1-1-2018

On the Mixtures of Weibull and Pareto (IV) Distribution: An Alternative to Pareto Distribution

I. Ghosh
University of North Carolina

Gholamhossein G. Hamedani
Marquette University, gholamhoss.hamedani@marquette.edu

Naveen K. Bansal
Marquette University, naveen.bansal@marquette.edu

Mehdi Maadooliat
Marquette University, mehdi.maadooliat@marquette.edu

Accepted version. *Communications in Statistics - Theory and Methods*, Vol. 47, No. 9 (2018): 2073-2084. DOI.

Communications in Statistics, Vol. 47, No. 9 (2018): 2073-2084. DOI.

On the Mixtures of Weibull and Pareto (IV) Distribution: An Alternative to Pareto Distribution

I. Ghosh
Department of Mathematics and Statistics, University of North Carolina, USA;

G.G. Hamedani
Department of Mathematics, Statistics and Computer Science, Marquette University, USA

N. Bansal
Department of Mathematics, Statistics and Computer Science, Marquette University, USA

M. Maadooliat
Department of Mathematics, Statistics and Computer Science, Marquette University, USA

Abstract

Finite mixture models have provided a reasonable tool to model various type of observed phenomena, specially those which are random in nature. In this paper, a finite mixture of Weibull and Pareto (IV) distribution is considered and studied. Some structural properties of the resulting model are discussed including estimation of

the model parameters via EM algorithm. A real life data application exhibits the fact that in certain situations, this mixture model might be a better alternative than the rival popular models.

1 Introduction

The mixture distributions over the years have provided a mathematical based way to model a wide variety of random phenomena statistically. The mixture distributions are effective and flexible models to analyze and interpret random durations in a possibly heterogenous population. In many situations, observed data may be assumed to have come from a mixture population of two or more distributions. Application of finite mixture models are in medicine, economics, psychology, botany, fisheries research, life testing and reliability among others. As an indirect application such a mechanism (finite mixture models) one may cite cluster analysis, latent structure models, empirical Bayes method and nonparametric density estimation.

In this paper we consider a finite mixture of two absolutely continuous distribution, a two parameter Weibull distribution and a three parameter Pareto (IV) distribution. In a finite mixture model, the distribution of random quantity of interest is modeled as a mixture of a finite number of component distributions in varying proportions. A mixture model is, thus, able to model quite complex situations through an appropriate choice of its components to represent accurately the local areas of support of the true distribution. It can handle situations where a single parametric family is unable to provide a satisfactory model for local variation in the observed data. The flexibility and high degree of accuracy of finite mixture models have been the main reason for their successful applications in a wide range of fields. The concept of finite mixture distribution was pioneered by Newcomb (1886) as a model for outliers. However, the credit for the introduction of statistical modeling using finite mixtures of distributions goes to Pearson (1894) while applying the technique in an analysis of crab morphometry data provided by Weldon (1892, 1893). For a comprehensive survey, readers are referred to Titterington et al. (1985), Lindsay (1995), Bohning (2000) and McLachlan & Peel (2000) and the references therein. The main objective here is to induce greater flexibility in modeling various types of data, specially in situations where these individual distributions fail to adequately fit the data separately.

Insurance companies need to investigate claims experience and apply mathematical techniques for many purposes, including, but not limited to rating, reserving, reinsurance arrangements solvency. In order to get an idea about loss distributions, which are a mathematical method of modeling individual claims, one needs to examine what distributions can be fitted to observed claims. Claims distributions tend to be positively skewed and long- tailed. This is where heavy tailed distributions, such as Weibull & Pareto appears to be a good fit. Insurance claims are generally modeled with Weibull or Pareto (IV) distribution. Weibull distribution has a heavier left tail and Pareto (IV) distribution has a heavier right tails. Thus, although Weibull distribution fits well for the lower insurance claims, it does not fit well for the higher insurance claims. On the other hand, the Pareto (IV) distribution fits well for the higher insurance claims but not the lower claims. A Truncated Composite Weibull-Pareto distribution has been used by some authors (see, for example, Teodorescu and Panaitescu, 2009), but estimating the parameters is extremely difficult. The proposed mixture of Weibull and Pareto (IV) is computationally easier and conceptually it makes sense since if there are two types of claims with one type fitting the Weibull and the other type fitting the Pareto (IV), the mixture model is an appropriate model. We will also demonstrate that truncated Weibull-Pareto is a special case of mixture model. Suppose we have the following model $f(x) = pf_1(x) + (1-p)f_2(x)$, where $f_i(x), i = 1,2$ are densities. Clearly, p and $1-p$ are mixture weights ($0 < p < 1$), and $f(x)$ is indeed a valid density.

$$f_1(x) = \frac{\alpha}{\sigma\delta}\left(\frac{x}{\sigma}\right)^{1/\delta-1}\left[1+\left(\frac{x}{\sigma}\right)^{1/\delta}\right]^{-(\alpha+1)}, x > 0, \text{ and}$$

$$f_2(x) = \frac{\beta}{\gamma}\left(\frac{x}{\gamma}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\gamma}\right)^{\beta}\right], \quad x \geq 0.$$

Without loss of generality we consider the location parameters to be zero. So, our weighted distribution will have density

$$f(x) = p\frac{\alpha}{\sigma\delta}\left(\frac{x}{\sigma}\right)^{1/\delta-1}\left[1+\left(\frac{x}{\sigma}\right)^{1/\delta}\right]^{-(\alpha+1)} + (1-p)\frac{\beta}{\gamma}\left(\frac{x}{\gamma}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\gamma}\right)^{\beta-1}\right]. \tag{1}$$

The density in (1) is called the mixture of Weibull and Pareto (IV) (hereafter MWP(IV) in short) distribution.

2 Properties of the MWP(IV) distribution

The hazard function associated with the MWP(IV) distribution is

$$\begin{aligned} h_f(x) &= \frac{f(x)}{1\text{-}F(x)} \\ &= \left\{p\frac{\alpha}{\sigma\delta}\left(\frac{x}{\sigma}\right)^{1/\delta-1}\left[1+\left(\frac{x}{\sigma}\right)^{1/\delta}\right]^{-(\alpha+1)}\right. \\ &\qquad \left. + (1-p)\frac{\beta}{\gamma}\left(\frac{x}{\gamma}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\gamma}\right)^{\beta}\right]\right\} \times \left\{p\left[1+\left(\frac{x}{\sigma}\right)^{1/\delta}\right]^{-\alpha}\right. \\ &\qquad \left. + (1-p)\left\{\exp\left[-\left(\frac{x}{\gamma}\right)^{\beta}\right]\right\}\right\}^{-1} \end{aligned} \tag{2}$$

The limiting behaviors of the pdf (probability density function) and the hazard function of MWP(IV) are given in the following theorem.

Theorem 1. The limit of the pdf and the hazard function of MWP(IV) as $x \to \infty$ is 0 and the limit as $x \to 0^+$ is given by

$$\lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} h_f(x) = \begin{cases} 0, & \delta > 0, \quad \beta > 1 \\ \infty, & \delta < 0\,, \quad \beta < 1 \end{cases} \tag{3}$$

Proof. The fact $x > 0$ and $\int_0^\infty f(x)dx = 1$, imply that $\lim_{x\to\infty} h_f(x) = 0$, L'Hôpital's rule implies $\lim_{x\to\infty} h_f(x) = -\lim_{x\to\infty}\frac{\partial}{\partial x}(\log f(x))$ which can be shown to be equal to zero.

For $\lim_{x\to 0+} h_f(x) = \lim_{x\to 0+} f(x)$, the result in (3) follows directly from the de_nition of the hazard function. □

Let X be a non-negative random variable with pdf, $f(x)$ given by (1). The reliability function $S(x)$ corresponding to the finite mixture of 2 components of (1) is given by

$$S(x) = pS_1(x) + (1-p)S_2(x)$$

$$= p\left[1+\left(\frac{x}{\sigma}\right)^{1/\delta}\right]^{-\alpha} + (1-p)\exp\left[-\left(\frac{x}{\gamma}\right)^{\beta}\right],$$

where $S_j(x)$, is the reliability function corresponding to the j-th component in the mixture, $j = 1, 2$. One can write the hazard rate function of a mixture in terms of the hazard rate functions of the two components as follows.

$h_f(x) = B(x)h_1(x) + [1 - B(x)]h_2(x)$, where $B(x) = \frac{pS_1(x)}{pS_1(x)+1(1-p)S_2(x)}$ and $h_j(x)$ are hazard rate function for the j-th component, $j = 1,2$.

On differentiating the hazard function, we get

$$h_f'(x) = B(x)h_1'(x) + \big(1 - B(x)\big)h_2'(x) - B(x)\big(1 - B(x)\big)\big(h_1(x) - h_2(x)\big)^2,$$

where prime denote the derivatives with respect to x.

Now, from the above, it follows that if $h_j'(x) < 0$, for all x, ($j = 1;\ 2$), then $h'(x) < 0$, for all x. Therefore, a mixture with decreasing hazard rate components has decreasing hazard rate. However, if the components have increasing hazard rates, their mixture need not have increasing hazard rate.

2.1 Reliability parameter

The reliability parameter R is defined by $R = P(X > Y)$, where X and Y are independent random variables. Numerous applications of the reliability parameter have appeared in the liter-ature such as the area of classical stress-strength model and the break down of a system having two components. Other applications of the reliability parameter can be found in Hall (1984) and Weerahandi and Johnson (1992).

If X and Y are two continuous and independent random variables with the cdf (cumulative distri-bution function) $F_1(x)$ and $F_2(y)$ and their pdfs $f_1(x)$ and $f_2(y)$ respectively, then the reliability parameter R can be written as

$$R = P(Y > X) = \int_{-\infty}^{\infty} F_1(t) f_2(t)\, dt \tag{4}$$

Theorem 1. Suppose that $X \sim \mathrm{MWP(IV)}(\alpha, \sigma_1, \delta_1, \gamma_1, \beta)$ and $Y \sim \mathrm{MWP(IV)}(\alpha, \sigma_2, \delta_2, \gamma_2, \beta)$, then

$$P(Y > X) = p^2(1 - A_1) + p(1-p)(2 - A_2 - A_3) + (1-p)^2(1 - A_4), \tag{5}$$

where

$$A_1 = \alpha \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} \left(\frac{\sigma_1}{\sigma_2}\right)^{j/\delta_2} B(j/\delta_2 + 1, \alpha - j/\delta_2)$$

$$A_2 = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{\sigma_1}{\gamma_2}\right)^{j\beta} B(j\delta_1\beta + 1, \alpha - j\delta_1\beta),$$

$$A_3 = \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} \left(\frac{\gamma_1}{\sigma_2}\right)^{j/\delta_2} \Gamma\left(\frac{j}{\delta_2 \beta} + 1\right),$$

$$A_4 = \left[1 + \left(\frac{\gamma_1}{\gamma_2}\right)^{\beta}\right]^{-1}.$$

Proof: From (3) and (4), we have

$$\begin{aligned} P(Y > X) &= \int_0^{\infty} \left\{ p \frac{\alpha}{\sigma_1 \delta_1} \left(\frac{t}{\sigma_1}\right)^{1/\delta_1 - 1} \left[1 + \left(\frac{t}{\sigma_1}\right)^{1/\delta_1}\right]^{-(\alpha+1)} + (1-p) \frac{\beta}{\gamma} \left(\frac{t}{\gamma_1}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\gamma_1}\right)^{\beta}\right] \right\} \\ &\quad \times \left\{ p \left\{ 1 - \left[1 + \left(\frac{t}{\sigma_2}\right)^{1/\delta_2}\right]^{-\alpha} \right\} + (1-p) \left\{ 1 - \exp\left[-\left(\frac{t}{\gamma_2}\right)^{\beta}\right] \right\} \right\} dt \\ &= I_1 + I_2 + I_3 + I_4, \end{aligned} \tag{6}$$

where

$$\begin{aligned} I_1 &= p^2 \int_0^{\infty} \left(\frac{\alpha}{\sigma_1 \delta_1}\right) \left(\frac{t}{\sigma_1}\right)^{1/\delta_1 - 1} \left(1 + \left(\frac{t}{\sigma_1}\right)^{1/\delta_1}\right)^{-(\alpha+1)} \left\{ p \left\{ 1 - \left[1 + \left(\frac{t}{\sigma_2}\right)^{1/\delta_2}\right]^{-\alpha} \right\} \right\} dt \\ &= p^2 (1 - I_{12}) \end{aligned} \tag{7}$$

where

$$\begin{aligned} I_{12} &= \int_0^{\infty} \left\{ \frac{\alpha}{\sigma_1 \delta_1} \left(\frac{t}{\sigma_1}\right)^{1/\delta_1 - 1} \left[1 + \left(\frac{t}{\sigma_1}\right)^{1/\delta_1}\right]^{-(\alpha+1)} \left[1 + \left(\frac{t}{\sigma_2}\right)^{1/\delta_2}\right]^{-\alpha} \right\} dt \\ &= \alpha \int_0^{\infty} (1+u)^{-(\alpha+1)} \left[1 + \left(\sigma_1 \frac{u^{\delta_1}}{\sigma_2}\right)^{1/\delta_2}\right]^{-\alpha} du, \;\; u = \left(\frac{t}{\sigma_1}\right)^{1/\delta_1} \\ &= \alpha \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} \left(\frac{\sigma_1}{\sigma_2}\right)^{j/\delta_2} \int_0^{\infty} (1+u)^{-(\alpha+1)} u^{j\delta_1/\delta_2} du \\ &= \alpha \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} \left(\frac{\sigma_1}{\sigma_2}\right)^{j/\delta_2} B(j/\delta_2 + 1, \alpha - j/\delta_2) \\ &= A_1, \end{aligned} \tag{8}$$

using generalized binomial expansion.

So, $I_1 = p^2(1 - A_1)$. Following similar logic other terms can be easily evaluated. Hence the proof. □

2.2 Moments

For any $r \geq 1$,

$$E(X^r) = p\int_0^\infty x^r f_1(x)\,dx + (1-p)\int_0^\infty x^r f_2(x)dx$$
$$= pJ_1 + (1-p)J_2 \tag{9}$$

where

$$J_1 = \int_0^\infty x^r f_1(x)\,dx$$
$$= \int_0^\infty x^r \frac{\alpha}{\sigma\delta}\left(\frac{x}{\sigma}\right)^{1/\delta-1}\left[1+\left(\frac{x}{\sigma}\right)^{1/\delta}\right]^{-(\alpha+1)} dx$$
$$= \alpha\sigma^r B(r\delta+1, \alpha - r\delta). \tag{10}$$

and

$$J_2 = \int_0^\infty x^r f_2(x)\,dx$$
$$= \int_0^\infty x^r \frac{\beta}{\gamma}\left(\frac{x}{\gamma}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\gamma}\right)^{\beta}\right] dx$$
$$= \gamma^r \Gamma\left(\frac{r}{\beta}+1\right), \tag{11}$$

Substituting (15) and (16) in (14) we get,

$$E(X^r) = p\alpha\sigma^r B(r\delta+1, \alpha - r\delta) + (1-p)\gamma^r\Gamma\left(\frac{r}{\beta}+1\right).$$

Skewness & Kurtosis plays a vital role in explaining the shape and tail property of a distribution. The expressions of Skewness and Kurtosis (in terms of non-central moments) are given by

- Skewness: $\theta_1 = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$,
- Kurtosis: $\theta_2 = \frac{\mu^4}{\sigma^4}$

where $\sigma^2 = E(X^2) - \mu^2$ and $\mu_4 = E(X^4 - 4\mu^3 + 6\mu^2 E(X^2) - 4\mu E(X^3) + \mu^4$.

In our case, from the general expression of moments, we have the following:

- $E(X) = p\alpha\sigma^1(\delta+1, \alpha-\delta) + (1-p)\gamma^1\Gamma\left(\frac{1}{\beta}+1\right)$, exists iff $\alpha > \delta$.
- $E(X^2) = p\alpha\sigma^2 B(2\delta+1, \alpha-2\delta) + (1-p)\gamma^2\Gamma\left(\frac{2}{\beta}+1\right)$, exists iff $\alpha > 2\delta$.
- $E(X^3) = p\alpha\sigma^3(3\delta+1, \alpha-3\delta) + (1-p)\gamma^3\Gamma\left(\frac{3}{\beta}+1\right)$, exists iff $\alpha > 3\delta$.
- $E(X^4) = p\alpha\sigma^4(4\delta+1, \alpha-4\delta) + (1-p)\gamma^4\Gamma\left(\frac{4}{\beta}+1\right)$, exists iff $\alpha > 4\delta$.

Hence, on substitution of these quantities, one can get exact expressions for skewness & kurtosis respectively. It is to be noted that, for skewness, the index of inequality parameter, α has to be bigger than 3δ, while for kurtosis measure, we need to have $\alpha > 4\delta$ (for the existence of fourth order non central moment). This is quite a strong assumption, in the sense that, in actual data, we do not have any prior information as to whether such restrictions are valid or not. Alternatively, one may consider quantile measure of skewness & kurtosis, but that too in this case will be difficult to obtain, because of the unavailability of a proper quantile function.

2.3 Mean Deviation

The deviation from the mean and the deviation from the median are used to measure the dispersion and the spread in a population from the center. If we denote the median by M, then the mean deviation from the mean, $D(\mu)$, and the mean deviation from the median, $D(M)$, can be written as

$$\begin{aligned} D(\mu) &= E|X-\mu| \\ &= \int_{-\infty}^{\infty} |x-\mu| f(x)dx \\ &= \int_0^{\mu} (\mu - x) f(x)dx + \int_{\mu}^{\infty} (x-\mu) f(x)dx \\ &= \mu\left(\int_0^{\mu} f(x)dx - \int_{\mu}^{\infty} f(x)dx\right) - \int_0^{\mu} xf(x)dx + \int_{\mu}^{\infty} xf(x)dx \\ &= (K_1 - K_2) - K_3 + K_4, \end{aligned} \tag{12}$$

$$\begin{aligned} K_1 &= \int_0^{\mu} f(x)dx \\ &= p\int_0^{\mu} f_1(x)dx + (1-p)\int_0^{\mu} f_2(x)dx \\ &= p\left\{1 - \left[1 + \left(\frac{\mu}{\sigma}\right)^{1/\delta}\right]^{-\alpha}\right\} + (1-p)\left\{1 - \exp\left[-\left(\frac{\mu}{\gamma}\right)^{\beta}\right]\right\}. \end{aligned} \tag{13}$$

Similarly, one can get

$$K_2 = p\left[1 + \left(\frac{\mu}{\sigma}\right)^{1/\delta}\right]^{-\alpha} + (1-p)\left\{\exp\left[-\left(\frac{\mu}{\gamma}\right)^{\beta}\right]\right\}, I_3 = pB_{\left(\frac{\mu}{\sigma}\right)^{\delta}}(\delta+1, \alpha-\delta) + (1-p)\Gamma_{\left(\frac{\mu}{\gamma}\right)^{\beta}}(1/\beta+1,)$$

where $B_x(a,b)$ and $\Gamma_x(a,b)$ are incomplete beta and gamma functions respectively.

Since, $K_4 = E(X) - \int_0^{\mu} xf(x)dx$, and $E(X) = p\alpha\sigma B(\delta+1, \alpha-\delta) + (1-p)(1/\beta+1)$, (on substituting $r=1$ in the general expression), we can write

$$D(\mu) = p\alpha\sigma B(\delta+1,\alpha-\delta) + (1-p)\gamma\Gamma(1/\beta+1) \times \left(p\left\{1-2\left[1+\left(\frac{\mu}{\sigma}\right)^{1/\delta}\right]^{-\alpha}\right\} + (1-p)\left\{1-2\exp\left[-\left(\frac{\mu}{\gamma}\right)^{\beta}\right]\right\}\right) + p\alpha\sigma B(\delta+1,\alpha-\delta) + (1-p)\gamma\Gamma(1/\beta+1) - 2\left[pB_{\left(\frac{\mu}{\sigma}\right)^{\delta}}(\delta+1,\alpha-\delta) + (1-p)\Gamma_{\left(\frac{\mu}{\gamma}\right)^{\beta}}(1/\beta+1)\right].$$

3 Maximum Likelihood Estimation

In this section we address the parameter estimation of the MWP(IV) distribution. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n drawn from the density in (1). Let $\theta_1 = (\alpha, \delta, \sigma)'$ and $\theta_2 = (\beta, \gamma)'$ denote the parameter vectors associated with $f_1(x)$ and $f_2(x)$, respectively. Furthermore, suppose, $\theta = (\theta_1, \theta_2, p)'$denotes the full parameter vector, to be estimated given the observed data $\mathrm{x} = (x_1, x_2, \cdots, x_n)'$. For clarity, we write $f_1(x)$ as $f_1(x|\theta_1)$, and $f_2(x)$ as $f_2(x|\theta_2)$.

Maximizing the likelihood function $L(\theta)$ directly is difficult as pointed out by Redner and Waiker (1984). We thus use EM algorithm (Dempster et al., 1977), considering $u_i(=0,1), i = 1,2,,n$ as missing value with ith observation x_i drawn from $f_1(x|\theta_1)$ if $u_i = 1$, and $f_2(x|\theta_2)$ if $u_i = 0$. Thus the complete likelihood function is given by

$$L_c(\theta) = \prod_{i=1}^{n}\left(f_1(x_i|\theta_1)\right)^{u_i}\left(f_2(x_i|\theta_2)\right)^{1-u_i} p^{u_i}(1-p)^{u_i}.$$

Denoting $\theta^{(k)}$as the k^{th}iterative solution, the E-step is given by

$$E\left[\log L_c(\theta)|x,\theta^{(k)}\right] = \sum_{i=1}^{n} E\left(u_i|x,\theta^{(k)}\right)\log f_1(x_i|\theta_2) + \sum_{i=1}^{n}\left[1-E\left(u_i|x,\theta^{(k)}\right)\right]\log f_2(x_i|\theta_2) + \sum_{i=1}^{n} E\left(u_i|x,\theta^{(k)}\right)\log p + \sum_{i=1}^{n}\left[1-E\left(u_i|x,\theta^{(k)}\right)\right]\log(1-p) \qquad (14)$$

Note that

$$E\left(u_i|x,\theta^{(k)}\right) = \frac{p^{(k)} f_1\left(x_i|\theta^{(1)}\right)}{p^{(k)} f_1\left(x_i|\theta^{(k)}\right) + (1-p^{(k)}) f_2\left(x_i|\theta^{(k)}\right)} = p_i^{(k)} \text{ say.} \qquad (15)$$

Now following the M-step, after differentiating (19) with respect to p and equating it to 0, we get the updated estimate of p as

$$p^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} p_i^{(k)}.$$

(16)

Similarly, differentiating (19) with respect to $\theta_1 = (\alpha, \delta, \sigma)'$ and $\theta_2 = (\beta, \gamma)'$, and equating the derivatives to 0, we get the updated estimates of $\left(\alpha^{(k+1)}, \delta^{(k+1)}, \sigma^{(k+1)}\right)'$ and $\left(\beta^{(k+1)}, \gamma^{(k+1)}\right)$ as the solution of

$$\alpha = \frac{\sum_{i=1}^{n} p_i^{(k)}}{\sum_{i=1}^{n} p_i^{(k)} \log\left[1 + \left(\frac{x_i}{\sigma}\right)^{1/\delta}\right]},$$

$$\alpha = \frac{\sum_{i=1}^{n} p_i^{(k)} \left[1 + \left(\frac{x_i}{\sigma}\right)^{1/\delta}\right]^{-1}}{\sum_{i=1}^{n} p_i^{(k)} \left(\frac{x_i}{\sigma}\right)^{1/\delta} \left[1 + \left(\frac{x_i}{\sigma}\right)^{1/\delta}\right]^{-1}}$$

$$\delta \sum_{i=1}^{n} p_i^{(k)} = \alpha \sum_{i=1}^{n} p_i^{(k)} \log\left(\frac{x_i}{\sigma}\right) - (\alpha + 1) \sum_{i=1}^{n} p_i^{(k)} \log\left(\frac{x_i}{\sigma}\right) \left[1 + \left(\frac{x_i}{\sigma}\right)^{1/\delta}\right]^{-1},$$

$$\gamma^{\beta} = \frac{\sum_{i=1}^{n} \left(1 - p_i^{(k)}\right) x_i^{\beta}}{\sum_{i=1}^{n} \left(1 - p_i^{(k)}\right)},$$

$$\beta^{-1} = \frac{\sum_{i=1}^{n} \left(1 - p_i^{(k)}\right) \log(x_i) x_i^{\beta}}{\sum_{i=1}^{n} \left(1 - p_i^{(k)}\right) x_i^{\beta}} - \frac{\sum_{i=1}^{n} \left(1 - p_i^{(k)}\right) \log(x_i)}{\sum_{i=1}^{n} \left(1 - p_i^{(k)}\right)}.$$

(17) (18) (19) (20) (21)

The estimates can be obtained by the following algorithm:

The initial estimates are obtained as follows: Obtain$\left(\alpha^{(0)}, \delta^{(0)}, \sigma^{(0)}\right)$ assuming that the data is generated from $f_1(x|\theta_1)$; obtain $\left(\beta^{(0)}, \gamma^{(0)}\right)$ assuming that the data in generated from $f_2(x|\theta_2)$. For the initial estimates of p, prior information such as proportion of low claims, or in the absence of such information, $p^{(0)} = 1/2$can be used. Estimate $p_i^{(0)}$ equation (20).

Suppose the estimates at the k^{th} iteration are $p^{(k)}$, $\left(\hat{\alpha}^{(k)}, \hat{\delta}^{(k)}, \hat{\sigma}^{(k)}\right)$, and $\left(\hat{\beta}^{(k)}, \hat{\gamma}^{(k)}\right)$. Estimate $p_i^{(k)}$ using (20). Update $p^{(k+1)}$ from (21). The updated solution $\left(\hat{\alpha}^{(k+1)}, \hat{\delta}^{(k+1)}, \hat{\sigma}^{(k+1)}\right)$ is obtained by solving equations (22)-(24), and $\left(\hat{\beta}^{(k+1)}, \hat{\gamma}^{(k+1)}\right)$ is obtained by solving (25) and (26).

To solve (22)- (24), we consider the following steps:

- Step 1: Start with a choice of δ (say, $\delta^{(k)}$), _nd σ so that the values of α given by (22) and (23) are equal.
- Step 2: Solve for α using (22).
- Step 3: Solve δ using (24).
- Step 4: For using this new δ, go back to Step 1. Continue until the iterative sequence of δconverges.

To solve (25) and (26), start with a given choice of β (say, $\beta^{(k)}$), and solve β iteratively from (26). Then estimate γ from (25).

4 Numerical Results

In this section, we present a simulation result, and use a real life data to fit the MWP(IV) distribu-tion. Note that the mixture distribution MWP(IV) is applicable in situations when the population distribution is a mixture of low values and high values. Relatively, the Weibull component of MWP(IV) describes the distribution of low values, and the Pareto (IV) component describes the distribution of high values. For example, if one considers income inequality in a particular coun-try, the MWP(IV) distribution can be used to describe the income distribution of the country's population, where the Weibull component describes the distribution of low income group and the Pareto (IV) component describes the distribution of high income group.

4.1 Simulation study

The mixing parameter p can play an important role in describing the inequality between high and low values such as income inequality in the above example. Thus, in our simulations, we compare the results with several p values. We considered nine set of values, $p = 0.1 \cdots 0.9$. In each case, sample size of n = 1000 is used with 100 simulation runs. The rest of the parameters are set at the following values: $\alpha = 2, \delta = 0.2, \sigma = 1.5, \beta = 0.8$, and $\gamma = 1.2$. We fit the simulated data by (1) MWP(IV) distribution, (2) Weibull distribution, and (3) Pareto (IV) distribution. We use the AIC criteria (Akaike, 1974) as a measure of goodness of fit. The Figure 1(A) shows AIC values for all simulation runs with Figure 1(B) showing the fits for one particular run. It is clear from the Figure 1(A) that, when the mixing parameter p is high (≥ 0.5), the fit by Weibull distribution alone and the fit by the Pareto distribution alone are very poor. Thus, as in many practical cases, when the low values of the population is in higher proportion in comparison to high values, the Weibull distribution and the Pareto (IV) distribution will show very poor fits to the real data. Figure 1(B) shows how Weibull distribution and the Pareto (IV) distribution look way off from a simulated data when $p = 0.7$. The average AIC values with standard errors are given in Table 1, which also shows high discrepancy for high $p(\geq 0.5)$, when comparing the average AIC under Weibull distribution and under Pareto (IV) distribution against the true average AIC (under the MWP(IV) distribution). In Table 1, we also included the fits from mixtures of two gaussian (Mix-Normal fit) and two Gamma (Mix-Gamma fit) distributions obtained from 'mixtools' package in R (Benaglia et al. 2006).

Table 1: Outcome of the simulation study.

AIC	**Pareto(**	**IV) fit**	**Weibull fit**	**MWP(IV) fit**		**Mix-Normal fit**	**Mix-Gamma fit**
p=0.1	2646.15	(7.03)	2641.58 (7.13)	2538.28	(7.21)	3041.42 (7.03)	2641.86 (7.19)
p=0.2	2602.64	(6.93)	2594.20 (7.51)	2513.09	(7.73)	2936.07 (7.86)	2577.50 (8.39)
p=0.3	2575.02	(6.47)	2577.15 (6.61)	2456.81	(6.49)	2815.93 (6.57)	2513.43 (8.96)
p=0.4	2554.05	(5.70)	2565.37 (5.94)	2359.14	(6.25)	2688.01(11.57)	2419.06(11.11)
p=0.5	2493.81	(5.90)	2515.43 (6.76)	2224.88	(6.32)	2486.20 (9.92)	2293.28(12.83)
p=0.6	2399.24	(6.75)	2445.87 (7.82)	2060.05	(6.83)	2271.89 (7.89)	2126.28 (14.8)
p=0.7	2237.03	(6.36)	2325.94 (7.37)	1853.16	(5.55)	2020.86 (5.88)	1897.53(13.48)
p=0.8	1964.57	(6.77)	2121.58 (8.97)	1604.58	(7.02)	1713.94 (5.62)	1640.01 (9.01)
p=0.9	1535.50	(7.13)	1809.22(11.11)	1353.60	(8.47)	1392.84 (5.56)	1384.55 (6.40)

4.2 Real data application

We now consider a real data set from Brinbaum and Saunders (1969). The data set represents the fatigue life of 6061-Tg aluminum coupons. We _t the data by Pareto (IV) $(\alpha;\ \delta;\ \sigma)$, Weibull $(\beta;\ \gamma)$, Mix-Normal $(p, \mu_1, \sigma_1, \mu_1, \sigma_2)$, Mix-Gamma $(p, \alpha_1, \beta_1, \alpha_2, \beta_2)$, and MWP (IV) $(p, \alpha, \delta, \sigma, \beta, \gamma)$ distributions. The maximum likelihood estimates are given as follows:

- Pareto(IV) fit: $\hat{\alpha} = 1.37, \hat{\delta} = 0.11$, and $\hat{\sigma} = 138.02$.
- Weibull fit: $\hat{\beta} = 5.79$, and $\hat{\gamma} = 142.84$.

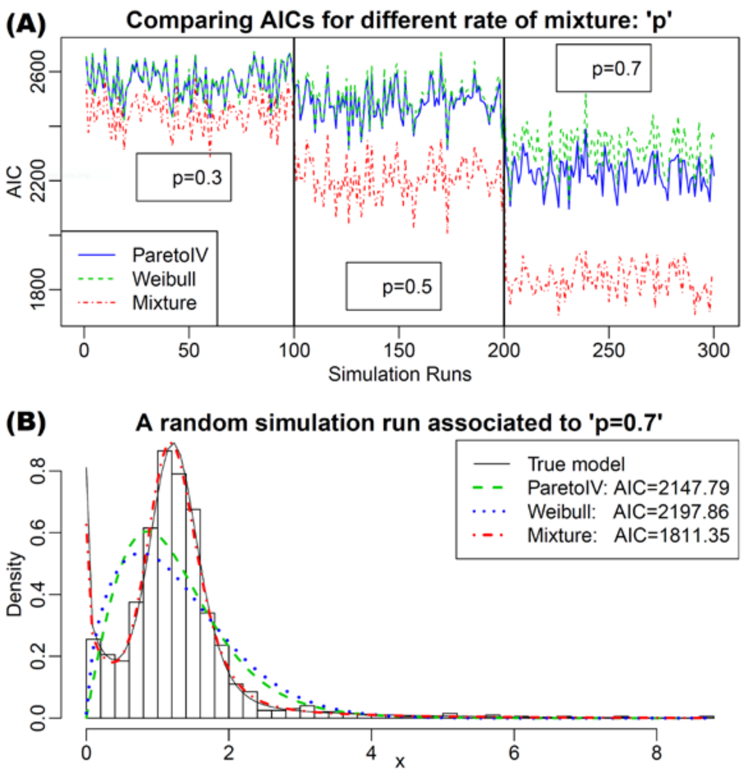

Figure 1: (A) Compares the AIC for 300 simulation runs associated to $p = 0.3$, 0.5 and 0.7. (B) For a randomly selected run, comparing the densities obtained from each model.

- Mix-Normal _t: $\hat{p} = 0.94, \hat{\mu}_1 = 131.36, \hat{\sigma}_1 = 22.92$ $\hat{\mu}_2 = 161.15$, and $\hat{\sigma}_2 = 4.46$.
- Mix-Gamma _t: $\hat{p} = 0.85, = \hat{\alpha}_1 = 42.76, \hat{\beta}_1 = 3.10, \hat{\alpha}_2 = 13.82$, and $\hat{\beta}_2 = 9.74$.
- MWP(IV) _t: $\hat{p} = 0.83, \hat{\alpha} = 0.94, \hat{\delta} = 0.09, \hat{\sigma} = 125.48, \hat{\beta} = 17.99$, and $\hat{\gamma} = 160.29$.

The estimates for MWP (IV) are obtained by the EM algorithm as described in Section 3. The maximum likelihood estimates for Pareto (IV) are obtained from the R package VGAM (vector generalized linear and additive models) (Yee, 2015). Figure 2 shows the fits of all three distributions together with histogram of the data. The data seems to be bi-model, thus a mixture distribution could be considered. Although the AIC criteria show a better fit for Pareto (IV), the MWP(IV) seems to be a better fit in terms of the bi-model structure of the data. Also note that although AIC of MWP(IV) is higher than the AIC of Pareto, they are nearly equal. Finally, with respect to AIC, MWP(IV) is better than the other two mixture models.

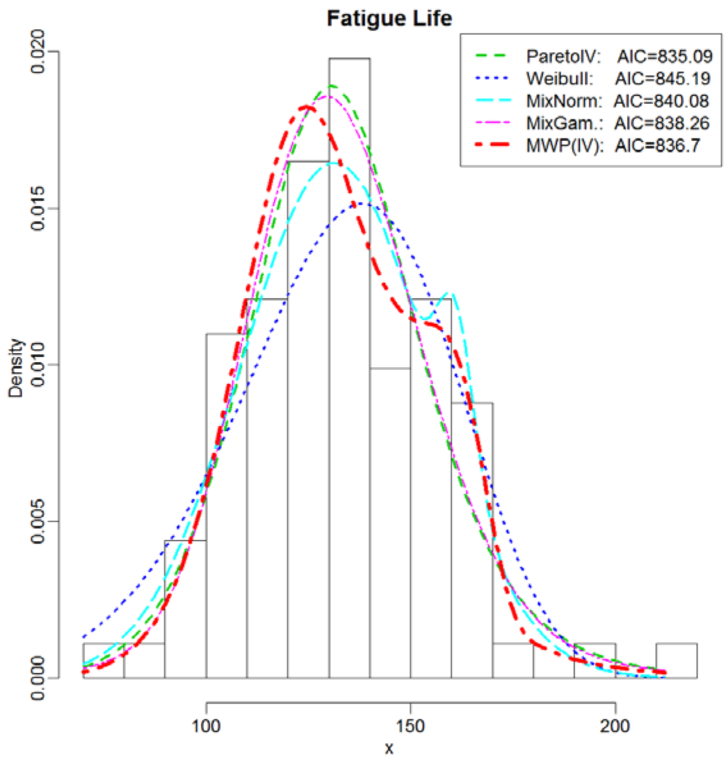

Figure 2: ML fits to the fatigue life data.

5 Conclusion

In this paper, we consider a simple mixture of two absolutely continuous distributions—Weibull and Pareto (IV) distribution. Some structural properties of the resulting distribution are discussed. The resulting model appears to be a reasonable choice in the sense of modeling insurance claims, in particular, where the popular choices (e.g., Weibull and/or Paretian distribution) fail to adequately model the observed phenomena. The highlight of this article lies in the fact that we have discussed in detail the problem of efficient estimation of the model parameters under maximum likelihood and how to efficiently induce the idea of EM algorithm. We sincerely hope that this particular mixture model will find many more applications in different spheres affecting human life.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716-723

[2] Benaglia, T., Chauveau, D., Hunter, D. R., & Young D. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software, 32(6), 1-29.

[3] Birnbaum, ZW, Saunders, SC (1969). A new family of life distributions. Journal of Applied probability, 6, 319-327.

[4] Bohning, D. (2000). Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others, Chapman & Hall, London.

[5] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1-38.

[6] Hall, I.J. (1984). Approximate one-sided tolerance limits for the difference or sum of two independent normal variates. Journal of Qualitative Technology, 16,15-19.

[7] Lindsay, B.G. (1995). Mixture Models. theory, geometry and applications. NSF-CBMS Regional Conference Series in Probability and Statistics, 5, Alexandria, Virginia: Institute of Mathematical Statistics and the American Statistical Association.

[8] McLachlan, G.J., Peel, D. (2000). Finite Mixture Models, Wiley, New York.
[9] Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. American Journal of Mathematics, 8, 343-366.
[10] Pearson, K. (1894). Contributions to the mathematical theory of evolution. Phil. Trans. A, 185, 7.
[11] Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. SIAM review, 26,195-239.
[12] Teodorescu, S. & Panattescu, E. (2009). On the truncated composite Weibull- Pareto model.Mathematical Reports- Bucharest, 11, 259-273.
[13] Titterington, D.M., Smith A.F.M, Markov, U.E. (1985). Statistical Analysis of Finite Mixture Distributions, Wiley, New York.
[14] Weldon, W.F.R. (1892). Certain correlated variations in Crangon Vulgaris. Proceedings of the Royal Society of London, 51, 2-21.
[15] Weldon, W.F.R. (1893). On certain correlated variations in carcinus moenas. Proceedings of the Royal Society of London, 54, 318-329.
[16] Weerahandi, S. and Johnson, R.A. (1992). Testing reliability in a stress-strength model when X and Y are normally distributed. Technometrics, 38, 83-91.
[17] Yee, W. Thomas. (2015). Vector Generalized Linear and Additive Models: With an Implementation in R. New York, USA: Springer.