STATISTICS IN TRANSITION new series, December 2025
Vol. 26, No. 4, pp. 55–68, https://doi.org/10.59139/stattrans-2025-039
Received – 20.01.2023; accepted – 22.05.2025

Exploring new mixtures of distribution to model skewed and heavy tailed data

Jitendra Kumar[1], **Anuj Nain**[2]

Abstract

The search for relevant models that can describe the loss data has been one of the main interests of researchers for decades. There is limited research on modeling such data using K-component mixture models. An example of that can be Miljkovic and Grün's study (2016) of six distributions where they proposed finite mixtures to model the data. In this paper we study two more distributions, namely log-logistic and inverse Weibull distribution in addition of all those proposed by Miljkovic and Grün (2016). We employed the EM algorithm for parameter estimation and then selected the best model using three model selection criteria, namely NLL, AIC and BIC. We also computed the risk measures such as VaR and TVaR and compared them with their empirical counterparts to assess the goodness-of-fit of our proposed models at the extreme quantiles. We found that K-component mixture distribution of log-logistic and inverse Weibull works better than competent models. To get a more generalized view on the theory of mixture distribution, a simulation was carried out, which gave satisfactory results.

Key words: K-component finite mixture models, EM algorithm, Danish fire insurance losses data set, log-logistic distribution, inverse Weibull distribution.

1. Introduction

Many times, in real-life situations random variables are not generated from a single but from a mixture of several distributions. The mixture distribution is a weighted sum of K distributions where the weights sum up to one. Each weight corresponds to the proportion or contribution of the corresponding component density. Each density in the mixture is characterized by an unknown parameter (or a parameter vector).

In actuarial domain, searching for distributions that can be used to model a data which is heavy tailed, positively skewed, non-Gaussian and multimodal has gained lot

[1] Department of Statistics, Central University of Rajasthan, Bandersindri, Ajmer, India.
E-mail: vjitendrav@gmail.com. ORCID: https://orcid.org/0000-0003-4473-4148.
[2] Department of Statistics, Central University of Rajasthan, Bandersindri, Ajmer, India.
E-mail: anuj.nain.stats@gmail.com. ORCID: https://orcid.org/0009-0003-3678-5647.

of attention of the researchers. Mixture distribution can be well used to model such data. Mixture models that can cover the humps and the tail provide a good fit for the data, but the literature on such models is limited. A mixture of exponential distributions was proposed by Keatinge (1999). In this model the maximum likelihood (ML) estimation was based on Newton's algorithm. Although the model was applicable in some areas it was not very useful in modelling heavy-tailed data.

Klugman and Rioux (2006) proposed a flexible model that included not only exponential components but also Gamma, Pareto and Log-normal with non-negative weights that sum up to one, provided that either weight associated with Log-normal or Gamma component must be zero. Further, some work regarding this type of modeling was done by Lee and Lin (2010) and Verbelen *et al.* (2015, 2016). They considered finite mixture of Erlang distribution.

Miljkovic and Grün (2016), extended the work beyond Erlang families. They developed K-component mixtures of six models with components from non-Gaussian families of distributions. The six families of distributions they used were Burr, Gamma, Inverse Burr, Inverse Gaussian, Log-normal, and Weibull. In total, there are more than thirty distributions in their paper. They compared their results with composite Weibull models developed by Bakar *et al.* (2015). Their results outperformed the models introduced by Bakar *et al.* (2015). In this paper, we have extended the work of Miljkovic and Grün (2016) by introducing K-component finite mixtures of two more distributions, log-logistic and inverse Weibull.

While the present work focuses on the development and application of finite mixture models using heavy-tailed distributions, it is also worth mentioning that considerable theoretical work has been carried out to address the asymptotic behavior and estimation challenges in mixture models. Traditional likelihood-based methods, such as the likelihood ratio test (LRT), are known to exhibit non-standard asymptotic distributions due to violations of regularity conditions in mixture settings. Pioneering studies by Ghosh and Sen (1985) and Chernoff and Lander (1995) revealed that the asymptotic distribution of LRT in such contexts may involve the supremum of a Gaussian process, rather than a standard chi-square distribution. To address these complexities, Chen and Chen (1998a, b) developed adjusted LRT methods, and Chen *et. al* (2001) further proposed a modified LRT with desirable asymptotic properties and enhanced power under local alternatives. In addition, Chen and Kalbfleisch (1996) introduced penalized minimum-distance estimators, which yield consistent estimation of both the mixing distribution and the number of components. While these works do not directly align with the specific mixture distributions explored in this paper, their contributions to model identifiability, penalization, and asymptotic inference provide valuable methodological context.

It is also to be noted that although the present study builds upon the framework established by Miljkovic and Grün (2016), it is important to acknowledge that further

research has been conducted in related modeling beyond 2016. For instance, Lachos Dávila et al. (2018) presented finite mixtures of skew-normal and skewed distributions, focusing on modeling skewness and kurtosis, rather than multimodality or tail-heaviness typical of actuarial data. Similarly, Gagnon and Wang (2024) developed robust heavy-tailed versions of generalized linear models, providing improved outlier resistance within a GLM framework rather than mixture modeling. Additionally, Marambakuyana and Shongwe (2024) explored a vast class of two-component composite and mixture models based on non-Gaussian distributions for insurance claims data; however, their emphasis was on empirical comparisons among predefined combinations, not on extending parametric families within finite mixtures.

While these studies contribute valuable insights, their modeling objectives and methodological approaches diverge from the specific structure of the finite mixture modeling considered in the present work. Therefore, the model proposed by Miljkovic and Grün (2016) remains the most relevant and appropriate foundation for the development pursued herein.

The chronology of the work presented in this study is as follows.

In Section 2.1 we have given a brief introduction to the mixture distribution and EM algorithm. In Section 2.2 and 2.3 we provide theory for the parameter estimates from log-logistic and inverse Weibull family respectively. In Section 2.4 we have discussed model selection criteria and risk measures.

In Section 3.1 we have given description of our data set and its characteristics. In Section 3.2 we provide our results. The mixture of log-logistic distribution is found to perform better than most of the models given by Miljkovic and Grün (2016) whereas those of inverse Weibull distribution completely outperform the best distributions given by Miljkovic and Grün (2016).

In Section 3.3, the K-S test is applied for checking the goodness-of-fit of the proposed mixture density. Plots of empirical density versus K-component mixture density are constructed to show how well the fitted density covers the empirical density, Q-Q plot and P-P Plot are also constructed to justify the goodness-of-fit. In Section 3.4, a simulation study is conducted on 50 samples of size 2492 each. The results of the simulation justify the fit of four component inverse Weibull distribution.

2. Mixture Distribution and EM Algorithm

2.1. Mixture Distribution

Let $X_1\ X_2 \ldots X_n$ be an independent and identically distributed random sample from the K-component finite mixture of probability distributions. This mixture distribution is represented as

$$f(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(x;\ \theta_K)\ , \tag{1}$$

subject to $\sum_{k=1}^{K} \pi_k = 1$,

where $\boldsymbol{\theta} = (\pi', \theta') = (\pi_1 , \pi_2 , \ldots , \pi_{k-1} , \theta_1 , \theta_2 , \ldots \theta_k)$ is the vector of unknown parameters and $0 < \pi_i \leq 1$. These K distributions may or may not be from the same family. In this paper we assume that for the mixture density given in (1) the component densities $f_k(.)$ are from the same family. Further, we implement the EM algorithm for parameter estimation. For more details one may refer Dempster (1997). Miljkovic and Grün (2016) obtained parameter estimates using EM algorithm for the K-component finite mixture with component densities from the same family. Using their notations and terminology we mention here an important function called Q-function. For the details of the Q-function one may refer to Miljkovic and Grün (2016).

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{old}) = \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik}[\log(\pi_k) + \log(f_k(x_i;\ \theta_k))] \tag{2}$$

where w_{ik} is the expected value of the i^{th} weight in the k^{th} component density. This function is maximized to obtain new estimates of $\boldsymbol{\theta}$ (θ and π). The estimates of π are updated in the s^{th} iteration by

$$\pi_k^{(s)} = \frac{\sum_{i=1}^{n} w_{ik}^{(s)}}{n} \tag{3}$$

For simplification we will write m component finite mixture distribution as m-K followed by the name of the distribution. For example, four component inverse Weibull distribution would be written as 4-K inverse Weibull distribution.

2.2. Parameter estimation in m-K log-logistic distribution.

The log-logistic distribution is widely used in actuarial and financial modelling. Its properties are such attractive that it acts as an alternative distribution to the lognormal and Weibull distribution. Finite mixture models of log-logistic distribution can be used for modelling heavy-tailed data on positive support. In this section we provide the parameter estimates of m-K log-logistic distribution through the E-M algorithm.

The log-logistic distribution has the following probability density function:

$$f(x) = \begin{cases} \dfrac{a\left(\frac{x}{b}\right)^{a-1}}{b\left(\left(\frac{x}{b}\right)^{a}+1\right)^{2}} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

With shape parameter a > 0, scale parameter b > 0. Introducing the notion of finite mixture, we see that each k^{th} , (k = 1,2 ... m) component density in (1) follows log-logistic distribution with parameters a_k and b_k . Maximization of the Q function (2) with respect to a_k gives the following expression:

$$\sum_{i=1}^{n} w_{ki}\left(\frac{-a_k\left(\frac{x}{b_k}\right)\log\left(\frac{x}{b_k}\right)+a\log\left(\frac{x}{b_k}\right)+\left(\frac{x}{b_k}\right)^{a_k}+1}{a_k\left(\left(\frac{x}{b_k}\right)^{a_k}+1\right)}\right)=0 \tag{4}$$

maximization of the Q function with respect to b_k gives the following expression:

$$\sum_{i=1}^{n} w_{ki}\left(\frac{a_k\left(\frac{x}{b_k}\right)^{a_k}-1}{b_k\left(\left(\frac{x}{b_k}\right)^{a_k}+1\right)}\right) = 0 \tag{5}$$

Solving equations (4) and (5) we get the estimates of a_k and b_k

These equations can be solved using uniroot function in R. The function is useful in searching the root of a nonlinear equation. For details on the function one may refer to Brent (1973).

2.3. Parameter estimation in m-K inverse Weibull distribution

The inverse Weibull distribution is another distribution used in actuarial and income modelling. It has gain interest of the researchers in modeling heavy-tailed data on positive support appearing in finance and actuaries. In this section we provide the parameter estimates of m-K inverse Weibull distribution. The pdf of inverse Weibull distribution is given by

$$f(x) = \begin{cases} \dfrac{a\left(\frac{b}{x}\right)^{a} e^{-\left(\frac{b}{x}\right)^{a}}}{x} & x > 0, \\ 0 & \text{otherwise} \end{cases}$$

shape parameter a > 0, scale parameter b > 0. Introducing the notion of the finite mixture, we see that each k^{th} , (k= 1 ,2 ... m) component density in (1) follows log-logistic distribution with parameters a_k and b_k. Maximization of the Q function with respect to a_k gives the following expression:

$$\sum_{i=1}^{n} w_{ki}\left(-\left(\frac{b_k}{x}\right)^{a_k}\log\left(\frac{b_k}{x}\right)+\log\left(\frac{b_k}{x}\right)+\frac{1}{a_k}\right)=0 \tag{6}$$

Solving it using the numerical technique we get the estimate $\widehat{a_k}$ of a_k.

Maximization of the Q function with respect to b_k gives the following expression:

$$\sum_{i=1}^{n} w_{ki} \left(a_k \left(\frac{1-\left(\frac{b_k}{x}\right)^{a_k}}{b_k} \right) \right) = 0 \tag{7}$$

Solving equations (6) and (7) we get the estimates of a_k and b_k.

2.4. Model Selection and Risk Measure

In order to access the goodness-of-fit of the proposed models, the following measures are considered: Negative log-likelihood (NLL), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Let us denote log-likelihood function by L(θ) for a given model, then NLL is computed as -L(θ). AIC is given by $-2L(\theta) + 2p$, where p is the number of parameters present in the model and BIC is given by $-2L(\theta) + p\log(n)$, where n is the number of observations.

Let X be the random variable denoting the loss and π_p is the $100p$th percentile of the distribution of X. Then according to the notations and definitions by Klugman *et al.* (2012), the Value at Risk of X at the $100p$% level, denoted by VaRp (X) is the $100pth$ percentile of the distribution of X. Let $F_X(x)$ be the cumulative distribution function of X, then for continuous distributions, VaRp (X) can be written as the value of π_p satisfying

$$F_X(\pi_p) = p, \tag{8}$$

VaRp (X) lacks additive property, therefore we define TVaRp (X), which is the average of all VaRp (X) values above the security level p, Again using the notations and definitions by Klugman *et al.* (2012) . TVaRp (X) is given by

$$\text{TVaR}p(X) = \frac{\int_p^1 \text{VaR}_u(X)du}{1-P}$$

When X is a continuous random variable, the formula simplifies to

$$\text{TVaR}p(X) = E\,[\,X\,|X > \text{VaR}_P(X)]$$

We can use convenient notations [see Miljkovic and Grün (2016)] as

$$\text{TVaR}p(X) = E\,[\,X\,|X > \pi_p\,] = \frac{\int_{\pi_p}^1 xf(x)dx}{1-F_X(\pi_p)} = \frac{\int_{\pi_p}^1 xf(x)dx}{1-P}$$

For the sake of convenience we write VaRp (X) as VaR as TVaR$p(X)$ as TVaR.

3. Analysis

3.1. Data

In this paper, we have used the Danish fire losses data set. This globally recognized dataset has been used by actuaries for actuarial modeling for more than two decades.

The Copenhagen Reinsurance company was responsible for collecting the data. The data set contains a record of 2492 observations (loss claims in Danish Krone) over the period of 1980-1990 [see Miljkovic and Grün (2016)].

One can easily access the Danish fire losses data set (available as *danish*) in R by using the package **SMpracticals** [Davison (2013)]. Figure 1 shows the histogram of the data and Table 1 gives the summary statistics.

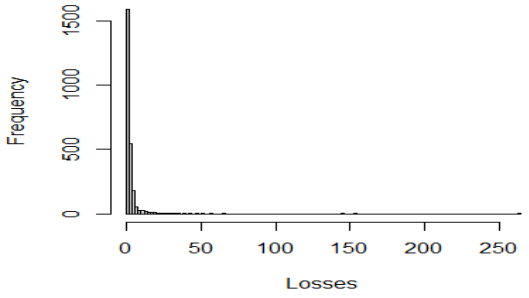

Figure 1. Histogram for Danish fire losses data set

Table 1. Summary Statistic

Min	Q1	Q2	Q3	Max	Mean
0.3134	1.1572	1.6339	2.6445	263.2504	3.0627

It can be observed that the mean is greater than the median and the third quartile (Q3), which indicates high skewness in the data. We consider modelling this data set using finite mixtures of various components for two distributions: log-logistic and inverse Weibull.

3.2. Results and discussions

In this section we have compared our work with work done by Miljkovic and Grün (2016). They considered six families of distributions and modeled their different K-component mixtures (in total there are 33 distributions). These six families were suitable for a data set having characteristics as present in the Danish fire losses data. However, two distributions, namely the log-logistic and the inverse Weibull remained unexplored in their study. We have extended their work by considering these two families, and modeling their K-component mixtures. The results are shown in the following table (Table 2).

Table 2. Model selection criteria for different component mixtures

Mixture	K(component)	NLL	AIC	BIC
log-logistic	1	4280.587	8565.175	8576.816
	2	3962.067	7934.133	7963.238
	3	**3850.580**	**7717.161**	**7753.728**
	4	3952.919	7927.838	7991.868
	5	3953.138	7936.276	8064.163
inverse Weibull	1	3966.83	7937.661	7949.302
	2	3842.314	7694.628	7723.733
	3	3793.387	7602.774	7649.341
	4	**3777.205**	**7576.409**	**7640.439**
	5	3773.208	7574.417	7655.908
	6	3771.172	3572.344	7663.482

It is observed that the proposed 3-K log-logistic distribution performs better than most of the models developed by Miljkovic and Grün (2016) (The parameter estimates have been presented in Figure 2). It performs better than 1-K, 2-K and 3-K models from the Gamma family, 1-K model from the Inverse Burr family, 1-K, 2-K and 3-K model from Inverse Gaussian, 1-K, 2-K and 3-K models from the Log-Normal family and 1-K, 2-K, 3-K, 4-K, 5-K models from the Weibull family.

Also, we observe that proposed 4-K inverse Weibull distribution performs better than all the distributions proposed by Miljkovic and Grün (2016) (The parameter estimates have been presented in Figure 4). (It is to be noted that we have reproduced the results for the six distributions studied by Miljkovic and Grün (2016). As these results were consistent with the original findings, we chose not to present them explicitly in the paper in order to avoid redundancy and to maintain focus on our novel contributions. However, for comparison one can refer to the table from Miljkovic and Grün (2016)).

The following Table 3 gives the VaR and TVaR risk measures.

Table 3. Risk measures

Specification	VaR(0.99)	TvaR(0.99)
Empirical Estimates	24.61	54.60
Proposed Models		
3-K log-logistic	20.33	31.50
4-K inverse Weibull	24.67	70.55

VaR from the proposed 3-K log-logistic (20.33) and that from 4-K inverse Weibull (23.57) are close to the empirical VaR (24.61). This adds to the goodness-of-fit of our proposed models. VaR from 4-K inverse Weibull is better than that from 2-K Burr

(25.02), 5-K Log Normal (26.75) and 3-K Inverse Burr (23.57), which were the best three models proposed earlier by Miljkovic and Grün (2016) [see Miljkovic and Grün (2016)]. This means that 4-K inverse Weibull covers tail of the data better than 2-K Burr, 5-K Log Normal and 3-K Inverse Burr. Due to presence of high skewness at the tail of the distribution it was obvious that TVaR from the distribution would be different from the empirical counterpart.

3.3. Goodness-of-fit

We conducted the Kolmogorov-Smirnov test (known as K-S test) for checking the goodness-of-fit of the proposed models. The null hypothesis states that the proposed distribution (either 3-K log-logistic or 4-K inverse Weibull) fits the data well. The significance level was set to .05. In the case of 3-K log-logistic distribution the test statistic was less than the critical value and hence the null hypothesis was accepted. In the case of 4-K inverse Weibull distribution also, the test statistic was less than the critical value and hence the null hypothesis was accepted. The results are summarized in Table 4.

Table 4. Goodness-of-fit test (H_0:The proposed distribution fits the data well)

Distribution	K-S Test Statistic	Critical Value (at α= 0.05)	Decision
3-K log-logistic	0.014	0.027	H_0 accepted
4-K inverse Weibull	0.009	0.027	H_0 accepted

Figure 2 and Figure 3 are for 3-K log-logistic distribution and show how well 3-K log-logistic distribution covers empirical density. Figure 2 shows the plot of empirical density against fitted density, Figure 3 shows the P-P plot and the Q-Q plot.

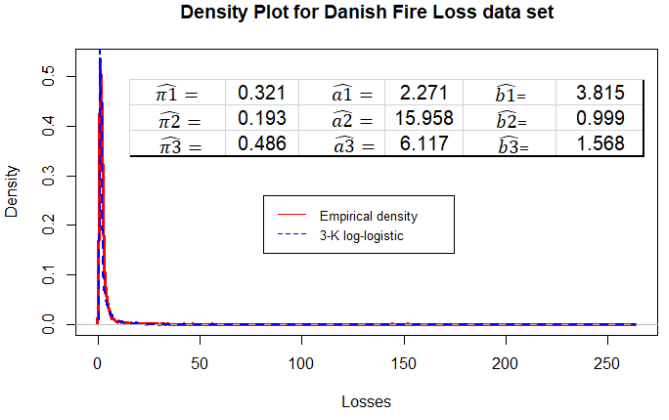

Figure 2. Plot of Empirical density and Fitted 3-K log-logistic density

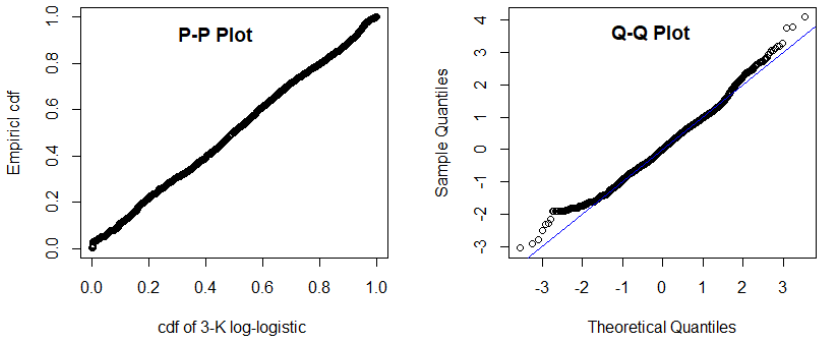

Figure 3. P-P plot and Q-Q Plot for the fitted 3-K log-logistic distribution

Figure 4 and Figure 5 are for 4-K inverse Weibull distribution and show how well the 4-K inverse Weibull density covers the empirical density. Figure 4 shows the plot of empirical density against fitted density, Figure 5 shows the P-P plot and the Q-Q plot.

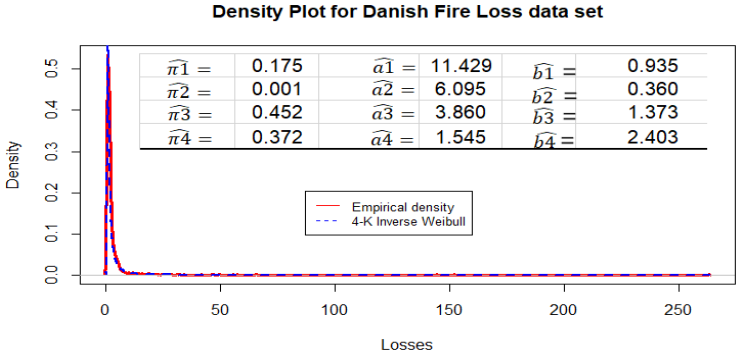

Figure 4. Plot of empirical density and fitted 4-K inverse Weibull density

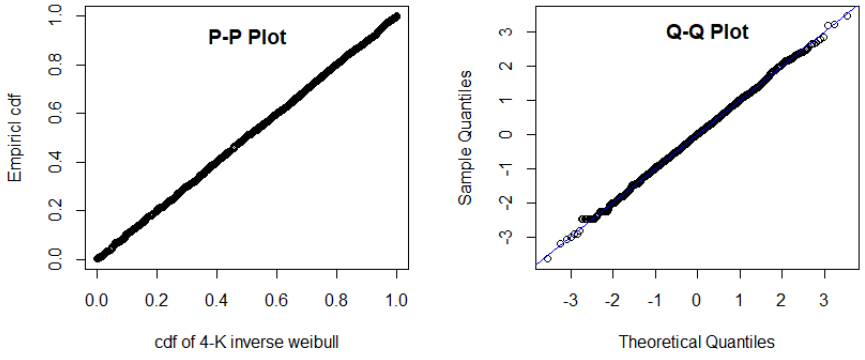

Figure 5. P-P plot and Q-Q Plot for the fitted 4-K inverse Weibull distribution

3.4. Simulation study

We conducted simulation for 4-K inverse Weibull distribution. We generated N=50 random samples from the given data of size n=2492 each, that can mimic the actual Danish fire losses data set. Then estimated the parameters, computed the VaR, constructed the density plot, conducted K-S test for goodness-of-fit and constructed P-P plot and Q-Q plot. We found that **4-K inverse Weibull** distribution fits each of the simulated data set well, as the Empirical VaR and fitted VaR from the distribution were very close to each other and the value of K-S test statistic was less than the critical value for each simulated sample. The table below (Table 5) shows the results for first ten simulated samples.

Table 5. Summary of results for the first ten simulated samples.

Sample No.	Empirical VaR(0.99)	Fitted VaR(0.99)	K-S Test statistic to be compared with 0.0272
1	25.288	23.59	0.012
2	21.989	23.1	0.010
3	25.059	24.33	0.013
4	26.309	29.34	0.020
5	20.873	22.12	0.011
6	25.954	26.03	0.012
7	22.465	21.54	0.012
8	23.900	23.661	0.013
9	20.963	23.09	0.012
10	22.351	21.7	0.011

As an example of simulation result we have shown Density plot, P-P Plot, Q-Q Plot for the first two simulated samples.

Simulated sample 1.

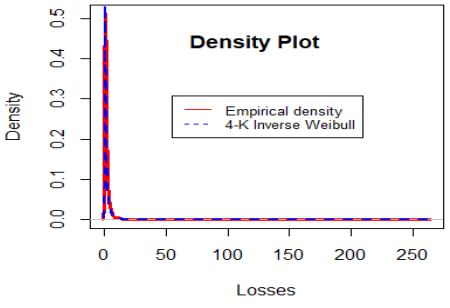

Figure 6. Plot of empirical density and simulated 4-K inverse Weibull density

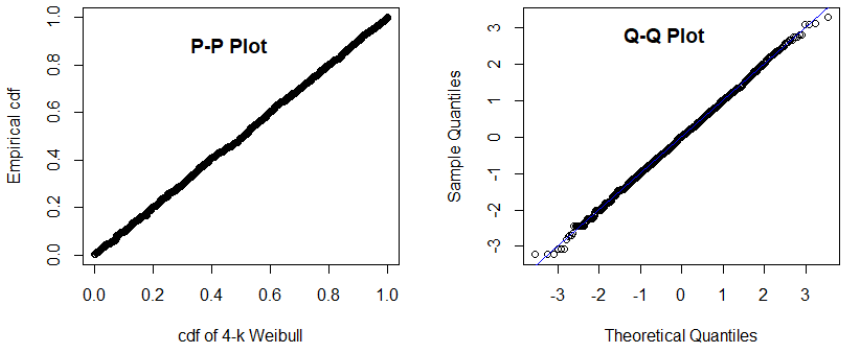

Figure 7. P-P plot and Q-Q Plot for the simulated 4-K inverse Weibull distribution

Simulated sample 2.

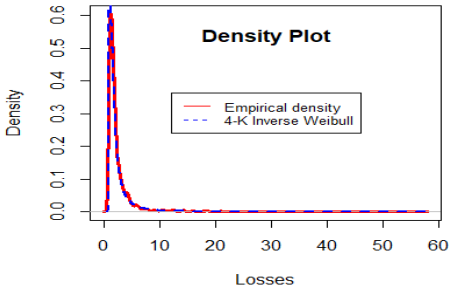

Figure 8. Plot of empirical density and simulated 4-K inverse Weibull density

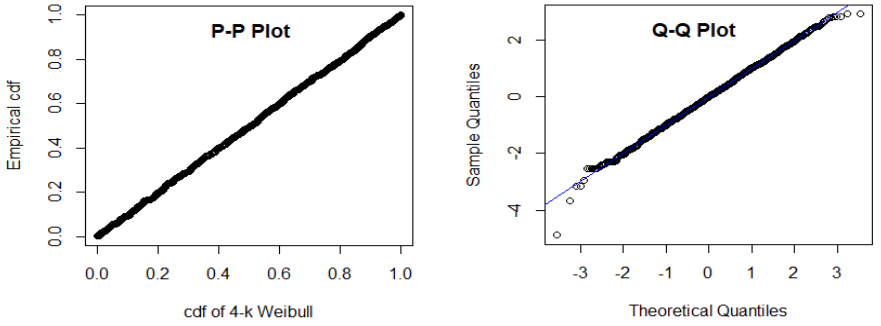

Figure 9. P-P plot and Q-Q Plot for the simulated 4-K inverse Weibull distribution

We conducted two sample t-tests to check the significance of difference between the Empirical VaR and VaR computed from the fitted 4-K inverse Weibull distribution. We formulate the null hypothesis, H_0: There is no significant difference between the empirical VaR and VaR computed from the fitted 4-K inverse Weibull distribution. We see that the null hypothesis is accepted at the significance level of 0.05. The result summary of the test is given in Table 9.

Table 9. Checking the significance of difference between the Empirical VaR and fitted VaR

Specification	
Test statistic	1.755
Critical Value	1.984
P-value	0.082
Level of significance	0.05

4. Conclusion

In this paper, we extended the work done by Miljkovic and Grün (2016) by adding and examining K-component finite mixtures of two more distributions: log-logistic and inverse Weibull.

We found that 3-K log-logistic distribution performed better than most of the distributions given earlier by Miljkovic and Grün (2016), whereas 4-K inverse Weibull distribution outperformed all the distributions given by Miljkovic and Grün (2016).

Hence, we conclude that for skewed and non-Gaussian data on positive support, finite mixtures of log-logistic and inverse Weibull distribution can successfully cover the heavy tail behavior and multimodal nature. To the best of our knowledge, we are among the recent after Miljkovic and Grün (2016) to work on K-component finite mixture model on such type of loss data.

We considered finite mixtures from the same families in our paper. Further work may be done by considering finite mixtures from different families. One may also use the Bayesian approach for parameter estimation.

References

Akaike, H., (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, Vol.19, pp. 716–723.

Bakar, S. A., Hamzah, N. A., Maghsoudi, M. and Nadarajah, S., (2015). Modeling loss data using composite models. *Insurance: Mathematics and Economics*, Vol. 61, pp. 146–154.

Brent, R., (1973). Algorithms for Minimization without Derivatives. Englewood Cliffs. NJ: *Prentice-Hall*.

Chen, H., Chen, J., (1998a). The likelihood ratio test for homogeneity in the finite mixture models. Technical Report STAT 98–08. Department of Statistics and Actuarial Science. *University of Waterloo*, Waterloo.

Chen, H., Chen, J., (1998b). Tests for homogeneity in normal mixtures with presence of a structural parameter. Technical Report STAT 98–09. Department of Statistics and Actuarial Science. *University of Waterloo*, Waterloo.

Chen, H., Chen, J. and Kalbfleisch, J. D., (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(1), pp. 19–29.

Chen, J., Kalbfleisch, J. D., (1996). Penalized minimum-distance estimates in finite mixture models. *Canadian Journal of Statistics*, *24*(2), pp. 167–175.

Chernoff, H., Lander, E., (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *Journal of Statistical Planning and Inference*, 43(1-2), pp. 19–40.

Dávila, V. H. L., Cabral, C. R. B. and Zeller, C. B., (2018). Finite mixture of skewed distributions. Germany: *Springer International Publishing*.

Davison, A.,(2013). SMPracticals: Practicals for Use with Davison (2003), Statistical Models. *R package version*, pp. 1–4.

Dempster, A. P., Laird, N. M. and Rubin, D. B., (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 19, pp. 1–22.

Gagnon, P., Wang, Y., (2024). Robust heavy-tailed versions of generalized linear models with applications in actuarial science. *Computational Statistics & Data Analysis*, 194, p. 107920.

Ghosh, J. K., Sen, P. K., (1984). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results.

Keatinge, C. L., (1999). Modeling losses with the mixed exponential distribution. In *Proceedings of the Casualty Actuarial Society*, Vol. 86, pp. 654–698.

Klugman, S. A., Panjer, H. H. and Willmot, G. E., (2012). Loss models: from data to decisions. *John Wiley & Sons*.

Klugman, S., Rioux, J., (2006). Toward a unified approach to fitting loss models. *North American Actuarial Journal*, Vol. 10, pp. 63–83

Lee, S. C., Lin, X. S., (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*, Vol. 14, pp. 107–130.

Marambakuyana, W. A., Shongwe, S. C., (2024). Composite and mixture distributions for heavy-tailed data—An application to insurance claims. *Mathematics*, 12(2), p. 335.

Miljkovic, T., Grün, B., (2016). Modeling loss data using mixtures of distributions. *Insurance: Mathematics and Economics*, Vol. 70, pp. 387–396.

Verbelen, R., Antonio, K. and Claeskens, G., (2016). Multivariate mixtures of Erlangs for density estimation under censoring. *Lifetime data analysis*, Vol. 22, pp. 429–455.

Verbelen, R., Gong, L., Antonio, K., Badescu, A. and Lin, S., (2015). Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. *ASTIN Bulletin: The Journal of the IAA*, Vol. 45, pp. 729–758.