in the paper that a reasonable range for σ and is the interval (0,6). As for μ, a reasonable interval is (-4,4).
This density admits log likelihood

$$n\left\{-\log\sigma + (k-\frac{1}{2})\log k - \log\Gamma(k) + \sqrt{k}\frac{\overline{Y}-\mu}{\sigma} - k\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)S_0\right\}$$

where $S_j \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i^j \exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right)$ Differentiating the log likelihood with respect to μ, σ and k, we get score equations

$$0 = \frac{n\sqrt{k}}{\sigma}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)S_0 - 1\right\} \tag{7}$$

$$0 = \frac{n\sqrt{k}}{\sigma^2}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)(S_1 - \mu S_0) - \frac{\sigma}{\sqrt{k}} - \overline{Y} + \mu\right\} \tag{8}$$

$$0 = n\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\left(\frac{S_1 - \mu S_0}{2\sigma\sqrt{k}} - S_0\right) + \log k + 1 - \frac{1}{2k} - \psi(k) + \frac{\overline{Y}-\mu}{2\sigma\sqrt{k}}\right\} \tag{9}$$

Solving the first equation, we get

$$\hat{\mu} = \sigma\sqrt{k}\log S_0 \tag{10}$$

Substituting this expression into Equation (8), we can write

$$R(\sigma) \equiv \frac{S_1}{S_0} - \overline{Y} - \frac{\hat{\sigma}}{\sqrt{k}} = 0 \tag{11}$$

The full estimation method can be described as follows:

1. Choose a range of values for k over which to estimate the other parameters μ and σ with a certain step size. I have chosen the interval [0.3,30] with step size 0.1.

2. For a given value k, estimate σ by using the bisection method for Equation (11).

3. Substitute the estimated value $\hat{\sigma}$ into Equation (10) to find $\hat{\mu}$

4. repeat this for all values k in the chosen range to obtain a set of vectors (σ,μ,k)

5. Within this set of vectors (σ,μ,k), choose the vector for which the loglikelihood logL($\hat{\sigma}$,$\hat{\mu}$,k) is highest. Let's denote this vector by $(\hat{\sigma}, \hat{\mu}, k)_{mid}$

6. For k_{mid}, choose interval $[k_{mid} - \varepsilon, k_{mid} + \varepsilon]$ with a chosen step size. I chose $\varepsilon = 2$, and step size 0.01.

7. Repeat step 1 until step 4 for this new range of values for k.

8. Within this new set of vectors (σ,μ,k), choose the vector for which the loglikelihood logL($\hat{\sigma}$,$\hat{\mu}$,k) is highest. This is the obtained estimator.

3.2.1 Results

Results will be given based on 500 estimations for every set of parameters. Every method will be used to estimate the following sets of parameters for (α, β, k): $(3, 0.5, 3)$, $(5, 2, 0.5)$, $(4, 0.5, 2)$, $(7, 3, 1)$, $(2, 5, 8)$. This will provide a varied range of values for the parameters and various shapes for the GGD to test the estimation methods on. Since this section and Section 3.3 contain a different parametrisation, these sets of parameters will correspond to $(1.15, 3.3, 3)$, $(0.71, 1.26, 0.5)$, $(1.41, 2.78, 2)$, $(0.33, 1.95, 1)$, $(0.07, 1.11, 8)$.

Lawless' method is computationally intensive since we have to find σ and μ for every single value of k in our range. Especially the repeated calculation of σ makes this a very costly computation. The estimations with Lawless' method took a significantly longer time to compute than the other methods. The first thing we notice by looking at the tables is that the estimates for k consistently have higher bias and MSE than the other two parameters. The MSE for k tends to be larger for larger values of k. This will be a recurring theme in the other estimation methods. The other two parameters however, are estimated quite nicely by Lawless' method. Even for small sample sizes the MSE for σ and μ tend to stay small.

input $(\sigma, \mu,$ **k)**	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.05, 3.23, 7.60)	(-0.11, -0.06, 4.60)	(0.07, 0.31, 60.13)
(0.71, 1.26, 0.5)	(0.65, 1.25, 1.51)	(-0.05, -0.01, 1.00)	(0.03, 0.05, 10.59)
(1.41, 2.78, 2)	(0.60, 2.24, 3.78)	(-0.81, -0.53, 1.78)	(0.91, 0.59, 32.89)
(0.33, 1.95, 1)	(0.29, 1.94, 4.40)	(-0.04, -0.00, 3.40)	(0.01, 0.02, 40.96)
(0.07, 1.11, 8)	(0.07, 1.11, 10.55)	(0.00, 0.00, 2.55)	(0.00, 0.00, 38.74)

Table 2: 500 estimations with Lawless method with data sets of size n = 20

input $(\sigma, \mu,$ **k)**	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.10, 3.35, 3.62)	(-0.05, 0.6, 0.62)	(0.02, 0.05, 4.74)
(0.71, 1.26, 0.5)	(0.64, 1.28, 1.84)	(-0.06, 0.01, 0.22)	(0.02, 0.03, 0.85)
(1.41, 2.78, 2)	(1.33, 2.82, 2.70)	(-0.08, 0.05, 0.70)	(0.05, 0.12, 4.36)
(0.33, 1.95, 1)	(0.33, 1.93, 1.55)	(0.00, -0.01, 0.55)	(0.00, 0.01, 2.14)
(0.07, 1.11, 8)	(0.07, 1.11, 8.96)	(0.00, 0.00, 0.96)	(0, 0, 13.94)

Table 3: 500 estimations with Lawless method with data sets of size n = 50

input $(\sigma, \mu,$ **k)**	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.15, 3.30, 3.67)	(-0.01, 0.01, 0.67)	(0.01, 0.03, 3.50)
(0.71, 1.26, 0.5)	(0.71, 1.24, 0.63)	(0.00, -0.02, 0.13)	(0.01, 0.03, 0.16)
(1.41, 2.78, 2)	(1.37, 2.76, 2.78)	(-0.04, -0.01, 0.78)	(0.02, 0.04, 4.42)
(0.33, 1.95, 1)	(0.32, 1.39, 1.02)	(-0.01, -0.56, 0.02)	(0.00, 0.31, 0.42)
(0.07, 1.11, 8)	(0.07, 1.11, 9.05)	(0.00, 0.00, 1.05)	(0, 0, 26.29)

Table 4: 500 estimations with Lawless method with data sets of size n = 100

input (σ, μ, k)	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.15, 3.28, 3.49)	(0.00, -0.01, 0.49)	(0.00, 0.00, 2.04)
(0.71, 1.26, 0.5)	(0.70, 1.26, 0.51)	(-0.01, 0.01, 0.01)	(0.01, 0.00, 0.01)
(1.41, 2.78, 2)	(1.41, 2.76, 2.08)	(-0.01, 0.01, 0.08)	(0.03, 0.01, 0.41)
(0.33, 1.95, 1)	(0.33, 1.95, 0.99)	(0.00, 0.00, -0.01)	(0.00, 0.00, 0.07)
(0.07, 1.11, 8)	(0.07, 1.11, 8.74)	(0.00, 0.00, 0.74)	(0.00, 0.00, 19.46)

Table 5: 500 estimations with Lawless method with data sets of size n = 500

3.3 Noufaily and Jones method

Noufaily and Jones solve the issues of maximum likelihood estimation by using the same parametrisation used by Lawless and finding new estimators for the resulting distribution. This method extends the work of Lawless. Where Lawless estimates the parameters σ and μ with k as a given, Noufaily & Jones iteratively solves for all three parameters.

Returning to the equations

$$\mu = \sigma\sqrt{k}\log S_0$$

$$R(\sigma) \equiv \frac{S_1}{S_0} - \overline{Y} - \frac{\sigma}{\sqrt{k}} = 0$$

Substituting these two expressions, we can reduce Equation (9) to

$$T(k) \equiv \log k - \psi(k) - \frac{L}{\sqrt{k}} = 0 \tag{12}$$

where $L = (\mu - \overline{Y})/\sigma$

Noufaily and Jones showed that Equation (11) and (12) have a unique root given that $L > 0$. This proof will be given in 3.2.1. Solving Equation (10), (11) and (12) iteratively will give a closer and closer estimates for μ, σ and k. Thus the estimation proposed by Noufaily & Jones can be explained as follows. We obtain an initial value for L, with this we can obtain an initial value for k by solving Equation (12). After this We can substitute this value into Equation (11) to obtain an estimate for μ from Equation (10). Now that we have new values for μ and σ, we can again obtain a new value for k by way of Equation (12). We can repeat this process again and again, and the fact that Equation (12) and (11) have unique roots will guarantee that we get closer and closer to the actual values of μ, σ and k. More formally, the procedure is given as follows:

1. Set i = 0, set $L = L_0 > 0$
2. Set i = i + 1
3. With L_{i-1}, compute $\hat{k}_i$ by solving Equation (12) by the bisection method or Newton Raphson.
4. Substitute $\hat{k}_i$ into Equation (11) and solve for σ by the bisection method or Newton Raphson.
5. Substitute $\hat{k}_i$ and $\hat{\sigma}_i$ into Equation (10) and compute $\hat{\mu}_i$
6. Now compute L_i to obtain the value of the log likelihood function
7. Repeat steps 2-6 until desired accuracy for the likelihood. That is, stop the process once the change in likelihood is smaller than a chosen ε

To initialize the iterative process, we have to obtain initial values for σ, μ and k. The authors of the paper suggested the initial values for σ and μ to be simulated by a gamma distribution with scale parameter 1, shape parameter 2. And to simulate the initial value for μ by a standard random normal. In the case that with these values $L_0 < 0$, we set $k = 1/L_0^2$ and continue to Step 4. For time reasons, the ε in Step 7 is chosen rather large, at $\varepsilon = 0.1$

3.3.1 Proof of unique solutions for $R(\sigma)$ and $T(k)$

Noufaily & Jones showed that the Equation (11) and (12) have unique solutions. The proofs will be provided below. First we show that the function $R(\sigma)$ is monotone decreasing:

$$\frac{\partial R(\sigma)}{\partial \sigma} = \frac{1}{\sqrt{k}} \left(\frac{S_1^2 - S_0 S_2}{\sigma^2 S_0^2} - 1 \right)$$

Both of the terms inside brackets are negative. For the first term this can be seen by applying the Cauchy-Schwartz inequality:

$$(\sum_{i=1}^{n} a_i b_i)^2 \leq (\sum_{i=1}^{n} a_i^2)(\sum_{i=1}^{n} b_i^2)$$

Taking $a_i = \frac{1}{n} Y_i \text{exp}\left(\frac{Y_i}{2\sigma\sqrt{k}}\right), b_i = \frac{1}{n}\text{exp}\left(\frac{Y_i}{2\sigma\sqrt{k}}\right)$, we obtain

$$\left[\frac{1}{n}\sum_{i=1}^{n} Y_i \text{exp}\left(\frac{Y_i}{\sigma\sqrt{k}}\right)\right]^2 \leq \left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \text{exp}\left(\frac{Y_i}{\sigma\sqrt{k}}\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n} \text{exp}\left(\frac{Y_i}{\sigma\sqrt{k}}\right)\right]$$

which shows $S_1^2 \leq S_0 S_2$ and consequently, $S_1^2 - S_0 S_2 \leq 0$. We also have that

$$\lim_{\sigma \to 0} R(\sigma) = Y_{max} - \bar{Y} > 0$$

and

$$\lim_{\sigma \to \infty} R(\sigma) = -\infty$$

So we know that $R(\sigma)$ is a continuous, monotone decreasing function with values greater than 0 and values less than zero. From this we can conclude (by the intermediate value theorem), that there is a unique value σ for which $R(\sigma) = 0$.
As for the function $T(k)$, we will try to prove uniqueness of a solution by the intermediate value theorem as well. We know that for small values of k we have $\psi(k) \sim -k^{-1}$. Thus for small values of k this function will dominate the other two. Thus we have $\lim_{k\to 0} T(k) = -\infty$. For large values of k, we can write $\psi(k) \sim log(k) - (2k)^{-1}$ and thus $\lim_{k\to\infty} T(k) = 0$. To know whether this 0 is reached from 0^+ or 0^-, we first rewrite (10) as follows: $\mu = \sigma\sqrt{k}\log S_0 \implies \frac{\mu}{\sigma\sqrt{k}} = \log S_0$. Then we have

$$\text{exp}\left(\frac{\mu}{\sqrt{k}\sigma}\right) = \text{average}\left(\text{exp}\left(\frac{Y}{\sigma\sqrt{k}}\right)\right) > \text{exp}\left(\text{average}\left(\frac{Y}{\sigma\sqrt{k}}\right)\right) = \text{exp}\left(\frac{\bar{Y}}{\sqrt{k}\sigma}\right)$$

The inequality can be shown with Jensen's inequality, which states that $E[g(x)] \geq g(E[x])$ given that $g(x)$ is a concave function. Which is true for the exponential function.

From this we can conclude that $\mu > \bar{Y}$. This means that L $= (\mu - \overline{Y})/\sigma > 0$. Since $-L/\sqrt{k}$ is negative and is the dominant term for large k, we approach zero from the negative side. Now the only step that is left is to show that $T(k)$ is a monotone decreasing function. Let us rewrite T(k) to

$$T(k) \equiv \sqrt{k}(\log(k) - \psi(k))$$

and solve for $T(k) = L$. This is an equivalent problem. Theorem 1 of (Alzer, 1997) states that $k^\alpha(\log(k) - \psi(k))$ is completely monotonic on $(0, \infty)$ for any $\alpha \leq 1$. Thus, our function $T(k)$ is monotone decreasing. At last, we conclude that Equations (11) and (12) have unique solutions.

3.3.2 Results

Even for small sample sizes, the estimator seems to perform quite well for the parameters σ and μ. The same can be said for when k is small. However, we can see that as k gets larger, the estimation for k becomes more off-target. This is not surprising, as we saw that the pdf of this parametrisation does not change that much as k gets larger and larger. It is noteworthy, however, to see that the estimates for k have larger error than Lawless' way of estimation.

input (σ, μ, k)	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(0.97, 3.54, 1.63)	(-0.18, 0.25, -1.37)	(0.08, 0.22, 4.82)
(0.71, 1.26, 0.5)	(0.75, 1.12, 1.84)	(0.05, -0.14, 1.34)	(0.04, 0.11, 5.57)
(1.41, 2.78, 2)	(1.2, 3.07, 1.55)	(-0.21, 0.3, -0.45)	(0.13, 0.38, 2.79)
(0.33, 1.95, 1)	(0.31, 1.95, 1.53)	(-0.02, 0, 0.53)	(0.01, 0.02, 2.89)
(0.07, 1.11, 8)	(0.06, 1.12, 3.01)	(-0.01, 0.01, -4.99)	(0, 0, 30.24)

Table 6: 500 estimations with Noufaily method with data sets of size n = 20

input (σ, μ, k)	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.07, 3.44, 1.95)	(-0.08, 0.15, -1.05)	(0.03, 0.08, 3.26)
(0.71, 1.26, 0.5)	(0.72, 1.19, 1.12)	(0.01, -0.07, 0.62)	(0.02, 0.04, 2.28)
(1.41, 2.78, 2)	(1.33, 2.91, 1.81)	(-0.08, 0.13, -0.19)	(0.04, 0.12, 2.08)
(0.33, 1.95, 1)	(0.32, 1.94, 1.9)	(-0.01, -0.01, 0.9)	(0, 0.01, 4.78)
(0.07, 1.11, 8)	(0.07, 1.12, 3.25)	(0, 0.01, -4.75)	(0, 0, 26.83)

Table 7: 500 estimations with Noufaily and Jones method with data sets of size n = 50

input (σ, μ, k)	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.1, 3.41, 1.97)	(-0.05, 0.11, -1.03)	(0.01, 0.04, 2.49)
(0.71, 1.26, 0.5)	(0.72, 1.21, 0.79)	(0.01, -0.05, 0.29)	(0.01, 0.03, 0.68)
(1.41, 2.78, 2)	(1.35, 2.87, 1.87)	(-0.06, 0.1, -0.13)	(0.02, 0.07, 2.05)
(0.33, 1.95, 1)	(0.33, 1.95, 1.47)	(-0.01, 0, 0.47)	(0, 0, 2.32)
(0.07, 1.11, 8)	(0.07, 1.12, 3.4)	(0, 0.01, -4.6)	(0, 0, 24.92)

Table 8: 500 estimations with Noufaily and Jones method with data sets of size n = 100

input (σ, μ, k)	**estimates**	**bias**	**MSE**
(1.15, 3.3, 3)	(1.1, 3.41, 1.97)	(-0.05, 0.11, -1.03)	(0.01, 0.04, 2.49)
(0.71, 1.26, 0.5)	(0.72, 1.21, 0.79)	(0.01, -0.05, 0.29)	(0.01, 0.03, 0.68)
(1.41, 2.78, 2)	(1.35, 2.87, 1.87)	(-0.06, 0.1, -0.13)	(0.02, 0.07, 2.05)
(0.33, 1.95, 1)	(0.33, 1.95, 1.47)	(-0.01, 0, 0.47)	(0, 0, 2.32)
(0.07, 1.11, 8)	(0.07, 1.12, 2.67)	(0.00, 0.02, -5.32)	(0, 0, 32)

Table 9: 500 estimations with Noufaily and Jones method with data sets of size n = 500

3.4 Gao method

Gao et al. uses yet another parametrisation of the GGD, namely

$$p(x) = \frac{|\beta| k^k}{\sigma \Gamma(k)} \left(\frac{x}{\sigma}\right)^{k\beta-1} \exp\left\{-k\left(\frac{x}{\sigma}\right)^{\beta}\right\}$$

The original parametrisation can be obtained by making the substitution $\sigma = \alpha \exp\left(\frac{\log k}{\alpha}\right)^{\beta}$ The estimator proposed is based on log cumulants of the GGD and expands on the work of (Li, Hong, Wu, & Fan, 2011). Gao et al. (2016) proposes an estimator that can be used in situations where the original method of log-cumulants (MoLC) fails analytical conditions for application. The first three theoretical log cumulants are given below by c_1, c_2 and c_3 and their empirical counterparts are given by $\hat{c}_1, \hat{c}_2$ and $\hat{c}_3$

$$c_1 = \log(\alpha) + (\psi(k) - \log(k))/\beta \qquad \hat{c}_1 = \frac{1}{n}\sum_{i=1}^{n} \log(x_i) \tag{13}$$

$$c_2 = \psi(1,k)/\beta^2 \qquad \hat{c}_2 = \frac{1}{n}\sum_{i=1}^{n} (\log(x_i - \hat{c}_1))^2 \tag{14}$$

$$c_3 = \psi(2,k)/\beta^3 \qquad \hat{c}_3 = \frac{1}{n}\sum_{i=1}^{n} (\log(x_i - \hat{c}_1))^3 \tag{15}$$

where $\psi(\cdot)$ denotes the digamma function and $\psi(m,\cdot)$ denotes polygamma function, the mth order derivative of the digamma function. These quantities can be estimated by the following equations: $\hat{c}_1 = \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$ $\hat{c}_2 = \frac{1}{n}\sum_{i=1}^{n} (\log(x_i - \hat{c}_1))^2$ $\hat{c}_3 = \frac{1}{n}\sum_{i=1}^{n} (\log(x_i - \hat{c}_1))^3$

From these log-cumulants and their estimators, we obtain a function $g(\hat{k})$. This function will be used to estimate the parameter k.

$$\lambda = \frac{\hat{c}_2^3}{\hat{c}_3^2} = \frac{\psi^3(1,\hat{k})}{\psi^2(2,\hat{k})} = g(\hat{k}) \tag{16}$$

(Li et al., 2011) showed that the remaining two parameters can be estimated by

$$\hat{\beta} = \operatorname{sgn}(-\hat{c}_3)\sqrt{\psi(1,\hat{k})/\hat{c}_2}$$

$$\hat{\alpha} = \exp\left\{\hat{c}_1 - (\psi(\hat{k}) - \log(\hat{k}))/\hat{\beta}\right\}$$

For large values of k, we can approximate the polygamma functions by $\psi(1,k) \approx \frac{1}{k} + \frac{1}{2k^2}$ and $\psi(2,k) \approx -\frac{1}{k^2} - \frac{1}{k^3}$ with this we can rewrite

$$g(\hat{k}) = \frac{\hat{c}_2^3}{\hat{c}_3^2} = \lambda \approx \frac{(\frac{1}{\hat{k}} + \frac{1}{2\hat{k}^2})^3}{(-\frac{1}{\hat{k}^2} - \frac{1}{\hat{k}^3})^2}$$

using yet another approximation $k^2 + k \approx (k + \frac{1}{2})^2$ we can further rewrite the equation to $\lambda = \frac{\hat{k}^2}{\hat{k}+\frac{1}{2}}$ $0 \approx \hat{k}^2 - k\lambda - \frac{1}{2}\lambda$. From this we get the final approximation for k:

$$\hat{k} = \frac{\lambda + \sqrt{\lambda^2 + 2\lambda}}{2} \tag{17}$$

The proposed scheme for estimating parameters is given by the following:

1. For a given sample of data, compute the sample log-cumulants given in (7), (8) and (9)
2. If $\lambda < 0.25$, estimate $\hat{k}$ by equation (11). If not, go to the next step.
3. If $\lambda \geq 2.3052$, the proposed estimator in (11) is still accurate. Otherwise, use the original MoLC.
4. use the estimated $\hat{k}$ to find estimates $\hat{\beta}$ and $\hat{\sigma}$.

As seen from these equations, Gao's estimation method is a series of simple calculations as opposed to iterative methods which can take multiple thousands of iterations before admitting approximations.

3.4.1 Results

For Gao's estimation method, we estimate parameters based on sample sizes $n = 20$, $n = 50$ and $n = 100$ and $n = 500$. For each size we repeat the estimation procedure M = 500 times and calculate the sample mean, bias and MSE.

Looking at the bias we see that the β is consistently estimated to be lower than the actual value. We also see that the MSE is highest for $k = 8$, especially in the cases where the sample size is smaller.

The first thing that should be noted is that for almost all sample sizes, the estimates for the third parameter k are off by a large margin. This can be seen most clearly when looking at the MSE of each table. It can be seen that the error in the estimated value gets larger as the real value of k grows. Though as the sample size grows, the estimated values get closer and closer to the real values. It should be noted that the authors showed reasonable results for this estimator for sample sizes of $n = 50000$.

The opposite seems to hold for the estimates for α and β, where the bias for these two parameters is less than 1 in every table. The MSE for these two parameters stays under 1 as well, for the most part. With this it seems that the estimation method proposed by (Gao et al., 2016) performs quite well under low sample sizes for the first two parameters.

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.79, 0.34, 1873.48)	(-0.21, -0.16, 1870.48)	(1.04, 0.2, 9.81 x 10^8)
(5, 2, 0.5)	(4.02, 0.94, 344.26)	(-0.98, -1.06, 343.76)	(2.12, 1.45, 2.71 x 10^7)
(4, 0.5, 2)	(3.72, 0.36, 6421.94)	(-0.28, -0.14, 6419.94)	(2.78, 0.15, 1.21 x 10^{10})
(7, 3, 1)	(6.53, 1.72, 248.06)	(-0.47, -1.28, 247.06)	(0.73, 3.84, 1.00 x 10^7)
(2, 5, 8)	(1.99, 3.53, 936.1)	(-0.01, -1.47, 928.1)	(0, 57.99, 8.21 x 10^7)

Table 10: 500 estimations with Gao method for data sets of size n = 20

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.92, 0.43, 10473700)	(-0.08, -0.07, 10473700)	(0.58, 0.11, 5.48 x 10^{16})
(5, 2, 0.5)	(4.29, 1.17, 1.67)	(-0.71, -0.83, 1.17)	(1.18, 0.93, 4.54)
(4, 0.5, 2)	(3.75, 0.43, 162404)	(-0.25, -0.07, 162402)	(1.45, 0.08, 1.31 x 10^{13})
(7, 3, 1)	(6.77, 2.3, 20.69)	(-0.23, -0.7, 19.69)	(0.35, 1.87, 8.61 x 10^4)
(2, 5, 8)	(2, 4.81, 505.84)	(0, -0.19, 497.84)	(0, 24.3, 1.85 x 10^7)

Table 11: 500 estimations with Gao method for data sets of size n = 50

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.95, 0.47, 362.36)	(-0.05, -0.03, 359.36)	(0.32, 0.07, 2.34 x 10^7)
(5, 2, 0.5)	(4.51, 1.33, 1.06)	(-0.49, -0.67, 0.56)	(0.71, 0.66, 0.68)
(4, 0.5, 2)	(3.85, 0.45, 20.07)	(-0.15, -0.05, 18.07)	(0.89, 0.05, 5.48 x 10^4)
(7, 3, 1)	(6.85, 2.56, 2.01)	(-0.15, -0.44, 1.01)	(0.2, 1.17, 6.72)
(2, 5, 8)	(2, 4.9, 3045.01)	(0, -0.1, 3037.01)	(0, 13.43, 2.02 x 10^9)

Table 12: 500 estimations with Gao method for data sets of size n = 100

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.99, 0.5, 3.63)	(-0.01, 0, 0.63)	(0.08, 0.02, 4.47)
(5, 2, 0.5)	(4.64, 1.46, 0.78)	(-0.36, -0.54, 0.28)	(0.26, 0.38, 0.12)
(4, 0.5, 2)	(3.93, 0.48, 2.4)	(-0.07, -0.02, 0.4)	(0.21, 0.01, 1.16)
(7, 3, 1)	(6.91, 2.7, 1.27)	(-0.09, -0.3, 0.27)	(0.05, 0.39, 0.24)
(2, 5, 8)	(2, 5.1, 15.43)	(0, 0.1, 7.43)	(0, 2.99, 5309.04)

Table 13: 500 estimations with Gao method for data sets of size n = 500

3.5 Zhang method

In (Zhang et al., 2020) the proposed estimator makes use of the maximum likelihood equations. Equation (2) is rewritten to

$$\hat{k} = \left[\hat{\beta} \left(\frac{\sum_{i=1}^{n} t_i^{\hat{\beta}} \log t_i}{\sum_{i=1}^{n} t_i^{\hat{\beta}}} - \frac{\sum_{i=1}^{n} \log t_i}{n} \right) \right]^{-1} \tag{18}$$

and the estimator for α remains unchanged,

$$\hat{\alpha} = \left(\frac{1}{n\hat{k}} \sum_{i=1}^{n} t_i^{\hat{\beta}} \right)^{1/\hat{\beta}} \tag{19}$$

Noting that both of these estimators are functions of β (and can be denoted $k(\beta) and \alpha(\beta)$), we only need a way of estimating β and the parameters α and will follow. In this paper an estimator for β is found based on log-cumulants. Making use of the Mellin transform, the equation for the second log-cumulant is obtained,

$$\xi_2 = \frac{1}{\beta^2} \psi(1, k)$$

Since we have an estimator for k, we can rewrite the second log-cumulant as

$$\xi_2(\beta) = \frac{1}{\beta^2} \psi \left(1, \left[\hat{\beta} \left(\frac{\sum_{i=1}^{n} t_i^{\hat{\beta}} \log t_i}{\sum_{i=1}^{n} t_i^{\hat{\beta}}} - \frac{\sum_{i=1}^{n} \log t_i}{n} \right) \right]^{-1} \right)$$

The sample log-cumulant can be calculated with formulas

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \log(t_i), \quad \hat{\xi}_2 = \frac{1}{n-1} \sum_{i=1}^{n} (\log t_i - \hat{\mu})^2$$

With this, we should have

$$\xi_2(\beta) - \hat{\xi}_2 = 0 \tag{20}$$

3.5.1 Details of the estimation method

Given Equation (20), the paper gives details on how to find an interval from which to start the bisection method to find the root. The main idea is to find values β_1 and β_2 such that $\xi_2(\beta_1) \leq \hat{\xi}_2$ and $\xi_2(\beta) > \hat{\xi}_2$. In the first part of the procedure we simulate random values β_{sim} in the parameter space of β and find values of β such that $\xi_2(\beta_{sim}) > \hat{\xi}_2$. The second part finds values β_{min} such that $\xi_2(\beta_{sim}) \leq \hat{\xi}_2$ by finding a local minimum of $\xi_2(\beta)$ by way of Gradient Descent.

1. Simulate N_β values of $\beta_{initial}$ within a preset parameter space. I chose $N_\beta = 30$ values to be simulated in interval $[0, 7]$
2. From the N_β values select N'_β values of $\beta_{initial}$ for which $\xi_2(\beta_{initial}) > \hat{\xi}_2$
3. Generate N'_β sets of vectors $(\beta_{initial}, \beta_{min}, \xi_2(\beta_{min}))$, where β_{min} and $\xi_2(\beta_{min})$ are obtained for each $\beta_{initial}$ as follows:
   (a) Initialise $\beta^{(0)} = \beta_{initial}$
   (b) While $|\beta^{(T)-\beta^{(T-1)}}| > \varepsilon$: do
   $\beta^{(T+1)} = \beta(T) + \eta[-\xi'_2(\beta^{(T)})]$,
   where I chose $\varepsilon = 0.1$, T is the current iteration number, $\xi'_2(\beta)$ is the first derivative of $\xi'_2(\beta)$, η is the learning rate which is found by a linear search in the interval $\eta \in [0,\tilde{\eta}]$ by $\hat{\eta} = \arg\min_{\eta}(|\xi'_2(\eta)|)$. Given that $\beta \in [0, 7]$, we have
   $$\tilde{\eta} = \begin{cases} (\beta^{(T)} - 7)/\xi'_2(\eta) & \text{for } \xi'_2(\eta) < 0 \\ \beta^{(T)}/\xi'_2(\eta) & \text{for } \xi'_2(\eta) > 0 \end{cases}$$
   (c) Return to Step 3(a) for the next $\beta_{initial}$
   (d) Once Steps 3(a)-3(c) terminate, obtain $\beta_{min} = \beta^{(T)}$ and $\xi'_2(\beta_{min})$
4. From the N'_β sets of vectors, select N''_β sets of vectors $(\beta_{initial}, \beta_{min}, \xi_2(\beta_{min}))$ according to the criterion $\xi_2(\beta_{min}) \leq \hat{\xi}_2$
5. From the N''_β sets of vectors, select vector for which β_{min} is largest. Denote this vector by $(\beta^*_{initial}, \beta^*_{min}, \xi_2(\beta^*_{min}))$.
6. Find $\hat{\beta}$ by performing the bisection method on Equation (20) with the interval for the bisection method given by $[\min(\beta^*_{initial}, \beta^*_{min}), \max(\beta^*_{initial}, \beta^*_{min})]$.
7. Estimate $\hat{k} = \left[\hat{\beta}\left(\frac{\sum_{i=1}^n t_i^{\hat{\beta}} \log t_i}{\sum_{i=1}^n t_i^{\hat{\beta}}} - \frac{\sum_{i=1}^n \log t_i}{n}\right)\right]^{-1}$
8. Estimate $\hat{\alpha} = \left(\frac{1}{n\hat{k}} \sum_{i=1}^n t_i^{\hat{\beta}}\right)^{1/\hat{\beta}}$

In step 3 we have

$$\xi'_2(\beta) = \frac{1}{\beta^2}\left(\frac{\partial k}{\partial \beta}\psi[2, k(\beta)] - 2\psi[1, k(\beta)]\right)$$

$$\frac{\partial k}{\partial \beta} = -\frac{1}{\beta^2\left(\frac{\sum_{i=1}^n t_i^\beta \log t_i}{\sum_{i=1}^n t_i} - \frac{\sum_{i=1}^n \log t_i}{n}\right)} - \frac{\frac{\sum_{i=1}^n t_i^\beta (\log t_i)^2}{\sum_{i=1}^n t_i} - \frac{(\sum_{i=1}^n t_i^\beta \log t_i)^2}{(\sum_{i=1}^n t_i}\big)^2}{\beta\left(\frac{\sum_{i=1}^n t_i^\beta \log t_i}{\sum_{i=1}^n t_i} - \frac{\sum_{i=1}^n \log t_i}{n}\right)}$$

3.5.2 Results

It is clear that Zhang's method gives very accurate estimates for β. However, in all sample sizes the estimates for $\beta = 5$ are not accurate. Future results for $\beta = 5$ might improve by increasing the number N_β for initial samples of β. What can also be seen is that across all sample sizes, the parameter α tends to get underestimated, similarly to the results in Section 3.4.

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.41, 0.46, 3.30)	(-0.58, -0.04, 0.30)	(3.82, 0.02, 2.14)
(5, 2, 0.5)	(4.69, 1.96, 0.55)	(-0.31, -0.04, 0.05)	(1.24, 0.08, 0.03)
(4, 0.5, 2)	(3.73, 0.46, 2.11)	(-0.27, -0.04, 0.12)	(8.13, 0.02, 0.74)
(7, 3, 1)	(6.64, 3.00, 1.24)	(-0.36, 0.00, 0.24)	(0.85, 0.00, 0.23)
(2, 5, 8)	(0.69, 1.80, 3.82)	(-1.31, -3.20, -4.18)	(2.57, 16.00, 47.33)

Table 14: 500 estimations with Zhang method for data sets of size n = 20

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.86, 0.45, 2.80)	(-0.14, -0.05, -0.20)	(3.24, 0.025, 1.23)
(5, 2, 0.5)	(4.72, 1.92, 0.51)	(-0.28, -0.08, 0.01)	(1.58, 0.16, 0.02)
(4, 0.5, 2)	(3.39, 0.42, 1.78)	(-0.61, -0.08, -0.22)	(4.63, 0.04, 0.77)
(7, 3, 1)	(6.78, 3.00, 1.07)	(-0.22, 0.00, 0.07)	(0.34, 0.00, 0.04)
(2, 5, 8)	(0.58, 1.50, 2.76)	(-1.42, -3.50, -5.24)	(2.80, 17.50, 45.78)

Table 15: 500 estimations with Zhang method for data sets of size n = 50

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(3.07, 0.47, 2.83)	(0.07, -0.03, -0.17)	(1.59, 0.02, 0.71)
(5, 2, 0.5)	(4.92, 2.00, 0.51)	(-0.08, 0.00, 0.01)	(0.16, 0.00, 0.00)
(4, 0.5, 2)	(3.61, 0.44, 1.81)	(-0.39, -0.06, -0.19)	(3.75, 0.03, 0.55)
(7, 3, 1)	(6.95, 3.00, 1.04)	(-0.05, 0.00, 0.04)	(0.12, 0.00, 0.02)
(2, 5, 8)	(0.71, 1.80, 3.16)	(-1.29, -3.20, -4.84)	(2.56, 16.00, 41.49)

Table 16: 500 estimations with Zhang method for data sets of size n = 100

input (α, β, k)	**estimates**	**bias**	**MSE**
(3, 0.5, 3)	(2.61, 0.45, 2.75)	(-0.39, -0.05, -0.25)	(1.02, 0.03, 0.94)
(5, 2, 0.5)	(4.82, 1.92, 0.48)	(-0.18, -0.08, -0.02)	(1.04, 0.16, 0.01)
(4, 0.5, 2)	(3.44, 0.44, 1.79)	(-0.56, -0.06, -0.21)	(2.20, 0.03, 0.50)
(7, 3, 1)	(6.96, 3.00, 1.02)	(-0.04, 0.00, 0.02)	(0.04, 0.00, 0.01)
(2, 5, 8)	(0.72, 1.80, 2.92)	(-1.28, -3.2, -5.08)	(2.56, 16.00, 41.01)

Table 17: 500 estimations with Zhang method for data sets of size n = 500

4 Comparison

4.1 Computational speed

The estimation method with the fastest runtime is Gao's method. For all parameter values and sample sizes I tried the runtime was no longer than a few seconds to calculate 500 estimations. This is to be expected as Gao's method contains no iterative processes and only relies on various formulas to provide an estimate. After Gao's method, Noufaily and Zhang's estimation method alternate in runtime. In Noufaily's algorithm, the long runtime comes from the fact that each iteration demands the execution of two bisection methods, one for finding σ and one for k and each estimate needs anywhere from 3 to 50 iterations. In Zhang's method the long runtime is caused by the gradient descent part of the code. How many iterations are needed is strongly dependent on the value of β to be estimated. By far the slowest running program is Lawless' method. As mentioned in Section 3.2, computing one estimate can demand 150 executions of the bisection method. Heuristically, computing 500 estimates for the Noufaily algorithm took a similar amount of time for to compute 50 estimates for the Lawless method.

4.2 Errors

Noufaily's algorithm admitted errors most often. In roughly 25% of the runs the Noufaily algorithm was unable to produce an estimate before running into an error. The Zhang method follows, where the calculation of $\xi_2'(\beta)$ sometimes produces "NaN" results and it can also happen that in Step 4 there are no vectors that fulfill the criterion, hence terminating that estimation attempt . The other two algorithms showed little to no errors.

4.3 Results

Comparing all four estimation methods will be hard given the two different parametrisations. We compare estimates for the parameter k and provide general comments on the estimates of the other parameters, so as to not flood this section with tables.

We can see that Zhang performs the best for the first 4 values, having the lowest MSE for all sample sizes. For $k = 8$, Lawless and Noufaily have better performance. We can see that all methods tend to have higher MSE for higher k. Furthermore, Gao has the worst MSE of all methods for sample sizes $n = 20, 50, 100$, but is better than Noufaily for $n = 500$

k	0.5	1	2	3	8
Lawless	10.59	40.96	32.89	60.1	38.74
Noufaily	5.57	2.89	2.79	4.82	30.24
Gao	2.71 x 10^7	1.00 x 10^7	1.21 x 10^{10}	9.81 x 10^8	8.21 x 10^7
Zhang	0.03	0.23	0.74	2.14	47.33

Table 18: MSE of all methods, n = 20

k	0.5	1	2	3	8
Lawless	0.85	2.14	4.36	4.74	13.94
Noufaily	2.28	4.78	2.08	3.26	26.83
Gao	4.54	8.61 x 10^4	1.31 x 10^{13}	5.48 x 10^{16}	1.85 x 10^7
Zhang	0.02	0.04	0.77	1.23	45.78

Table 19: MSE of all methods, n = 50

k	0.5	1	2	3	8
Lawless	0.16	0.42	3.50	4.42	26.29
Noufaily	0.68	2.32	2.05	2.49	24.92
Gao	0.68	6.72	5.48 x 10^4	2.34 x 10^7	2.02 x 10^9
Zhang	0.00	0.02	0.55	0.71	41.49

Table 20: MSE of all methods, n = 100

k	0.5	1	2	3	8
Lawless	0.01	0.07	0.41	2.04	19.46
Noufaily	0.68	2.32	2.05	2.49	32
Gao	0.12	0.24	1.16	4.47	5309.04
Zhang	0.01	0.01	0.50	0.94	41.01

Table 21: MSE of all methods, n = 500

Looking at α and β, Zhang has very good results for β, even for $n = 20$ the MSE is close to zero for all values except for $\beta = 5$. For α, Gao has lower MSE than Zhang. For σ and μ, Noufaily and Lawless provide similar results, which is to be expected as the only difference between the methods is in estimating the k.

4.4 Practical suggestions

If runtime for the estimation of the GGD is no issue (for example if the parameters of the GGD only need to be calculated once), it is clear that Lawless method performs best for small sample sizes (in this case meaning sample sizes between $n = 20$ and $n = 500$). In the case that runtime becomes more important and errors become more tolerable, one might opt to implement the other methods in the following order: Noufaily and Jones, Zhang, Gao. Gao's method was shown to perform quite poorly for the sample sizes in this report. However, as sample sizes increase, the computation time needed for Gao's method becomes a significant factor. For sample sizes greater than $n = 2000$, the estimates become similar to the estimates shown in this report for the other methods, while only needing a fraction of the time. This property might be helpful when GGD parameters need to be estimated repeatedly for large sample sizes.

Another consideration is the ease of implementation for the methods. For every method except for Gao's method one needs to adjust various values in the code, for example the tolerance level ε in the iterative processes needs to be chosen, or a predetermined range for possible values of the parameter to search through (this is the case for Lawless, Noufaily and Jones, and Zhang). The choices for these values greatly impact the runtime for the methods as well. For example. if one has a notion for the range of the β parameter (in the (α, β, k) parametrisation), and no idea of a range for the other parameters, Zhang's method can be advantageous. Similarly, if one has knowledge of a possible range for k, the runtime for Noufaily and Jones' method as well as Lawless' method can be sped up rapidly. Thus these two are excellent choices. On the other hand, if no knowledge is available of the parameter space, one has to choose a wide range for the implementations of the methods. This is no problem if the parameters need to be estimated only once, but in the case that multiple estimations are necessary, again the runtime can be an issue.

5 New findings

As part of this project I would like to add to the existing estimation methods for the GGD. To this end, I will be looking to extend an estimation method for the full conditional for gamma shape parameters (Miller, 2019). In this Section I will introduce the proposed method by Miller and will give an overview of the problem that it solves. We will take a look at the theoretical results that produced this estimator before trying to extend the idea to a method that can estimate the parameters of the GGD.

5.1 Miller's method for the full conditional for gamma shape parameters

The gamma distribution often comes up in Bayesian modelling. it is desirable to use a conjugate prior for the shape parameters. However, there is no such prior that is easy to use. The conjugate priors that do exist do not have a closed form, making it hard to implement in modelling. This problem can be bypassed by way of Markov Chain Monte Carlo sampling methods, namely the Metropolis-Hastings method, in which the value of the shape parameter is updated iteratively until the value converges to a true value. The downside to using Metropolis-Hastings however is that it can take a lot of computation before an accurate estimate can be obtained. In modern applications, some models contain a large number of parameters that need to be estimated, which makes it so that efficient and accurate computation is necessary.

To solve this problem, Miller proposes to use a gamma distribution as a prior for the shape parameter of a gamma. It turns out that the full conditional distribution can be approximated well with again a gamma distribution. Miller goes on to propose an algorithm that approximates this gamma distribution for the full conditional.

The basic idea is to find a gamma distribution g and match the first and second derivatives $\log g$ to the first and second derivative of $\log f$ at a point near the mean of f, where f is our full conditional. Since the mean of f is not known in closed form, the approximation is done by iteratively matching the functions at the mean of the current g.

5.2 Miller's algorithm

Let $X_1, ..., X_n|\alpha, \mu \sim \text{Gamma}(\text{shape} = a, \text{rate} = a/\mu), \text{with} a, \mu > 0$. Assume that the shape parameter has a gamma prior, $a \sim \text{Gamma}(\text{shape} = a_0, \text{rate} = b_0)$. The following algorithm will produce A and B such that $p(a|x_1, ..., x_n, \mu, a_0, b_0) \approx \text{Gamma}(a|\text{shape} = A, \text{rate} = B)$. Note that the mean of $\text{Gamma}(a|A, B)$ is given by A/B

Algorithm 1: Approximating the full conditional of the shape parameter.

Data: input $\text{x}_1, ..., x_n > 0$, parameters $\mu, a_0, b_0 > 0$,
tolerance $\varepsilon > 0$, maximum number of iterations M
Result: A and B
begin;
$R \leftarrow \sum_{i=1}^{n} \log(x_i)$;
$S \leftarrow \sum_{i=1}^{n} x_i$;
$T \leftarrow S/\mu - R + n\log(\mu) - n$;
$A \leftarrow a_0 + n/2$;
$B \leftarrow b_0 + T$;
for *j = 1,...,M* **do**
  $a \leftarrow A/B$;
  $A \leftarrow a_0 - na + na^2\psi'(a)$;
  $B \leftarrow b_0 + (A - a_0)/a - n\log(a) + n\psi(a) + T$;
  if $|a/(A/B) - 1| < \varepsilon$ **then**
    return A,B
  end
end
return A,B

The author recommends to set $\varepsilon = 10^{-8}$ and $M = 10$. In the following subsection I will give the derivation to this algorithm. This derivation will inform the extension of Miller's work to an estimation method for the parameters of the GGD.

5.3 Derivation of Miller's algorithm

Let $f(a) = p(a|x_1, ..., x_n, \mu, a_0, b_0)$ denote the full conditional density of parameter a, where a_0, b_0 can be chosen to depend on μ. Let $g(a) = \text{Gamma}(a|A, B)$ for some A,B. Put simply the idea of the algorithm is to choose A and B such that for a point a near the mean of f, we have that $\frac{\partial}{\partial a} \log f(a) = \frac{\partial}{\partial a} \log g(a)$ and $\frac{\partial^2}{\partial a^2} \log f(a) = \frac{\partial^2}{\partial a^2} \log g(a)$. Since $f(a)$ is not known in closed form, we approximate the point of the mean of f by taking the mean of g. Then A and B are refined iteratively by setting A equal to the second derivative of $\log f$, setting B equal to the first derivative of $\log f$, and then setting $a = A/B$ (where A/B is the mean of g) in the next iteration. The formulas for the first and second derivative will be derived in Section 5.3.1. The iterative process is initialized by choosing A and B based on an approximation of the Gamma function. The derivation of this will be given in Section 5.3.2.

5.3.1 Derivatives of $\log f$ and $\log g$

We have

$$g(a) = \text{Gamma}(a|A, B) = \frac{a^{A-1} e^{-Ba} B^A}{\Gamma(A)}$$

Then

$$\begin{aligned}
\log g(a) &= A \log(B) - \log \Gamma(A) + (A-1) \log(a) - Ba \\
\frac{\partial}{\partial a} \log g(a) &= \frac{A-1}{a} - B \\
\frac{\partial^2}{\partial a^2} \log g(a) &= -\frac{A-1}{a^2}
\end{aligned}$$

We also have

$$f(a) \propto p(x_1, ..., x_n|a, \mu, a_0, b_0) p(a|\mu, a_0, b_0)$$

and

$$\begin{aligned}
p(x_1, ..., x_n|a, \mu, a_0, b_0) &= \prod_{i=1}^{n} \frac{(a/\mu)^a}{\Gamma(a)} x_i^{a-1} \exp(-(a/\mu) x_i) && (21) \\
&= \frac{(\alpha/\mu)^{na}}{\Gamma(a)^n} \exp((a-1)R) \exp(-(a/\mu)S) && (22) \\
&= \frac{a^{na}}{\Gamma(a)^n} \exp(-R - (T+n)a) && (23)
\end{aligned}$$

where we define $R = \sum_{i=1}^{n} \log(x_i), S = \sum_{i=1}^{n} x_i$ and $T = S/\mu - R + n \log(\mu) - n$. Remember that we put a gamma prior on a so that $p(a|\mu, a_0, b_0) = \text{Gamma}(a|a_0, b_0)$

$$\begin{aligned}
\log f(a) &= \text{const} + \log p(x_1, ..., x_n|a, \mu, a_0, b_0) + \log p(a|\mu, a_0, b_0) \\
&= \text{const} + na \log(a) - n \log \Gamma(a) - (T+n)a + (a_0 - 1) \log(a) - b_0 a
\end{aligned}$$

where any term not proportional to a was lumped into "const". From this we further derive

$$\begin{aligned}
\frac{\partial}{\partial a} \log f(a) &= n \log(a) + n - n\psi(a) - (T+n) + \frac{a_0 - 1}{a} - b_0 && (24) \\
\frac{\partial^2}{\partial a^2} \log f(a) &= n/a - n\psi'(a) - \frac{a_0 - 1}{a^2} && (25)
\end{aligned}$$

Setting $\frac{\partial^2}{\partial a^2} \log f(a) = \frac{\partial^2}{\partial a^2} \log g(a)$, we obtain

$$A = a_0 - na + na^2\psi'(a)$$

setting $\frac{\partial}{\partial a} \log f(a) = \frac{\partial}{\partial a} \log g(a)$ gives us

$$B = b_0 + \frac{A - a_0}{a} - n\log(a) + n\psi(a) + T$$

5.3.2 Initialization by approximating $\Gamma(a)$

In order to find good initial values for A and B, we approximate the gamma function by way of Stirling's formula. Two situations should be discerned, when a is large and when a is small. The second case is omitted as it will not be used in the algorithm.

Stirling's approximation to $\Gamma(a)$ is given by $\Gamma(a) \sim \sqrt{2\pi}a^{-1/2}\left(\frac{a}{e}\right)^a$ as $a \to \infty$. Here $h_1(a) \sim h_2(a)$ as $a \to a^*$ means that $h_1(a)/h_2(a) \to 1$ as $a \to a^*$. From this we get that $\Gamma(a)/a^a \sim \sqrt{2\pi}a^{-1/2}e^{-a}$. Plugging this into Equation (23) we get

$$p(x_1, ..., x_n|a, \mu, a_0, b_0) \approx (2\pi)^{-n/2}a^{n/2}e^{na}\exp(-R - (T+n)a) \propto a^{n/2}\exp(-Ta)$$

Adding that $p(a|\mu, a_0, b_0) \propto a^{a_0-1}e^{-b_0 a}$, we get

$$f(a) \propto a^{a_0+n/2-1}e^{-(b_0+T)a}$$

In other words, $f(a)$ is proportional to $\text{Gamma}(a|a_0 + n/2, b_0 + T)$. For this reason, we initialize with $A = a_0 + n/2$ and $B = b_0 + T$

5.4 Extension to the GGD

My main idea for the extension will rely on the same principles as Miller's algorithm in order to estimate the α parameter. We put a GGD prior on α and then try to approximate the full conditional posterior f by another GGD g. The parameters will be estimated by matching the first and second derivative of $\log f$ and $\log g$ at the mean. This will give us two formulas, one for α and one for k. This means that we need another way of estimating β. Taking the third derivative of both log functions does not bear any useful formulas. Instead we opt to use Zhang's method for estimating β, as it is a standalone procedure for estimating the parameter and thus can be used in our procedure without any problem.

5.4.1 Derivation

Let $f(\alpha) = p(\alpha|x_1, ..., x_n, \beta, k, \alpha_0, \beta_0, k_0)$ denote the full conditional density of parameter α. Let $g(\alpha) = GGD(\alpha|A, B, K)$ for some A, B, K. We have

$$g(\alpha) = \frac{B}{A\Gamma(K)} \left(\frac{\alpha}{A}\right)^{BK-1} \exp\left[-\left(\frac{\alpha}{A}\right)^{B}\right]$$

This gives us

$$\begin{aligned}
\log g(\alpha) &= \log B + (BK-1)\log\alpha - BK\log A - \log\Gamma(K) - \left(\frac{a}{A}\right)^{B} \\
\frac{\partial}{\partial\alpha}\log g(\alpha) &= \frac{BK-1}{\alpha} - B\frac{\alpha^{B-1}}{A^{B}} \\
\frac{\partial^2}{\partial\alpha^2}\log g(\alpha) &= -\frac{BK-1}{\alpha^2} - (B^2-B)\frac{\alpha^{B-2}}{A^{B}}
\end{aligned}$$

We also have

$$f(\alpha) \propto p(x_1, ..., x_n|\alpha, \beta, k)p(\alpha|\alpha_0, \beta_0, k_0)$$

This gives us

$$\begin{aligned}
\log f(\alpha) = {} & n\log\beta - n\log\Gamma(k) + (\beta k - 1)\log\left(\sum_{i=1}^{n} t_i\right) - (n\beta k - n)\log\alpha - \frac{\sum_{i=1}^{n} t_i^{\beta}}{\alpha^{\beta}} \\
& + \log\left(\frac{\beta_0}{\alpha_0 k_0}\right) + (\beta_0 k_0 - 1)\log\alpha - (\beta_0 k_0 - 1)\log\alpha_0 - \left(\frac{\alpha}{\alpha_0}\right)^{\beta_0} \\
\frac{\partial}{\partial\alpha}\log f(\alpha) = {} & -\frac{n\beta k}{\alpha} + \beta\frac{1}{\alpha^{\beta+1}}\sum_{i=1}^{n} t_i^{\beta} + \frac{\beta_0 k_0 - 1}{\alpha} - \beta_0\frac{\alpha^{\beta_0-1}}{\alpha_0} \\
\frac{\partial^2}{\partial\alpha^2}\log f(\alpha) = {} & \frac{n\beta k}{\alpha^2} - (\beta^2+\beta)\frac{\sum_{i=1}^{n} t_i^{\beta}}{\alpha^{\beta+2}} - \frac{\beta_0 k_0 - 1}{\alpha^2} - (\beta_0^2 - \beta_0)\frac{\alpha^{\beta_0-2}}{\alpha_0}
\end{aligned}$$

Setting the first derivatives equal to one another gives us

$$A = \left\{ B\alpha^{B-1} \left[\frac{BK-1}{\alpha} + \frac{n\beta k}{\alpha} + \beta_0 \frac{\alpha^{\beta_0 - 1}}{\alpha_0} - \beta \frac{1}{\alpha^{\beta+1}} - \frac{\beta_0 k_0 - 1}{\alpha} \right]^{-1} \right\}^{1/B}$$

Setting the second derivatives equal to one another gives us

$$K = -[n\beta k - (\beta^2 + \beta)\frac{\sum_{i=1}^{n} t_i^{\beta}}{\alpha^{\beta}} - \beta_0 k_0 - 1 - (\beta_0^2 - \beta_0)\frac{\alpha_0^{\beta}}{\alpha_0} + (B^2 - B)\frac{\alpha^B}{A^B} - 1]/B$$

and B is estimated by making use of Zhang's method. Now we only need to find initial values A and K. I was unable to find a good way of calculating these initial values from the data. So these values will be assumed known. Practically speaking, this will mean that values close to the real values will be used to initiate this estimation method.

5.5 Procedure

Let $Z(x_1, ..., x_n)$ denote the estimation of β by Zhang's method. Then the following produces A, B, K such that $p(\alpha|x_1, ..., x_n, \beta, k, \alpha_0, \beta_0, k_0) \approx g(\alpha|A, B, K)$, where g is the GGD with parameters A, B, K After obtaining A, B, K, one can infer the value

Algorithm 2: Approximating the full conditional of the parameter α.

Data: input $x_1, ..., x_n > 0$, parameters β, k, α_0, β_0, k_0, $A_{init}, B_{init}, K_{init}$
tolerance $\varepsilon > 0$, maximum number of iterations M
Result: A and B
begin;
$A \leftarrow A_{init}$;
$B \leftarrow B_{init}$;
$K \leftarrow K_{init}$;
for *j = 1,...,M* **do**
  $\alpha \leftarrow A\frac{\Gamma(K+\beta^{-1})}{\Gamma(K)}$;
  $A \leftarrow \left\{ B\alpha^{B-1} \left[\frac{BK-1}{\alpha} + \frac{n\beta k}{\alpha} + \beta_0 \frac{\alpha^{\beta_0-1}}{\alpha_0} - \beta \frac{1}{\alpha^{\beta+1}} - \frac{\beta_0 k_0 - 1}{\alpha} \right]^{-1} \right\}^{1/B}$;
  $B \leftarrow Z(x_1, ..., x_n)$;
  $K \leftarrow -[n\beta k - (\beta^2+\beta)\frac{\sum_{i=1}^{n} t_i^{\beta}}{\alpha^{\beta}} - \beta_0 k_0 - 1 - (\beta_0^2 - \beta_0)\frac{\alpha_0^{\beta}}{\alpha_0} + (B^2 - B)\frac{\alpha^B}{A^B} - 1]/B$
  if $|a/(A\frac{\Gamma(K+\beta^{-1})}{\Gamma(K)}) - 1| < \varepsilon$ **then**
    return A, B, K
  end
end
return A, B, K

of α by simulation methods (Metropolis Hastings for example), or from the formula $\alpha = A\frac{\Gamma(K+\beta^{-1})}{\Gamma(K)}$. Unfortunately, during implementation of this prodcedure the program continuously produced errors or results that were not a number NaN. Further investigation into how best to implement this idea is needed.

6 Conclusions

The Generalized Gamma distribution can be used in various fields in statistics. The GGD encompasses distributions oftentimes used in, among others, lifetime analysis. Despite its flexibility, parameter estimation for the GGD through maximum likelihood estimation did not produce good results. This is due to the fact that very different values for the parameters can produce very similar distribution functions. In this project I analysed four estimators that were proposed to circumvent these problems with the MLE. Lawless' and Noufaily and Jones' method proved to give very accurate results for low sample sizes in exchange for runtime. Zhang's method gives very accurate estimates for β in the (α, β, k) parametrisation. Gao's method showed poor performance for the low sample sizes used in this project, but can be shown to have good performance for sample sizes greater than $n = 2000$.

After comparing these four estimation methods. I aimed to extend Miller's algorithm for the gamma distribution into a method for the generalized gamma. This method would give a way to estimate the first parameter of the GGD, α. Unfortunately, implementation of this idea was not succesful.

References

Alzer, H. (1997). On some inequalities for the gamma and psi functions. *Mathematics of computation*, *66*(217), 373–389.

Gao, G., Ouyang, K., Luo, Y., Liang, S., & Zhou, S. (2016). Scheme of parameter estimation for generalized gamma distribution and its application to ship detection in sar images. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(3), 1812–1832.

Hager, H. W., & Bain, L. J. (1970). Inferential procedures for the generalized gamma distribution. *Journal of the American Statistical Association*, *65*(332), 1601–1609.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. John Wiley & Sons.

Lawless, J. F. (1980). Inference in the generalized gamma and log gamma distributions. *Technometrics*, *22*(3), 409–419.

Li, H.-C., Hong, W., Wu, Y.-R., & Fan, P.-Z. (2011). On the empirical-statistical modeling of sar images with generalized gamma distribution. *IEEE Journal of selected topics in signal processing*, *5*(3), 386–397.

Manning, W. G., Basu, A., Mullahy, J., & Manning, W. (2002). Modeling costs with generalized gamma regression. *Chicago: The University of Chicago*.

Miller, J. W. (2019). Fast and accurate approximation of the full conditional for gamma shape parameters. *Journal of Computational and Graphical Statistics*, *28*(2), 476–480.

Noufaily, A., & Jones, M. (2013). On maximization of the likelihood for the generalized gamma distribution. *Computational Statistics*, *28*(2), 505–517.

Pham, T., & Almhana, J. (1995). The generalized gamma distribution: its hazard rate and stress-strength model. *IEEE transactions on reliability*, *44*(3), 392–397.

Ramos, P. L., Achcar, J. A., Moala, F. A., Ramos, E., & Louzada, F. (2017). Bayesian analysis of the generalized gamma distribution using non-informative priors. *Statistics*, *51*(4), 824–843.

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*, 1187–1192.

Zhang, P., Li, B., Boudaren, M. E. Y., Yan, J., Li, M., & Wu, Y. (2020). Parameter estimation of generalized gamma distribution toward sar image processing. *IEEE Transactions on Aerospace and Electronic Systems*, *56*(5), 3701–3717.