Statistical Inferences on Different Types of Bivariate Pareto Distributions

by

BIPLAB PAUL

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI-781039, INDIA

April, 2019

Statistical Inferences on Different Types of Bivariate Pareto Distributions

A thesis submitted

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

by

BIPLAB PAUL

(Roll Number: 156123014)

to the

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

April, 2019

Declaration

I hereby declare that the work contained in this thesis entitled **"Statistical Inferences on Different Types of Bivariate Pareto Distributions"** was done by me, under the supervision of **Dr. Arabin Kumar Dey**, Assistant Professor, Department of Mathematics, Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy and this work has not been submitted elsewhere for a degree.

April, 2019

Biplab Paul
Roll No. 156123014
Department of Mathematics
Indian Institute of Technology Guwahati

Certificate

It is certified that the work contained in this thesis entitled **"Statistical Inferences on Different Types of Bivariate Pareto Distributions"** by **Biplab Paul,** a student in Department of Mathematics, Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

April, 2019

Dr. Arabin Kumar Dey
Assistant Professor
Department of Mathematics
Indian Institute of Technology Guwahati

To

All of My Teachers

Acknowledgements

This thesis work is one of the most wonderful experiences to me. It is not only about writing a thesis and presenting it, but the process helps to understand that it is a fascinating continuous process which always tries to push the boundaries beyond its limit. I would like to express my sincere gratitude to my thesis supervisor Dr. Arabin Kumar Dey for his continuous support, motivation and immense knowledge.

I wish to express my sincere thanks to the Head of Department of Mathematics, IIT Guwahati, for providing the necessary facilities during my stay at the department. My gratitude also goes to all the faculty members of the Department of Mathematics who taught me invaluable knowledge.

Last but not least, I am grateful to my family and friends who have endlessly been supportive and kind.

IIT Guwahati Biplab Paul

Abstract

In this thesis work, we discuss different types of Marshal-Olkin form of bivariate Pareto distribution. We use EM algorithm and Bayesian estimation through slice cum Gibbs sampler to estimate the parameters. The thesis covers some innovative solutions for different problems related to implementation of EM algorithm and Bayesian estimation. We also analyze the interval estimation of the parameters both in frequentist and bayesian set up. Numerical results are shown to verify the performance of the algorithms. A real-life data analysis is also shown in each study for illustrative purposes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Thinking is the capital, Enterprise is the way, Hard work is the solution."

- Dr. A.P.J. Abdul Kalam

1.1 Motivation and Objective of the Thesis

The main objective of this thesis is to explore efficient estimation techniques through expectation and maximization algorithm (EM algorithm) of the parameters of different bivariate distributions. We consider a few higher dimensional distributions which are not well studied in the literature. We focus mainly on a few variants of the Marshall-Olkin bivariate Pareto in this thesis. Marginals of these bivariate distributions are heavy-tail in nature. Pareto distribution plays an important role in the Extreme Value Theory. These distributions have wide applications in finance, computer network, biological systems and many more areas of science and technology. We observe that the inclusion of location and scale parameters in such distributions complicates the estimation procedure. Also direct maximum likelihood estimate is difficult to handle and many times they can not be obtained in explicit form. They can only be obtained by solving a system of non-linear equations. In this situation, finding initial values and the convergence of the algorithm are important issues. To avoid this problem, we discuss the implementation of the EM algorithm in computing MLEs of unknown parameters.

Bayesian estimation of these higher dimensional distributions is another popular research direction. Bayesian analysis contains more information than maximum likelihood estimation, which is better for further statistical analysis. For instance, not only the mean, median or mode of posterior distribution can be computed, but also the performances of these estimators (through their variance and higher-order moments) are available. Maximum likelihood estimators can be quite unstable, i.e., they may vary widely for small variations of the observations. Moreover, the knowledge of the posterior distribution also allows for the derivation of confidence regions through highest posterior density (HPD) regions in both univariate and multivariate cases. Therefore, working on Bayesian set up with such a complicated distribution has its own advantages.

Given a higher dimensional data set, it is difficult to identify the parametric distribution of the data set. We can use graphical tools up to dimension two. But beyond dimension two, the problem is severe. Therefore, non-parametric approaches are more popular in higher dimensional case. In this connection we should mention some popular recent modeling techniques for higher dimensional data which is modeling through copula. However, we took up the challenges in parametric models, as somehow or whenever we can establish the assumption, any inference on parametric approach will beat non-parametric approaches. We discuss issues in validating the distributional assumption in this thesis, but mainly work on different statistical inferences when we already know the data follows a fixed parametric distribution. Recently copula techniques for modeling higher dimensional distributions has become another interesting domain. Finding an efficient technique for estimation of parameters in this set up opens a different research direction.

In this thesis we consider mainly four types of bivariate distributions and they are : Marshall Olkin bivariate Pareto with location and scale parameters, Absolute continuous Marshall-Olkin bivariate Pareto, Geometric Marshall-Olkin bivariate Pareto, Absolute Continuous Geometric Marshall-Olkin bivariate Pareto. In Bayesian estimation we use Lindley approximation, Slice sampling, Gibbs sampler etc. A short description of the tools and techniques used throughout this thesis is discussed in the next section.

1.2 Description of Tools and Techniques

In this section we introduce some technique for estimating parameters. First we discuss about point estimation procedure like Maximum likelihood estimation and Bayesian estimation, then interval estimation using bootstrap technique.

1.2.1 Maximum Likelihood Estimation

From theoretical point of view, the most general method of estimation is the method of maximum likelihood estimations (MLE) which was initially formulated by C. F. Gauss but as a general method of estimation was first introduced by Prof. R. A. Fisher and later on developed by him in a series of papers. Before introducing the method we will define likelihood function.

Let $\mathbf{x}=\{x_1, x_2, \cdots, x_n\}$ is a sample drawn from a population with density function $f(x;\theta)$ or $f(x|\theta)$. Then the likelihood function of the given sample is denoted as $L = L(\theta|\mathbf{x})$ defined as

$$L(\theta|\mathbf{x}) = f(x_1|\theta)f(x_2|\theta)f(x_3|\theta)\cdots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Then the maximum likelihood estimates of θ can be written as,

$$\hat{\theta} = \arg\ \max_\theta \log(L(\theta|\mathbf{x}))$$

since $L(\theta|\mathbf{x}) > 0$, and $\log(L(\theta|\mathbf{x}))$ is non-decreasing function of $L(\theta|\mathbf{x}) > 0$. So both of $L(\theta|\mathbf{x}) > 0$ and $\log(L(\theta|\mathbf{x}))$ attain their maximum value at $\hat{\theta}$.

Some of the cases we are unable to obtain the MLEs of parameters in explicit form. They can be obtained only by solving a system of non-linear equations of unknown parameters. We treat this problem as a missing value problem. To solve this type of problem we use EM Algorithm.

EM Algorithm: EM (Expectation- Maximization) [12] algorithm was introduced by Dempster et al. in 1977. It is one of the most commonly used algorithm in statistic.

Suppose $\mathbf{X} = \{X_1, X_2, \cdots, X_n\}$ is a random sample of size n taken from the distribution with pdf $g(x|\theta)$ and we would like to compute the maximum likelihood estimator of θ i.e. $\hat{\theta}=$ arg max $L(\theta|\mathbf{x})$. We augment the data with $\mathbf{z}$, where $\mathbf{X}$, $\mathbf{Z}\sim f(\mathbf{x}, \mathbf{z}|\theta)$ to avoid the difficulties in maximizing likelihood function. Here $\mathbf{Z}$ is known as hidden or missing variable.

Suppose $\theta^{(0)}$ be the initial value for the parameter θ and want to find θ such that $L(\theta|\mathbf{x}) > L(\theta^0|\mathbf{x})$. This can be done by considering the difference between observed-data log-likelihood

$$\Delta \log L = logL(\theta|\mathbf{x}) - logL(\theta^{(0)}|\mathbf{x}) = log(\frac{L(\theta|\mathbf{x})}{L(\theta^{(0)}|\mathbf{x})}). \tag{1.2.1}$$

Now we introduce the hidden variable Z, then (1.2.1) can be written as

$$\begin{aligned}
logL(\theta|\mathbf{x}) - logL(\theta^{(0)}|\mathbf{x}) &= log\frac{\int f(\mathbf{x}, \mathbf{z}|\theta)d\mathbf{z}}{f(\mathbf{x}|\theta^{(0)})} \\
&= log\frac{\int f(\mathbf{z}|\theta^{(0)}, \mathbf{x})f(\mathbf{x}, \mathbf{z}|\theta)}{f(\mathbf{z}|\theta^{(0)}, \mathbf{x})f(\mathbf{x}|\theta^{(0)})}d\mathbf{z} \\
&\geq \int f(\mathbf{z}|\theta^{(0)}, \mathbf{x})log\frac{f(\mathbf{x}, \mathbf{z}|\theta)}{f(\mathbf{z}|\theta^{(0)}, \mathbf{x})f(\mathbf{x}|\theta^{(0)})}d\mathbf{z} \\
&= \Delta \log L(\theta, \theta^{(0)}),
\end{aligned}$$

where the last inequality is due to Jensens inequality and the fact that $\log(\cdot)$ is concave. We can maximize $\Delta \log L(\theta, \theta^{(0)})$ w.r.t. θ to obtain a sequence $\{\theta^i\}$(every time $\theta^{(i)}$ replaced by $\theta^{(i+1)}$) and as long as $\log L(\theta^{(i+1)}, \theta^{(i)}) \geq 0$, where

$$\begin{aligned}
\theta^{(i+1)} &= \arg\ \max_\theta \Delta \log L(\theta, \theta^{(i)}) \quad \text{where i= 0,1,2,3,}\cdots \\
&= \arg\ \max_\theta \int f(\mathbf{z}|\theta^{(i)}, \mathbf{x})log\frac{f(\mathbf{x}, \mathbf{z}|\theta)}{f(\mathbf{z}|\theta^{(i)}, \mathbf{x})f(\mathbf{x}|\theta^{(i)})}d\mathbf{z} \\
&= \arg\ \max_\theta \int f(\mathbf{z}|\theta^{(i)}, \mathbf{x})log f(\mathbf{x}, \mathbf{z}|\theta)d\mathbf{z}.
\end{aligned}$$

Let us define

$$\begin{aligned} Q(\theta,\theta^{(i)}) &= f(\mathbf{z}|\theta^{(i)},\mathbf{x}) log f(\mathbf{x},\mathbf{z}|\theta) \\ &= E_{\theta^{(i)}|\theta,\mathbf{x}}[log L(\theta|\mathbf{X},\mathbf{Z})]. \end{aligned}$$

So it is clear that the EM algorithm has two steps

(i)**E-step:**Compute $Q(\theta,\theta^{(i)})$, which is the expectation of complete-data log-likelihood logL($\theta|\mathbf{X},\mathbf{Z}$) and expectation is w.r.t conditional pdf of missing data $\mathbf{Z}$, $f(\mathbf{z}|\theta^{(i)},\mathbf{x})$.

(ii)**M-step:** Maximize $Q(\theta,\theta^{(i)})$ w.r.t θ to obtain $\theta^{(i+1)}$.

1.2.2 Bayesian Estimation

In Bayesian statistics the parameters are treated as random variable and unknown to researcher. The researcher has some initial ideas about the value of these parameters and collects data to improve this understanding. The distribution of this parameter based on those initial ideas is known as prior distribution. Let $f(x|\theta)$ be the density function of the population. Here our aim to find a point estimate of the parameter θ. Let us denote $\pi(\theta)$ as the prior density function of θ and $\mathbf{x}=(x_1,x_2,.......x_n)$ be a random sample of size n.

The distribution of θ, given sample $\mathbf{x}$, is called the posterior distribution which is defined as

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{g(\mathbf{x})}, \tag{1.2.2}$$

where $g(\mathbf{x})$ is the marginal probability distribution of $\mathbf{x}$, over θ.

For discrete case $g(\mathbf{x}) = \sum_\theta f(\mathbf{x}|\theta)\pi(\theta)$ and

for continuously case $g(\mathbf{x}) = \int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta$.

Actually $g(\mathbf{x})$ is a constant with respect to θ and $g(\mathbf{x})$ is also known as normalizing constant. Using these fact we say that posterior distribution is proportional to the prior distribution times the likelihood function:

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta). \tag{1.2.3}$$

Under the squared-error loss function the mean of the posterior distribution is called the Bayes estimate of θ and it expressed as

$$\hat{\theta} = \int \theta\pi(\theta|\mathbf{x})d\theta. \tag{1.2.4}$$

We can approximate the integral by Lindley approximation method [33]. Alternatively we can use a suitable Monte Carlo methods to find the Bayes estimate.

Description of the basics of a few widely used Monte Carlo techniques which are relevant for this thesis are described below :

Inverse Transform Algorithm: Let X be a continuous random variable with the cumulative distribution function F_X and $U \sim U(0,1)$, then probability integral transform theorem [49] implies that $F_X^{-1}(U) \sim F_X$. Corresponding algorithm steps are
Step-1: generate a random number U from $U(0,1)$.
Step-2: set $X = F_X^{-1}(U)$.

For practical purpose, this algorithm only applies when both of cdf and inverse of cdf are available. But this kind of situation only covers a small number of distributions (for example exponential distribution, Weibull distribution etc.). Other methods, like Accept/Reject method [49] are more general and do not have any such kind of problems.

Accept/Reject Algorithm: When we are unable to generate the desired random variable using direct transformation method, an extremely powerful method Accept/Reject algorithm can often provide a solution. Suppose our goal is to generate X from a distribution with the pdf $f_X(x)$ and we are able to generate V from a distribution with $f_V(v)$, where f_X and f_V have common support with

$$M = \sup_x \frac{f_X(x)}{f_V(x)} < \infty.$$

We can generate $X \sim f_X$ using the following steps,
Step-1: generate a random number U from $U(0,1)$ and $V \sim f_V$, independent.
Step-2: if $U < \frac{f_X(V)}{M f_V(V)}$, set $X = V$; otherwise return to step-1.

Here the density of X, f_X and the the density of V, f_V are known as target density and the candidate density respectively. Sometime it is observed the that the target density has heavy tails in nature, it is difficult to find the candidate density that will provide a finite value of M. In such cases Accept/Reject algorithm will no longer apply, and one led to another class of methods known as Markov Chain Monte Carlo (MCMC) methods. Some special cases of these methods are Metropolis-Hasting (M-H) algorithm, Gibbs Sampler and Slice sampler etc.

Metropolis-Hasting Algorithm: Metropolis-Hasting algorithm is a Markov Chain Monte Carlo (MCMC) [49] algorithm which approximately generates the random numbers from the target stationary distribution. This algorithm was introduced by Metropolis et at. [36] (1953) in a setup of optimization on a discrete state-space. Later it is generalized by Hastings (1970). MCMC techniques are often applied to solve integration and optimization problems in large dimensional spaces. These two types of problem play a fundamental role in Bayesian analysis described in this thesis.

Suppose we wish to generate random number from the distribution $\pi(\cdot)$ (also known as target distribution), on some continuous state spaces χ. Suppose we have a density that can generate the candidates, which is known as the proposal distribution or candidate-generating distribution say $q(\cdot|x)$ given a current candidate x. Let $k_{M-H}(\cdot|x)$ be a transition kernel density related to the proposal distribution $q(\cdot|x)$. The M-H algorithm produces a Markov chain through the following transition,

Step-1: given current candidate X_n, generates $Y_{n+1} \sim q(y|X_n)$.

Step-2: set

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{with probability} \quad \alpha(X_n, Y_{n+1}) \\ X_n & \text{with probability} \quad 1-\alpha(X_n, Y_{n+1}) \end{cases},$$

where $\alpha(x,y)$ is known as MH acceptance probability and defined as $\alpha(x,y) = min\{\frac{\pi(y)}{\pi(x)}\frac{q(x|y)}{q(y|x)}, 1\}$. The transition kernel density for the M-H algorithm is

$$k_{M-H}(x_{i+1}|x_i) = q(x_{i+1}|x_i)\alpha(x_i, x_{i+1}) + \delta_{x_i}(x_{i+1})r(x_i), \tag{1.2.5}$$

where $\delta_{x_i}(x_{i+1}) = 1$ if $x_i = x_{i+1}$ and 0 otherwise, and $r(x_i)$ is the term associated with rejection

$$r(x_i) = \int_\chi q(y|x_i)(1-\alpha(x_i, y))dy.$$

One can easily prove that the samples generated by this algorithm will mimic samples drawn from the target distribution asymptotically. By construction, k_{M-H} satisfies the detailed balance condition

$$\pi(x_i)k_{M-H}(x_{i+1}|x_i) = \pi(x_{i+1})k_{M-H}(x_i|x_{i+1}),$$

and consequently, the M-H algorithm admits $\pi(\cdot)$ as an invariant distribution.

Multi-Stage Gibbs Sampler: Suppose $X = (X_1, X_2, \cdots, X_m)$ be a m dimensional ($m > 1$) random variable with joint probability density function $f(x_1, x_2, \cdots, x_m)$. Let $x^{(0)}$ be the initial realization for X. The multi-stage Gibbs sampler generates a Markov chain $\{X^{(t)}\}$ according to the following steps. Take $X^{(0)} = x^{(0)}$, for t= 1,2,3, $\cdots$

Step-1: $X_1^{(t+1)} \sim f_{X_1|X_2,X_3\cdots,X_m}(.|x_2^{(t)}, x_3^{(t)}, \cdots, x_m^{(t)})$, conditional distribution of X_1 given $(X_2, X_3, \cdots, X_m)$.

Step-i: $X_i^{(t+1)} \sim f_{X_i|X_1,X_2\cdots,X_{i-1},X_{i+1},\cdots,X_m}(.|x_1^{(t+1)}, x_2^{(t+1)}, \cdots, x_{i-1}^{(t+1)}, x_{i+1}^{(t)}, \cdots, x_m^{(t)})$, conditional distribution of X_i given $(X_1, X_2\cdots, X_{i-1}, X_{i+1}, \cdots, X_m)$ where $i = 1, 2, 3, \cdots, m$.

Slice Sampling: In 2003 R. M. Neal introduced the slice sampling technique [38][49] that can be applied to a wide variety of distributions. Slice sampling is simplest when the distribution of interest is univariate, we use this method to sample from a multivariate distribution for $\mathbf{x}$=$(x_1, x_2..., x_n)$

by sampling repeatedly for each variable in turn. We update x_i by using a function of x_i, which is proportional to the actual conditional distribution of x_i, $f(x_i|\{x_j\}_{j\neq i})$, where $\{x_j\}_{j\neq i}$ are the values of the other variables. Here we will describe the slice sampling for the univariate case (also known as a 2D slice sampler). Suppose f(x) is a function which is proportional to the probability density function of x and $x^{(0)}$ be an initial value for x.

step-1: Draw a real value $u^{(1)}$, uniformly from the interval $(0, f(x^{(0)}))$. Now we define a horizontal slice $S^{(1)} = \{x : u^{(1)} \leq f(x)\}$. Note that $x^{(0)}$ is always within $S^{(1)}$.

step-2: Calculate an interval $I^{(1)} = (L^{(1)}, R^{(1)})$ which contains $x^{(0)}$ and also contains all or much of the slice $S^{(1)}$.

step-3: Draw a new value $x^{(1)}$, uniformly from $I^{(1)}$ with checking option, $f(x^{(1)})$ exists.

For drawing next value $x^{(2)}$, putting $x^{(1)}$ in place of $x^{(0)}$ and follow all steps. According to above steps it generates a Markov chain $\{X^{(t)}\}$ where $X^{(t)}$ takes the value $x^{(t)}$.

Lindley Approximation: This is the approximate evaluation of integral of the forms

$$\frac{\int w(\theta)e^{L(\theta)}d\theta}{\int v(\theta)e^{L(\theta)}d\theta},$$

where $\theta = (\theta_1, \theta_2, \theta_3, \cdots, \theta_n)$ is a parameter and

$$L(\theta|\underline{x}) = \ln \prod_{i=1}^{m} f(x_i|\theta) = \sum_{i=1}^{m} \ln f(x_i|\theta)$$

is the log-likelihood function based on the random variable X, whose PDF is $f(x|\theta)$ and $\underline{x} = \{x_1, x_2, x_3, \cdots, x_m\}$ is a sample of size m taken from this distribution. Here $w(\cdot)$ and $v(\cdot)$ are any arbitrary functions of θ. Let $\pi(\theta)$ is a prior distribution of θ and $g(\theta)$ is any arbitrary function of θ. We also choose $v(\theta) = \pi(\theta)$ and $w(\theta) = g(\theta)\pi(\theta)$. We know that the Bayes estimate of $g(\theta)$ is the posterior mean of $g(\theta)$. The Bayes estimate of $g(\theta)$ can be written as,

$$\hat{g}_B = \frac{\int_\theta g(\theta)l(\theta|\underline{x})\pi(\theta)d\theta}{\int_\theta l(\theta|\underline{x})\pi(\theta)d\theta}, \tag{1.2.6}$$

where $l(\theta|\underline{x})$ is the likelihood function i.e. $L(\theta|\underline{x}) = \ln l(\theta|\underline{x})$. Let us assumed that $\rho(\theta) = \ln \pi(\theta)$. So (1.2.6) can be written as,

$$\begin{aligned}\hat{g}_B =& \frac{\int_\theta g(\theta)e^{[L(\theta)+\rho(\theta)]}d\theta}{\int_\theta e^{[L(\theta)+\rho(\theta)]}d\theta}\\ =& g + \frac{1}{2}\sum(g_{ij} + 2g_i\rho_j)\sigma_{ij} + \frac{1}{2}\sum L_{ijk}g_l\sigma_{ij}\sigma_{kl},\end{aligned} \tag{1.2.7}$$

where $i, j, k, l = 1, 2, 3, \cdots, n$. Many partial derivatives occur in RHS of (1.2.7) and we write, for example $\frac{\delta^3 L}{\delta_i\delta_j\delta_k}$ as L_{ijk}. Similar for g and ρ also. Here σ_{ij} is the $(i, j)th$ element of the inverse of the matrix $\{L_{ij}\}$. All term in RHS (1.2.7) are calculated at MLE of θ $(= \hat{\theta}, \text{say})$.

1.2.3 Interval Estimation

We have so far estimated a population parameter using a single statistic. Such estimation by a single statistic is known as point estimation, and a point estimate, which is computed from an observed sample gives us a value somewhat close to the true value of the estimated parameter. In the method of the interval estimation we use a pair of statistic, which forming a random interval for the parameter. In practical point of view, it is more useful to give an interval of the parameter in place of point estimation. Here a probability statement associated with it, is known as confidence level. It indicates what is the probability that the random interval contains the exact value of the parameter. The idea of interval estimation is precisely stated as follows.

Suppose $\mathbf{X}$=$\{X_1, X_2, \cdots X_n\}$ is a random sample of size n drawn from the population with pdf $f(x;\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Theta \subseteq \mathbf{R}^k$. A family of subsets s($\mathbf{x}$) of Θ is said to be a family of confidence sets at confidence level $(1-\alpha)$ (sometimes $100(1-\alpha)\%$) if

$$P(\boldsymbol{\theta} \in s(\mathbf{x})) \geqslant (1-\alpha) \quad \forall \quad \boldsymbol{\theta} \in \Theta.$$

For k=1 , s($\mathbf{x}$)= (a($\mathbf{x}$),b($\mathbf{x}$)) is said to be $100(1-\alpha)\%$ level of confidence interval for θ if

$$P(a(\boldsymbol{X}) < \boldsymbol{\theta} < b(\boldsymbol{X})) \geqslant (1-\alpha) \quad \forall \quad \boldsymbol{\theta} \in \Theta.$$

Most of time we use bootstrap methods for estimating confidence interval.

Confidence Interval by Bootstrap Method:

Bootstrap methods[14] are based on re-sampling of the observed data. It was introduced by Efron in 1979. Suppose $\mathbf{X}$=$\{X_1, X_2, \cdots, X_n\}$ and $\mathbf{x}$=$\{x_1, x_2, x_3, \cdots, x_n\}$ denote the random sample and its observed realization respectively from the distribution with the parameter θ. The basic idea of the bootstrap is to evaluate the properties of an arbitrary estimator $\hat{\theta}(\mathbf{x})$ through the empirical distribution of the random sample $\mathbf{X}$,

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_j \leq x}$$

in place of actual distribution function F of the sample. There are mostly two different ways to obtain and evaluate bootstrap confidence intervals: non parametric and parametric.

Non-Parametric:

step-1: Generates a random sample (with replacement) of size n from the sample $\mathbf{X}$, denote it by $\boldsymbol{X^*} = \{X_1^*, X_2^*, \cdots, X_n^*\}$ actually $\boldsymbol{X^*} \sim F_n$.

step-2: Calculate $\hat{\theta}^* = \hat{\theta}(\mathbf{x}^*)$

step-3: Repeat step-1, step-2, B times and calculate $\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \cdots, \hat{\theta}_B^*$ *(generally* $1000 < B < 10000$)

step-4: Arrange $\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \cdots, \hat{\theta}_B^*$ in increasing order so that

$$\hat{\theta}_{(1)}^* \leq \hat{\theta}_{(2)}^* \leq \hat{\theta}_{(3)}^*, \cdots, \leq \hat{\theta}_{(B)}^*$$

step-5: Then a $100(1-\alpha)\%$ bootstrap confidence interval is given by

$$\left[\hat{\theta}_{(i)}^*, \hat{\theta}_{\left(i+(1-\alpha)*B\right)}^*\right] \quad \text{for} \quad i = 1, 2, 3, \cdots, [\alpha * B],$$

where $[\cdot]$ represents the corresponding greatest integer. We have multiple confidence interval, generally we choose shortest length interval or symmetric interval. This technique also known as Efron's Simple Percentile Confidence Intervals.

Parametric Bootstrap:

Non-parametric bootstrap makes no assumptions about how our observations are distributed, and re-sample our original sample, whereas parametric bootstrap re-sample a known distribution function, whose parameters are estimated from our sample. Here we find an estimator $\hat{\theta}(\boldsymbol{X})$ for the parameter θ. Then generates a random sample $\boldsymbol{X^*} = \{X_1^*, X_2^*, \cdots, X_n^*\}$ from $f(x|\hat{\theta}(\boldsymbol{x}))$ all other step are same as non-parametric.

1.3 Literature Survey

Extensive work has been done in this area over the past several decades. Here we describe some bivariate distributions similar to the Marshall Olkin Bivariate Exponential (MOBVE) distribution [35], the Block Basu Bivariate Exponential (BBBVE) distribution [9] and the Geometric univariate also bivariate distribution [25] etc.

Marshall and Olkin [35] introduced the Bivariate Exponential distribution(BVE) in 1967, the latter which is known as the MOBVE Distribution. The survival function of MOBVE distribution is given as

$$P(X > x, Y > y) = e^{-\lambda_1 x - \lambda_2 y - \lambda_0 max(x,y)} \quad \text{where} \quad x, y > 0. \tag{1.3.1}$$

An interesting fact of this distribution is that it has both an absolutely continuous and a singular part. Using this fact, in 1974 Block and Basu [9] introduced an absolute continuous Bivariate Exponential distribution by removing singular part from MOBVE distribution, which is known as BBBVE distribution. The survival function of BBBVE distribution is given as,

$$P(X > x, Y > y) = \frac{\lambda}{\lambda_1 + \lambda_2} e^{-\lambda_1 x - \lambda_2 y - \lambda_0 max(x,y)} - \frac{\lambda_0}{\lambda_1 + \lambda_2} \times e^{-\lambda max(x,y)},$$

where x,y >0 and $\lambda = \lambda_0 + \lambda_1 + \lambda_2$.

In 1990 Pena and Gupta [44] calculated the Bayes estimates of the parameters of MOBVE distribution. For the Bayes estimator of the parameters, they used the quadratic loss function and prior distribution for

$$\lambda \propto \lambda^{(\alpha-1)} e^{-\beta\lambda} \quad \text{where} \quad \alpha, \beta > 0.$$

Also, they used the Gamma-Dirichlet prior for $(\lambda_0, \lambda_1, \lambda_2)$ with the parameters satisfy $\alpha > 0$, $\beta > 0$ and $\alpha_i > 0 (i = 0, 1, 2)$ pdf as follows

$$\pi^*(\lambda_0, \lambda_1, \lambda_2 | \alpha, \beta, \alpha_0, \alpha_1, \alpha_2) = \frac{\Gamma(\Delta)}{\Gamma(\alpha)} (\beta\lambda)^{(\alpha-\Delta)} \prod_{i=0}^{2} \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta\lambda_i} \tag{1.3.2}$$

and denoted it as $GD(\lambda_0, \lambda_1, \lambda_2 | \alpha, \beta, \alpha_0, \alpha_1, \alpha_2)$. Under the quadratic loss function, Bayes estimators are the mean of the posterior distribution of the parameters given the data.

In 1999 Gupta and Kundu [17] introduced the Generalized Exponential(GE) distribution with three parameters (location, scale and shape). A random variable X is said to have a generalized exponential distribution if X has the distribution function

$$F(x; \alpha, \lambda, \mu) = \left(1 - e^{-\frac{(x-\mu)}{\lambda}}\right)^{\alpha} \quad \text{where} \quad x > \mu, \lambda > 0 \quad \text{and} \quad \alpha > 0. \tag{1.3.3}$$

Here α is a shape parameter, λ is a scale parameter and μ is a location parameter and denoted as $GE(\mu, \lambda, \alpha)$.

In 2003 Karlis [19] calculated estimators for the parameters of the MOBVE distribution using an MLE estimation technique via an EM algorithm. He defined MOBVE distribution, whose pdf as follows

$$f(x, y) = \begin{cases} \lambda_1(\lambda_0 + \lambda_2) e^{-\lambda_1 x - \lambda_2 y - \lambda_0 max(x,y)} & \text{where} \quad 0 < x < y \\ \lambda_2(\lambda_0 + \lambda_1) e^{-\lambda_1 x - \lambda_2 y - \lambda_0 max(x,y)} & \text{where} \quad x > y > 0 \\ \lambda_0 e^{-\lambda_1 x - \lambda_2 y - \lambda_0 max(x,y)} & \text{where} \quad x = y > 0. \end{cases}$$

Here the system of the likelihood equations cannot be solved in closed form expressions, for that reason he used the EM algorithm. In first step he calculates the expectation of the complete log-likelihood (log- likelihood of data and missing variable) w.r.t. the conditional pdf of missing variable given the data or sample, then maximize the expectation w.r.t. the parameters and calculates the estimator (actually realization of an estimator).

In 2008 Kundu and Gupta [26] estimate the parameters of the two parameters Generalized Exponential (GE) distribution as an alternative to the Gamma and Weibull distribution. The two-parameter GE distribution has the following density function

$$f(x; \lambda, \alpha) = \alpha\lambda \left(1 - e^{-x\lambda}\right)^{\alpha-1} e^{-x\lambda} \quad \text{where} \quad x > 0, \lambda > 0 \quad \text{and} \quad \alpha > 0. \tag{1.3.4}$$

Here α is a shape parameter, λ is a scale parameter. In that paper they consider the Bayes estimator of the unknown parameters under the assumptions of gamma prior on both α and λ. It is assumed that λ and α has the following gamma prior distributions

$$\pi_1(\lambda) \propto \lambda^{b-1}e^{-a\lambda}$$
$$\pi_2(\alpha) \propto \alpha^{d-1}e^{-c\alpha}.$$

Here all the hyper parameters a, b, c, d are assumed to be known and non-negative. The log-likelihood function based on the observed data $\mathbf{x} = \{x_1, x_2, ..., x_n\}$;

$$logL(\lambda, \alpha|\boldsymbol{x}) = \alpha^n\lambda^n e^{-\lambda\ \sum_{i=1}^n x_i}\prod_{i=1}^{n}(1-e^{-\lambda x_i})^{\alpha-1}.$$

Hence the joint posterior pdf of α and λ can be written as;

$$\pi^*(\alpha, \lambda|\boldsymbol{x}) = \frac{logL(\lambda, \alpha|\boldsymbol{x})\pi_1(\lambda)\pi_2(\alpha)}{\int_0^\infty\int_0^\infty logL(\lambda, \alpha|\boldsymbol{x})\pi_1(\lambda)\pi_2(\alpha)d\lambda d\alpha}.$$

So the Bayes estimator of any function of α and λ, say $g(\alpha, \lambda)$ under the squared error loss function is

$$\hat{g} = E_{\alpha,\lambda|\boldsymbol{x}}(g(\alpha, \lambda)) = \frac{\int_0^\infty\int_0^\infty g(\lambda, \alpha)logL(\lambda, \alpha|\boldsymbol{x})\pi_1(\lambda)\pi_2(\alpha)d\lambda d\alpha}{\int_0^\infty\int_0^\infty logL(\lambda, \alpha|\boldsymbol{x})\pi_1(\lambda)\pi_2(\alpha)d\lambda d\alpha}.$$

The Bayes estimators can not be obtained in explicit form. They proposed Lindley approximation technique for computing approximate Bayes estimators. In that paper they also proposed Markov Chain Monte Carlo (MCMC) specially Gibbs sampling technique to generate the samples from the posterior distributions and in turn computing the approximate Bayes estimators.

In 2009 Kundu and Pradhan [32] computed the MLE estimators of two parameters GE distribution via the EM algorithm in the presence of hybrid censoring. In this paper they assumed censored data as the missing data and calculate the estimators as in the paper [19] and also obtain the Bayes estimates of the unknown parameters under the assumption of independent gamma prior using the importance sampling procedure.

In 2009 Kundu and Gupta [27] introduced the Bivariate Generalized Exponential (BVGE) distribution using the technique provided by Arnold[1] (one can assume as the technique provided by Marshall and Olkin[35]). Here they used three independent random variables $U_1 \sim GE(\alpha_1, \lambda)$, $U_2 \sim GE(\alpha_2, \lambda)$, and $U_3 \sim GE(\alpha_3, \lambda)$ and defined $X_1 = max\{U_1, U_3\}$ and $X_2 = max\{U_2, U_3\}$ and also, used $\lambda = 1$. Then the distribution of (X_1, X_2) is known as the BVGE distribution and the cdf is

$$F_{X_1,X_2}(x_1, x_2) = (1-e^{-x_1})^{\alpha_1}(1-e^{-x_2})^{\alpha_2}(1-e^{-z})^{\alpha_3},$$

where $x_1 > 0$ $x_2 > 0$ and $z = min\{x_1, x_2\}$. They also used scale parameter $\lambda = 1$. So it's pdf is

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} f_1(x_1, x_2), & \text{if} \quad 0<x_1<x_2<\infty \\ f_2(x_1, x_2), & \text{if} \quad 0<x_2<x_1<\infty \\ f_0(x), \quad \text{if} \quad 0<x_1 = x_2 = x<\infty \end{cases}$$

, where

$$f_1(x_1, x_2) = (\alpha_1 + \alpha_3)\alpha_2(1 - e^{-x_1})^{\alpha_1+\alpha_3-1}(1 - e^{-x_2})^{\alpha_2-1}e^{-x_1-x_2}$$

$$f_2(x_1, x_2) = (\alpha_2 + \alpha_3)\alpha_1(1 - e^{-x_1})^{\alpha_1-1}(1 - e^{-x_2})^{\alpha_2+\alpha_3-1}e^{-x_1-x_2}$$

$$f_0(x) = \alpha_3(1 - e^{-x})^{\alpha_1+\alpha_2+\alpha_3-1}e^{-x},$$

They also provide an absolute continuous BGE distribution using the technique provided by Block and Basu [9] and it's pdf function is

$$f_a(x_1, x_2) = \begin{cases} \frac{1}{p} f_1(x_1, x_2) & \text{if} \quad x_1<x_2 \\ \frac{1}{p} f_2(x_1, x_2) & \text{if} \quad x_2<x_1 \end{cases} \quad \text{where} \quad p = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3},$$

They used the MLE technique when α_3 is known and used the EM algorithm when it is not known. For the EM algorithm (Δ_1, Δ) is the missing variable, defined as

$$\Delta_1 = \begin{cases} 1 & \text{if} \quad U_1>U_3 \\ 3 & \text{if} \quad U_1<U_3 \end{cases}$$
$$\Delta_2 = \begin{cases} 2 & \text{if} \quad U_2>U_3 \\ 3 & \text{if} \quad U_2<U_3 \end{cases}$$

then it follow all the steps of the EM algorithm similarly as Karlis[19].

In 2009 Kundu and Dey [23] introduced the Marshall Olkin Bivariate Weibull Distribution(MOBW) using the technique provided by Marshall and Olkin[35]. They assumed $U_0 \sim WE(\alpha, \lambda_0)$, $U_1 \sim WE(\alpha, \lambda_1)$ and $U_2 \sim WE(\alpha, \lambda_2)$ and they are mutually independent and defined $X_1 = min\{U_0, U_1\}$ and $X_2 = min\{U_0, U_2\}$, then the joint distribution function of (X_1, X_2) is called the Marshall-Olkin bivariate Weibull distribution and denoted it by $MOBW(\alpha, \lambda_0, \lambda_1, \lambda_2)$. The joint survival function of (X_1, X_2) can be written for $z = max\{x_1, x_2\}$ as

$$S_{X_1,X_2}(x_1, x_2) = e^{-\lambda_1 x_1^\alpha} e^{-\lambda_2 x_2^\alpha} e^{-\lambda_0 z^\alpha} \quad \text{where} \quad x_1 > 0, x_2 > 0. \tag{1.3.5}$$

So it's joint pdf is;

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} f_1(x_1, x_2), & \text{if} \quad 0<x_1<x_2<\infty \\ f_2(x_1, x_2), & \text{if} \quad 0<x_2<x_1<\infty \\ f_0(x), \quad \text{if} \quad 0<x_1 = x_2 = x<\infty \end{cases}, \tag{1.3.6}$$

where

$$f_1(x_1,x_2) = \alpha^2\lambda_1(\lambda_0+\lambda_2)x_1^{\alpha-1}e^{-\lambda_1 x_1^\alpha}x_2^{\alpha-1}e^{-(\lambda_0+\lambda_2)x_2^\alpha}$$

$$f_2(x_1,x_2) = \alpha^2\lambda_2(\lambda_0+\lambda_1)x_1^{\alpha-1}e^{-(\lambda_0+\lambda_1)x_1^\alpha}x_2^{\alpha-1}e^{-\lambda_2 x_2^\alpha}$$

$$f_0(x) = \alpha\lambda_0 x^{\alpha-1}e^{-(\lambda_0+\lambda_1+\lambda_2)x^\alpha}$$

They suggested to use the EM algorithm to compute the MLEs of the unknown parameters of the MOBW distribution model and used (Δ_1, Δ_2) as the missing random vector, where (Δ_1, Δ_2) is defined as follows

$$\Delta_1 = \begin{cases} 0 & \text{if} \quad X_1 = U_0 \\ 1 & \text{if} \quad X_1 = U_1 \end{cases}$$

$$\Delta_2 = \begin{cases} 0 & \text{if} \quad X_2 = U_0 \\ 2 & \text{if} \quad X_2 = U_2 \end{cases}$$

then it follows the steps of the EM algorithm similarly as Karlis[19]. In this case the 'pseudo log-likelihood' function ('E' step) is formed from the log-likelihood function by replacing the log-likelihood contribution of (X_1, X_2) by its expected value, then maximized the 'pseudo log-likelihood' w.r.t. the unknown parameters.

In 2010 Kundu and Gupta [28] introduced the Block Basu Bivariate Weibull (BBBW) distribution using the technique provided by Block and Basu [9]. They wrote the joint survival function (1.3.5) as a mixture of an absolutely continuous part and a singular part as follows;

$$S_{X_1,X_2}(x_1,x_2) = \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3}S_a(x_1,x_2) + \frac{\lambda_0}{\lambda_1+\lambda_2+\lambda_3}S_s(x_1,x_2),$$

where $S_a(\cdot,\cdot)$ is the absolutely continuous part and $S_s(\cdot,\cdot)$ is the singular part. Note that for $z = max\{x_1, x_2\}$,

$$S_s(x_1,x_2) = e^{-(\lambda_0+\lambda_1+\lambda_2)z^\alpha}$$

and $S_a(x_1, x_2)$ can be obtained by subtraction as

$$S_a(x_1,x_2) = \frac{\lambda_0+\lambda_1+\lambda_2}{\lambda_1+\lambda_2}e^{-\lambda_1 x_1^\alpha}e^{-\lambda_2 x_2^\alpha}e^{-\lambda_0 z^\alpha} - \frac{\lambda_0}{\lambda_1+\lambda_2}e^{-(\lambda_0+\lambda_1+\lambda_2)z^\alpha} \tag{1.3.7}$$

and it's joint pdf function is;

$$f_a(x_1,x_2) = \begin{cases} \frac{1}{p}\alpha^2\lambda_1(\lambda_0+\lambda_2)x_1^{\alpha-1}e^{-\lambda_1 x_1^\alpha}x_2^{\alpha-1}e^{-(\lambda_0+\lambda_2)x_2^\alpha} & \text{if} \quad x_1<x_2 \\ \frac{1}{p}\alpha^2\lambda_2(\lambda_0+\lambda_1)x_1^{\alpha-1}e^{-(\lambda_0+\lambda_1)x_1^\alpha}x_2^{\alpha-1}e^{-\lambda_2 x_2^\alpha} & \text{if} \quad x_2<x_1, \end{cases} \tag{1.3.8}$$

where $p = \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$. Then they used the EM algorithm for estimating parameters, which is similar to ([23]). The Multivariate Pareto is used in this thesis was also introduced by Asmit et al [2] in 2010. The joint distribution used in the paper is given by,

$$f(x_1,x_2) = \begin{cases} f_1(x_1,x_2) & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1,x_2) & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x) & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x, \end{cases}$$

where

$$\begin{aligned} f_1(x_1,x_2) &= \frac{\alpha_1}{\sigma_1\sigma_2}(\alpha_0+\alpha_2)(1+\frac{x_2-\mu_2}{\sigma_2})^{-(\alpha_0+\alpha_2+1)}(1+\frac{x_1-\mu_1}{\sigma_1})^{-(\alpha_1+1)} \\ f_2(x_1,x_2) &= \frac{\alpha_2}{\sigma_1\sigma_2}(\alpha_0+\alpha_1)(1+\frac{x_1-\mu_1}{\sigma_1})^{-(\alpha_0+\alpha_1+1)}(1+\frac{x_2-\mu_2}{\sigma_2})^{-(\alpha_2+1)} \\ f_0(x) &= \alpha_0(1+x)^{-(\alpha_1+\alpha_2+\alpha_0+1)} \end{aligned}$$

In 2013 Kundu and Gupta [24] estimated the parameters of the MOBW distribution, whose pdf function in equation (1.3.6) using the Bayesian estimation technique. They proposed $Gamma(a,b)$, say $\pi(\cdot|a,b)$, as the prior for $\lambda(=\lambda_0+\lambda_1+\lambda)$;

$$\pi_1(\lambda|a,b) = \begin{cases} \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} & \text{if } \lambda{>}0 \\ 0 & \text{otherwise} \end{cases}$$

and Dirichlet prior, say $\pi_1(\cdot|a_0,a_1,a_2)$ for the joint prior of $\left(\frac{\lambda_1}{\lambda},\frac{\lambda_2}{\lambda}\right)$, i.e.

$$\pi_2\left(\frac{\lambda_1}{\lambda},\frac{\lambda_2}{\lambda}|\lambda,a_0,a_1,a_2\right) = \frac{\Gamma(a_0+a_1+a_2)}{\Gamma(a_0)\Gamma(a_1)\Gamma(a_2)}\left(\frac{\lambda_0}{\lambda}\right)^{a_0-1}\left(\frac{\lambda_1}{\lambda}\right)^{a_1-1}\left(\frac{\lambda_2}{\lambda}\right)^{a_2-1},$$

where all hyper parameters a, b, a_0, a_1, and a_2 are positive. Then the joint prior can be written as

$$\pi(\lambda_0,\lambda_1,\lambda_2|a,b,a_0,a_1,a_2) = \frac{\Gamma(\underline{a})}{\Gamma(a)}(b\lambda)^{a-\underline{a}}\prod_{i=0}^{2}\frac{b^{a_i}}{\Gamma(a_i)}\lambda_i^{a_i-1}e^{-b\lambda_i}, \tag{1.3.9}$$

where $\underline{a} = a_0+a_1+a_2$. The posterior distribution of $(\lambda_0,\lambda_1,\lambda_2)$ based on the prior and the data D can be written as;

$$\pi^*(\lambda_0,\lambda_1,\lambda_2|\alpha,D) \propto logL(\lambda_0,\lambda_1,\lambda_2|\alpha,D)\pi(\lambda_0,\lambda_1,\lambda_2|a,b,a_0,a_1,a_2),$$

where α is known and L is the likelihood function. They proposed to use, the importance sampling technique, to compute the Bayes estimator of $\theta=\theta(\lambda_0,\lambda_1,\lambda_2)$, any function of λ_0, λ_1 and λ_2 as a Bayes estimator of θ under squared error loss function is the posterior mean of θ.

In 2014 Kundu and Gupta [25] introduced the bivariate Weibull-Geometric distribution. Let us consider the joint random variable (X_1,X_2) is said have a bivariate Weibull-Geometric distribution with the parameters θ, α, λ_0, λ_1 and λ_2 if the joint survival function (X_1,X_2) can be written as follows,

$$\bar{G}(x_1,x_2) = \begin{cases} \frac{\theta e^{-\lambda_1 x_1^{\alpha}} e^{-(\lambda_0+\lambda_2)x_2^{\alpha}}}{1-(1-\theta)e^{-\lambda_1 x_1^{\alpha}} e^{-(\lambda_0+\lambda_2)x_2^{\alpha}}} & \text{if } x_1 < x_2 \\ \frac{\theta e^{-(\lambda_0+\lambda_1)x_1^{\alpha}} e^{-\lambda_2 x_2^{\alpha}}}{1-(1-\theta)e^{-(\lambda_0+\lambda_1)x_1^{\alpha}} e^{-\lambda_2 x_2^{\alpha}}} & \text{if } x_1 > x_2 \\ \frac{\theta e^{-(\lambda_0+\lambda_1+\lambda_2)x^{\alpha}}}{1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)x^{\alpha}}} & \text{if } x_1 = x_2 = x. \end{cases}$$

It is denoted as $BWG(\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$. They proposed to use the EM algorithm to obtain the maximum likelihood estimators.

In 2016 Asimit et al. [3] used the EM algorithm to estimate the parameters of the Marshall-Olkin bivariate Pareto distribution. Some research on Bayesian analysis of the BBBW distribution can be found in Pradhan and Kundu [46]. Equation (1.3.8) represents it's pdf function. They used the importance sampling technique to estimate the parameters.

1.4 Outline of the Thesis

In chapter-2, we discuss about the problem of estimation of the parameters of the Marshal-Olkin bivariate Pareto distribution with location and scale parameters using the EM algorithm. This is a singular distribution. In 2010 Asimit et al. [2] used an EM algorithm to estimate the parameters. In this context, we not only propose the different innovative variations of the EM algorithm, which is different from the existing algorithm by Asimit et. al.[2], but also make a comparison of all the methods with respect to the average number of iterations and the mean squared errors. Numerical simulation has been done to judge the performance of the different proposed algorithms. A real life data analysis has been performed for all the algorithms for the illustrative purposes, which was missing in Asimit et al. [2] paper. We also formulate approximate asymptotic confidence intervals for all the parameters. However, more research is needed in that direction. The content of the work opens the door of research in the implementation of the EM algorithm for other complicated higher dimensional distributions with location and scale parameters.

Chapter-3 addresses the problem of estimation of the parameters of the absolute continuous bivariate Pareto distribution with or without the presence of location and scale parameters. This distribution is also known as the Block-Basu bivariate Pareto distribution. We study the estimation of the parameters by the EM algorithm both with and without the presence of location and scale parameters. Some innovative solutions are provided for different problems which arise during implementation of the EM algorithm. Some numerical simulation has been carried out to show the effectiveness of the proposed algorithms. It is very difficult to construct asymptotic confidence intervals for the parameters as the log-likelihood function is discontinuous with respect to location and scale parameters. We propose the use of parametric bootstrap confidence interval for the shape parameters without the presence of location and scale parameters. Since, the location parameter behaves like a threshold, we are still working to construct the confidence interval in the presence of location and scale parameters. We illustrate the procedures through a real life data analysis.

Chapter-4 contains the Bayesian analysis of the absolute continuous Marshal-Olkin bivariate Pareto distribution without the presence of location and scale parameters using both Lindley and slice cum Gibbs sampler approach. In this context, we study two types of prior - reference prior and gamma prior. For reference prior, sometimes the conditional log-posterior becomes discontinuous in some cases. We modify the usual slice sampling technique in such a way that it works for the discontinuous setup. Credible intervals are also provided for all the methods and all the prior distributions. Numerical simulation is performed for all the methods. An analysis of real life data has been shown for illustrative purpose.

Chapter-5 deals with the problem of the Bayesian estimation of the absolute continuous Marshal-Olkin bivariate Pareto distribution with the presence of location and scale parameters using slice cum Gibbs sampler approach. In this context, we provide two different novel approaches of slice sampling based on gamma prior for shape and scale parameters and truncated normal for location parameters. We also estimate the credible intervals and the coverage probabilities of the intervals for all the methods. An extensive numerical simulation has been studied here to show the performance of the proposed methods. A real-life data analysis is also shown for illustrative purposes.

In Chapter-6, we address the Bayesian estimation of both of the Marshal-Olkin bivariate Pareto distribution with and without the presence of location and scale parameters. Here, we use gamma prior for both shape and scale parameters and truncated normal prior for location parameters. The Bayes estimate of the parameters is obtained based on the slice cum Gibbs sampler. In this context, we implement the empirical Bayes estimation for tuning hyper-parameters. It is observed that the algorithm works for any data set, but it requires some particular choice of starting values of the hyper-parameters. We construct the credible intervals based on generated samples and also calculate the corresponding coverage probabilities. Some numerical simulation has been shown here to judge the performance of the proposed algorithm. A data analysis is kept for illustrative purpose.

In Chapter-7, we propose the Geometric Marshal-Olkin bivariate Pareto distribution without the presence of location and scale parameters and also estimate the parameters using both the EM algorithm and the Bayesian estimation procedure through slice cum Gibbs sampler approach. We choose gamma prior to conduct the Bayesian analysis. An extensive simulation based on numerical results is shown here to judge the effectiveness of the proposed procedure. One real life data analysis has been performed for illustrative purpose.

Chapter-8 addresses the problem of the EM algorithm and the Bayesian estimation of the absolute continuous Geometric Marshal-Olkin bivariate Pareto distribution without the presence of location and scale parameters. We call the distribution as four parameter Geometric Block-Basu bivariate Pareto distribution. Bayesian estimation has been done through the slice cum Gibbs sampler approach based on gamma prior. Numerical results are shown to verify the performance of the algorithms. We illustrate the procedures through a real life data analysis.

Finally, we conclude the thesis and provide some future research problems in Chapter-9.

Chapter 2

Some variations of EM algorithms for Marshall-Olkin bivariate Pareto distribution with location and scale

2.1 Introduction

Bivariate Pareto has several forms. In this chapter we study the Marshall-Olkin formulation of this distribution Marshall and Olkin (1967) [35], which includes both location and scale parameters. In a very recent paper Asimit et al. (2016) [3] used the EM algorithm to estimate the parameters of this distribution. We adapt a few more variations of the same. We observe that some of our proposed algorithms either work equally well or outperform their algorithms. We also present some real life data analysis, which is absent in Asimit et al. (2016) [3] paper.

In this chapter [1], we deal with **singular** the bivariate Pareto (MOBVPA) distribution, whose both the marginals have the Pareto type-II distribution. We can obtain bivariate Pareto distribution considering peak over threshold method in a bivariate data. The distribution has wide application in modeling, data related to finance, insurance, environmental sciences and internet network. Any analysis based on this distribution requires an efficient technique for estimating parameters of the distribution. We consider a more generalized form of the Marshal-Olkin bivariate Pareto distribution, which includes location and scale parameters. We propose a few innovative ways to implement the EM algorithm.

A random variable X is said to have a Pareto of second kind, i.e. $X \sim Pa(II)(\mu, \sigma, \alpha)$ if it has the survival function

$$\bar{F}_X(x; \mu, \sigma, \alpha) = P(X > x) = (1 + \frac{x - \mu}{\sigma})^{-\alpha}$$

and the probability density function (pdf)

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma}(1 + \frac{x - \mu}{\sigma})^{-\alpha - 1}$$

[1] this work got accepted by the journal *Journal of Statistical Theory and Practice*

with $x > \mu \in \mathcal{R}$, $\sigma > 0$ and $\alpha > 0$.

Kundu and Gupta (2009) [27], Kundu and Gupta (2010) [28], Kundu and Dey (2009) [23] used the EM algorithm for estimating parameters of different bivariate distributions, e.g. bivariate generalized exponential, bivariate Weibull etc. Dey and Kundu (2012) [13] performed the discrimination between the bivariate Weibull and the bivariate generalized exponential distributions, where they used parameter estimation through the EM algorithm. Sarhan and Balakrishnan (2007) [52] considered the estimation issues in Sarhan and Balakrishnan bivariate distribution with extra scale parameter in their model. Many works have been done on multivariate Pareto distribution to Hanagal (1996) [18], Yeh (2000) [55], Yeh (2004) [56], Asimit et al. (2010)[2]. The statistical inference of the multivariate Pareto distribution through the EM algorithm is attempted by Asimit et al. (2016)[3]. This chapter also deals with seven parameters including location and scale as its parameters. Our current work, not only proposes some innovative approaches, but also make comparison among all the methods.

We arrange this chapter in the following way. In Section-2.2, we keep the Marshall-Olkin bivariate Pareto formulation and some of its properties. In Section-2.3, we describe our proposed EM algorithms. We discuss about the confidence interval in Section-2.4. Some simulation results show the performance of the algorithm in Section-2.5. In Section-2.6, we show the data analysis. Finally, we conclude this chapter in Section-2.7.

2.2 Formulation of Marshal-Olkin bivariate Pareto

Let U_0, U_1 and U_2 be three independent univariate type-II Pareto distributions $Pa(II)(0, 1, \alpha_0)$, $Pa(II)(\mu_1, \sigma_1, \alpha_1)$ and $Pa(II)(\mu_2, \sigma_2, \alpha_2)$ respectively.

We define $X_1 = \min\{\sigma_1 U_0 + \mu_1, U_1\}$ and $X_2 = \min\{\sigma_2 U_0 + \mu_2, U_2\}$. We can show that (X_1, X_2) jointly follow the Marshal-Olkin bivariate Pareto (MOBVPA) distribution of second kind, we call it as $BVPA(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$

The joint distribution is given by

$$f(x_1, x_2) = \begin{cases} f_1(x_1, x_2) & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1, x_2) & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x) & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x, \end{cases}$$

where

$$\begin{aligned} f_1(x_1, x_2) &= \frac{\alpha_1}{\sigma_1\sigma_2}(\alpha_0 + \alpha_2)(1 + \frac{x_2 - \mu_2}{\sigma_2})^{-(\alpha_0+\alpha_2+1)}(1 + \frac{x_1 - \mu_1}{\sigma_1})^{-(\alpha_1+1)} \\ f_2(x_1, x_2) &= \frac{\alpha_2}{\sigma_1\sigma_2}(\alpha_0 + \alpha_1)(1 + \frac{x_1 - \mu_1}{\sigma_1})^{-(\alpha_0+\alpha_1+1)}(1 + \frac{x_2 - \mu_2}{\sigma_2})^{-(\alpha_2+1)} \\ f_0(x) &= \alpha_0(1 + x)^{-(\alpha_1+\alpha_2+\alpha_0+1)}. \end{aligned}$$

It can be shown that,

1. The distribution of X_j is $Pa(II)(\mu_j, \sigma_j, \alpha_{0j})$, $\alpha_{0j} = \alpha_0 + \alpha_j$, where $j = 1, 2$.

2. The distribution of the minimum of X_1 and X_2, is also Pareto distribution, when U_1 and U_2 has same location and scale parameter.

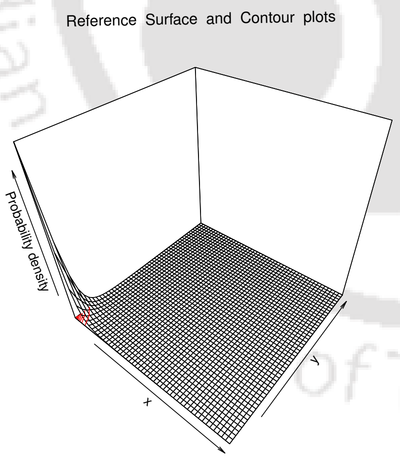
3. Maximum Likelihood estimate of μ, σ, α, the parameters of univariate Pareto can be given based on the data set $Y_1, Y_2, \cdots, Y_n$ as : $\mu = \min\{Y_1, Y_2, \cdots, Y_n\}$, whereas estimates of α and σ can be obtained by solving the fixed point iterations :

$$\sigma = \frac{\alpha + 1}{n} \sum_{i=1}^{n} \frac{(x_i - \mu)}{(1 + \frac{x_i - \mu}{\sigma})},$$

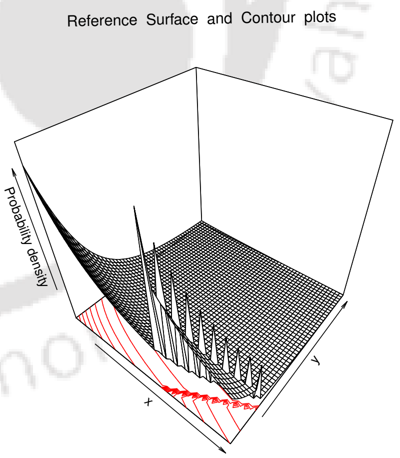
where

$$\alpha = \frac{n}{\sum_{i=1}^{n} \ln\left(1 + \frac{x_i - \mu}{\sigma}\right)}.$$

The surface and contour plots of the pdf of this distribution are shown in Figure-2.1 and Figure-2.2. The following four different sets of parameters provide four subfigures in each figure. $\xi_1 : \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 0.3, \alpha_2 = 1.4$; $\xi_2 : \mu_1 = 1, \mu_2 = 2, \sigma_1 = 0.4, \sigma_2 = 0.5, \alpha_0 = 2, \alpha_1 = 1.2, \alpha_2 = 1.4$; $\xi_3 : \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.4, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1.4$; $\xi_4 : \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.4, \sigma_2 = 0.5, \alpha_0 = 2, \alpha_1 = 0.4, \alpha_2 = 0.5$.

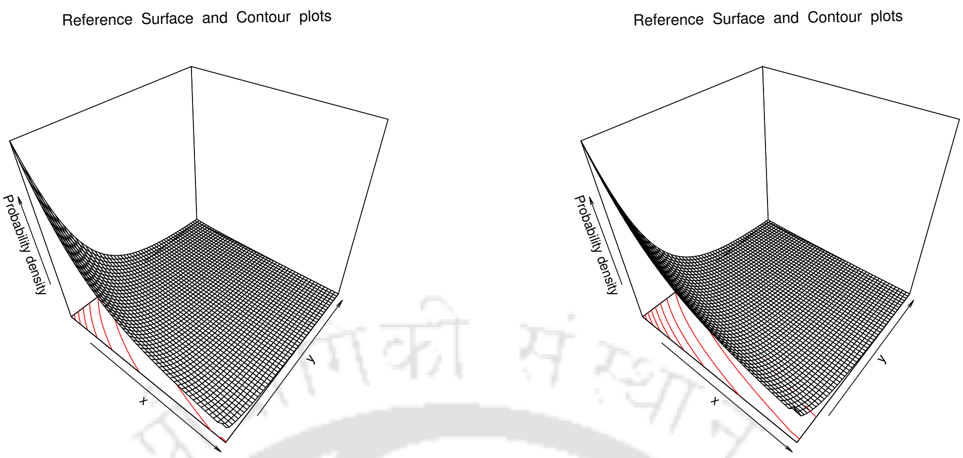


(a) ξ_1



(b) ξ_2

Figure 2.1: Surface and Contour plots of pdf of MOBVPA wrt parameter set ξ_1 and ξ_2



(a) ξ_3 (b) ξ_4

Figure 2.2: Surface and Contour plots for pdf of MOBVPA wrt parameter set ξ_3 and ξ_4

2.3 EM-algorithm

Let us denote the data set, $\chi = \{(x_{11}, x_{21}), (x_{12}, x_{22}), \cdots, (x_{1n}, x_{2n})\}$, We can divide the data into following three parts :- $I_1 = \{(x_{1i}, x_{2i}) : \frac{x_{1i}-\mu_1}{\sigma_1} < \frac{x_{2i}-\mu_2}{\sigma_2}\}$, $I_2 = \{(x_{1i}, x_{2i}) : \frac{x_{1i}-\mu_1}{\sigma_1} > \frac{x_{2i}-\mu_2}{\sigma_2}\}$ and $I_0 = \chi \backslash \{I_1 \cup I_2\}$. We define n_0, n_1, n_2 as : $n_0 = |I_0|, n_1 = |I_1|, n_2 = |I_2|$ where $|I_j|$ for $j = 0, 1, 2$ denotes the number of elements in the set I_j.

Usual log-likelihood function for the parameters of the MOBVPA can be written as

$$\begin{aligned}
& L(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2) \\
= \;& n_1 \ln \alpha_1 + n_1 \ln(\alpha_0 + \alpha_2) - n_1 \ln \sigma_1 - n_1 \ln \sigma_2 \\
- \;& (\alpha_0 + \alpha_2 + 1) \sum_{i \in I_1}^{n} \ln\left(1 + \frac{x_{2i} - \mu_2}{\sigma_2}\right) - (\alpha_1 + 1) \sum_{i \in I_1} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) \\
- \;& n_2 \ln \sigma_1 - n_2 \ln \sigma_2 + n_2 \ln \alpha_2 + n_2 \ln(\alpha_0 + \alpha_1) \\
- \;& (\alpha_0 + \alpha_1 + 1) \sum_{i \in I_2}^{n} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) - (\alpha_2 + 1) \sum_{i \in I_2}^{n} \ln\left(1 + \frac{x_{2i} - \mu_2}{\sigma_2}\right) \\
+ \;& n_0 \ln \alpha_0 - n_0 \ln \sigma_1 - (\alpha_0 + \alpha_1 + \alpha_2 + 1) \sum_{i \in I_0}^{n} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right).
\end{aligned}$$

Direct maximum likelihood estimate of the parameters based on χ may not be simple. We implement the EM algorithm for some given μ_1, μ_2, σ_1 and σ_2. In other words, we form the EM algorithm to estimate the parameters α_0, α_1 and α_2 based on the transformed data set $\{(y_{1i}, y_{2i}) :$

$1, \ldots, n\}$, where $y_{ji} = \frac{x_{ji}-\mu_j}{\sigma_j}; j = 1, 2$, for some given known values of μ_1, μ_2, σ_1 and σ_2. It is not quite sure how the performance of the algorithm will change if we vary all parameters together with the EM formulation instead of estimating the location and scale parameters in the beginning and fixed it to the next level of the EM algorithm. We describe several variations of the EM algorithms in this line and explore the best algorithm in terms of the minimum number of iterations and mean square error.

Usual EM implementation requires the identification of some missing structure within the problem. Here we do not know whether X_1 is $\sigma_1 U_0 + \mu_1$ or U_1 and similarly we also have no information regarding X_2 whether it is $\sigma_2 U_0 + \mu_2$ or U_2. Therefore, we introduce two new random variables Δ_1 and Δ_2 as

$$\Delta_1 = \begin{cases} 0 & \text{if } X_1 = \sigma_1 U_0 + \mu_1 \\ 1 & \text{if } X_1 = U_1 \end{cases}$$

and

$$\Delta_2 = \begin{cases} 0 & \text{if } X_2 = \sigma_2 U_0 + \mu_2 \\ 2 & \text{if } X_2 = U_2 \end{cases}$$

Here Δ_1 and Δ_2 are the hidden random variables of the EM algorithm. To calculate the E-step we need the conditional distribution of Δ_1 and Δ_2. The problem is more complicated here, as the distribution of Δ_1 and Δ_2 depend on unknown values of location and scale parameters.

Let's assume that first we form EM algorithm for α_0, α_1 and α_2 by using Y_1, Y_2, Δ_1 and Δ_2 when location and scale parameters are known. Therefore, the sets I_0, I_1 and I_2 can be expressed based on Y_1 and Y_2 only. We can easily show $\triangleright$ for group I_0, both Δ_1 and Δ_2 are known,

$$\Delta_1 = \Delta_2 = 0$$

.

$\triangleright$ for group I_1, Δ_1 is known, Δ_2 is unknown,

$$\Delta_1 = 1, \Delta_2 = 0 \text{ or } 2$$

Therefore we need to find out $u_1 = P(\Delta_2 = 0|I_1)$and $u_2 = P(\Delta_2 = 2|I_1)$
$\triangleright$ For group I_2, Δ_2 is known, Δ_1 is unknown,

$$\Delta_1 = 0 \text{ or } 1, \Delta_2 = 2$$

Moreover we need $w_1 = P(\Delta_1 = 0|I_2)$ and $w_2 = P(\Delta_1 = 1|I_2)$

Since, each posterior probability corresponds to one of the ordering from Table 2.1, we calculate u_1, u_2, w_1, w_2 using the probability of appropriate ordering, where $U_1^* = \frac{(U_1-\mu_1)}{\sigma_1}$, $U_2^* = \frac{(U_2-\mu_2)}{\sigma_2}$.

Ordering	(X_1, X_2)	Group
$U_0 < U_1^* < U_2^*$	(U_0, U_0)	I_0
$U_0 < U_2^* < U_1^*$	(U_0, U_0)	I_0
$U_1^* < U_0 < U_2^*$	(U_1^*, U_0)	I_1
$U_1^* < U_2^* < U_0$	(U_1^*, U_2^*)	I_1
$U_2^* < U_0 < U_1^*$	(U_0, U_2^*)	I_2
$U_2^* < U_1^* < U_0$	(U_1^*, U_2^*)	I_2

Table 2.1: Groups and corresponding orderings of hidden random variables U_0, U_1^* and U_2^*.

We have the following expressions for u_1, u_2, w_1, w_2

$$u_1 = \frac{P(U_1^* < U_0 < U_2^*)}{P(U_1^* < U_0 < U_2^*) + P(U_1^* < U_2^* < U_0)}$$

$$u_2 = \frac{P(U_1^* < U_2^* < U_0)}{P(U_1^* < U_0 < U_2^*) + P(U_1^* < U_2^* < U_0)}$$

$$w_1 = \frac{P(U_2^* < U_0 < U_1^*)}{P(U_2^* < U_0 < U_1^*) + P(U_2^* < U_1^* < U_0)}$$

$$w_2 = \frac{P(U_2^* < U_1^* < U_0)}{P(U_2^* < U_0 < U_1^*) + P(U_2^* < U_1^* < U_0)}$$

We can easily evaluate the probabilities to get values of u_1, u_2, w_1, w_2 as : $u_1 = \frac{\alpha_0}{\alpha_0+\alpha_2}$ and $u_2 = (1 - u_1)$ $w_1 = \frac{\alpha_0}{\alpha_0+\alpha_1}$ and $w_2 = (1 - w_1)$

2.3.1 Pseudo-likelihood expression

The Pseudo likelihood function is obtained by multiplying all the log-likelihood contributions by its posterior probabilities of hidden random variables. We have seen each of these posterior probabilities depend on some the particular ordering of U_0, U_1^* and U_2^*. Therefore, we can write the likelihood contributions based on the likelihood of U_0, U_1^* and U_2^*. For example, likelihood contribution for $U_1^* < U_0 < U_2^*$ can be given by $f_{U_1^*}(x_1) f_{U_0}(x_2) S_{U_2^*}(x_2)$.

The final expression of the pseudo log-likelihood can be written down, skipping all in between algebras as

$$\begin{aligned} Q &= L(\alpha_0, \alpha_1, \alpha_2) \\ &= -\alpha_0 \Big(\sum_{i \in I_0} \ln(1 + y_i) + \sum_{i \in I_2} \ln(1 + y_{1i}) + \sum_{i \in I_1} \ln(1 + y_{2i})\Big) \\ &+ (n_0 + u_1 n_1 + w_1 n_2) \ln \alpha_0 - \alpha_1 \Big(\sum_{i \in I_0} \ln(1 + y_i) + \sum_{i \in I_1 \cup I_2} \ln(1 + y_{1i})\Big) \\ &+ (n_1 + w_2 n_2) \ln \alpha_1 - \alpha_2 \Big(\sum_{i \in I_0} \ln(1 + y_i) + \sum_{i \in I_1 \cup I_2} \ln(1 + y_{2i})\Big) \\ &+ (n_2 + u_2 n_1) \ln \alpha_2. \end{aligned} \tag{2.3.1}$$

Therefore M-step involves maximizing (3.3.1) with respect to α_0, α_1, α_2 at

$$\hat{\alpha}_0^{(t+1)} = \frac{n_0 + u_1^{(t)} n_1 + w_1^{(t)} n_2}{\sum_{i \in I_0} \ln(1+y_i) + \sum_{i \in I_2} \ln(1+y_{1i}) + \sum_{i \in I_1} \ln(1+y_{2i})}. \tag{2.3.2}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(n_1 + w_2^{(t)} n_2)}{\sum_{i \in I_0} \ln(1+y_i) + \sum_{i \in (I_1 \cup I_2)} \ln(1+y_{1i})}. \tag{2.3.3}$$

$$\hat{\alpha}_2^{(t+1)} = \frac{n_2 + u_2^{(t)} n_1}{(\sum_{i \in I_0} \ln(1+y_i) + \sum_{i \in (I_1 \cup I_2)} \ln(1+y_{2i}))}. \tag{2.3.4}$$

Therefore the algorithm can be given as

Algorithm 2.1 EM procedure for bivariate Pareto distribution.

1: Calculate the estimates of μ_1, μ_2, σ_1 and σ_2 from marginals of bivariate distribution.
2: Fix I_0, I_1 and I_2.
3: **while** $\Delta Q/Q < tol$ **do**
4: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
5: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (3.3.2), (3.3.3) and (3.3.4).
6: Calculate Q for the new iterate.
7: **end while**

2.3.2 Algorithm-2.2 : Modified Algorithm 1

The above algorithm does not work even if all the I_0, I_1 and I_2 are non-empty. In our previous approach, we calculate the estimates of the location and the scale parameters and construct the above three sets based on the transformed variables using those estimated location and scale parameters. Therefore, it is highly likely that there will be almost no element in I_0 for most of the generated sample. Consequently, it should provide the estimate of α_0 as zero most of the time.

We propose to make a few modifications in the above EM algorithm. Since, we have observed the normalized data with respect to the estimated location and scale parameters, the transformation is not going to provide the distribution of the normalized data exactly, as a Pareto distribution with location parameters zero and scale parameters one. The transformation rather from some distribution close to Pareto with location zero and scale one. It is very difficult to know the exact distribution. Approximately, even if we assume some Pareto distribution with location parameters near zero and scale parameters near one, calculation of probability that an observation will come from any one of I_0, I_1 or I_2, becomes another problem.

To avoid such instances, we assume that information of all n_0, n_1 and n_2 are missing because of the small perturbation inherited by estimating the location and scale parameters. In an ideal situation, we should get the EM algorithm steps as described in algorithm-2.3. Since, we have now three more latent information n_0, n_1 and n_2, we try to approximate them and incorporate in the original EM algorithm.

It is clear that (n_0, n_1, n_2) will jointly follow the multinomial distribution with parameters $n = n_0 + n_1 + n_2$ and $(\frac{\alpha_0}{\alpha_0+\alpha_1+\alpha_2}, \frac{\alpha_1}{\alpha_0+\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_0+\alpha_1+\alpha_2})$.

We approximate $\tilde{n_i} = n \cdot \frac{\alpha_i}{\alpha_0+\alpha_1+\alpha_2}, \qquad i = 1, 2.$

Therefore M-step involves maximizing (3.3.1) with respect to α_0, α_1, α_2 at

$$\hat{\alpha}_0^{(t+1)} = \frac{\tilde{n_0} + u_1^{(t)}\tilde{n_1} + w_1^{(t)}\tilde{n_2}}{\sum_{i \in I_0} \ln(1 + y_i) + \sum_{i \in I_2} \ln(1 + y_{1i}) + \sum_{i \in I_1} \ln(1 + y_{2i})}. \tag{2.3.5}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(\tilde{n_1} + w_2^{(t)}\tilde{n_2})}{\sum_{i \in I_0} \ln(1 + y_i) + \sum_{i \in (I_1 \cup I_2)} \ln(1 + y_{1i})}. \tag{2.3.6}$$

$$\hat{\alpha}_2^{(t+1)} = \frac{\tilde{n_2} + u_2^{(t)}\tilde{n_1}}{(\sum_{i \in I_0} \ln(1 + y_i) + \sum_{i \in (I_1 \cup I_2)} \ln(1 + y_{2i}))}. \tag{2.3.7}$$

Finally modified algorithm would be

Algorithm 2.2 EM procedure for bivariate Pareto distribution.

1: Calculate the estimates of μ_1, μ_2, σ_1 and σ_2 from marginals of bivariate distribution.
2: Fix I_0, I_1 and I_2.
3: **while** $\Delta Q/Q < tol$ **do**
4: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n_0}$, $\tilde{n_1}$ and $\tilde{n_2}$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
5: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (2.3.5), (2.3.6) and (2.3.7).
6: Calculate Q for the new iterate.
7: **end while**

2.3.3 Algorithm-2.3 : Modified Algorithm 2

In this case, we update σ_1 and σ_2 along with the EM iterations similar to what Asimit et al. has done in their paper. But the main difference of our update is that it is based on one step ahead gradient descent instead of fixed point iteration on scale parameters. Similar Approaches are available in the EM algorithm used for estimating parameters of the absolute continuous bivariate Pareto by Paul, Dey and Kundu[43]. Surely gradient descent with respect to the bivariate likelihood will not work as the bivariate likelihood is a discontinuous function with respect to location and scale parameters. Therefore, we estimate the scale parameters using the density function of the marginals. At every iteration we use one step ahead gradient descent for σ_1 and σ_2 based on the likelihood of its marginal density combined with the usual EM steps for other parameters to solve the problem. The idea is similar to stochastic gradient descent method. We begin the algorithm estimating the sets I_0, I_1 and I_2, which involves location and scale parameters. Given the location and scale parameters, we can update previous EM steps of α_0, α_1 and α_2. The EM steps will ensure to take the correct direction of α_0, α_1 and α_2 starting from any values of these three parameters and also it will enable gradient descent steps of σ_1 and σ_2 to converge faster starting from any values.

This algorithm works even for moderate sample sizes. However, it takes a lot of time to converge or roam around the actual value for a very few samples. We calculate the estimates both with full iterations until convergence and with at most 2000 iterations. We observe MSEs and the number of iterations in both the cases. We expect that within 2000 iterations estimated points should be sufficiently close to the actual value or it should take some value which can be used as a good estimate for the starting point of some other optimization algorithms.

Algorithm 2.3 EM procedure for bivariate Pareto distribution.

Take the estimates of μ_1, μ_2 by $\mu_1 = \min\{X_{1i}; i = 1, \cdots, n\}$ and $\mu_2 = \min\{X_{2i}; i = 1, \cdots, n\}$.
Start with some initial choice of σ_1, σ_2, α_0, α_1, α_2.
while $|\Delta Q/Q| > tol$ **do**
  Fix I_0, I_1 and I_2 with estimated μ_1, μ_2, $\sigma_1^{(i)}$ and $\sigma_2^{(i)}$.
  Compute updates of $\sigma_1^{(i)}$, $\sigma_2^{(i)}$ through one step ahead gradient descend, using $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
  Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n_0}$, $\tilde{n_1}$ and $\tilde{n_2}$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
  Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (2.3.5), (2.3.6) and (2.3.7).
  Calculate Q for the new iterate.
end while

2.3.4 Algorithm-2.4 : Modified Algorithm 3

We can make more variations on it. This is similar to what Asimit et al. has done in their paper. At each iteration we update σ_1 and σ_2 from the marginal density using one step ahead fixed point iteration instead of Gradient decent algorithm and carry on the same EM algorithm. However approach of Asimit et al is different, therefore the EM steps are also differing.

Algorithm 2.4 EM procedure for bivariate Pareto distribution

Take the estimates of μ_1, μ_2 by $\mu_1 = \min\{X_{1i}; i = 1, \cdots, n\}$ and $\mu_2 = \min\{X_{2i}; i = 1, \cdots, n\}$.
Start with some initial choice of σ_1, σ_2, α_0, α_1, α_2.
while $|\Delta Q/Q| > tol$ **do**
  Fix I_0, I_1 and I_2 with estimated μ_1, μ_2, $\sigma_1^{(i)}$ and $\sigma_2^{(i)}$.
  Compute one step ahead update of $\sigma_1^{(i)}$, $\sigma_2^{(i)}$ through fixed point iteration, using $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
  Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n_0}$, $\tilde{n_1}$ and $\tilde{n_2}$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
  Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (2.3.5), (2.3.6) and (2.3.7).
  Calculate Q for the new iterate.
end while

2.3.5 Algorithm-2.5 : Modified Algorithm 4

This variation is of two fold. We update σ_1 and σ_2 from the marginal density until the convergence using fixed point iteration instead of the gradient descent algorithm and carry forward the same EM algorithm. However, at the time of using the EM iterations, we update the σ_1 and σ_2 using one step ahead fixed point iteration along with other parameters. The detailed algorithmic steps are shown below.

Algorithm 2.5 EM procedure for bivariate Pareto distribution

1: Start with some initial choice of σ_1, σ_2, α_0, α_1, α_2.
2: **while** $|(\Delta\sigma)| > tol$ **do**
3: Compute one step ahead fixed point iteration of $\sigma_1^{(i+1)}$, $\sigma_2^{(i+1)}$, using $\sigma_1^{(i)}$, $\sigma_2^{(i)}$, $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
4: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n}_0$, $\tilde{n}_1$ and $\tilde{n}_2$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
5: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (2.3.5), (2.3.6) and (2.3.7).
6: **end while**
7: **while** $|\Delta Q/Q| > tol$ **do**
8: Fix I_0, I_1 and I_2 with last updated estimates of μ_1, μ_2, $\sigma_1^{(i)}$, $\sigma_2^{(i)}$, $\alpha_0^{(i)}$, $\alpha_1^{(i)}$ and $\alpha_2^{(i)}$.
9: Compute one step ahead fixed point iteration of $\sigma_1^{(i+1)}$, $\sigma_2^{(i+1)}$, using $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
10: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n}_0$, $\tilde{n}_1$ and $\tilde{n}_2$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
11: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (2.3.5), (2.3.6) and (2.3.7).
12: Calculate Q for the new iterate.
13: **end while**

All the algorithm should work under the assumption that likelihood is uni-modal. However, that may not be the case. The theoretical proof is very difficult. One common practice is to take many different initial points and run the algorithms. In case of a multimodal set up, the global maximum would be the maximum of all the convergent points. However, fortunately we have observed in our simulation that the point of convergence is always same no matter what initial value we choose.

2.4 Confidence Interval

In this chapter we propose to use the confidence interval in two ways.

1. We obtain the confidence intervals (CI) for the parameters σ_1, σ_2, α_0, α_1 and α_2 by using the parametric bootstrap technique. The parametric bootstrap confidence intervals are simply the 2.5^{th} percentile and 97.5^{th} percentile quantiles of the 1000 simulated estimates of the corresponding parameter in the 1000 simulations. However, the parametric bootstrap confidence interval for μ_1 and μ_2 does not exist. Therefore, we can find the confidence interval for μ_1 and μ_2 separately using the distribution of the estimates.

2. We obtain the asymptotic confidence interval for σ_1, σ_2, α_0, α_1 and α_2. For the confidence interval of μ_1 and μ_2, we use the same distribution of the estimates technique.

Confidence intervals for location parameters can be obtained using the distribution of MLE estimate of the location parameters, $\hat{\mu}_j = X_{(j)} = \min\{X_{j1}, X_{j2}, \cdots, X_{jn}\}$, $j = 1,2$ from marginals. Since, $(X_{(j)} - \mu_j)/\sigma_j$ follows the Pareto type II distribution, PA(II)$(0, 1, n(\alpha_0 + \alpha_j)))$, where the scale parameters and the shape parameters are known, 95% approximate confidence intervals can be written as,

$$X_{(j)} - b_j\hat{\sigma}_j \leq \mu_j \leq X_{(j)} - a_j\hat{\sigma}_j \tag{2.4.1}$$

where $a_j = 0.9750^{-\frac{1}{n(\hat{\alpha}_0+\hat{\alpha}_j)}} - 1$ and $b_j = 0.0250^{-\frac{1}{n(\alpha_0+\alpha_j)}} - 1$. In our case, we use the EM estimate for $\hat{\sigma}_j,\ j = 1,2$ and $\hat{\alpha}_j,\ j = 0,1,2$.

For finding confidence interval for σ_1 and σ_2, we use marginal likelihood, as they are continuous function. We calculate the observed Fisher information matrix for both scale parameters σ_1 and σ_2 respectively. The marginal observed Fisher information matrix is

$$B_{X_j} = \begin{bmatrix} -\frac{n}{\hat{\sigma_j}^2} + (\hat{\alpha_0} + \hat{\alpha_j} + 1)\sum_{i=1}^{n} \frac{(x_{ji}-\hat{\mu_j})(2\hat{\sigma_j}+x_{ji}-\hat{\mu_j})}{\hat{\sigma_j}^2(\hat{\sigma_j}+x_{ji}-\hat{\mu_j})^2} & -\sum_{i=1}^{n} \frac{(x_{ji}-\hat{\mu_j})}{\hat{\sigma_j}(\hat{\sigma_j}+x_{ji}-\hat{\mu_j})} \\ -\sum_{i=1}^{n} \frac{(x_{ji}-\hat{\mu_j})}{\hat{\sigma_j}(\hat{\sigma_j}+x_{ji}-\hat{\mu_j})} & \frac{n}{(\hat{\alpha}_0+\hat{\alpha}_j)^2} \end{bmatrix},$$

where $j = 1,2$. Hence the confidence intervals for σ_1 and σ_2 are

$$\hat{\sigma_j} \pm c\sqrt{(B_{X_j})^{-1}_{jj}}; \quad j = 1,2.$$

We propose to find the confidence interval of α_0, α_1 and α_2 from the Fisher Information matrix calculated based on the bivariate likelihood function. Since estimates of α_0, α_1 and α_2 involves the EM algorithm, we get the Fisher Information matrix as

$$F_{obs} = B - SS',$$

where B is the negative of second derivatives of the pseudo log-likelihood function and S is the corresponding derivative vectors. Below we provide B and S.

$$B = \begin{bmatrix} \frac{n_0+u_1n_1+w_1n_2}{\hat{\alpha_0}^2} & 0 & 0 \\ 0 & \frac{n_1+w_2n_2}{\hat{\alpha_1}^2} & 0 \\ 0 & 0 & \frac{n_2+u_2n_1}{\hat{\alpha_2}^2} \end{bmatrix}$$

and

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix},$$

where

$$S_0 = -\Big(\sum_{I_0}\log(1+y_i) + \sum_{I_2}\log(1+y_{1i}) + \sum_{I_1}\log(1+y_{2i})\Big) + \frac{n_0+u_1n_1+w_1n_2}{\hat{\alpha_0}}$$

$$S_1 = -\Big(\sum_{I_0}\log(1+y_i) + \sum_{I_1\cup I_2}\log(1+y_{1i})\Big) + \frac{n_1+w_2n_2}{\hat{\alpha_1}}$$

$$S_2 = -\Big(\sum_{I_0}\log(1+y_i) + \sum_{I_1\cup I_2}\log(1+y_{2i})\Big) + \frac{n_2+u_2n_1}{\hat{\alpha_2}}.$$

The confidence intervals for shape parameters are

$$\hat{\alpha_j} \pm c\sqrt{(F_{obs}^{-1})_{jj}}; \quad j = 0, 1, 2.$$

where c is the standard normal critical value, for example, $c = 1.96$ for 95% confidence interval. The obtained confidence interval can further be improved by finding asymptotic distribution of $\log(\hat{\alpha}_i)$ and we know that

$$\sqrt{n}(\log(\hat{\alpha}_j) - \log(\alpha_j) \longrightarrow N(0, [\frac{1}{\alpha_j^2}]\sqrt{(F_{obs}^{-1})_{jj}}).$$

Therefore, the alternative approximate asymptotic confidence interval can be suggested as

$$[\hat{\alpha}_j e^{-z_{\frac{\alpha}{2}}\frac{\sqrt{(F_{obs}^{-1})_{jj}}}{\sqrt{n}\hat{\alpha_j}}}, \hat{\alpha}_j e^{z_{\frac{\alpha}{2}}\frac{\sqrt{(F_{obs}^{-1})_{jj}}}{\sqrt{n}\hat{\alpha_j}}}].$$

This procedure is adapted to match the domain of normal distribution and the domain of random variable $\log(\hat{\alpha}_j)$, both of which is $(-\infty, \infty)$.

However, the above suggested procedure is a naive approximate asymptotic confidence interval. It is observed that the coverage probabilities of this approximate asymptotic confidence interval for scale and shape parameters is very low. More research is needed to explore better confidence interval than the above suggested one.

2.5 Numerical Results

We use the programming software R 3.2.3 to perform the estimation procedures. All the programs will be available from the author on request. First, we take four different sets of parameters and observe the average iteration calculated over different sample sizes. We take our sample size as $n = 150, 250, 350, 450$. We can find the results in Table-2.2 based on 1000 replications. Tolerance limit (denoted as "tol") of stopping criteria is taken as 0.00001. We have used the stopping criteria as the absolute value of log-likelihood changes with respect to previous pseudo log-likelihood at each iteration. Results are shown in Table-2.2 provide average iteration by six different approaches. They are (a) Modified Approach 1 (b) Modified Approach 2, which we truncate after 2000 iterations and denote this as Modified Approach 2 (T) (c) Modified Approach 2 without truncating the iteration and denote this as Modified Approach 2 (WT) (d) Modified Approach 3 (e) Modified Approach 4 (f) Asimit's approach. We choose four different sets of parameters to compare the results. These are $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 0.3, \alpha_2 = 1.4$; $\mu_1 = 1, \mu_2 = 2, \sigma_1 = 0.4, \sigma_2 = 0.5, \alpha_0 = 2, \alpha_1 = 1.2, \alpha_2 = 1.4$; $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.4, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1.4$; $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.4, \sigma_2 = 0.5, \alpha_0 = 2, \alpha_1 = 0.4, \alpha_2 = 0.5$. The proposed EM algorithms work for any initial value. However the estimates of μ_1 and μ_2 are always $\hat{\mu}_1 = \min\{X_{1i}; 1, \cdots, n\}$ and $\hat{\mu}_2 = \min\{X_{2i}; 1, \cdots, n\}$ respectively. The original paper of Asimit et al. (2016) [3] uses different stopping criteria and results are provided for large sample sizes. To compare our results, we keep stopping criteria and sample size same across all algorithms. To choose an initial value we have drawn a value uniformly over a certain range for each of the parameters. The algorithms do not depend on the initial choice in all the algorithms except the places where gradient descend cum EM is applied. In that case smaller values of σ_1 and σ_2 (below say 2) as initial value is desired, however, these choices will work for any value as original parameter set. For α_0, α_1 and α_2, there is no restriction in choices.

Average estimates are provided in Table-2.3, when samples are simulated from the bivariate Pareto distribution with the parameters $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 0.3, \alpha_2 = 1.4$. Results of average estimates are shown only by two best approaches (best in the sense of minimum average of iteration) i.e. Modified Approach 1 and Modified Approach 4. Table-2.4 and Table-2.5 show the MSEs for all the procedures and Table-2.6 represents the MSEs based on the log-likelihood values for all the procedures, where samples are generated from the following parameter sets : $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 0.3, \alpha_2 = 1.4$. MSE is an important criteria for selecting the best algorithm. Therefore, MSEs are calculated for all the procedures. We simulate 1000 samples and estimate of the parameters based on the simulated samples. The confidence intervals by two procedures described above are provided in Table-2.7 and Table-2.8 for sample size 450 only. Modified Algorithm 1 comes out as best approach in term of MSE, whereas Modified Algorithm 2 with truncation is the best approach in term of MSEs based on log-likelihood values. Table-2.9 shows the coverage probabilities of the approximate asymptotic confidence interval for the sample size 450 and 1000 based on Modified Algorithm 1.

Important Comments :

1. The mean square error for α_2 is a little higher for all methods.

2. MSEs are more or less same for all methods. In the above results, we have seen Modified Approach 1 provides minimum MSE among all procedures second best in term of average iteration among all the procedures. For large sample size Modified Approach 1 comes out

as the winner for any chosen parameter sets. Performance of Modified Approach 2 with truncation is also worth mentioning, as it provides best performance with respect to minimum MSE based on log-likelihood values and second best performance with respect to minimum MSE.

3. Modified Approach 4 is the best among other approaches in terms of the average iteration. Modified Approach 1 appears to be the second best performer. However, average iteration for Modified Approach 2 with truncation is much higher as compared to Modified Approach 1 and 4.

4. As expected, average iteration is very high in case of Modified Approach 2 (WT).

5. If we compare the algorithms based on minimum MSE and average iteration together, Modified approach 1 stands either winner or closer to the winner.

parameter set		n	150	250	350	450
$\mu_1 = 0$, $\mu_2 = 0$ $\sigma_1 = 1$ $\sigma_2 = 0.5$ $\alpha_0 = 1$, $\alpha_1 = 0.3$, $\alpha_2 = 1.4$	The average number of Iterations (AI)	Modified Approach 1	244	231	223	220
		Modified Approach 2 (WT)	3402	2163	1542	1220
		Modified Approach 2 (T)	1771	1631	1414	1194
		Modified Approach 3	266	249	238	234
		Modified Approach 4	191	162	151	147
		Asimit's Approach	1002	945	909	902
$\mu_1 = 1$, $\mu_2 = 2$ $\sigma_1 = 0.4$ $\sigma_2 = 0.5$ $\alpha_0 = 2$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$	The average number of Iterations (AI)	Modified Approach 1	174	165	159	157
		Modified Approach 2 (WT)	7734	4137	2428	1633
		Modified Approach 2 (T)	1704	1553	1378	1184
		Modified Approach 3	764	551	458	424
		Modified Approach 4	243	185	153	128
		Asimit's Approach	2345	1747	1464	1309
$\mu_1 = 0$, $\mu_2 = 0$ $\sigma_1 = 1.4$ $\sigma_2 = 0.5$ $\alpha_0 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1.4$	The average number of Iterations (AI)	Modified Approach 1	168	155	150	147
		Modified Approach 2 (WT)	7230	4372	3077	2423
		Modified Approach 2 (T)	1617	1738	1732	1697
		Modified Approach 3	265	237	220	209
		Modified Approach 4	123	101	91	83
		Asimit's Approach	897	547	484	453
$\mu_1 = 0$, $\mu_2 = 0$ $\sigma_1 = 1.4$, $\sigma_2 = 0.5$ $\alpha_0 = 2$, $\alpha_1 = 0.4$, $\alpha_2 = 0.5$	The average number of Iterations (AI)	Modified Approach 1	138	129	124	122
		Modified Approach 2 (WT)	2172	2707	2237	1821
		Modified Approach 2 (T)	397	591	773	881
		Modified Approach 3	330	261	238	219
		Modified Approach 4	117	100	96	95
		Asimit's Approach	1145	977	919	875

Table 2.2: Average number of iterations (AI) by six approaches: (a) Modified Approach 1 (b) Modified Approach 2 (WT) (c) Modified Approach 2 (T) (d) Modified Approach 3 (e) Modified Approach 4 (f) Asimit's Approach.

n	parameters	μ_1	μ_2	σ_1	σ_2	α_0	α_1	α_2
150	Average (Mod. 1) estimates (AE)	0.0056	0.0014	0.9031	0.5154	0.8993	0.3379	1.5660
	Average (Mod. 4) estimates (AE)	0.0056	0.0014	1.0521	0.5755	0.9372	0.4122	1.7416
250	Average (Mod. 1) Estimates (AE)	0.0031	0.0008	0.8943	0.4989	0.9013	0.3226	1.5046
	Average (Mod. 4) estimates (AE)	0.0031	0.0008	1.0386	0.5457	0.9408	0.3909	1.6292
350	Average (Mod. 1) Estimates (AE)	0.0022	0.0006	0.8854	0.4878	0.9086	0.3107	1.4550
	Average (Mod. 4) estimates (AE)	0.0022	0.0006	1.0252	0.5283	0.9497	0.3745	1.5543
450	Average (Mod. 1) Estimates (AE)	0.0017	0.0005	0.8719	0.4799	0.9099	0.2955	1.4270
	Average (Mod. 4) estimates (AE)	0.0017	0.0005	1.0077	0.5187	0.9492	0.3586	1.5221

Table 2.3: The average estimates (AE) for $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$ through two best approches i.e. Modified Approach 1 and Modified Approach 4.

	parameters	μ_1	μ_2	σ_1
n=150	MSE (Mod. 1)	6.14e-05	3.69e-06	0.0832
	MSE (Mod. 2 (T))	0.0000614	0.0000037	0.0543
	MSE (Mod. 2 (WT))	0.0000614	0.0000370	0.1043
	MSE (Mod. 3)	0.0000614	0.0000037	0.1216
	MSE (Mod. 4)	0.0000614	0.0000037	0.1179
	MSE (Asimit)	0.0000614	0.0000037	0.1218
n = 150	σ_2	α_0	α_1	α_2
MSE (Mod. 1)	0.0477	0.0402	0.0375	0.5178
MSE (Mod. 2 (T))	0.0368	0.0386	0.0342	0.4154
MSE (Mod. 2 (WT))	0.0682	0.0433	0.0609	0.7969
MSE (Mod. 3)	0.1028	0.0439	0.0737	1.2035
MSE (Mod. 4)	0.1090	0.0429	0.0691	1.2536
MSE (Asimit)	0.1011	0.0433	0.0726	1.1860
	parameters	μ_1	μ_2	σ_1
n = 250	MSE (Mod. 1)	0.000020	1.4757e-06	0.0581
	MSE (Mod. 2 (T))	0.0000614	0.00000370	0.0543
	MSE (Mod. 2 (WT))	0.0000198	0.0000014	0.0682
	MSE (Mod. 3)	0.00001978	0.00000147	0.0745
	MSE (Mod. 4)	0.00001978	0.00000147	0.0724
	MSE (Asimit)	0.00001978	0.00000148	0.0746
n = 250	σ_2	α_0	α_1	α_2
MSE (Mod. 1)	0.0278	0.0296	0.0205	0.2817
MSE (Mod. 2 (T))	0.0368	0.0386	0.0342	0.4153
MSE (Mod. 2 (WT))	0.0410	0.0294	0.0386	0.4441
MSE (Mod. 3)	0.0461	0.0296	0.0431	0.5033
MSE (Mod. 4)	0.0460	0.0289	0.0401	0.4978
MSE (Asimit)	0.0462	0.0292	0.0427	0.5638

Table 2.4: Mean Square Error (MSE) through all approaches when samples are generated from $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$.

	parameters	μ_1	μ_2	σ_1
n = 350	MSE (Mod. 1)	9.236e-06	6.992e-07	0.0427
	MSE (Mod. 2 (T))	0.00000197	0.000000699	0.0389
	MSE (Mod. 2 (WT))	0.00000924	0.000000699	0.0428
	MSE (Mod. 3)	0.00000923	0.000000699	0.0456
	MSE (Mod. 4)	0.00000924	0.000000699	0.0440
	MSE (Asimit)	0.00000923	0.000000699	0.0454
n=350	σ_2	α_0	α_1	α_2
MSE (Mod. 1)	0.0155	0.0211	0.0143	0.1475
MSE (Mod. 2 (T))	0.02001	0.01907	0.02573	0.2068
MSE (Mod. 2 (WT))	0.0213	0.0192	0.0280	0.2204
MSE (Mod. 3)	0.0221	0.0194	0.0302	0.2290
MSE (Mod. 4)	0.0220	0.0188	0.0275	0.2252
MSE (Asimit)	0.0222	0.0194	0.0301	0.2290
	parameters	μ_1	μ_2	σ_1
n = 450	MSE (Mod. 1)	5.589e-06	4.398e-07	0.0366
	MSE (Mod. 2 (T))	0.0000055	0.000000440	0.0283
	MSE (Mod. 2 (WT))	0.00000559	0.00000044	0.0288
	MSE (Mod. 3)	0.00000559	0.00000047	0.03008
	MSE (Mod. 4)	0.00000558	0.000000439	0.0292
	MSE (Asimit)	0.00000559	0.000000440	0.0300
n=450	σ_2	α_0	α_1	α_2
MSE (Mod. 1)	0.0107	0.0176	0.0097	0.1140
MSE (Mod. 2 (T))	0.0143	0.015	0.0181	0.1655
MSE (Mod. 2 (WT))	0.0147	0.0151	0.0184	0.1717
MSE (Mod. 3)	0.0151	0.0152	0.0196	0.1759
MSE (Mod. 4)	0.0150	0.0148	0.0179	0.1736
MSE (Asimit)	0.0151	0.0151	0.0195	0.1756

Table 2.5: Mean Square Error (MSE) through all approaches when samples are generated from $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$.

Sample Size	Mod. Algorithm 1	Mod. Algorithm 2 (T)	Mod. Algorithm 2 (WT)	Mod. Algorithm 3	Mod. Algorithm 4	Asimit Approach
150	2712.568	2387.443	2435.494	2589.999	2679.931	2647.673
250	6878.606	6085.97	6137.539	6573.093	6689.585	7009.098
350	12769.37	11487.95	11513.61	11832.11	11781.32	12748.90
450	20491.31	17925.16	17932.16	18087.06	18341.27	19940.82

Table 2.6: Mean Square Error (MSE) based on log-likelihood values through all approaches when samples are generated from $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$.

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
μ_1	[-0.0044, 0.0003]	[-0.0040, 0.0003]	[-0.0039, 0.0003]	[-0.0039, 0.0003]
μ_2	[-0.0017, 0.00004]	[-0.0015, 0.00004]	[-0.0013, 0.00004]	[-0.0013, 0.00004]
σ_1	[0.6257, 1.1722]	[0.6119, 1.2963]	[0.6520, 1.3849]	[0.7136, 1.3764]
σ_2	[0.3183, 0.7207]	[0.3043, 0.7987]	[0.3158, 0.8060]	[0.3373, 0.8037]
α_0	[0.7331, 1.1046]	[0.7674, 1.1759]	[0.7620, 1.1791]	[0.7563, 1.1792]
α_1	[0.1658, 0.5370]	[0.1155, 0.5792]	[0.1508, 0.6585]	[0.1893, 0.6740]
α_2	[0.9771, 2.2410]	[0.8716, 2.4551]	[0.9058, 2.4801]	[1.0284, 2.4710]

Table 2.7: 95% Confidence Intervals (CI) for $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$ by first approach (includes parametric bootstap).

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
μ_1	[-0.0054, 0.0014]	[-0.0049, 0.0003]	[-0.0048, 0.0014]	[-0.0048, 0.0014]
μ_2	[-0.0029, 0.0003]	[-0.0027, 0.0003]	[-0.0027, 0.0003]	[-0.0027, 0.0003]
σ_1	[0.6296, 1.2371]	[0.7019, 1.4664]	[0.7083, 1.4894]	[0.7064, 1.4822]
σ_2	[0.2874, 0.7484]	[0.2559, 0.8588]	[0.2959, 0.8583]	[0.2958, 0.8556]
α_0	[0.8502, 1.0426]	[0.9110, 1.1185]	[0.9174, 1.1263]	[0.9182, 1.1270]
α_1	[0.1968, 0.2972]	[0.2267, 0.3387]	[0.2291, 0.3419]	[0.2252, 0.3371]
α_2	[1.4428, 1.7372]	[1.5769, 1.8830]	[1.5689, 1.8749]	[1.5631, 1.8689]

Table 2.8: 95% Confidence Intervals (CI) for $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$ by second procedure (does not include parametric bootstap).

Parameters \ Sample Size	450	1000
μ_1	0.957	0.951
μ_2	0.961	0.955
σ_1	0.589	0.566
σ_2	0.417	0.423
α_0	0.638	0.614
α_1	0.58	0.564
α_2	0.414	0.442

Table 2.9: Coverage probabilities of 95% approximate asymptotic confidence interval for $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\alpha_2 = 1.4$ based on Modified Algorithm 1.

2.6 Data Analysis

We analyze a data set of the indemnity payments (Loss) and the allocated loss adjustment expense (ALAE) relating to 1500 general liability claims from insurance companies are available in the R package evd ([54]). From Falk and Guillou (2008) [15], we know that the peak over threshold method on random variable U (say) provides polynomial generalized Pareto distribution for any x_0 with $1 + log(G(x_0)) \in (0, 1)$ i.e. $P(U > tx_0|U > x_0) = t^{-\alpha}$, $t \geq 1$, where $G(\cdot)$ is the distribution function of U. We choose an appropriate t and x_0 on the marginals of the bivariate data, so that data should behave like near Pareto distribution. A brief summary of the transformed data can be provided as

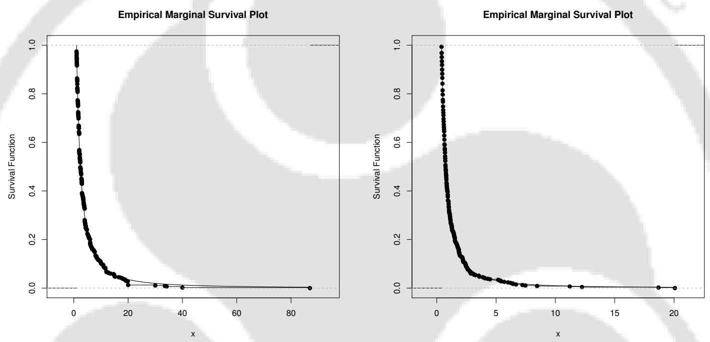
	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
Marginal of X_1	1.000	1.590	2.400	4.550	4.407	86.940
Marginal of X_2	0.400	0.530	0.745	1.279	1.272	20.070

Table 2.10: Brief summary of transformed data for MOBVPA distribution.

We assume the data follows near equal to singular Marshall Olkin bivariate Pareto and try to verify our assumption. We fit the empirical survival functions with the marginals of this bivariate Pareto, whose parameters can be obtained from the EM algorithm that we have developed. The

Figure-2.3 shows a good fit for both the marginals. The original data has sample size 1500, whereas after transformation size of the data is 468. The final number of data points in each cell I_0, I_1 and I_2 depend on what methods we choose. For example, $n_0 = 145$, $n_1 = 146$ and $n_2 = 177$ for the modified approach 1; For approach 2, $n_0 = 133$, $n_1 = 138$, $n_2 = 197$; approach 3 estimates provide $n_0 = 132$, $n_1 = 163$ and $n_2 = 173$, whereas approach 4 sets $n_0 = 139$, $n_1 = 160$, $n_2 = 169$.

Empirical two dimensional density plot in Figure-2.4 verifies that the Marshall-Olkin bivariate Pareto can be an alternative model for the transformed data set. Parameter estimates of this bivariate distribution are provided in Table-2.14 and Table-2.15. The mean square errors are also calculated using the parametric bootstrap technique and provided in Table-2.16 and Table-2.17. During the parametric bootstrap, sample sizes are kept same as the size of the data. The confidence intervals by using both of the procedures are provided in Table-2.18 and Table-2.19.



(a) Marginal for X_1 (b) Marginal for X_2

Figure 2.3: Survival plots for two marginals of MOBVPA of the transformed dataset.

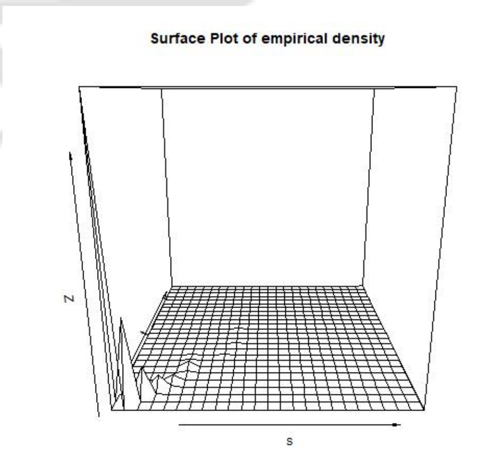


Figure 2.4: Two dimensional density plots for MOBVPA of the transformed dataset.

Apart from graphical diagnostics, we can also show the goodness of fit test for the distribution of marginals and distribution of a minimum of the components as univariate Pareto distribution. It is observed that $W = \min\{\frac{X_1-\mu_1}{\sigma_1}, \frac{X_2-\mu_2}{\sigma_2}\}$ has univariate Pareto distribution, i.e. $W \sim PA(II)(0, 1, \alpha_0+\alpha_1+\alpha_2)$. We can use this property to validate the actual distribution as the Marshall-Olkin singular bivariate Pareto. For the reason, we find the Kolmogorov-Smirnov (K-S) distance between the empirical distribution function of W based on transformed data and fitted reference distribution function, which is given in Table 2.11. Table-2.12 and Table-2.13 represent the K-S distance and p-value of the marginals of X_1 and X_2 respectively.

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
K-S distance	0.074786	0.07906	0.083333	0.081197
p-value	0.1459	0.1073	0.08754	0.09141

Table 2.11: K-S distance (or K-S statistic value) and p-value related to W based on data.

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
K-S distance	0.047009	0.059829	0.057692	0.057692
p-value	0.6793	0.3721	0.4173	0.4173

Table 2.12: K-S distance (or K-S statistic value) and p-value for marginal of X_1 based on data.

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
K-S distance	0.053419	0.059829	0.059829	0.049145
p-value	0.5165	0.3721	0.3721	0.6242

Table 2.13: K-S distance (or K-S statistic value) and p-value for marginal of X_2 based on data.

Parameters \ procedures	μ_1	μ_2	σ_1
Mod. 1	1	0.4	2.5594
Mod. 2 (T)	1	0.4	2.4936
Mod. 2 (WT)	1	0.4	2.9459
Mod. 3	1	0.4	3.2459
Mod. 4	1	0.4	3.1802

Table 2.14: EM estimates of the parameters - Data Analysis.

Parameters \ procedures	σ_2	α_0	α_1	α_2
Mod. 1	0.7176	0.7557	0.8162	0.9792
Mod. 2 (T)	0.8164	0.759	0.7878	1.1267
Mod. 2 (WT)	0.8162	0.7960	0.9209	1.0895
Mod. 3	0.8163	0.8171	1.0100	1.0684
Mod. 4	0.7947	0.8330	0.9667	1.0158

Table 2.15: EM estimates of the parameters - Data Analysis.

Parameters \ procedures	μ_1	μ_2	σ_1
Mod. 1	2.34×10^{-5}	1.398×10^{-6}	0.301
Mod. 2 (T)	2.296×10^{-5}	1.513×10^{-6}	0.120
Mod. 2 (WT)	2.598×10^{-5}	1.522×10^{-6}	0.236
Mod. 3	2.777×10^{-5}	1.529×10^{-6}	0.436
Mod. 4	2.756×10^{-5}	1.525×10^{-6}	0.413

Table 2.16: MSEs of the parameters through parametric bootstrap - Data Analysis.

Parameters \ procedures	σ_2	α_0	α_1	α_2
Mod. 1	0.018	0.012	0.020	0.031
Mod. 2 (T)	0.027	0.010	0.012	0.066
Mod. 2 (WT)	0.028	0.012	0.023	0.068
Mod. 3	0.028	0.013	0.057	0.062
Mod. 4	0.026	0.012	0.053	0.056

Table 2.17: MSEs of the parameters through parametric bootstrap - Data Analysis.

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
μ_1	[0.9962, 1.0001]	[0.9993, 1.0036]	[0.9975, 1.0001]	[0.9974, 1.0001]
μ_2	[0.3970, 0.4001]	[0.3976, 0.40001]	[0.3979, 0.4001]	[0.3977, 0.4001]
σ_1	[1.5793, 2.9953]	[1.7689, 2.7684]	[2.3573, 4.9268]	[2.3343, 4.7869]
σ_2	[0.4554, 0.9018]	[0.5711, 1.1713]	[0.5731, 1.2015]	[0.5659, 1.1673]
α_0	[0.5424, 0.8405]	[0.5482, 0.8790]	[0.5870, 0.9988]	[0.6090, 1.0025]
α_1	[0.5489, 1.0409]	[0.5769, 0.9780]	[0.7776, 1.6502]	[0.7526, 1.5866]
α_2	[0.6926, 1.3674]	[0.8699, 1.7688]	[0.7995, 1.7044]	[0.7770, 1.6220]

Table 2.18: Type-I procedure (does not include parametric bootstrap) : 95% confidence intervals (CI) for μ_1, μ_2, σ_1, σ_2, α_0, α_1 and α_2 based on data.

	Mod. Algorithm 1	Mod. Algorithm 2	Mod. Algorithm 3	Mod. Algorithm 4
μ_1	[0.9950, 1.0000]	[0.9949, 1.0000]	[0.9957, 1.0000]	[0.9956, 1.0000]
μ_2	[0.3954, 0.4000]	[0.3958, 0.4000]	[0.39582, 0.4000]	[0.3957, 0.4000]
σ_1	[1.6959, 3.4230]	[1.6641, 3.3232]	[1.9704, 4.5205]	[1.9493, 4.4110]
σ_2	[0.4719, 0.9634]	[0.5116, 1.1213]	[0.5115, 1.1211]	[0.5036, 1.0858]
α_0	[0.6895, 0.8659]	[0.6710, 0.8472]	[0.7214, 0.9128]	[0.7371, 0.9290]
α_1	[0.6894, 0.8913]	[0.6878, 0.8879]	[0.8868, 1.1329]	[0.8471, 1.0863]
α_2	[0.8343, 1.0730]	[1.0023, 1.2512]	[0.9440, 1.1928]	[0.8925, 1.1392]

Table 2.19: Type-II procedure (includes parametric bootstrap): 95% confidence intervals (CI) for μ_1, μ_2, σ_1, σ_2, α_0, α_1 and α_2 based on the data set.

2.7 Conclusion

We use different variations of the EM algorithm for estimating the parameters of the singular Marshall-Olkin bivariate Pareto distribution. All variations work quite well, even for moderate sample size (say, 150, 250 etc.). Some of our proposed algorithms outperform the current state-of-art algorithm by Asimit et al. The data analysis makes this work more interesting. We discuss the implementation of all the algorithms in real-life data, which was absent in earlier works. We have formulated the approximate asymptotic confidence interval for scale and shape parameters. However, more research is needed in that direction. The content of this work opens the door of research in the implementation of the EM algorithm for other complicated higher dimensional distributions with location and scale parameters. It can also be used in discrimination of several distributions, which includes MOBVPA distribution. More work is needed in this direction too. The implementation of the EM algorithm on the absolute continuous part of this distribution is an another interesting problem which is discussed in the next chapter.

Chapter 3

An EM algorithm for the absolute continuous bivariate Pareto distribution

3.1 Introduction

Recently, Asimit, A. V., Furman, E. and Vernic, R (2016)[3] used the EM algorithm to estimate the singular Marshall-Olkin bivariate Pareto distribution. However, there is no paper available for formulation and estimation of the absolute continuous version of this distribution. A few recent works (Kundu D. and Gupta RD. (2009) [27], Kundu D. and Gupta RD. (2010) [28], Kundu D, Kumar A and Gupta AK. (2015)[30], Mirhosseini SM, Amini M, Kundu D and Dolati A. (2015)[37] include the Marshall-Olkin bivariate distribution. Estimation through the EM algorithm can be seen there. But in the estimation of the absolute continuous version of the bivariate Pareto through the EM algorithm is not straight forward. The problem becomes more complicated when we consider the location and scale parameters in our problem. Usual gradient descent does not work as the bivariate likelihood is a discontinuous function with respect to the location and scale parameters. We suggest a novel way to handle all related computational problems in the most efficient way. The first contribution of this chapter is on three parameter implementation of EM where a crucial modification is suggested to make the algorithm work for any range of parameter. The other contribution is an exploration of an efficient EM algorithm which will work in case of the seven parameter setup. The suggested structure is a combination of several ideas, which definitely includes previous suggested modification of three parameter setup. The distribution can be used as an alternative model for the data transformed via peak over threshold method. Bivariate Pareto has a wide application in modeling data related to finance, insurance, environmental sciences and internet network, etc. The dependence structure of the absolute continuous version of the bivariate Pareto can also be described by well-known Marshall-Olkin copula [Nelsen, R. B. (2007) [39], Marshall, A. W. (1996)[34]]. The methodology proposed in this chapter works for moderately large sample.

Block H. and Basu AP (1974) [9] proposed an absolute continuous bivariate exponential distribution. Later Kundu D. and Gupta RD. (2011) [29] introduced an absolute continuous bivariate generalized exponential distribution, whose marginals are not the generalized exponential distributions. Many papers related to multivariate extreme value theory talk about absolute continuous/ singular bivariate and multivariate Pareto distribution [Yeh, H. C. (2000) [55], Yeh, H. C.(2004) [56], Smith, R. L. (1994) [53], Rakonczai, P. and Zempl eni, A. (2012) [47]]. None of the mentioned paper discussed anything about **Marshall-Olkin form** of the distribution or its estimation through the EM algorithm in the presence of location and scale parameters. A novel EM cum Gradient descent is the major contribution of this chapter. However, we address a few modifications in the implementation of EM even in case of three parameter setup. The chapter contains a mix of several issues and implementation issues which together provides our final proposal algorithm. We show that our proposed algorithm works quite well for the moderately large sample size. We also propose to use a naive construction of confidence interval for the parameters. A formulation of trivariate extension is discussed at the end of this chapter.

We arrange this chapter in the following way. In section-3.2, we keep the formulation of the Marshall-Olkin bivariate Pareto and its absolute continuous version. In section-3.3, we describe the EM algorithm for singular case. Different extension of the EM algorithm for the absolute continuous version is available in section-3.4. We discuss about the confidence interval in Section-3.5. Some simulation results show the performance of the algorithm in section-3.6. In section-3.7, we show data analysis. We give a brief introduction about singular and absolute continuous Marshall-Olkin trivariate Pareto distribution in section-3.8. Finally, we conclude this chapter in Section-3.9.

3.2 Formulation of Marshal-Olkin bivariate Pareto

A random variable X is said to have Pareto of second kind, i.e. $X \sim Pa(II)(\mu, \sigma, \alpha)$ if it has the survival function

$$\bar{F}_X(x; \mu, \sigma, \alpha) = P(X > x) = (1 + \frac{x - \mu}{\sigma})^{-\alpha}$$

and the probability density function (pdf)

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma}(1 + \frac{x - \mu}{\sigma})^{-\alpha - 1}$$

with $x > \mu \in \mathbb{R}$, $\sigma > 0$ and $\alpha > 0$.

Let us consider U_0, U_1 and U_2 be three independent univariate pareto distributions $Pa(II)(0, 1, \alpha_0)$, $Pa(II)(\mu_1, \sigma_1, \alpha_1)$ and $Pa(II)(\mu_2, \sigma_2, \alpha_2)$.

We define $X_1 = \min\{\sigma_1 U_0 + \mu_1, U_1\}$ and $X_2 = \min\{\sigma_2 U_0 + \mu_2, U_2\}$. We can show that (X_1, X_2) jointly follow the Marshall-Olkin bivariate Pareto distribution of second kind and we denote it as

$BVPA(\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$.

The joint distribution can be given by

$$f(x_1, x_2) = \begin{cases} f_1(x_1, x_2) & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1, x_2) & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x) & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x, \end{cases}$$

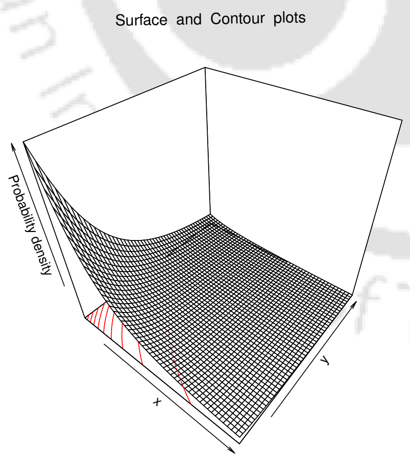
where

$$f_1(x_1, x_2) = \frac{\alpha_1}{\sigma_1\sigma_2}(\alpha_0 + \alpha_2)(1 + \frac{(x_2 - \mu_2)}{\sigma_2})^{-(\alpha_0+\alpha_2+1)}(1 + \frac{(x_1 - \mu_1)}{\sigma_1})^{-(\alpha_1+1)}$$

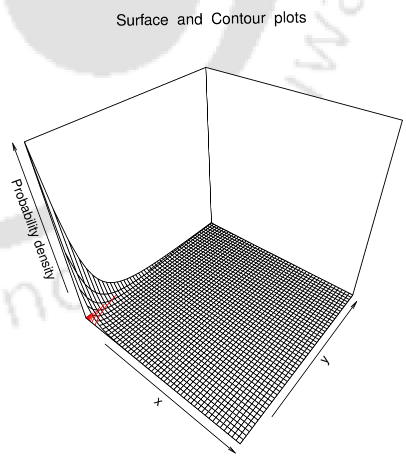
$$f_2(x_1, x_2) = \frac{\alpha_2}{\sigma_1\sigma_2}(\alpha_0 + \alpha_1)(1 + \frac{x_1 - \mu_1}{\sigma_1})^{-(\alpha_0+\alpha_1+1)}(1 + \frac{x_2 - \mu_2}{\sigma_2})^{-(\alpha_2+1)}$$

$$f_0(x) = \alpha_0(1 + x)^{-(\alpha_1+\alpha_2+\alpha_0+1)}.$$

We denote the absolute continuous part as BVPAC. Surface and contour plots of the absolute continuous part of the pdf are shown in Figure-3.1 and Figure-3.2 respectively. The following four different sets of parameters provide four subfigures in each figure. $\xi_1 : \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 0.3, \alpha_2 = 1.4$; $\xi_2 : \mu_1 = 1, \mu_2 = 2, \sigma_1 = 0.4, \sigma_2 = 0.5, \alpha_0 = 2, \alpha_1 = 1.2, \alpha_2 = 1.4$; $\xi_3 : \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.4, \sigma_2 = 0.5, \alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1.4$; $\xi_4 : \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.4, \sigma_2 = 0.5, \alpha_0 = 2, \alpha_1 = 0.4, \alpha_2 = 0.5$.

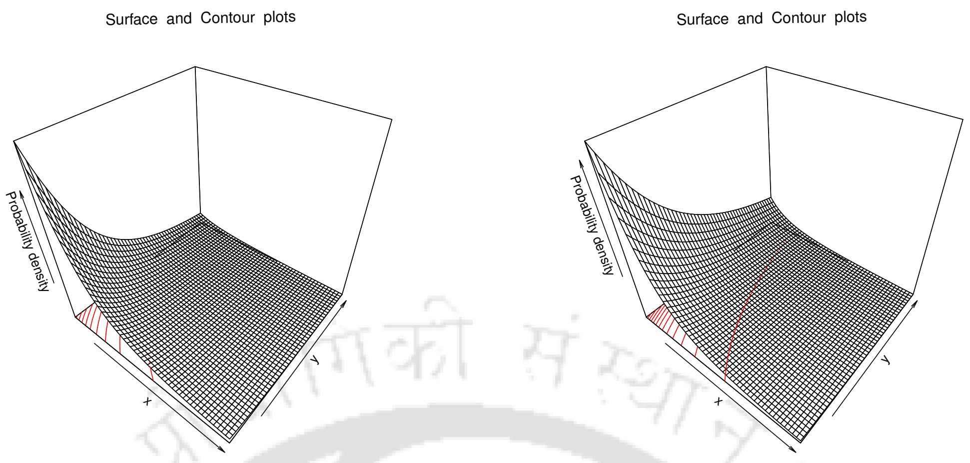


(a) ξ_1



(b) ξ_2

Figure 3.1: Surface and Contour plots of pdf of BVPAC wrt parameter set ξ_1 and ξ_2



(a) ξ_3 (b) ξ_4

Figure 3.2: Surface and Contour plots for pdf of BVPAC wrt parameter set ξ_3 and ξ_4

3.2.1 Absolute Continuous Extension

We assume that the random vector (Z_1, Z_2) follows the absolute continuous bivariate Pareto distribution (as described in Kundu D. and Gupta RD. (2010) [28]).

The joint pdf of BVPAC can be written as

$$f_{BVPAC}(z_1, z_2) = \begin{cases} cf_1(z_1, z_2) = cf_{PA}(z_1; \mu_1, \sigma_1, \alpha_1) f_{PA}(z_2; \mu_2, \sigma_2, \alpha_0 + \alpha_2) & \text{if } \frac{(z_1-\mu_1)}{\sigma_1} < \frac{(z_2-\mu_2)}{\sigma_2} \\ cf_2(z_1, z_2) = cf_{PA}(z_1; \mu_1, \sigma_1, \alpha_0 + \alpha_1) f_{PA}(z_2; \mu_2, \sigma_2, \alpha_2) & \text{if } \frac{(z_1-\mu_1)}{\sigma_1} > \frac{(z_2-\mu_2)}{\sigma_2}, \end{cases}$$

where c is a normalizing constant and $c = \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}$ and $f_{PA}(\cdot)$ denotes the pdf of univariate distribution.

Log-likelihood function for the parameters of the BVPAC can be written as

$$\begin{aligned} L(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2) &= n \ln(\alpha_0 + \alpha_1 + \alpha_2) - n \ln(\alpha_1 + \alpha_2) + n_1 \ln \alpha_1 + n_1 \ln(\alpha_0 + \alpha_2) \\ &- n_1 \ln \sigma_1 - n_1 \ln \sigma_2 - (\alpha_0 + \alpha_2 + 1) \sum_{i \in I_1} \ln\left(1 + \frac{z_{2i} - \mu_2}{\sigma_2}\right) \\ &- (\alpha_1 + 1) \sum_{i \in I_1} \ln\left(1 + \frac{z_{1i} - \mu_1}{\sigma_1}\right) - n_2 \ln \sigma_1 - n_2 \ln \sigma_2 + n_2 \ln \alpha_2 \\ &+ n_2 \ln(\alpha_0 + \alpha_1) - (\alpha_0 + \alpha_1 + 1) \sum_{i \in I_2} \ln\left(1 + \frac{z_{1i} - \mu_1}{\sigma_1}\right) \\ &- (\alpha_2 + 1) \sum_{i \in I_2} \ln\left(1 + \frac{z_{2i} - \mu_2}{\sigma_2}\right) \end{aligned}$$

The log-likelihood function for the three parameters case, taking $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ corresponding to this pdf can be given by

$$\begin{aligned}
L(\alpha_0, \alpha_1, \alpha_2) &= (n_1 + n_2) \ln(\alpha_0 + \alpha_1 + \alpha_2) - (n_1 + n_2) \ln(\alpha_1 + \alpha_2) \\
&+ \sum_{i \in I_1} \ln f_{PA}(z_{1i}; 0, 1, \alpha_1) + \sum_{i \in I_1} \ln f_{PA}(z_{2i}; 0, 1, \alpha_0 + \alpha_2) \\
&+ \sum_{i \in I_2} \ln f_{PA}(z_{1i}; 0, 1, \alpha_0 + \alpha_1) + \sum_{i \in I_2} \ln f_{PA}(z_{2i}; 0, 1, \alpha_2) \\
&= (n_1 + n_2) \ln(\alpha_0 + \alpha_1 + \alpha_2) - (n_1 + n_2) \ln(\alpha_1 + \alpha_2) \\
&+ n_1 \ln \alpha_1 - (\alpha_1 + 1) \sum_{i \in I_1} \ln(1 + z_{1i}) + n_1 \ln(\alpha_0 + \alpha_2) \\
&- (\alpha_0 + \alpha_2 + 1) \sum_{i \in I_1} \ln(1 + z_{2i}) + n_2 \ln(\alpha_0 + \alpha_1) \\
&- (\alpha_0 + \alpha_1 + 1) \sum_{i \in I_2} \ln(1 + z_{1i}) + n_2 \ln \alpha_2 \\
&- (\alpha_2 + 1) \sum_{i \in I_2} \ln(1 + z_{2i})
\end{aligned}$$

Marginal distribution of Z_1 and Z_2 can be obtained by routine calculation. Expressions are as follows :

$$\begin{aligned}
& f_{Z_1}(z_1; \mu_1, \sigma_1, \alpha_0 + \alpha_1) \\
= \; & c f_{PA}(z_1; \mu_1, \sigma_1, \alpha_0 + \alpha_1) - c \frac{\alpha_0}{(\alpha_0 + \alpha_1 + \alpha_2)} f_{PA}(z_1; \mu_1, \sigma_1, \alpha_0 + \alpha_1 + \alpha_2)
\end{aligned}$$

$$\begin{aligned}
& f_{Z_2}(z_2; \mu_2, \sigma_2, \alpha_0 + \alpha_2) \\
= \; & c f_{PA}(z_2; \mu_2, \sigma_2, \alpha_0 + \alpha_2) - c \frac{\alpha_0}{(\alpha_0 + \alpha_1 + \alpha_2)} f_{PA}(z_2; \mu_2, \sigma_2, \alpha_0 + \alpha_1 + \alpha_2)
\end{aligned}$$

respectively, where $c = \frac{(\alpha_0 + \alpha_1 + \alpha_2)}{(\alpha_1 + \alpha_2)}$.

3.3 EM-algorithm for singular three parameter BVPA distribution

Now we divide our data into three parts :-

$I_0 = \{i : x_{1i} = x_{2i}\}$,$I_1 = \{i : x_{1i} < x_{2i}\}$, $I_2 = \{i : x_{1i} > x_{2i}\}$.

We do not know if X_1 is U_0 or U_1. We also do not know if X_2 is U_0 or U_2. So we introduce two new random variables $(\Delta_1; \Delta_2)$ as

$$\Delta_1 = \begin{cases} 0 & \text{if } X_1 = U_0 \\ 1 & \text{if } X_1 = U_1 \end{cases} \quad \text{and} \quad \Delta_2 = \begin{cases} 0 & \text{if } X_2 = U_0 \\ 2 & \text{if } X_2 = U_2 \end{cases}$$

The Δ_1 and Δ_2 are the missing value of the E-M algorithm. To calculate the E-step we need to calculate the conditional distribution of Δ_1 and Δ_2.

Using the definition of X_1, X_2 and Δ_1, Δ_2 we have :

$\triangleright$ For group I_0, both Δ_1 and Δ_2 are known,

$$\Delta_1 = \Delta_2 = 0$$

$\triangleright$ For group I_1, Δ_1 is known,Δ_2 is unknown,

$$\Delta_1 = 1, \Delta_2 = 0 \text{ or } 2$$

Therefore, we need to find out $u_1 = P(\Delta_2 = 0|I_1)$and $u_2 = P(\Delta_2 = 2|I_1)$.

$\triangleright$ For group I_2, Δ_2 is known, Δ_1 is unknown,

$$\Delta_1 = 0 \text{ or } 1, \Delta_2 = 2$$

Moreover, we need $w_1 = P(\Delta_1 = 0|I_2)$ and $w_2 = P(\Delta_1 = 1|I_2)$.

Since, each of the posterior probability corresponds to one of the ordering from Table-3.1, we calculate u_1, u_2, w_1 and w_2 using the probability of appropriate ordering.

Ordering	(X_1, X_2)	Group
$U_0 < U_1 < U_2$	(U_0, U_0)	I_0
$U_0 < U_2 < U_1$	(U_0, U_0)	I_0
$U_1 < U_0 < U_2$	(U_1, U_0)	I_1
$U_1 < U_2 < U_0$	(U_1, U_2)	I_1
$U_2 < U_0 < U_1$	(U_0, U_2)	I_2
$U_2 < U_1 < U_0$	(U_1, U_2)	I_2

Table 3.1: Groups and corresponding orderings of hidden random variables U_0, U_1 and U_2.

We have the following expressions for u_1, u_2, w_1, w_2

$$u_1 = \frac{P(U_1 < U_0 < U_2)}{P(U_1 < U_0 < U_2) + P(U_1 < U_2 < U_0)}$$

$$u_2 = \frac{P(U_1 < U_2 < U_0)}{P(U_1 < U_0 < U_2) + P(U_1 < U_2 < U_0)}$$

$$w_1 = \frac{P(U_2 < U_0 < U_1)}{P(U_2 < U_0 < U_1) + P(U_2 < U_1 < U_0)}$$

$$w_2 = \frac{P(U_2 < U_1 < U_0)}{P(U_2 < U_0 < U_1) + P(U_2 < U_1 < U_0)}$$

Now we have

$$P(U_1 < U_0 < U_2) = \int_0^\infty [1 - (1+x)^{-\alpha_1}]\alpha_0(1+x)^{-(\alpha_0+1)}(1+x)^{-(\alpha_2)}$$

$$= \alpha_0 \int_0^{\infty} (1+x)^{-(\alpha_0+\alpha_2+1)} - (1+x)^{-(\alpha_0+\alpha_1+\alpha_2+1)}$$
$$= \frac{\alpha_0\alpha_1}{(\alpha_0+\alpha_2+1)(\alpha_0+\alpha_1+\alpha_2+1)}.$$

Using this above result we evaluate the other probabilities to get values of u_1, u_2, w_1, w_2 as : $u_1 = \frac{\alpha_0}{\alpha_0+\alpha_2}$ and $u_2 = \frac{\alpha_2}{\alpha_0+\alpha_2}$ $w_1 = \frac{\alpha_0}{\alpha_0+\alpha_1}$ and $w_2 = \frac{\alpha_1}{\alpha_0+\alpha_1}$.

3.3.1 Pseudo-likelihood expression

We define n_0, n_1, n_2 as : $n_0 = |I_0|, n_1 = |I_1|, n_2 = |I_2|$. where $|I_j|$ for $j = 0, 1, 2$ denotes the number of elements in the set I_j. Now the pseudo log-likelihood can be written down as

$$\begin{aligned} L^*(\alpha_0,\alpha_1,\alpha_2) &= -\alpha_0(\sum_{i\in I_0}\ln(1+x_i)+\sum_{i\in I_2}\ln(1+x_{1i})+\sum_{i\in I_1}\ln(1+x_{2i})) \\ &+ (n_0+u_1n_1+w_1n_2)\ln\alpha_0-\alpha_1(\sum_{i\in I_0}\ln(1+x_i)+\sum_{i\in I_1\cup I_2}\ln(1+x_{1i})) \\ &+ (n_1+w_2n_2)\ln\alpha_1-\alpha_2(\sum_{i\in I_0}\ln(1+x_i)+\sum_{i\in I_1\cup I_2}\ln(1+x_{2i})) \\ &+ (n_2+u_2n_1)\ln\alpha_2. \end{aligned} \tag{3.3.1}$$

Therefore, M-step involves maximizing (3.3.1) with respect to α_0, α_1, α_2 at

$$\hat{\alpha}_0^{(t+1)} = \frac{n_0+u_1^{(t)}n_1+w_1^{(t)}n_2}{\sum_{i\in I_0}\ln(1+x_i)+\sum_{i\in I_2}\ln(1+x_{1i})+\sum_{i\in I_1}\ln(1+x_{2i})}, \tag{3.3.2}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(n_1+w_2^{(t)}n_2)}{\sum_{i\in I_0}\ln(1+x_i)+\sum_{i\in(I_1\cup I_2)}\ln(1+x_{1i})}, \tag{3.3.3}$$

$$\hat{\alpha}_2^{(t+1)} = \frac{n_2+u_2^{(t)}n_1}{(\sum_{i\in I_0}\ln(1+x_i)+\sum_{i\in(I_1\cup I_2)}\ln(1+x_{2i}))}. \tag{3.3.4}$$

Therefore, the algorithm can be given as

Algorithm 3.1 EM procedure for bivariate Pareto distribution

1: **while** $\Delta Q/Q < tol$ **do**
2: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
3: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (3.3.2), (3.3.3) and (3.3.4).
4: Calculate Q for the new iterate.
5: **end while**

3.4 EM-algorithm for absolute continuous case

We adopt similar existing methods described in Kundu D. and Gupta RD.(2010) [28], we replace the missing observations n_0 (number of observations in I_0) and each observation U_0 falling within I_0 by its estimate $\tilde{n_0}$ and $a = E(U_0|U_0 < \min\{U_1, U_2\})$ respectively. Note that in this case, n_0 is a random variable which has the negative binomial with parameters $(n_1 + n_2)$ and $\frac{\alpha_1+\alpha_2}{\alpha_0+\alpha_1+\alpha_2}$.

Clearly, $\tilde{n_0} = (n_1 + n_2)\frac{\alpha_0}{\alpha_1+\alpha_2}$, $a = E(U_0|U_0 < \min\{U_1, U_2\}) = \frac{1}{(\alpha_0+\alpha_1+\alpha_2-1)}$

An important restriction for this approximation is that we have to ensure $\alpha_0 + \alpha_1 + \alpha_2 > 1$. The restriction will ensure the existence of the above expectation.

Therefore, under the above mentioned restriction, we write the pseudo log-likelihood function by replacing the missing observations with its expected value.

$$\begin{aligned}
L^*(\alpha_0, \alpha_1, \alpha_2) \;&=\; -\alpha_0(\tilde{n_0}\ln(1+a) + \sum_{i\in I_2}\ln(1+x_{1i}) + \sum_{i\in I_1}\ln(1+x_{2i})) \\
&+\; (\tilde{n_0} + u_1 n_1 + w_1 n_2)\ln\alpha_0 - \alpha_1(\tilde{n_0}\ln(1+a) + \sum_{i\in I_1\cup I_2}\ln(1+x_{1i})) \\
&+\; (n_1 + w_2 n_2)\ln\alpha_1 - \alpha_2(\tilde{n_0}\ln(1+a) + \sum_{i\in I_1\cup I_2}\ln(1+x_{2i})) \\
&+\; (n_2 + u_2 n_1)\ln\alpha_2.
\end{aligned} \tag{3.4.1}$$

Therefore, M-step involves maximizing (3.4.1) with respect to α_0, α_1, α_2 at

$$\hat{\alpha}_0^{(t+1)} = \frac{\tilde{n_0} + u_1^{(t)} n_1 + w_1^{(t)} n_2}{\tilde{n_0}\ln(1+a) + \sum_{i\in I_2}\ln(1+x_{1i}) + \sum_{i\in I_1}\ln(1+x_{2i})}, \tag{3.4.2}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(n_1 + w_2^{(t)} n_2)}{\tilde{n_0}\ln(1+a) + \sum_{i\in(I_1\cup I_2)}\ln(1+x_{1i})} \tag{3.4.3}$$

and

$$\hat{\alpha}_2^{(t+1)} = \frac{n_2 + u_2^{(t)} n_1}{(\tilde{n_0}\ln(1+a) + \sum_{i\in(I_1\cup I_2)}\ln(1+x_{2i}))}. \tag{3.4.4}$$

Therefore, the modified EM algorithm steps would be according to Algorithm-3.2.

Algorithm 3.2 Modified EM procedure for three parameter absolute continuous bivariate Pareto distribution

1: **while** $\Delta Q/Q < tol$ **do**
2: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n_0}$, a from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
3: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (3.4.2), (3.4.3) and (3.4.4).
4: Calculate Q for the new iterate.
5: **end while**

Important Remark : The above method fails when $\alpha_0 + \alpha_1 + \alpha_2 < 1$.

3.4.1 Proposed Modification :

To make the algorithm valid for any range of parameters, instead of estimating U_0 we estimate $\ln(1+U_0)$ conditional on $U_0 < \min\{U_1, U_2\}$. Since $\ln(x)$ is an increasing function of x, our condition is equivalent to $\ln(1+U_0) < \min\{\ln(1+U_1), \ln(1+U_2)\}$. Therefore, we replace the unknown missing information n_0 and $\ln(1+U_0)$ by

$$\tilde{n}_0 = (n_1+n_2)\frac{\alpha_0}{\alpha_1+\alpha_2}$$

and

$$a^* = E(\ln(1+U_0)|\ln(1+U_0) < \min\{\ln(1+U_1), \ln(1+U_2)\}) = \frac{1}{(\alpha_0+\alpha_1+\alpha_2)}$$

respectively.

Therefore, M-step involves maximizing (3.4.1) with respect to α_0, α_1, α_2 at

$$\hat{\alpha}_0^{(t+1)} = \frac{\tilde{n}_0 + u_1^{(t)}n_1 + w_1^{(t)}n_2}{\tilde{n}_0 a^* + \sum_{i\in I_2}\ln(1+x_{1i}) + \sum_{i\in I_1}\ln(1+x_{2i})}. \tag{3.4.5}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(n_1 + w_2^{(t)}n_2)}{\tilde{n}_0 a^* + \sum_{i\in(I_1\cup I_2)}\ln(1+x_{1i})} \tag{3.4.6}$$

$$\hat{\alpha}_2^{(t+1)} = \frac{n_2 + u_2^{(t)}n_1}{(\tilde{n}_0 a^* + \sum_{i\in(I_1\cup I_2)}\ln(1+x_{2i}))} \tag{3.4.7}$$

Therefore, the final EM steps for three parameters case can be kept according to Algorithm-3.3.

Algorithm 3.3 Final modified EM procedure for three parameter absolute continuous bivariate Pareto distribution

1: **while** $\Delta Q/Q < tol$ **do**
2: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{n}_0$, a^* from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$.
3: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ using Equation (3.4.5), (3.4.6) and (3.4.7).
4: Calculate Q for the new iterate.
5: **end while**

3.4.2 Algorithm for seven parameters absolute continuous Pareto distribution

For seven parameter case, we first estimate the location parameters from the marginal distributions and keep it fixed. Estimates of the location parameters can be replaced by minimum of the marginals of the data. In our EM algorithm, we update the scale parameters and α_0, α_1, α_2 at every iteration. We use one step ahead gradient descent of σ_1 and σ_2 based on the likelihood of its marginal density. We can not use the bivariate likelihood function in gradient descent updates of σ_1 and σ_2, as the likelihood function may not be differentiable with respect to σ_1 and σ_2.

Once, we get an update of the scale parameters, we first fix I_1 and I_2 and try to mimic the EM steps for α_0, α_1 and α_2, which is discussed in the previous section. We define $z_{1i} = \frac{(x_{1i}-\hat{\mu_1})}{\hat{\sigma_1}}$ and

$z_{2i} = \frac{(x_{2i} - \hat{\mu_2})}{\hat{\sigma_2}}$. If the estimates of location and scale parameters are correct, then (Z_{1i}, Z_{2i}) follows a bivariate Pareto distribution with parameter sets $(0, 0, 1, 1, \alpha_0, \alpha_1, \alpha_2)$. But one step ahead scale parameter will not be a correct estimate of scale parameters. However, we assume approximate distribution of (Z_{1i}, Z_{2i}) would be the bivariate Pareto with parameter sets $(0, 0, 1, 1, \alpha_0, \alpha_1, \alpha_2)$.

Therefore, M-step expression would be same as before :

$$\hat{\alpha}_0^{(t+1)} = \frac{n_0 + u_1^{(t)} n_1 + w_1^{(t)} n_2}{\sum_{i \in I_0} \ln(1 + z_{1i}) + \sum_{i \in I_2} \ln(1 + z_{1i}) + \sum_{i \in I_1} \ln(1 + z_{2i})}, \tag{3.4.8}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(n_1 + w_2^{(t)} n_2)}{\sum_{i \in I_0} \ln(1 + z_{1i}) + \sum_{i \in (I_1 \cup I_2)} \ln(1 + z_{1i})} \tag{3.4.9}$$

and

$$\hat{\alpha}_2^{(t+1)} = \frac{n_2 + u_2^{(t)} n_1}{(\sum_{i \in I_0} \ln(1 + z_{1i}) + \sum_{i \in (I_1 \cup I_2)} \ln(1 + z_{2i}))}. \tag{3.4.10}$$

We make the similar modifications of the above expressions. As usual observations within I_0 and cardinality of I_0 would be unknown in this case. We estimate $\sum_{i \in I_0} \ln(1 + z_i)$ by $\tilde{n}_0 a^*$, where $a_0^* = E(\ln(1 + U_0) | \ln(1 + U_0) < \min\{\ln(1 + U_1), \ln(1 + U_2)\}) = \frac{1}{(\alpha_0 + \alpha_1 + \alpha_2)}$ and n_0 by $\tilde{n_0} = (n_1 + n_2)\frac{\alpha_0}{\alpha_1 + \alpha_2}$.

Therefore, final M-step expression would be same as before :

$$\hat{\alpha}_0^{(t+1)} = \frac{\tilde{n_0^*} + u_1^{(t)} n_1 + w_1^{(t)} n_2}{\tilde{n_0} a_0^* + \sum_{i \in I_2} \ln(1 + z_{1i}) + \sum_{i \in I_1} \ln(1 + z_{2i})}, \tag{3.4.11}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{(n_1 + w_2^{(t)} n_2)}{\tilde{n_0} a_0^* + \sum_{i \in (I_1 \cup I_2)} \ln(1 + z_{1i})} \tag{3.4.12}$$

and

$$\hat{\alpha}_2^{(t+1)} = \frac{n_2 + u_2^{(t)} n_1}{(\tilde{n_0} a_0^* + \sum_{i \in (I_1 \cup I_2)} \ln(1 + z_{2i}))}. \tag{3.4.13}$$

The key intuition of this algorithm is very similar to the stochastic gradient descent. As we increase the number of iteration, the EM steps try to force the three parameters $\alpha_0, \alpha_1, \alpha_2$ to pick up the right direction starting from any value and gradient descent steps of σ_1 and σ_2 gradually ensure them to roam around the actual values. One of the drawbacks with this approach is that it considers too many approximations. However, this algorithm works even for moderately large sample sizes. Sometimes it takes a lot of time to converge or roam around the actual value for some really bad sample. The probability of such events are very low. We stop the calculation after 2000 iteration in such cases. Our numerical result in the next section shows that this does not make any significant effect on the calculation of mean square error.

Therefore, algorithmic steps of our proposed algorithm for seven parameter absolute continuous bivariate Pareto distribution can be kept according to Algorithm-3.4.

Algorithm 3.4 EM procedure for seven parameter absolute continuous bivariate Pareto distribution

```
Take minimum of the marginals as estimates of the location parameters. Initialize other parameters.
while ΔQ/Q < tol do
    Compute one step ahead gradient descent update of scale parameters based on likelihood function of the marginals.
    Fix I_1 and I_2 based on the estimated location and scale parameters.
    Compute u_1^(i), u_2^(i), w_1^(i), w_2^(i), ñ_0^*, a_0^* from α_0^(i), α_1^(i), α_2^(i).
    Update α_0^(i+1), α_1^(i+1), α_2^(i+1) using Equation (3.4.11), (3.4.12) and (3.4.13).
    Calculate Q for the new iteration.
end while
```

3.5 Confidence Interval

In this section, we propose to use the confidence interval. Here, the location parameters μ_1 and μ_2 behaves like a threshold. So parametric bootstrap confidence interval for μ_1 and μ_2 does not exist. Therefore, we can find the confidence interval for μ_1 and μ_2 separately using the distribution of the estimates. For others parameters σ_1, σ_2, α_0, α_1 and α_2, we have obtained the confidence intervals (CI) by using the parametric bootstrap technique. The parametric bootstrap confidence intervals are simply the 2.5^{th} percentile and 97.5^{th} percentile quantiles of the 200 simulated estimates of the corresponding parameter in the 200 simulations.

The confidence intervals for the location parameters can be obtained by using the distribution of MLE estimate of the location parameters, $\hat{\mu}_j = X_{(j)} = \min\{X_{j1}, X_{j2}, \cdots, X_{jn}\}$, $j = 1, 2$ from the marginals. Now the survival function of $Z_j = (X_{(j)} - \mu_j)/\sigma_j$ is given as follows,

$$S_{Z_j}(z_j) = \left\{\frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_1 + \alpha_2}(1 + z_j)^{-\alpha_0 - \alpha_j} - \frac{\alpha_0}{\alpha_1 + \alpha_2}(1 + z_j)^{-\alpha_0 - \alpha_1 - \alpha_2}\right\}^n. \tag{3.5.1}$$

Then 95% approximate confidence intervals can be written as,

$$X_{(j)} - b_j\hat{\sigma}_j \leq \mu_j \leq X_{(j)} - a_j\hat{\sigma}_j. \tag{3.5.2}$$

We calculate a_j and b_j using the survival functions. In our case, we use the EM estimate for $\hat{\sigma}_p,\ p = 1, 2$ and $\hat{\alpha}_p,\ p = 0, 1, 2$.

However, the above suggested procedure is just an approximate asymptotic confidence interval. More research is needed to explore better confidence interval than the above suggested one.

3.6 Numerical Results

We use the programming software R 3.2.2 to perform the estimation procedure. All the programs will be available to the authors on request. We generate samples of different sizes to calculate the average estimates based on 1000 simulations. We also calculate the mean square error and the parametric bootstrap confidence interval for all parameters.

The results are shown in Table-3.2 deals only with three unknown parameters using the Algorithm-3.3. We start the EM algorithm taking initial guesses for the parameters as $\alpha_0 = 1, \alpha_1 = 0.2, \alpha_2 = 0.2$. We try other initial guesses, but average estimates and MSEs are same. Estimates are calculated based on the sample sizes $n = 50, 150, 250, 350, 450$ respectively. We use stopping criteria as the absolute value of the log-likelihood changes with respect to the previous log-likelihood at each iteration. In stopping criteria we use tolerance value as 0.00001. The results are shown here are based on modified algorithm, which works for any choice of the parameters within its usual range. The average estimates (AE), mean squared errors (MSE) are obtained based on 1000 replications.

Table-3.3 provides the estimates in the presence of the location and the scale parameters using the Algorithm-3.4. In this case, we take sample size $n = 450, 550, 1000, 1500$. However, the algorithm works even for smaller sample sizes, although mean square errors are little higher. Since the EM algorithm starts after plug-in the estimates of location parameters as minimum of the marginals, we take the initial values of other parameters as $\alpha_0 = 1, \alpha_1 = 0.2, \alpha_2 = 0.2, \sigma_1 = 0.4$ and $\sigma_2 = 0.4$. We find the average estimates which are closer to the true values of the parameters. However, the parametric bootstrap confidence interval does not work for location parameters as it never contains the true parameters. Table-3.4 represent the confidence intervals by the procedures described above for the sample size 450 and 1000 based on 200 replications.

n	parameters	α_0	α_1	α_2
50	(AE) MSE Parametric Bootstrap Confidence Interval (CI)	1.9549 0.3882 [0.6381, 2.7367]	0.4640 0.1699 [0.0003, 1.4601]	0.5699 0.2292 [0.0003, 1.6711]
150	(AE) MSE Parametric Bootstrap Confidence Interval (CI)	1.9802 0.1797 [1.1441, 2.7102]	0.4229 0.0700 [0.0029, 1.0054]	0.5252 0.0998 [0.0033, 1.2025]
250	(AE) MSE Parametric Bootstrap Confidence Interval (CI)	1.9876 0.1102 [1.3788, 2.6498]	0.4141 0.0421 [0.0230, 0.8423]	0.5155 0.0620 [0.0326, 1.0031]
350	(AE) MSE Parametric Bootstrap Confidence Interval (CI)	1.9981 0.0851 [1.4356, 2.5851]	0.4064 0.0312 [0.0638, 0.7576]	0.5059 0.0458 [0.0770, 0.9228]
450	(AE) MSE Parametric Bootstrap Confidence Interval (CI)	2.0023 0.0601 [1.5085, 2.5053]	0.4028 0.0216 [0.1257, 0.7030]	0.5028 0.0329 [0.1537, 0.8709]

Table 3.2: The average estimates (AE), the Mean Square Error (MSE) and parametric bootstrap confidence interval for $\alpha_0 = 2$, $\alpha_1 = 0.4$ and $\alpha_2 = 0.5$

n	parameters		μ_1	μ_2	σ_1
450	AE		0.1013	0.1011	0.8377
	MSE		3.0698e-6	2.5216de-6	0.0242
450		σ_2	α_0	α_1	α_2
	AE	0.8269	1.9138	0.4931	0.6127
	MSE	0.0346	0.1005	0.0558	0.1187
550	parameters		μ_1	μ_2	σ_1
	AE		0.1012	0.1009	0.8336
	MSE		2.7214e-6	1.6559e-6	0.0202
550		σ_2	α_0	α_1	α_2
	AE	0.8300	1.9286	0.4756	0.6038
	MSE	0.0276	0.0740	0.0410	0.1032
1000			μ_1	μ_2	σ_1
	AE		0.1006	0.1006	0.8313
	MSE		7.1534e-7	6.5660e-7	0.0131
1000		σ_2	α_0	α_1	α_2
	AE	0.8072	1.9461	0.4613	0.5613
	MSE	0.0184	0.0452	0.0272	0.0605
1500	parameters		μ_1	μ_2	σ_1
	AE		0.1004	0.1004	0.8217
	MSE		4.0696e-7	2.4771e-7	0.0089
1500		σ_2	α_0	α_1	α_2
	AE	0.7968	1.9744	0.4330	0.5210
	MSE	0.0122	0.0288	0.0162	0.0338

Table 3.3: The average estimates (AE), the Mean Square Error (MSE) for $\mu_1 = 0.1$, $\mu_2 = 0.1$, $\sigma_1 = 0.8$, $\sigma_2 = 0.8$, $\alpha_0 = 2$, $\alpha_1 = 0.4$ and $\alpha_2 = 0.5$

Parameters \ Sample Size	450	1000
μ_1	[0.0959, 0.1013]	[0.0982, 0.1004]
μ_2	[0.0966, 0.1012]	[0.0992, 0.1012]
σ_1	[0.6039, 1.1443]	[0.6399, 1.0632]
σ_2	[0.5415, 1.2182]	[0.5962, 1.1278]
α_0	[1.3315, 2.5556]	[1.5772, 2.3192]
α_1	[0.1510, 0.9446]	[0.1954, 0.7821]
α_2	[0.1672, 1.4202]	[0.2258, 1.2264]

Table 3.4: 95% Confidence Intervals (CI) for $\mu_1 = 0.1$, $\mu_2 = 0.1$, $\sigma_1 = 0.8$, $\sigma_2 = 0.8$, $\alpha_0 = 2$, $\alpha_1 = 0.4$ and $\alpha_2 = 0.5$ by the suggested procedure.

3.7 Data Analysis

The data set is taken from *https://archive.ics.uci.edu/ml/machine-learning-databases*. The age of the abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope. The data set contains related measurements. We extract a part of the data for this bivariate modeling. We consider only measurements related to the female population where two of the variables are "Shucked weight" and "Viscera weight". They represent the weight of meat and the gut weight (after bleeding) of the abalones which are measured in grams. We use peak over threshold method on this data set. A brief summary of the transformed data has been shown in Table-3.5.

	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
Marginal X_1	6.644	7.227	7.896	8.356	9.009	18.258
Marginal X_2	3.423	3.727	4.061	4.277	4.663	7.239

Table 3.5: Brief summary of transformed data for BVPAC distribution.

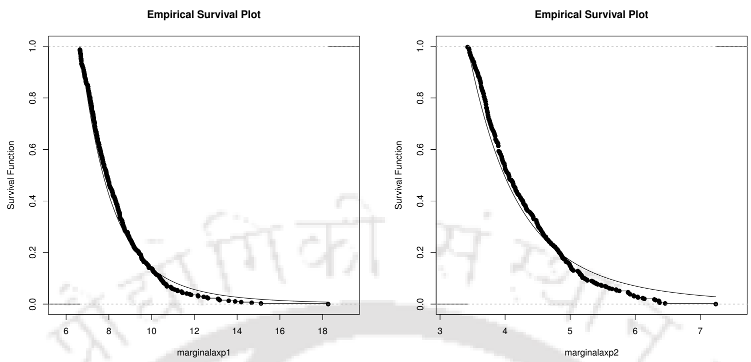
From Falk, M. and Guillou, A. (2008) [15], we know that the peak over threshold method on a random variable U (say) provides polynomial generalized Pareto distribution for any x_0 with $1 + log(G(x_0)) \in (0, 1)$, i.e., $P(U > tx_0|U > x_0) = t^{-\alpha}, \quad t \geq 1$, where $G(\cdot)$ is the distribution function of U. We choose an appropriate t and x_0 on the marginals of the bivariate data, so that the data should behave more like Pareto distribution. The transformed data set does not have any singular component. Therefore, one possible assumption can be the absolute continuous Marshall Olkin bivariate Pareto.

We do not have any proper goodness of fit test for the bivariate distribution like the univariate case. For that reason, we can show the goodness of fit test for the distribution of marginals and the distribution of minimum of components as univariate Pareto distribution. It is observed that $W = \min\{\frac{X_1-\mu_1}{\sigma_1}, \frac{X_2-\mu_2}{\sigma_2}\}$ has same univariate Pareto distribution, i.e., $W \sim PA(II)(0, 1, \alpha_0 + \alpha_1 + \alpha_2)$. We can use this property to validate the actual distribution as the Marshall-Olkin singular bivariate Pareto. The Kolmogorov-Smirnov (K-S) distance and p-value for W, marginal X_1 and marginal X_2 are provided in Table-3.6.

	For W	For marginal X_1	For marginal X_2
K-S distance	0.0869	0.0435	0.0843
p-value	0.104	0.8536	0.1234

Table 3.6: K-S distance (or K-S statistic value) and p-value based on data.

Apart from numerical diagnostics, we also verify our assumption by plotting empirical two dimensional density plot in Figure-3.4 which resembles closer to the surface of the absolute continuous Marshall-Olkin bivariate Pareto distribution.



(a) Marginal for X_1 (b) Marginal for X_2

Figure 3.3: Survival plots for two marginals of BVPAC of the transformed dataset.

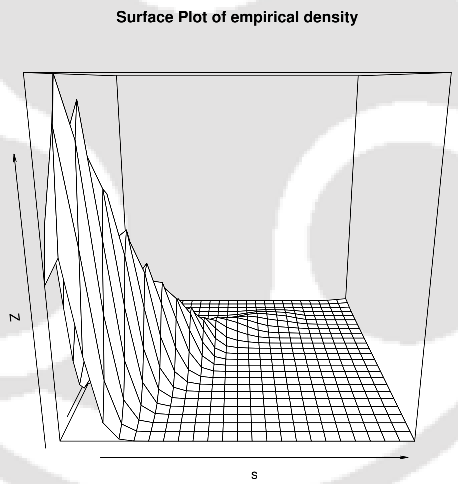


Figure 3.4: Two dimensional density plots for BVPAC of the transformed dataset.

The estimated parameters are $\mu_1 = 6.644$, $\mu_2 = 3.423$, $\sigma_1 = 4.0375$, $\sigma_2 = 2.4821$, $\alpha_0 = 2.9622$, $\alpha_1 = 1.1400$ and $\alpha_2 = 1.4552$. We can also calculate the mean square error by using the parametric bootstrap method. We calculate the MSEs for the sample size 391 based on 1000 bootstrap samples. The corresponding MSEs are 5.1861×10^{-5}, 7.0141×10^{-6}, 0.9314, 0.2222, 0.1955, 0.1778 and 0.2342 respectively. The confidence intervals of the data set are provided in Table-3.7 for the sample size 391 and 450 based on 200 bootstrap samples.

Parameters \ Sample Size	391	450
μ_1	[6.6306, 6.6478]	[6.6332, 6.6478]
μ_2	[3.4162, 3.4243]	[3.4172, 3.4243]
σ_1	[2.9512, 3.4506]	[2.9437, 3.4397]
σ_2	[1.8677, 2.3925]	[1.7927, 2.4069]
α_0	[1.9608, 3.5330]	[1.9485, 3.4084]
α_1	[0.3613, 1.3206]	[0.4219, 1.1794]
α_2	[0.5156, 1.7035]	[0.6008, 1.7706]

Table 3.7: 95% confidence intervals (CI) for μ_1, μ_2, σ_1, σ_2, α_0, α_1 and α_2 based on data

3.8 Formulating the Marshall-Olkin Multivariate Pareto Distribution

Let U_i, for $i = 0, 1, 2, \cdots n$, are mutually independent random variable, where $U_0 \sim PA(II)(0, 1, \alpha_0)$, $U_i \sim PA(II)(\mu_i, \sigma_i, \alpha_1)$, for $i = 1, 2, \cdots n$. We define $X_i = \min\{\mu_i + \sigma_i U_0, U_1\}$, for $i = 1, 2, \cdots, n$, then the joint distribution of $(X_1, X_2, \cdots, X_n)$ is called the Marshall-Olkin Multivariate Pareto (MOMVPA) distribution or singular multivariate Pareto distribution. For notational complexity, we discuss only three dimensional case here, the same procedure can be used for any dimensions. Then the joint survival function of (X_1, X_2, X_3) can be written, for $z = \max\{\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}, \frac{x_3-\mu_3}{\sigma_3}\}$ as

$$S(x_1, x_2, x_3) = (1+z)^{-\alpha_0} \frac{1}{\left(1+\frac{x_1-\mu_1}{\sigma_1}\right)^{\alpha_1}\left(1+\frac{x_2-\mu_2}{\sigma_2}\right)^{\alpha_2}\left(1+\frac{x_3-\mu_3}{\sigma_3}\right)^{\alpha_3}}.$$

We denote this distribution as $MOTVPA(\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3, \alpha_0, \alpha_1, \alpha_2, \alpha_3)$. Hence the pdf is

$$f(x_1,x_2,x_3) = \begin{cases} f_1(x_1,x_2,x_3), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} > \frac{x_3-\mu_3}{\sigma_3} \text{ or } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_3-\mu_3}{\sigma_3} > \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1,x_2,x_3), & \text{if } \frac{x_2-\mu_2}{\sigma_2} > \frac{x_1-\mu_1}{\sigma_1} > \frac{x_3-\mu_3}{\sigma_3} \text{ or } \frac{x_3-\mu_3}{\sigma_3} > \frac{x_2-\mu_2}{\sigma_2} > \frac{x_1-\mu_1}{\sigma_1} \\ f_3(x_1,x_2,x_3), & \text{if} \frac{x_3-\mu_3}{\sigma_3} > \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \text{ or } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} < \frac{x_3-\mu_3}{\sigma_3} \\ f_4(x_1,x_2,x_3), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} = \frac{x_3-\mu_3}{\sigma_3} \\ f_5(x_1,x_2,x_3), & \text{if } \frac{x_2-\mu_2}{\sigma_2} < \frac{x_1-\mu_1}{\sigma_1} = \frac{x_3-\mu_3}{\sigma_3} \\ f_6(x_1,x_2,x_3), & \text{if } \frac{x_3-\mu_3}{\sigma_3} < \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x_1,x_2,x_3), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = \frac{x_3-\mu_3}{\sigma_3}, \end{cases}$$

where

$$f_1(x_1,x_2,x_3) = \frac{\alpha_2(\alpha_0+\alpha_1)\alpha_3}{\sigma_1\sigma_2\sigma_3} \frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_0+\alpha_1+1}(1+\frac{x_2-\mu_2}{\sigma_2})^{\alpha_2+1}(1+\frac{x_3-\mu_3}{\sigma_3})^{\alpha_3+1}},$$

$$f_2(x_1,x_2,x_3) = \frac{\alpha_1(\alpha_0+\alpha_2)\alpha_3}{\sigma_1\sigma_2\sigma_3} \frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_1+1}(1+\frac{x_2-\mu_2}{\sigma_2})^{\alpha_0+\alpha_2+1}(1+\frac{x_3-\mu_3}{\sigma_3})^{\alpha_3+1}},$$

$$f_3(x_1,x_2,x_3) = \frac{\alpha_1(\alpha_0+\alpha_3)\alpha_2}{\sigma_1\sigma_2\sigma_3}\frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_1+1}(1+\frac{x_2-\mu_2}{\sigma_2})^{\alpha_2+1}(1+\frac{x_3-\mu_3}{\sigma_3})^{\alpha_0+\alpha_3+1}},$$
$$f_4(x_1,x_2,x_3) = \frac{\alpha_1(\alpha_0+\alpha_2+\alpha_3)}{\sigma_1\sigma_2}\frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_1+1}(1+\frac{x_2-\mu_2}{\sigma_2})^{\alpha_0+\alpha_2+\alpha_3+1}},$$

$$f_5(x_1,x_2,x_3) = \frac{\alpha_2(\alpha_0+\alpha_1+\alpha_3)}{\sigma_1\sigma_2}\frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_0+\alpha_1+\alpha_3+1}(1+\frac{x_2-\mu_2}{\sigma_2})^{\alpha_2+1}},$$
$$f_6(x_1,x_2,x_3) = \frac{\alpha_3(\alpha_0+\alpha_1+\alpha_2)}{\sigma_1\sigma_3}\frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_0+\alpha_1+\alpha_2+1}(1+\frac{x_3-\mu_3}{\sigma_3})^{\alpha_3+1}},$$
$$f_0(x_1,x_2,x_3) = \frac{\alpha_0-p'}{\sigma_1}\frac{1}{(1+\frac{x_1-\mu_1}{\sigma_1})^{\alpha_0+\alpha_1+\alpha_2+\alpha_3+1}},$$

also $p' = \left\{\frac{\alpha_3(\alpha_1+\alpha_2)}{(\alpha_0+\alpha_1+\alpha_2)} + \frac{\alpha_2(\alpha_1+\alpha_3)}{\alpha_0+\alpha_1+\alpha_3} + \frac{\alpha_1(\alpha_2+\alpha_3)}{(\alpha_0+\alpha_2+\alpha_3)}\right\}$ and $p = \frac{p'}{(\alpha_0+\alpha_1+\alpha_2+\alpha_3)}$. From the Lebesgue decomposition theorem the joint pdf $f(x_1,x_2,x_3)$ say, can be written as follows,

$$f(x_1,x_2,x_3) = pf_{ac}(x_1,x_2,x_3) + (1-p)f_s(x_1,x_2,x_3), \tag{3.8.1}$$

where $f_{ac}(x_1,x_2,x_3)$ and $f_s(x_1,x_2,x_3)$ are the absolute continuous part and the singular part of MOTVPA distribution. Here, we are interested in the absolute continuous part only. We call this absolute continuous as Block-Basu Trivariate Pareto (BBTVPA) distribution and denote it as $BBTVPA(\mu_1,\mu_2,\mu_3,\sigma_1,\sigma_2,\sigma_3,\alpha_0,\alpha_1,\alpha_2,\alpha_3)$. Hence the pdf of BBTVPA distribution can be written as,

$$f_{ac}(x_1,x_2,x_3) = \begin{cases} \frac{1}{p}f_1(x_1,x_2,x_3), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} > \frac{x_3-\mu_3}{\sigma_3} \text{ or } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_3-\mu_3}{\sigma_3} > \frac{x_2-\mu_2}{\sigma_2} \\ \frac{1}{p}f_2(x_1,x_2,x_3), & \text{if } \frac{x_2-\mu_2}{\sigma_2} > \frac{x_1-\mu_1}{\sigma_1} > \frac{x_3-\mu_3}{\sigma_3} \text{ or } \frac{x_3-\mu_3}{\sigma_3} > \frac{x_2-\mu_2}{\sigma_2} > \frac{x_1-\mu_1}{\sigma_1} \\ \frac{1}{p}f_3(x_1,x_2,x_3), & \text{if } \frac{x_3-\mu_3}{\sigma_3} > \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \text{ or } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} < \frac{x_3-\mu_3}{\sigma_3} \end{cases}$$

and the pdf of singular part is,

$$f_s(x_1,x_2,x_3) = \begin{cases} f_4(x_1,x_2,x_3), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} = \frac{x_3-\mu_3}{\sigma_3} \\ f_5(x_1,x_2,x_3), & \text{if } \frac{x_2-\mu_2}{\sigma_2} < \frac{x_1-\mu_1}{\sigma_1} = \frac{x_3-\mu_3}{\sigma_3} \\ f_6(x_1,x_2,x_3), & \text{if } \frac{x_3-\mu_3}{\sigma_3} < \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} \\ \frac{\alpha_0-p'}{\sigma_1}(1+\frac{x_1-\mu_1}{\sigma_1})^{-\alpha_0-\alpha_1-\alpha_2-\alpha_3-1}, & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = \frac{x_3-\mu_3}{\sigma_3}. \end{cases}$$

3.9 Conclusion

We have observed the successful implementation of the EM algorithm for the absolute continuous bivariate Pareto distribution. The approximation process runs in multiple stages. Therefore, the algorithm works better for moderately larger sample size. The estimation procedure, even in case of three parameters is not straightforward. This work also shows some innovative approach to handle the estimation in case of location and scale parameters. The work of this chapter can be further used for discrimination of several models. The Bayesian estimation of this distribution even without location and scale parameters using Gamma prior and Reference prior can also be a challenging problem. We address this problem in the next chapter.

Chapter 4

Bayesian analysis of three parameters absolute continuous Marshall-Olkin bivariate Pareto distribution

4.1 Introduction

In this chapter, [1] we consider the Bayesian analysis of the absolute continuous version of the Marshall-Olkin bivariate Pareto distribution, whose marginals are not type-II univariate Pareto distributions. We use the notation BB-BVPA for absolute continuous Marshall-Olkin bivariate Pareto. This form of Marshall-Olkin bivariate Pareto is similar to absolute continuous bivariate exponential distribution as proposed by Block and Basu (1974) [9]. Finding an efficient estimation technique to estimate the parameters of BB-BVPA distribution was a major challenge in last few decades. Parameter estimation by the EM algorithm for the absolute continuous bivariate Pareto distribution is also available in a recent work by Paul, Dey and Kundu (2017) [43]. There is no work in Bayesian set up for BB-BVPA distribution. There is very less work using the reference prior as posterior becomes extremely complicated, while dealing with such complicated prior. We make an easy formulation for the same through the slice cum Gibbs sampler method, which is not available in the literature. The formulation makes sense as slice sampling is also a type of the Gibbs sampler. In this chapter, we restrict ourselves only up to three parameter BB-BVPA distribution. BB-BVPA distribution can be very useful in modeling the data related to finance, climate and network-security, etc. This is one of the higher dimensional distribution which is heavy-tail in nature. A variety of bivariate (multivariate) extensions of the bivariate Pareto distribution also have been studied in the literature. These include the distributions described in the following works : [51], [55], [56], [3].

Bayesian analysis contains more information than the maximum likelihood estimation, which

[1] this work got accepted by the journal *Communications in Statistics - Case Studies and Data Analysis*

is better for further statistical analysis. For instance, not only the mean, mode, or median of the posterior distribution can be computed, but also the performances of these estimators (through their variance and higher order moments) are available. The maximum likelihood estimators can be quite unstable, i.e., they may vary widely for small variations of the observations. Moreover, the knowledge of the posterior distribution also allows for the derivation of the confidence regions through the highest posterior density (HPD) regions in both univariate and multivariate cases. Therefore, working on Bayesian set up with such a complicated distribution has its own advantages. In this work, both informative prior like gamma prior and non-informative prior like reference prior is used. The Bayesian estimator can not be obtained in closed form. Therefore, we propose to use two methods. (1) Lindley approximation [33] (2) Slice cum Gibbs Sampler [[38], [10]]. However, we can use other Monte Carlo methods for the same. Here we made slight modification in the calculation of the Lindley approximation. We use the EM algorithms instead of direct optimization method to obtain the MLEs. We also calculate the credible intervals for the parameters.

The rest of the chapter is organized as follows. Section-4.2 is kept in the Bayesian analysis of BB-BVPA distribution. Section-4.3 deals with the construction of the credible interval. Numerical results are discussed in section-4.4. Data analysis is shown in section-4.5. Finally, we conclude this chapter in Section-4.6.

4.2 Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution

BB-BVPA distribution is introduced and parameter estimation through the EM algorithm is available at [43]. In this section, we discuss the Bayesian estimate of the three parameters absolute continuous bivariate Pareto through the Lindley and the slice cum Gibbs sampler technique. Lindley approximation is done based on the EM estimates instead of direct MLE.

The survival function of three parameter MOBVPA distribution as we derived in previous section is

$$S_{X_1,X_2}(x_1,x_2) = \begin{cases} (1+x_1)^{-\alpha_1}(1+x_2)^{-(\alpha_0+\alpha_2)} & \text{if } x_1 < x_2 \\ (1+x_1)^{-(\alpha_0+\alpha_1)}(1+x_2)^{-\alpha_2} & \text{if } x_1 > x_2 \\ (1+z)^{-(\alpha_0+\alpha_1+\alpha_2)} & \text{if } x_1 = x_2 = z. \end{cases} \tag{4.2.1}$$

The above expression can be decomposed as follows,

$$S_{X_1,X_2}(x_1,x_2) = pS_{ac}(x_1,x_2) + (1-p)S_s(x_1,x_2), \tag{4.2.2}$$

where $S_{ac}(x_1,x_2)$ is the absolute continuous part and $S_s(x_1,x_2)$ is the singular part of $S_{X_1,X_2}(x_1,x_2)$. Clearly,

$$S_s(x_1,x_2) = (1+z)^{-(\alpha_0+\alpha_1+\alpha_2)}, \tag{4.2.3}$$

when $x_1 = x_2 = z$.

$$\frac{\partial^2}{\partial x_1 \partial x_2} S_{X_1,X_2}(x_1, x_2) = \begin{cases} f_1(x_1, x_2) & \text{if } x_1 < x_2 \\ f_2(x_1, x_2) & \text{if } x_1 > x_2. \end{cases} \tag{4.2.4}$$

Therefore,

$$f_1(x_1, x_2) = \alpha_1(\alpha_0 + \alpha_2)(1 + x_1)^{-(\alpha_1+1)}(1 + x_2)^{-(\alpha_0+\alpha_2+1)}, \tag{4.2.5}$$

$$f_2(x_1, x_2) = \alpha_2(\alpha_0 + \alpha_1)(1 + x_1)^{-(\alpha_0+\alpha_1+1)}(1 + x_2)^{-(\alpha_2+1)} \tag{4.2.6}$$

and

$$\iint_{x_1<x_2} f_1(x_1, x_2)dx_1dx_2 + \iint_{x_1>x_2} f_2(x_1, x_2)dx_1dx_2 = p. \tag{4.2.7}$$

Consider,

$$\iint_{x_1<x_2} f_1(x_1, x_2)dx_1dx_2 \tag{4.2.8}$$

$$= \alpha_1(\alpha_0 + \alpha_2) \int_0^\infty (1 + x_1)^{-\alpha_1-1} \left(\int_{x_1}^\infty (1 + x_2)^{-(\alpha_0+\alpha_2+1)} dx_2 \right) dx_1$$

$$= \frac{\alpha_1}{\alpha_0 + \alpha_1 + \alpha_2}. \tag{4.2.9}$$

Similarly,

$$\iint_{x_1>x_2} f_2(x_1, x_2)dx_1dx_2 = \frac{\alpha_2}{\alpha_0 + \alpha_1 + \alpha_2}. \tag{4.2.10}$$

Therefore, from (4.2.7), we get,

$$\frac{\alpha_1 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2} = p. \tag{4.2.11}$$

We denote the absolute continuous part of the Marshal-Olkin bivariate Pareto as BB-BVPA distribution. Therefore, if (X_1, X_2) follows the absolute continuous bivariate Pareto distribution, then the joint pdf can be written as

$$f_{BVPAC}(x_1, x_2) = \begin{cases} cf_1(x_1, x_2) & \text{if } x_1 < x_2 \\ cf_2(x_1, x_2) & \text{if } x_1 > x_2, \end{cases} \tag{4.2.12}$$

where c is a normalizing constant and

$$c = \frac{1}{p} = \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_1 + \alpha_2}. \tag{4.2.13}$$

4.2.1 Likelihood Function

The likelihood function corresponding to this pdf is given by,

$$l(x_1, x_2; \alpha_0, \alpha_1, \alpha_2) = \prod_i^n (cf_1(x_1, x_2))^{\mathrm{I}_{(x_1,x_2)\in I_1}} \prod_i^n (cf_2(x_1, x_2))^{\mathrm{I}_{(x_1,x_2)\in I_2}}$$

$$\begin{aligned} &= c^{n_1+n_2}\alpha_1^{n_1}(\alpha_0+\alpha_2)^{n_1}\alpha_2^{n_2}(\alpha_0+\alpha_1)^{n_2} \\ &\quad \prod_{i\in I_1}(1+x_{1i})^{-(\alpha_1+1)}(1+x_{2i})^{-(\alpha_0+\alpha_2+1)} \\ &\times \prod_{i\in I_2}(1+x_{1i})^{-(\alpha_0+\alpha_1+1)}(1+x_{2i})^{-(\alpha_2+1)}, \end{aligned} \tag{4.2.14}$$

where $I_1 = \{i \mid x_{1i} < x_{2i}\}$ and $I_2 = \{i \mid x_{1i} > x_{2i}\}$ and $|I_1| = n_1$, $|I_2| = n_2$, also $n_1 + n_2 = n$.

Therefore the log-likelihood function takes the form,

$$\begin{aligned} L(x_1, x_2; \alpha_0, \alpha_1, \alpha_2) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1 \ln\alpha_1 - (\alpha_1+1)\sum_{i\in I_1}\ln(1+x_{1i}) + n_1\ln(\alpha_0+\alpha_2) \\ &- (\alpha_0+\alpha_2+1)\sum_{i\in I_1}\ln(1+x_{2i}) + n_2\ln(\alpha_0+\alpha_1) \\ &- (\alpha_0+\alpha_1+1)\sum_{i\in I_2}\ln(1+x_{1i}) + n_2\ln(\alpha_2) \\ &- (\alpha_2+1)\sum_{i\in I_2}\ln(1+x_{2i}). \end{aligned} \tag{4.2.15}$$

4.2.2 Prior Assumption

Gamma Prior Assumption

We assume that α_0, α_1, and α_2 are distributed according to the gamma distribution with shape parameter k_i and scale parameter θ_i, where $i = 0, 1, 2$, i.e.,

$$\begin{aligned} \alpha_0 &\sim \Gamma(k_0, \theta_0) \equiv \text{Gamma}(k_0, \theta_0) \\ \alpha_1 &\sim \Gamma(k_1, \theta_1) \equiv \text{Gamma}(k_1, \theta_1) \\ \alpha_2 &\sim \Gamma(k_2, \theta_2) \equiv \text{Gamma}(k_2, \theta_2). \end{aligned} \tag{4.2.16}$$

The probability density function of the gamma distribution is given by,

$$f_\Gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}. \tag{4.2.17}$$

Here $\Gamma(k)$ is the gamma function evaluated at k.

Reference Prior Assumption

We have taken the reference prior as prior for each to the parameters conditional on the others. We calculate the expression using Bernardo's reference Prior [[7], [8]] in this context.

The key idea is to derive reference prior described by Berger and Bernando (1992) is to get prior $\pi(\theta)$ that maximizes the expected posterior information about the parameters. Let $I(\theta)$ is an expected information about θ given the data $X = \underline{x}$. Then

$$I(\theta) = E_X(K(p(\theta|\underline{x})||\pi(\theta))),$$

where $K(p(\theta|\tilde{x})||\pi(\theta))$ is the Kullback-Leibler distance between posterior $p(\theta|\underline{x})$ and prior $p(\theta)$ can be given by

$$K(\pi(\theta|\underline{x})||\pi(\theta)) = \int_{\Omega_\theta} \ln \frac{\pi(\theta|\underline{x})}{\pi(\theta)} \pi(\theta|\underline{x}).$$

Let $\theta = (\theta_1, \theta_2)$, where θ_1 is $p_1 \times 1$ and θ_2 is $p_2 \times 1$. We define $p = p_1 + p_2$. Let $I(\theta) = I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix}$

Suppose that θ_1 is the parameter of interest and θ_2 is a nuisance parameter (meaning that it's not really of interest to us in the model).

Begin with

$$\pi(\theta_2|\theta_1) = |I_{22}(\theta)|^{\frac{1}{2}} c(\theta_1), \tag{4.2.18}$$

where $c(\theta_1)$ is the constant that makes this distribution a proper density. Now try to maximize $E(\frac{\log(\pi(\theta_1|x))}{\pi(\theta_1)})$ to find out the marginal prior $\pi(\theta)$. We write

$$\ln \frac{\pi(\theta_1|x)}{\pi(\theta_1)} = \ln \frac{\pi(\theta_1, \theta_2|x)/\pi(\theta_2|\theta_1, x)}{\pi(\theta_1, \theta_2)/\pi(\theta_2|\theta_1)} = \ln \frac{\pi(\theta|x)}{\pi(\theta)} - \ln \frac{\pi(\theta_2|\theta_1, x)}{\pi(\theta_2|\theta_1)}.$$

We can write

$$E(\log \frac{\pi(\theta|x)}{\pi(\theta)}) = \frac{p}{2} \log n - \frac{p}{2} \log(2\pi e) + \int \pi(\theta) \log \frac{|I(\theta)|^{\frac{1}{2}}}{\pi(\theta)} d\theta + O(n^{-\frac{1}{2}}). \tag{4.2.19}$$

Similarly,

$$E(\log \frac{\pi(\theta_2|\theta_1, x)}{\pi(\theta_2|\theta_1)}) = \frac{p_2}{2} \log n - \frac{p_2}{2} \log(2\pi e) + \int \pi(\theta) \log \frac{|I_{22}(\theta)|^{\frac{1}{2}}}{\pi(\theta_2|\theta_1)} d\theta + O(n^{-\frac{1}{2}}). \tag{4.2.20}$$

From (4.2.19) and (4.2.20), we find

$$E(\log \frac{\pi(\theta_1|x)}{\pi(\theta_1)}) = \frac{p_1}{2} \log n - \frac{p_1}{2} \log 2\pi e + \int \pi(\theta) \log \frac{|I_{11.2}(\theta)|^{\frac{1}{2}}}{\pi(\theta_1)} d\theta + O(n^{-\frac{1}{2}}), \tag{4.2.21}$$

where $I_{11.2}(\theta) = I_{11}(\theta) - I_{12}(\theta) I_{22}^{-1}(\theta) I_{21}(\theta)$

We now break up the integral in (4.2.21) and we define

$$\log \psi(\theta_1) = \int \pi(\theta_2|\theta_1) \log |I_{11.2}(\theta)|^{\frac{1}{2}} d\theta_2.$$

We find that

$$
\begin{aligned}
E[\log \frac{\pi(\theta_1|x)}{\pi(\theta_1)}] &= \frac{p_1}{2}\log n - \frac{p_1}{2}\log 2\pi e + \int \pi(\theta)\log|I_{11.2}(\theta)|^{\frac{1}{2}}d\theta \\
&- \int \pi(\theta)\log \pi(\theta_1)d\theta + O(n^{-\frac{1}{2}}) \\
&= \frac{p_1}{2}\log n - \frac{p_1}{2}\log(2\pi e) + \int \pi(\theta_1)[\int \pi(\theta_2|\theta_1)\log|I_{11.2}|^{\frac{1}{2}}d\theta_2]d\theta_1 \\
&- \int \pi(\theta_1)\log \pi(\theta_1)d\theta_1 + O(n^{-\frac{1}{2}}) \\
&= \frac{p_1}{2}\log n - \frac{p_1}{2}\log 2\pi e + \int \pi(\theta_1)\log\frac{\psi(\theta_1)}{\pi(\theta_1)}d\theta_1 + O(n^{-\frac{1}{2}}).
\end{aligned}
$$

To maximize the above integral, we choose $\pi(\theta_1) = \psi(\theta_1)$. Note that $I_{11.2}^{-1}(\theta) = I_{11}(\theta)$, where
$I^{-1}(\theta) = \begin{pmatrix} I_{11}(\theta) & -I_{11}(\theta)I_{12}^{-1}(\theta)I_{22}(\theta) \\ -I_{22}(\theta)I_{21}^{-1}(\theta)I_{11}(\theta) & I_{22}(\theta) \end{pmatrix}$
Writing out our prior, we find that
$\pi(\theta_1) = \exp\{\int \pi(\theta_2|\theta_1))\log|I_{11.2}(\theta)|^{\frac{1}{2}}d\theta_2\} = \exp\{\int |I_{22}(\theta)|^{\frac{1}{2}}\log|I_{11.2}|^{\frac{1}{2}}d\theta_2\}$.

Finally, the expression of the full prior can be obtained as $\pi(\theta) = \pi(\theta_2|\theta_1)\pi(\theta_1)$. In our case, we just need to use $\pi(\theta_2|\theta_1)$ due to Gibbs sampler step.

In bivariate Pareto set up, $\theta = (\alpha_0, \alpha_1, \alpha_2)$. The expressions are as follows :

$$
\begin{aligned}
R_1 = \pi(\alpha_0|\alpha_1,\alpha_2) &\propto \sqrt{-\left(E(\frac{\partial^2 L}{\partial \alpha_0^2})\right)} \\
&= \sqrt{\frac{(n_1+n_2)}{(\alpha_0+\alpha_1+\alpha_2)^2} + \frac{n_1}{(\alpha_0+\alpha_2)^2} + \frac{n_2}{(\alpha_0+\alpha_1)^2}},
\end{aligned}
$$

$$
\begin{aligned}
R_2 = \pi(\alpha_1|\alpha_0,\alpha_2) &\propto \sqrt{-\left(E(\frac{\partial^2 L}{\partial \alpha_1^2})\right)} \\
&= \sqrt{\frac{n_1+n_2}{(\alpha_0+\alpha_1+\alpha_2)^2} - \frac{n_1+n_2}{(\alpha_1+\alpha_2)^2} + \frac{n_1}{(\alpha_1)^2} + \frac{n_2}{(\alpha_0+\alpha_1)^2}}
\end{aligned}
$$

and

$$
\begin{aligned}
R_3 = \pi(\alpha_2|\alpha_0,\alpha_1) &\propto \sqrt{-\left(E(\frac{\partial^2 L}{\partial \alpha_2^2})\right)} \\
&= \sqrt{\frac{n_1+n_2}{(\alpha_0+\alpha_1+\alpha_2)^2} - \frac{n_1+n_2}{(\alpha_1+\alpha_2)^2} + \frac{n_2}{(\alpha_2)^2} + \frac{n_1}{(\alpha_0+\alpha_2)^2}}.
\end{aligned}
$$

Note that we can replace all the expected Fisher information matrix by the observed Fisher information matrix to form an approximate likelihood. However, in our case the conditional prior distributions consisting of the expected Fisher information matrix and the respective observed Fisher information matrix and both are same, as they are an exactly same function of the parameters and independent of the data.

4.2.3 Bayesian estimates for BB-BVPA

In this section, we provide the Bayesian estimates of the unknown parameters, namely α_0, α_1, and α_2 using Lindley approximation. The posterior PDF of $(\alpha_0, \alpha_1, \alpha_2)$ given data D_2 based on the prior $\pi(\cdot)$ can be written as,

$$\begin{aligned}
\pi(\alpha_0, \alpha_1, \alpha_2 | D_2) \propto \quad & l(\alpha_0, \alpha_1, \alpha_2 | D_2) \pi(\alpha_0, \alpha_1, \alpha_2 | a, b, a_0, a_1, a_2) \\
=& \left(\frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_1 + \alpha_2} \right)^n \alpha_1^{n_1} \alpha_2^{n_2} (\alpha_0 + \alpha_1)^{n_2} (\alpha_0 + \alpha_2)^{n_1} \\
& \prod_{i \in I_1} (1 + x_{1i})^{-\alpha_1 - 1} (1 + x_{2i})^{-\alpha_0 - \alpha_2 - 1} \\
& \prod_{i \in I_2} (1 + x_{1i})^{-\alpha_0 - \alpha_1 - 1} (1 + x_{2i})^{-\alpha_2 - 1} \\
& \times \alpha_0^{k_0 - 1} \alpha_1^{k_1 - 1} \alpha_2^{k_2 - 1} \\
& e^{-(\frac{\alpha_0}{\theta_0} + \frac{\alpha_1}{\theta_1} + \frac{\alpha_2}{\theta_2})} \\
=& \pi_1(\alpha_0, \alpha_1, \alpha_2 | D_2) \quad \text{(say)}.
\end{aligned} \tag{4.2.22}$$

If we want to compute the Bayesian estimate of some function of α_0, α_1 and α_2, say $g(\alpha_0, \alpha_1, \alpha_2)$, the Bayesian estimate of g, say $\hat{g}$ under the squared error loss function is the posterior mean of g, i.e.,

$$\hat{g} = \frac{\int_0^\infty \int_0^\infty \int_0^\infty g(\alpha_0, \alpha_1, \alpha_2) \pi_1(\alpha_0, \alpha_1, \alpha_2 | D_2) d\alpha_0 d\alpha_1 d\alpha_2}{\int_0^\infty \int_0^\infty \int_0^\infty \pi_1(\alpha_0, \alpha_1, \alpha_2 | D_2) d\alpha_0 d\alpha_1 d\alpha_2}. \tag{4.2.23}$$

Lindley Approximation in BB-BVPA

Let α_0, α_1 and α_2 be the three parameters of joint the prior distribution $\pi(\alpha_0, \alpha_1, \alpha_2)$ (say). Then the Bayesian estimate of any function of α_0, α_1 and α_2, says $g = g(\alpha_0, \alpha_1, \alpha_2)$ under the squared error loss function is,

$$\hat{g}_B = \frac{\int_{(\alpha_0, \alpha_1, \alpha_2)} g(\alpha_0, \alpha_1, \alpha_2) e^{[L(\alpha_0, \alpha_1, \alpha_2) + \rho(\alpha_0, \alpha_1, \alpha_2)]} d(\alpha_0, \alpha_1, \alpha_2)}{\int_{(\alpha_0, \alpha_1, \alpha_2)} e^{[L(\alpha_0, \alpha_1, \alpha_2) + \rho(\alpha_0, \alpha_1, \alpha_2)]} d(\alpha_0, \alpha_1, \alpha_2)}, \tag{4.2.24}$$

where $L(\alpha_0, \alpha_1, \alpha_2)$ is the log-likelihood function and $\rho(\alpha_0, \alpha_1, \alpha_2)$ is the logarithm of the joint prior of α_0, α_1 and α_2, i.e, $\rho(\alpha_0, \alpha_1, \alpha_2) = \log \pi(\alpha_0, \alpha_1, \alpha_2)$. By the Lindley approximation (7.3.9), it can be written as,

$$\begin{aligned}
\hat{g}_B =& g(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2) + (g_0 b_0 + g_1 b_1 + g_2 b_2 + b_3 + b_4) + \frac{1}{2}[A(g_0 \sigma_{00} + g_1 \sigma_{01} + g_2 \sigma_{02}) \\
& + B(g_0 \sigma_{10} + g_1 \sigma_{11} + g_2 \sigma_{12}) + C(g_0 \sigma_{20} + g_1 \sigma_{21} + g_2 \sigma_{22})],
\end{aligned}$$

where $\hat{\alpha}_0, \hat{\alpha}_1$ and $\hat{\alpha}_2$ are the MLE of α_0, α_1 and α_2 respectively and

$$b_i = \rho_0 \sigma_{i0} + \rho_1 \sigma_{i1} + \rho_2 \sigma_{i2}, \quad i = 0, 1, 2$$

$$
\begin{aligned}
b_3 =& g_{01}\sigma_{01} + g_{02}\sigma_{02} + g_{12}\sigma_{12} \\
b_4 =& \frac{1}{2}(g_{00}\sigma_{00} + g_{11}\sigma_{11} + g_{22}\sigma_{22}) \\
A =& \sigma_{00}L_{000} + 2\sigma_{01}L_{010} + 2\sigma_{02}L_{020} + 2\sigma_{12}L_{120} + \sigma_{11}L_{110} + \sigma_{22}L_{220} \\
B =& \sigma_{00}L_{001} + 2\sigma_{01}L_{011} + 2\sigma_{02}L_{021} + 2\sigma_{12}L_{121} + \sigma_{11}L_{111} + \sigma_{22}L_{221} \\
C =& \sigma_{00}L_{002} + 2\sigma_{01}L_{012} + 2\sigma_{02}L_{022} + 2\sigma_{12}L_{122} + \sigma_{11}L_{113} + \sigma_{22}L_{222}.
\end{aligned}
$$

Also,

$$
\rho_i = \left[\frac{\partial \rho}{\partial \alpha_i}\right]_{\text{at}(\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2)}, \quad g_i = \left[\frac{\partial g(\alpha_0,\alpha_1,\alpha_2)}{\partial \alpha_i}\right]_{\text{at}(\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2)}, \quad i = 0,1,2
$$

$$
g_{ij} = \left[\frac{\partial^2 g(\alpha_0,\alpha_1,\alpha_2)}{\partial \alpha_i \partial \alpha_j}\right]_{\text{at}(\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2)}, \quad L_{ij} = \left[\frac{\partial^2 L(\alpha_0,\alpha_1,\alpha_2)}{\partial \alpha_i \partial \alpha_j}\right]_{\text{at}(\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2)}, i = 0,1,2
$$

$$
L_{ijk} = \left[\frac{\partial^3 L(\alpha_0,\alpha_1,\alpha_2)}{\partial \alpha_i \partial \alpha_j \partial \alpha_k}\right]_{\text{at}(\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2)} \quad i = 0,1,2.
$$

Here σ_{ij} is the $(i,j)th$ element of the inverse of the matrix $\{L_{ij}\}$ all evaluated at the MLE of α_0, α_1 and α_2, i.e., at $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$. We have already mentioned the log-likelihood function for BB-BVPA distribution as,

$$
\begin{aligned}
L(\alpha_0,\alpha_1,\alpha_2|D_2) =& (n_1+n_2)\log(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\log(\alpha_1+\alpha_2) \\
& + n_1\log(\alpha_1) + n_2\log(\alpha_2) + n_2\log(\alpha_0+\alpha_1) + n_1\log(\alpha_0+\alpha_2) \\
& -(\alpha_1+1)\sum_{i\in I_1}\log(1+x_{1i}) - (\alpha_0+\alpha_2+1)\sum_{i\in I_1}\log(1+x_{2i}) \\
& -(\alpha_0+\alpha_1+1)\sum_{i\in I_2}\log(1+x_{1i}) - (\alpha_2+1)\sum_{i\in I_2}\ln(1+x_{2i}).
\end{aligned}
$$

Let's assume the MLE of parameters α_0, α_1 and α_2 of the BB-BVPA or absolute continuous bivariate Pareto distribution are $\hat{\alpha}_0, \hat{\alpha}_1$ and $\hat{\alpha}_2$ respectively. Now $\rho = \log\pi(\alpha_0,\alpha_1,\alpha_2)$ then $\rho_0 = \frac{k_0-1}{\alpha_0} - \frac{1}{\theta_0}$, $\rho_1 = \frac{k_1-1}{\alpha_1} - \frac{1}{\theta_1}$ and $\rho_2 = \frac{k_2-1}{\alpha_2} - \frac{1}{\theta_2}$. Again,

$$
\begin{aligned}
L_{00} =& -\frac{n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^2} - \frac{n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^2} - \frac{n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^2} \\
L_{11} =& -\frac{n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^2} + \frac{n}{(\hat{\alpha}_1+\hat{\alpha}_2)^2} - \frac{n_1}{(\hat{\alpha}_1)^2} - \frac{n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^2} \\
L_{22} =& -\frac{n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^2} + \frac{n}{(\hat{\alpha}_1+\hat{\alpha}_2)^2} - \frac{n_2}{(\hat{\alpha}_2)^2} - \frac{n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^2} \\
L_{01} =& -\frac{n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^2} - \frac{n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^2} = L_{10} \\
L_{02} =& -\frac{n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^2} - \frac{n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^2} = L_{20} \\
L_{12} =& -\frac{n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^2} + \frac{n}{(\hat{\alpha}_1+\hat{\alpha}_2)^2} = L_{21}.
\end{aligned}
$$

The values of L_{ijk} for $i,j,k = 0,1,2$ are given by

$$
\begin{aligned}
L_{000} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^3} + \frac{2n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^3} \\
L_{111} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} - \frac{2n}{(\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_1}{(\hat{\alpha}_1)^3} + \frac{2n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^3} \\
L_{222} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} - \frac{2n}{(\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_2}{(\hat{\alpha}_2)^3} + \frac{2n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^3} \\
L_{001} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^3} = L_{010} = L_{100} \\
L_{002} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^3} = L_{020} = L_{200} \\
L_{011} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_2}{(\hat{\alpha}_0+\hat{\alpha}_1)^3} = L_{101} = L_{110} \\
L_{012} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} = L_{021} = L_{102} = L_{120} = L_{201} = L_{210} \\
L_{022} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} + \frac{2n_1}{(\hat{\alpha}_0+\hat{\alpha}_2)^3} = L_{202} = L_{220} \\
L_{112} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} - \frac{2n}{(\hat{\alpha}_1+\hat{\alpha}_2)^3} = L_{121} = L_{211} \\
L_{122} =& \frac{2n}{(\hat{\alpha}_0+\hat{\alpha}_1+\hat{\alpha}_2)^3} - \frac{2n}{(\hat{\alpha}_1+\hat{\alpha}_2)^3} = L_{212} = L_{221}.
\end{aligned}
$$

Now, we can obtain the Bayesian estimates of α_0, α_1 and α_2 under squared error loss function (i) for α_0, choose $g(\alpha_0, \alpha_1, \alpha_2) = \alpha_0$. So Bayesian estimates of α_0 can be written as,

$$\hat{\alpha}_{0B} = \hat{\alpha}_0 + b_0 + \frac{1}{2}[A\sigma_{00} + B\sigma_{10} + C\sigma_{20}], \tag{4.2.25}$$

(ii) for α_1, choose $g(\alpha_0, \alpha_1, \alpha_2) = \alpha_1$. So Bayesian estimates of α_1 can be written as,

$$\hat{\alpha}_{1B} = \hat{\alpha}_1 + b_1 + \frac{1}{2}[A\sigma_{01} + B\sigma_{11} + C\sigma_{21}] \tag{4.2.26}$$

and

(iii) for α_2, choose $g(\alpha_0, \alpha_1, \alpha_2) = \alpha_2$. So Bayesian estimates of α_2 can be written as,

$$\hat{\alpha}_{2B} = \hat{\alpha}_2 + b_2 + \frac{1}{2}[A\sigma_{02} + B\sigma_{12} + C\sigma_{22}]. \tag{4.2.27}$$

Remark : In this calculation, we replace the MLEs of the parameters by the estimates of the parameters obtained through the EM algorithm ([43]).

The Full Conditional Posterior distributions

To obtain the Bayesian estimates through the slice cum Gibbs sampler, we need to calculate the conditional distribution of each of the parameters given the other parameters and the data. We provide the expressions of the logarithm of those conditional distributions for both reference and gamma prior. We propose to perform the standard step-out method in slice sampling ([38]) on each of the conditional distributions based on gamma prior to generate the respective parameters. Algorithmic steps for slice sampling via step out method can be provided as follows:

- Sample z and u uniformly from the area under the distribution, say $p(\cdot)$.
  1. Fix z, sample u uniformly from $[0, p(z)]$.
  2. Fix u, sample z uniformly from the slice through the region $\{z : p(z) > u\}$
- How to sample z from the slice.
  1. Start with the region of width w containing $z^{(t)}$.
  2. If end point in slice, then extend region by w in that direction.
  3. Sample $z^{'}$ uniformly from the region.
  4. If $z^{'}$ is in the slice, the accept it as $z^{(t+1)}$.
  5. If not : make $z^{'}$ new end point of the region, and resample $z^{'}$.

In case of gamma prior, we observe that the algorithm does not depend much on the choice of width w. However, it is not same for the reference prior as reference prior is not defined in certain interval for some set of parameter values. As a result, the conditional posterior functions become discontinuous in nature and therefore, slice sampling should not work.

However, we propose to use a modified slice sampling in that set up which works quite well, even though the function is discontinuous or not defined in some intervals. The modified algorithm is as follows:

- Sample z and u uniformly from the area under the distribution, say $p(\cdot)$.
  1. Fix z, sample u uniformly from $[0, p(z)]$.
  2. Fix u, sample z uniformly from the slice through the region $\{z : p(z) > u\}$
- How to sample z from the slice.
  1. Start with the region of width w containing $z^{(t)}$ **where it is defined**.
  2. If end point in slice and **in the region where the function is defined**, then extend region by w in that direction.
  3. If end point in slice and **in the region where the function is not defined**, then also extend region by w in that direction.
  4. Sample $z^{'}$ uniformly from the region **until the point is defined**.

5. If $z^{'}$ is in the slice, the accept it as $z^{(t+1)}$.
6. If not : make $z^{'}$ new end point of the region, and resample $z^{'}$.
7. If resample point is not defined, **perform resampling until it is defined**.

We can use the different other techniques, like importance sampling, HMC, etc. taking Gamma as a prior. However, we avoid the calculation of the full posterior using reference prior, as it is much easier to express the conditional posterior distribution of each of the parameters given the other parameters and the data.

The full log-conditional distributions in Gamma prior

The full logarithm of the conditional posterior distributions of α_0, α_1, and α_2 are given by,

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) + n_1\ln(\alpha_0+\alpha_2)\\ &- \alpha_0\sum_{i\in I_1}\ln(1+x_{2i}) + n_2\ln(\alpha_0+\alpha_1)\\ &- \alpha_0\sum_{i\in I_2}\ln(1+x_{1i}) + (k_0-1)\ln\alpha_0 - \frac{\alpha_0}{\theta_0},\end{aligned} \tag{4.2.28}$$

$$\begin{aligned}\ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2)\\ &+ n_1\ln\alpha_1 - \alpha_1\sum_{i\in I_1}\ln(1+x_{1i}) + n_2\ln(\alpha_0+\alpha_1)\\ &- \alpha_1\sum_{i\in I_2}\ln(1+x_{1i}) + (k_1-1)\ln\alpha_1 - \frac{\alpha_1}{\theta_1}\end{aligned} \tag{4.2.29}$$

and

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \alpha_0, \alpha_1, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2)\\ &+ n_1\ln(\alpha_0+\alpha_2) - \alpha_2\sum_{i\in I_1}\ln(1+x_{2i}) + n_2\ln\alpha_2\\ &- \alpha_2\sum_{i\in I_2}\ln(1+x_{2i}) + (k_2-1)\ln\alpha_2 - \frac{\alpha_2}{\theta_2}.\end{aligned} \tag{4.2.30}$$

Expression for full log-conditional distributions in case of Reference Prior

The full logarithm of the conditional distributions of α_0, α_1, and α_2 for reference prior are given by,

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) + n_1\ln(\alpha_0+\alpha_2)\\ &- \alpha_0\sum_{i\in I_1}\ln(1+x_{2i}) + n_2\ln(\alpha_0+\alpha_1)\end{aligned}$$

$$- \quad \alpha_0 \sum_{i \in I_2} \ln(1 + x_{1i}) + \log(R_1), \tag{4.2.31}$$

$$\begin{aligned}
\ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, x_1, x_2)) &= (n_1 + n_2)\ln(\alpha_0 + \alpha_1 + \alpha_2) - (n_1 + n_2)\ln(\alpha_1 + \alpha_2) \\
&+ n_1 \ln \alpha_1 - \alpha_1 \sum_{i \in I_1} \ln(1 + x_{1i}) + n_2 \ln(\alpha_0 + \alpha_1) \\
&- \alpha_1 \sum_{i \in I_2} \ln(1 + x_{1i}) + \log(R_2)
\end{aligned} \tag{4.2.32}$$

and

$$\begin{aligned}
\ln(\pi(\alpha_2 \mid \alpha_0, \alpha_1, x_1, x_2)) &= (n_1 + n_2)\ln(\alpha_0 + \alpha_1 + \alpha_2) - (n_1 + n_2)\ln(\alpha_1 + \alpha_2) \\
&+ n_1 \ln(\alpha_0 + \alpha_2) - \alpha_2 \sum_{i \in I_1} \ln(1 + x_{2i}) + n_2 \ln \alpha_2 \\
&- \alpha_2 \sum_{i \in I_2} \ln(1 + x_{2i}) + \log(R_3).
\end{aligned} \tag{4.2.33}$$

4.3 Constructing credible Intervals for Parameters

We find the credible intervals for the parameters as described by Chen and Shao (1999) [11]. Let's assume $\underline{\theta}$ is a vector. To obtain the credible intervals of first variable θ_{1i}, we order $\{\theta_{1i}\}$, as $\theta_{1(1)} < \theta_{1(2)} < \cdots < \theta_{1(M)}$. Then 100(1 - γ)% credible interval of θ_1 become

$$(\theta_{1(j)}, \theta_{1(j+M-M\gamma)}), \qquad for \quad j = 1, \cdots, M\gamma$$

.

Therefore, 100(1 - γ)% credible interval for θ_1 becomes $(\theta_{1(j^*)}, \theta_{1(j^*+M-M\gamma)})$, where j^* is such that

$$\theta_{1(j^*+M-M\gamma)} - \theta_{1(j^*)} \leq \theta_{1(j+M-M\gamma)} - \theta_{1(j)}$$

for all $j = 1, \cdots, M\gamma$. Similarly, we can obtain the credible interval for other co-ordinates of θ. In this case, the parameter vector $\theta = (\alpha_0, \alpha_1, \alpha_2)$.

We have scope to construct such intervals when the full posterior is not known and tractable. In this chapter, we calculate the Bayesian confidence interval for both gamma prior and reference prior. We skip working with the full expression of the posterior under the reference prior as it is not tractable. We use the R package 'coda 'to obtain the credible intervals described above.

4.4 Numerical Results

The numerical results are obtained by using programming software R 3.2.3. The codes are run at IIT Guwahati computers with model : Intel(R) Core(TM) i5-6200U CPU 2.30GHz. The codes will be available on request to the authors.

We use the following hyper parameters of gamma prior: $k_0 = 2,\ \theta_0 = 3,\ k_1 = 4,\ \theta_1 = 3,\ k_2 = 3,\ \theta_2 = 2$. Bayesian estimates, mean square errors, credible intervals are calculated for all the parameters α_0, α_1 and α_2 using both gamma prior and reference prior. The results do not depend much on the choice of hyper parameters. Just to verify the fact we have shown some results taking gamma as prior in Table-4.3. Table-4.1 and Table-4.2 show the results obtained by different methods, e.g. the Lindley and the slice sampling, etc. for the absolute continuous Marshall-Olkin bivariate Pareto distribution with two different parameter sets. In slice cum Gibbs sampling, we take burn in period as 500. The Bayesian estimates are calculated based on 2000 iterations after burn-in period. We made further investigation of sample size needed for all the methods to work. In case of BB-BVPA distribution, slice cum Gibbs sampling with gamma as prior demands 250 or more samples to converge for slightly larger parameter values, whereas for parameter values closer to zero, it is sufficient to work with sample size around 50. Required sample size is more (around 350), as expected in the case of reference prior. The Lindley approximation works for sample size around 250 in this case for almost all ranges of parameters. Note that slice sampling iterations are done based on width as 1. However, in this case study, we observe that it also works for moderately small or large choice of width. The mean square error is calculated based on 200 different samples drawn from the original set of parameters. For gamma prior straight forward step out slice sampling works, as posterior distribution is a continuous function. However, we use modified slice sampling algorithm for the posterior based on reference prior as posterior in this case is a discontinuous function within specific ranges of parameter values. We can easily see this thing in Figure-4.1. In this Figure-4.1, we plot the function square of the reference prior against α_2 for given α_1 and α_0 for some particular data set.

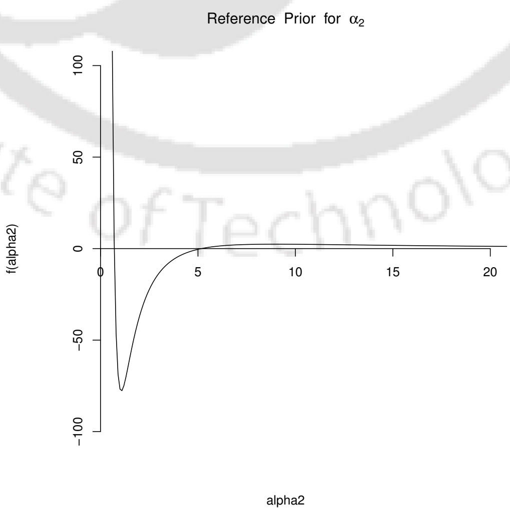


Figure 4.1: Square of the prior with respect to α_2 when α_1 and α_0 is given

Slice-cum-Gibbs			
Gamma Prior			
n = 450			
Original Parameter Sets	$\alpha_0 = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$
Starting Value	0.8947	0.1591	0.6613
Bayesian Estimates	0.1165	0.1910	0.3951
Mean Square Error	0.00026	0.000075	0.0000218
Credible Intervals	[0.0567, 0.1828]	[0.1442, 0.2350]	[0.3295, 0.4587]
n = 1000			
Original Parameter Sets	$\alpha_0 = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$
Starting Value	0.8312	0.2278	0.7127
Bayesian Estimates	0.0840	0.2100	0.4062
Mean Square Error	0.00032	0.00013	0.00006
Credible Intervals	[0.0384, 0.1298]	[0.1745, 0.2446]	[0.3601, 0.4517]
Reference Prior			
n = 450			
Original Parameter Sets	$\alpha_0 = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$
Starting Value	0.05	0.1	0.2
Bayesian Estimates	0.1214	0.1846	0.3877
Mean Square Error	0.00057	0.00027	0.00023
Credible Intervals	[0.0666, 0.1957]	[0.1354, 0.2264]	[0.3151, 0.4474]
n = 1000			
Original Parameter Sets	$\alpha_0 = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$
Starting Value	0.05	0.1	0.2
Bayesian Estimates	0.0838	0.2090	0.4052
Mean Square Error	0.00028	0.000094	0.00003
Credible Intervals	[0.0321, 0.1290]	[0.1750, 0.2488]	[0.3581, 0.4570]
Lindley			
n = 450			
Gamma Prior			
Original Parameter Sets	$\alpha_0 = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$
Bayesian Estimates	0.0751	0.2199	0.4187
Mean Square Error	0.1856	0.1070	0.1130
n = 1000			
Original Parameter Sets	$\alpha_0 = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$
Bayesian Estimates	0.1009	0.1997	0.3989
Mean Square Error	0.0007	0.0004	0.0007

Table 4.1: The Bayesian Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\alpha_0 = 0.1$, $\alpha_1 = 0.2$ and $\alpha_2 = 0.4$

Slice-cum-Gibbs			
Gamma Prior			
n = 450			
Original Parameter Sets	$\alpha_0 = 4$	$\alpha_1 = 5$	$\alpha_2 = 10$
Starting Value	2.5285	2.7894	4.0057
Bayesian Estimates	4.1589	4.8930	9.6534
Mean Square Error	0.0532	0.0248	0.1577
Credible Intervals	[2.417, 5.884]	[3.768, 6.248]	[7.974, 11.450]
n = 1000			
Original Parameter Sets	$\alpha_0 = 4$	$\alpha_1 = 5$	$\alpha_2 = 10$
Starting Value	0.5739	2.6358	2.2957
Bayesian Estimates	3.9486	4.8672	9.6471
Mean Square Error	0.0237	0.0045	0.0744
Credible Intervals	[2.6077, 5.1435]	[3.979, 5.7306]	[8.4117, 10.8747]
Reference Prior			
n = 450			
Original Parameter Sets	$\alpha_0 = 4$	$\alpha_1 = 5$	$\alpha_2 = 10$
Starting Value	2.5285	5.9298	5.6676
Bayesian Estimates	4.3734	4.7059	9.5367
Mean Square Error	0.1173	0.0716	0.2017
Credible Intervals	[2.5335, 6.3448]	[3.4889, 5.9824]	[7.5845, 11.4139]
n = 1000			
Original Parameter Sets	$\alpha_0 = 4$	$\alpha_1 = 5$	$\alpha_2 = 10$
Starting Value	0.5739	5.8786	5.3826
Bayesian Estimates	3.9283	4.8652	9.6974
Mean Square Error	0.0146	0.0124	0.0815
Credible Intervals	[2.6447, 5.1663]	[3.9432, 5.7327]	[8.2888, 10.9374]
Lindley			
n = 450			
Gamma Prior			
Original Parameter Sets	$\alpha_0 = 4$	$\alpha_1 = 5$	$\alpha_2 = 10$
Bayesian Estimates	3.7392	5.2327	10.4020
Mean Square Error	1.3090	0.6822	1.3167
n = 1000			
Original Parameter Sets	$\alpha_0 = 4$	$\alpha_1 = 5$	$\alpha_2 = 10$
Bayesian Estimates	3.7977	5.1595	10.2419
Mean Square Error	0.4278	0.2269	0.4682

Table 4.2: The Bayesian Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\alpha_0 = 4$, $\alpha_1 = 5$ and $\alpha_2 = 10$

Slice-cum-Gibbs			
Gamma Prior			
n = 1000			
Original Parameter Sets	$\alpha_0 = 3.808$	$\alpha_1 = 1.105$	$\alpha_2 = 1.006$
Starting Value	0.4794	0.8654	0.7781
Bayesian Estimates	4.063	0.892	0.809
Mean Square Error	0.212	0.118	0.0995
Credible Intervals	[4.579, 5.368]	[0.018, 0.387]	[0.018, 0.380]
Coverage Prob.	0.865	0.825	0.805

Table 4.3: The Bayesian Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\alpha_0 = 3.808$, $\alpha_1 = 1.105$ and $\alpha_2 = 1.006$ and hyper parameters as $k_0 = 0.2,\ \theta_0 = 0.3,\ k_1 = 0.4,\ \theta_1 = 0.3,\ k_2 = 0.3,\ \theta_2 = 0.2$.

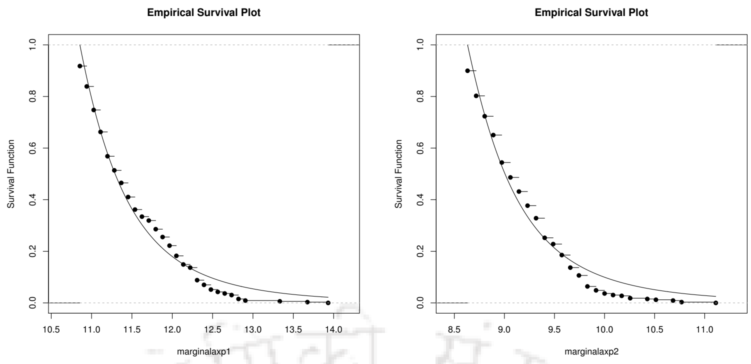
4.5 Data Analysis

The data set is taken from *https://archive.ics.uci.edu/ml/machine-learning-databases*. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope. The data set contains related measurements. We extract a part of the data for the bivariate modeling. We consider only the measurements related to the female population where one of the variable is the length as a longest shell measurement and another variable is diameter, which is perpendicular to the length. We use peak over threshold method on this data set

From [15], we know that peak over threshold method on random variable U (say) provides polynomial generalized Pareto distribution for any x_0 with $1 + log(G(x_0)) \in (0, 1)$, i.e., $P(U > tx_0|U > x_0) = t^{-\alpha},\ \ t \geq 1$, where $G(\cdot)$ is the distribution function of U. We choose an appropriate t and x_0 on the marginals of the bivariate data, so that the data should behave more like Pareto distribution. The transformed data set does not have any singular component. Therefore, one possible assumption can be the absolute continuous Marshall Olkin bivariate Pareto.

This data set are used to model the seven parameter BB-BVPA distribution. The EM estimates for the BB-BVPA distribution produces the values as $\mu_1 = 10.855$, $\mu_2 = 8.632$, $\sigma_1 = 2.124$, $\sigma_2 = 1.7110$, $\alpha_0 = 3.124$, $\alpha_1 = 1.743$ and $\alpha_2 = 1.602$. Figure-4.2 shows that the empirical marginals coincide with the marginals calculated from the estimated parameters.

We also verify our assumption by plotting the empirical two dimensional density plot in Figure-4.3, which resembles closer to the surface of the absolute continuous Marshall-Olkin bivariate Pareto distribution.



(a) ξ_1 (b) ξ_2

Figure 4.2: Survival plots for two marginals of Three parameter BB-BVPA of the transformed dataset

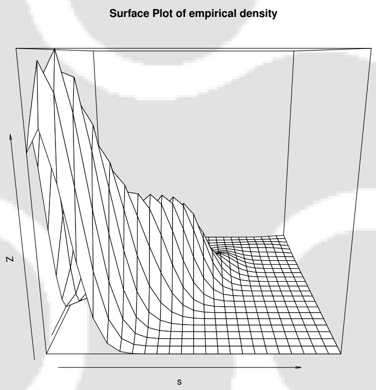


Figure 4.3: Two dimensional density plots for Three parameter BB-BVPA of the transformed dataset

Direct real life data which will model three parameter BB-BVPA distribution is not available. Therefore, we modify the data with location and scale transformation. We observe the number of singular observations after the transformation is zero. Therefore, this could be a reasonable data to work with. The sample size for the data set after transformation is 329. Note that even after location, scale transformation observed cardinalities of I_1 and I_2 are good representative of the actual ones. The Bayesian estimates are calculated and provided in Table-4.4 under the different methods and the different choice of priors based on this data set. We use the parametric bootstrap technique to generate different samples from this three parameter BB-BVPA distribution and then find out the mean square error and the coverage probabilities for the constructed credible interval. Results are available in Table-4.5 and Table-4.6. For the gamma prior, we observe that the mean square error is higher for α_0 as compared to α_1 and α_2 and the coverage probabilities are much lower than the desired 95% confidence level. In this case study, we observe that we cannot use the reference prior to form the posterior for this data set, since sample size is not sufficiently large. While making a simulation study, we see that with sample size 329, the Bayes estimates in case of reference prior, do not converge to the actual value for α_1 and α_2. However, if we increase the sample size above 750, say 1000, we get low mean square error values and the coverage probabilities closer to 95% using the reference as a prior.

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Bayesian Estimates	3.808	1.105	1.006
Reference Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	2	0.5	0.5
Bayesian Estimates	3.320	1.606	1.477
Lindley			
Gamma Prior			
Original Parameter Sets	α_0	α_1	α_2
Bayesian Estimates	3.572	1.4093	1.30002

Table 4.4: The Bayesian Estimates (BE) of the parameters based on Abalone data

Slice-cum-Gibbs			
Gamma Prior			
n = 329			
Original Parameter Sets	$\alpha_0 = 3.808$	$\alpha_1 = 1.105$	$\alpha_2 = 1.006$
Starting Value	0.4794	0.8654	0.7781
Mean Square Error	0.703	0.420	0.348
Credible Intervals	[2.879, 4.367]	[0.539, 1.449]	[0.392, 1.376]
Coverage Probability	0.705	0.670	0.695
n = 1000			
Original Parameter Sets	$\alpha_0 = 3.808$	$\alpha_1 = 1.105$	$\alpha_2 = 1.006$
Starting Value	0.4794	0.8654	0.7781
Mean Square Error	0.197	0.111	0.089
Credible Intervals	[3.235, 4.227]	[0.574, 1.244]	[0.544, 1.170]
Coverage Probability	0.845	0.805	0.84

Table 4.5: Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\alpha_0 = 3.808$, $\alpha_1 = 1.105$ and $\alpha_2 = 1.006$ and the hyper parameters of the priors as $k_0 = 2,\ \theta_0 = 3,\ k_1 = 4,\ \theta_1 = 3,\ k_2 = 3,\ \theta_2 = 2$

Slice-cum-Gibbs			
Reference Prior			
n = 1000			
Original Parameter Sets	α_0	α_1	α_2
Starting Value	2	0.5	0.5
Mean Square Error	0.162	0.087	0.070
Credible Intervals	[2.834, 4.066]	[0.310, 1.142]	[0.306, 1.078]
Coverage Probability	0.94	0.924	0.935

Table 4.6: Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\alpha_0 = 3.320$, $\alpha_1 = 1.606$ and $\alpha_2 = 1.477$ and the hyper parameters of the priors as $k_0 = 2,\ \theta_0 = 3,\ k_1 = 4,\ \theta_1 = 3,\ k_2 = 3,\ \theta_2 = 2$

4.6 Conclusion

The Bayesian estimation of the parameters of the three parameters absolute continuous bivariate Pareto under square error loss are obtained using both Lindley and slice cum Gibbs sampler approach. Both the methods are working quite well, even for moderately large sample size. Use of informative prior like Gamma and non-informative prior like reference prior is studied in this context. The algorithms have not worked always for the small sample size with reference as a prior. The coverage Probability for the credible interval is coming little low for the gamma prior. More work in needed that direction. The same study can be made using many other algorithms like importance sampling, HMC, etc. The Bayesian estimation in case of the seven parameter bivariate Pareto with location and scale parameters is a challenging problem. We deal with this problem in the next chapter.

Chapter 5

Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters

5.1 Introduction

In this chapter, we consider the Bayesian analysis of the parameters of the absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters through the slice sampling. Usual slice sampling will not work with the Monte Carlo set up as the posterior function in this case is discontinuous with respect to location and scale parameters. We propose to use two novel approaches in the implementation of slice sampling to perform the Bayesian analysis.

Recently, three parameter Bayesian analysis of the absolute continuous version of Marshall-Olkin bivariate Pareto distribution is studied by Paul, Dey and Kundu (2018)[41]. This absolute continuous version of the Marshall-Olkin bivariate Pareto distribution has marginals which are not type-II univariate Pareto distributions. We use the notation BB-BVPA for the absolute continuous Marshall-Olkin bivariate Pareto. This form of Marshall-Olkin bivariate Pareto is similar to the absolute continuous bivariate exponential distribution, proposed by Block and Basu (1974) [9]. A variety of bivariate (multivariate) extensions of the bivariate Pareto distribution also have been studied in the literature. These include the distributions described in the following works : [51], [55], [56], [2], [3]. Finding an efficient technique to estimate the parameters of BB-BVPA distribution, particularly in the presence of location and scale parameters was a major challenge in last few decades. There is no work available for seven parameters Bayesian analysis for BB-BVPA distribution.

Finding the confidence interval for location and scale parameters for BB-BVPA distribution is a difficult problem in frequentist set up. In this situation, Bayesian credible intervals can easily solve this problem. Sometimes, Bayes estimators exist when MLEs do not exist. Bayesian estimators may work reasonably well with the suitable choice of the prior even when MLE's performance is extremely poor. Therefore, working with Bayesian set up with such a complicated distribution has its own advantages. Here, we use the gamma prior for shape and scale parameters. However, for location parameters we use the truncated normal as prior.

The Rest of the chapter is organized as follows. Section-5.2 shows the formulation of the absolute continuous Marshall-Olkin bivariate Pareto distribution. Bayesian analysis through the slice sampling is described in Section-5.3. Section-5.4 deals with the construction of the credible interval. Numerical results are discussed in section-5.5. Data analysis is shown in section-5.6. Finally, we conclude this chapter in Section-5.7.

5.2 Marshall-Olkin bivariate Pareto Distribution

Let U_0, U_1 and U_2 are three mutually independent random variable, where $U_0 \sim PA(II)(0, 1, \alpha_0)$, $U_1 \sim PA(II)(\mu_1, \sigma_1, \alpha_1)$ and $U_2 \sim PA(II)(\mu_2, \sigma_2, \alpha_2)$. We define $X_1 = \min\{\mu_1 + \sigma_1 U_0, U_1\}$ and $X_2 = \min\{\mu_2 + \sigma_2 U_0, U_2\}$, then the joint distribution of (X_1, X_2) is called the **Marshall-Olkin bivariate Pareto (MOBVPA) distribution** or singular bivariate Pareto distribution. The joint survival function of (X_1, X_2) can be written for $z = \max\{\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}\}$ as;

$$
\begin{aligned}
S(x_1, x_2) &= (1+z)^{-\alpha_0}\Big(1 + \frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1}\Big(1 + \frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2} \\
&= \begin{cases} S_1(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ S_2(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ S_0(x), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x, \end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
S_1(x_1, x_2) &= \Big(1 + \frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2}\Big(1 + \frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1} \\
S_2(x_1, x_2) &= \Big(1 + \frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2}\Big(1 + \frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1} \\
S_0(x) &= \Big(1 + x\Big)^{-\alpha_0-\alpha_1-\alpha_2}.
\end{aligned}
$$

So its pdf can be written as

$$
f(x_1, x_2) = \begin{cases} f_1(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x. \end{cases}
$$

where

$$
f_1(x, x_2) = \frac{\alpha_1(\alpha_0+\alpha_2)}{\sigma_1\sigma_2}\Big(1 + \frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2-1}\Big(1 + \frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1-1}
$$

$$f_2(x_1,x_2) = \frac{\alpha_2(\alpha_0+\alpha_1)}{\sigma_1\sigma_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2-1}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1-1}$$
$$f_0(x) = \alpha_0\Big(1+x\Big)^{-\alpha_0-\alpha_1-\alpha_2-1}.$$

We denote this distribution as $MOBVPA(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha_0,\alpha_1,\alpha_2)$.

5.2.1 Block Basu bivariate Pareto Distribution

We know that joint survival function of (X_1,X_2) can be written as a mixture of an absolute continuous part and a singular part as follows;

$$S(x_1,x_2) = pS_a(x_1,x_2) + (1-p)S_s(x_1,x_2),$$

where $S_a(x_1,x_2)$ is the absolute continuous part and $S_s(x_1,x_2)$ is the singular part. Also $p = \frac{\alpha_1+\alpha_2}{\alpha_0+\alpha_1+\alpha_2}$. Note that for $z = \max\{\frac{x_1-\mu_1}{\sigma_1},\frac{x_2-\mu_2}{\sigma_2}\}$,

$$S_s(x_1,x_2) = (1+z)^{-\alpha_0-\alpha_1-\alpha_2}.$$

We define $S_a(x_1,x_2)$ as the joint survival function of the **Block Basu bivariate Pareto (BB-BVPA) distribution** or absolute continuous bivariate pareto distribution. It can then be expressed as

$$S_a(x_1,x_2) = \begin{cases} S_{a1}(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ S_{a2}(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2}, \end{cases}$$

where

$$\begin{aligned} S_{a1}(x_1,x_2) &= \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1} \\ &\quad -\frac{\alpha_0}{\alpha_1+\alpha_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_1-\alpha_2}, \\ S_{a2}(x_1,x_2) &= \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1} \\ &\quad -\frac{\alpha_0}{\alpha_1+\alpha_2}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1-\alpha_2}. \end{aligned}$$

So its pdf can be written as

$$f_a(x_1,x_2) = \begin{cases} \frac{1}{p}f_1(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ \frac{1}{p}f_2(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2}, \end{cases}$$

where

$$\frac{1}{p}f_1(x_1,x_2) = \frac{(\alpha_0+\alpha_1+\alpha_2)}{(\alpha_1+\alpha_2)}\frac{\alpha_1(\alpha_0+\alpha_2)}{\sigma_1\sigma_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2-1}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1-1},$$

$$\frac{1}{p}f_2(x_1,x_2) = \frac{(\alpha_0+\alpha_1+\alpha_2)}{(\alpha_1+\alpha_2)}\frac{\alpha_2(\alpha_0+\alpha_1)}{\sigma_1\sigma_2}\left(1+\frac{x_2-\mu_2}{\sigma_2}\right)^{-\alpha_2-1}\left(1+\frac{x_1-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-1}.$$

We denote this distribution as $BB-BVPA(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha_0,\alpha_1,\alpha_2)$. One can easily calculate the marginal distribution of X_1 and X_2. The pdf expressions are as follows:

$$f_{X_1}(x_1) = \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}\left\{\frac{\alpha_0+\alpha_1}{\sigma_1}\left(1+\frac{x_1-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-1} - \frac{\alpha_0}{\sigma_1}\left(1+\frac{x_1-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-\alpha_2-1}\right\},$$

$$f_{X_2}(x_2) = \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}\left\{\frac{\alpha_0+\alpha_2}{\sigma_2}\left(1+\frac{x_2-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_2-1} - \frac{\alpha_0}{\sigma_2}\left(1+\frac{x_2-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_1-\alpha_2-1}\right\}.$$

5.2.2 Likelihood Function

Now we divide our data set $I = \{(x_{11},x_{21}),(x_{12},x_{22})\cdots,(x_{1n},x_{2n})\}$ into three parts. They are as follow, $I_0 = \{i : \frac{x_{1i}-\mu_1}{\sigma_1} = \frac{x_{2i}-\mu_2}{\sigma_2}\}$, $I_1 = \{i : \frac{x_{1i}-\mu_1}{\sigma_1} < \frac{x_{2i}-\mu_2}{\sigma_2}\}$, $I_2 = \{i : \frac{x_{1i}-\mu_1}{\sigma_1} > \frac{x_{2i}-\mu_2}{\sigma_2}\}$ and $n_0 = |I_0|$, $n_1 = |I_1|$, $n_2 = |I_2|$, where $|I_i|$ denotes the number of observations in I_i. The total number of observations $n = n_0 + n_1 + n_2$.

Therefore, the log-likelihood function takes the form,

$$\begin{aligned}
& L(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha_0,\alpha_1,\alpha_2) \\
= \;& n\ln(\alpha_0+\alpha_1+\alpha_2) - n\ln(\alpha_1+\alpha_2) + n_1\ln\alpha_1 + n_1\ln(\alpha_0+\alpha_2) \\
- \;& n_1\ln\sigma_1 - n_1\ln\sigma_2 - (\alpha_0+\alpha_2+1)\sum_{i\in I_1}\ln\left(1+\frac{x_{2i}-\mu_2}{\sigma_2}\right) \\
- \;& (\alpha_1+1)\sum_{i\in I_1}\ln\left(1+\frac{x_{1i}-\mu_1}{\sigma_1}\right) - n_2\ln\sigma_1 - n_2\ln\sigma_2 + n_2\ln\alpha_2 \\
+ \;& n_2\ln(\alpha_0+\alpha_1) - (\alpha_0+\alpha_1+1)\sum_{i\in I_2}\ln\left(1+\frac{x_{1i}-\mu_1}{\sigma_1}\right) \\
- \;& (\alpha_2+1)\sum_{i\in I_2}\ln\left(1+\frac{x_{2i}-\mu_2}{\sigma_2}\right).
\end{aligned}$$

5.3 Bayesian Analysis

The first step of all Bayesian analysis follows the Bayesian theorem to calculate the posterior distribution. We use $posterior \propto likelihood \times prior$. Therefore, we need to choose a suitable choice of a set of priors in this case.

5.3.1 Prior Assumption

We assume that the shape parameters α_0, α_1 and α_2 are distributed according to the gamma distribution with shape parameters k_i and scale parameters θ_i, i.e.,

$$\pi_0(\alpha_0) = \text{Gamma}(k_0,\theta_0) = \frac{1}{\Gamma(k_0)\theta_0^{k_0}}\alpha_0^{k_0-1}e^{-\frac{\alpha_0}{\theta_0}}, \quad \alpha_0 > 0$$

$$\pi_1(\alpha_1) = \text{Gamma}(k_1, \theta_1) = \frac{1}{\Gamma(k_1)\theta_1^{k_1}} \alpha_1^{k_1-1} e^{-\frac{\alpha_1}{\theta_1}}, \quad \alpha_1 > 0$$
$$\pi_2(\alpha_2) = \text{Gamma}(k_2, \theta_2) = \frac{1}{\Gamma(k_2)\theta_2^{k_2}} \alpha_2^{k_2-1} e^{-\frac{\alpha_2}{\theta_2}}, \quad \alpha_2 > 0$$

where $k_i > 0$ and $\theta_i > 0$, $i = 0, 1, 2$. Here $\Gamma(k)$ is the gamma function evaluated at k. Suppose, f_N is the probability density function of the normal distribution. Let's assume that the location parameters μ_1 and μ_2 are distributed according to the truncated normal distribution with mean = $\mu_i^{'}$ and variance = $\sigma_i^{'}$, $i = 1, 2$,

$$\pi_3(\mu_1) = \frac{f_N(\mu_1; \mu_1^{'}, \sigma_1^{'})}{\Phi(\min\{x_{1i}\}; \mu_1^{'}, \sigma_1^{'})}$$
$$\pi_4(\mu_2) = \frac{f_N(\mu_2; \mu_2^{'}, \sigma_2^{'})}{\Phi(\min\{x_{2i}\}; \mu_2^{'}, \sigma_2^{'})}$$

Further, We assume that the scale parameters σ_1 and σ_2 are distributed according to the gamma distribution with the hyper-parameters c_1, c_2, d_1 and d_2,

$$\pi_5(\sigma_1) = \text{Gamma}(c_1, d_1) = \frac{1}{\Gamma(c_1)d_1^{c_1}} \sigma_1^{c_1-1} e^{-\frac{\sigma_1}{d_1}}, \quad \sigma_1 > 0$$
$$\pi_6(\sigma_2) = \text{Gamma}(c_2, d_2) = \frac{1}{\Gamma(c_2)d_2^{c_2}} \sigma_2^{c_2-1} e^{-\frac{\sigma_2}{d_2}}, \quad \sigma_2 > 0$$

where $c_i > 0$ and $d_i > 0$, i = 1, 2.

One important assumption here is prior distributions of all parameters are independent of each other.

5.3.2 Proposed Methodology

In this chapter, we obtain the Bayes estimates through the slice cum Gibbs sampler. To form the Gibbs sampler first, we need to find out the expressions for the conditional distribution of each parameter given the other parameters and the data. We provide the expressions of logarithm of those conditional distributions.

The log conditional posterior distributions of α_0, α_1, α_2, μ_1, μ_2, σ_1, and σ_2 are given by,

$$\begin{aligned} \ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, x_1, x_2)) &\propto n \ln(\alpha_0 + \alpha_1 + \alpha_2) + n_1 \ln(\alpha_0 + \alpha_2) \\ &- (\alpha_0 + \alpha_2 + 1) \textstyle\sum_{i \in I_1} \ln\left(1 + \frac{x_{2i} - \mu_2}{\sigma_2}\right) + n_2 \ln(\alpha_0 + \alpha_1) \\ &- (\alpha_0 + \alpha_1 + 1) \textstyle\sum_{i \in I_2} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) + (k_0 - 1) \ln \alpha_0 - \frac{\alpha_0}{\theta_0}, \end{aligned} \tag{5.3.1}$$

$$\begin{aligned} \ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, x_1, x_2)) &\propto n \ln(\alpha_0 + \alpha_1 + \alpha_2) - n \ln(\alpha_1 + \alpha_2) + n_1 \ln \alpha_1 \\ &- (\alpha_1 + 1) \textstyle\sum_{i \in I_1} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) + n_2 \ln(\alpha_0 + \alpha_1) \\ &- (\alpha_0 + \alpha_1 + 1) \textstyle\sum_{i \in I_2} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) + (k_1 - 1) \ln \alpha_1 - \frac{\alpha_1}{\theta_1}, \end{aligned} \tag{5.3.2}$$

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \alpha_0,\alpha_1,\mu_1,\mu_2,\sigma_1,\sigma_2,x_1,x_2)) &\propto n\ln(\alpha_0+\alpha_1+\alpha_2)-n\ln(\alpha_1+\alpha_2)\\ &+n_1\ln(\alpha_0+\alpha_2)-(\alpha_0+\alpha_2+1)\textstyle\sum_{i\in I_1}\ln\Big(1+\frac{x_{2i}-\mu_2}{\sigma_2}\Big)+n_2\ln\alpha_2\\ &-(\alpha_2+1)\textstyle\sum_{i\in I_2}\ln\Big(1+\frac{x_{2i}-\mu_2}{\sigma_2}\Big)+(k_2-1)\ln\alpha_2-\frac{\alpha_2}{\theta_2},\end{aligned} \tag{5.3.3}$$

$$\begin{aligned}\ln(\pi(\mu_1 \mid \alpha_0,\alpha_1,\alpha_2,\mu_2,\sigma_1,\sigma_2,x_1,x_2)) &\propto -(\alpha_1+1)\textstyle\sum_{i\in I_1}\ln\Big(1+\frac{x_{1i}-\mu_1}{\sigma_1}\Big)\\ &-(\alpha_0+\alpha_1+1)\textstyle\sum_{i\in I_2}\ln\Big(1+\frac{x_{1i}-\mu_1}{\sigma_1}\Big)-0.5\frac{(\mu_1-\mu_1')^2}{(\sigma_1')^2},\end{aligned} \tag{5.3.4}$$

$$\begin{aligned}\ln(\pi(\mu_2 \mid \alpha_0,\alpha_1,\alpha_2,\mu_1,\sigma_1,\sigma_2,x_1,x_2)) &\propto -(\alpha_0+\alpha_2+1)\textstyle\sum_{i\in I_1}\ln\Big(1+\frac{x_{2i}-\mu_2}{\sigma_2}\Big)\\ &-(\alpha_2+1)\textstyle\sum_{i\in I_2}\ln\Big(1+\frac{x_{2i}-\mu_2}{\sigma_2}\Big)-0.5\frac{(\mu_2-\mu_2')^2}{(\sigma_2')^2},\end{aligned} \tag{5.3.5}$$

$$\begin{aligned}\ln(\pi(\sigma_1 \mid \alpha_0,\alpha_1,\alpha_2,\mu_1,\mu_2,\sigma_2,x_1,x_2)) &\propto -n_1\ln\sigma_1-(\alpha_1+1)\textstyle\sum_{i\in I_1}\ln\Big(1+\frac{x_{1i}-\mu_1}{\sigma_1}\Big)\\ &-n_2\ln\sigma_1-(\alpha_0+\alpha_1+1)\textstyle\sum_{i\in I_2}\ln\Big(1+\frac{x_{1i}-\mu_1}{\sigma_1}\Big)+(c_1-1)\ln(\sigma_1)-\frac{\sigma_1}{d_1}\end{aligned} \tag{5.3.6}$$

and

$$\begin{aligned}&\ln(\pi(\sigma_2 \mid \alpha_0,\alpha_1,\alpha_2,\mu_1,\mu_2,\sigma_1,x_1,x_2)) \propto -n_1\ln\sigma_2-n_2\ln\sigma_2\\ &-(\alpha_0+\alpha_2+1)\textstyle\sum_{i\in I_1}\ln\Big(1+\frac{x_{2i}-\mu_2}{\sigma_2}\Big)-(\alpha_2+1)\textstyle\sum_{i\in I_2}\ln\Big(1+\frac{x_{2i}-\mu_2}{\sigma_2}\Big)\\ &+(c_2-1)\ln(\sigma_2)-\frac{\sigma_2}{d_2}.\end{aligned} \tag{5.3.7}$$

We also require the conditional distribution of μ_1, μ_2, σ_1 and σ_2 from the posterior distribution based on the univariate likelihood function.

$$\begin{aligned}\ln(\pi(\mu_1 \mid \alpha_0,\alpha_1,\alpha_2,\mu_2,\sigma_1,\sigma_2,x_1)) &\propto \sum_{i=1}^{n}\ln\left\{\frac{\alpha_0+\alpha_1}{\sigma_1}\left(1+\frac{x_{1i}-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-1}\right.\\ &-\left.\frac{\alpha_0}{\sigma_1}\left(1+\frac{x_{1i}-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-\alpha_2-1}\right\}-0.5\frac{(\mu_1-\mu_1')^2}{(\sigma_1')^2},\end{aligned} \tag{5.3.8}$$

$$\begin{aligned}\ln(\pi(\mu_2 \mid \alpha_0,\alpha_1,\alpha_2,\mu_1,\sigma_1,\sigma_2,x_2)) &= \sum_{i=1}^{n}\ln\left\{\frac{\alpha_0+\alpha_2}{\sigma_2}\left(1+\frac{x_{2i}-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_2-1}\right.\\ &-\left.\frac{\alpha_0}{\sigma_2}\left(1+\frac{x_{2i}-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_1-\alpha_2-1}\right\}-0.5\frac{(\mu_2-\mu_2')^2}{(\sigma_2')^2},\end{aligned} \tag{5.3.9}$$

$$\ln(\pi(\sigma_1 \mid \alpha_0,\alpha_1,\alpha_2,\mu_1,\mu_2,\sigma_2,x_1)) \propto \sum_{i=1}^{n}\ln\left\{\frac{\alpha_0+\alpha_1}{\sigma_1}\left(1+\frac{x_{1i}-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-1}\right.$$

$$- \frac{\alpha_0}{\sigma_1}\left(1+\frac{x_{1i}-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-\alpha_2-1}\Bigg\} + (c_1-1)\ln(\sigma_1) - \frac{\sigma_1}{d_1} \tag{5.3.10}$$

and

$$\ln(\pi(\sigma_2 \mid \alpha_0,\alpha_1,\alpha_2,\mu_1,\mu_2,\sigma_1,x_2)) \propto \sum_{i=1}^{n} \ln\Bigg\{\frac{\alpha_0+\alpha_2}{\sigma_2}\left(1+\frac{x_{2i}-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_2-1}$$

$$- \frac{\alpha_0}{\sigma_2}\left(1+\frac{x_{2i}-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_1-\alpha_2-1}\Bigg\} + (c_2-1)\ln(\sigma_2) - \frac{\sigma_2}{d_2}. \tag{5.3.11}$$

We propose to perform two variations of the algorithms to solve the problem using step-out method in the slice sampling ([38]) to generate the respective parameters.

Approach 1 : In this approach, we propose to use usual step out method for the slice sampling for each conditional distribution. Standard slice sampling works only for continuous posterior. The algorithm is as follows :

- Sample z and u uniformly from the area under the distribution, say $p(\cdot)$.
  1. Fix z, sample u uniformly from $[0, p(z)]$.
  2. Fix u, sample z uniformly from the slice through the region $\{z : p(z) > u\}$
- How to sample z from the slice.
  1. Start with the region of width w containing $z^{(t)}$.
  2. If end point in slice, then extend region by w in that direction.
  3. Sample $z^{'}$ uniformly from the region.
  4. If $z^{'}$ is in the slice, the accept it as $z^{(t+1)}$.
  5. If not : make $z^{'}$ new end point of the region, and resample $z^{'}$.

Since the above algorithm needs continuous posterior, it is not going to work in this case. Therefore, we suggest the following modifications. Instead of drawing sample from (5.3.4), (5.3.5), (5.3.6), (5.3.7), we draw μ_1, μ_2, σ_1 and σ_2 from (5.3.8), (5.3.9), (5.3.10), (5.3.11), which are based on univariate likelihood. Therefore, the steps of sampling are as follows :

- Start with some initial choice of parameters $\mu_1^{(0)}$, $\mu_2^{(0)}$, $\sigma_1^{(0)}$, $\sigma_2^{(0)}$, $\alpha_0^{(0)}$, $\alpha_1^{(0)}$ and $\alpha_2^{(0)}$.
- Sample $\alpha_0^{(t+1)}$, $\alpha_1^{(t+1)}$ and $\alpha_2^{(t+1)}$ from (5.3.1), (5.3.2) and (5.3.3) using standard step-out slice sampling.
- Sample $\mu_1^{(0)}$, $\mu_2^{(0)}$, $\sigma_1^{(0)}$, $\sigma_2^{(0)}$ from (5.3.8), (5.3.9), (5.3.10), (5.3.11) using standard step-out slice sampling.

- Go back to step-2.

Approach 2 : In this approach, we use directly the discontinuous conditional posterior based on the likelihood of the bivariate distribution. Using the slice sampling on each of these conditional distributions from (5.3.1) - (5.3.7) is not straight forward. We modify our slice sampling to fit for a discontinuous set up. Modified slice sampling steps are as follows :

- Sample z and u uniformly from the area under the distribution, say $p(\cdot)$.
  1. Fix z, sample u uniformly from $[0, p(z)]$.
  2. Fix u, sample z uniformly from the slice through the region $\{z : p(z) > u\}$
- How to sample z from the slice.
  1. Start with the region of width w containing $z^{(t)}$ **where it is defined**.
  2. If end point in slice and **in the region where the function is defined**, then extend region by w in that direction.
  3. If end point in slice and **in the region where the function is not defined**, then also extend region by w in that direction.
  4. Sample $z^{'}$ uniformly from the region **until the point is defined**.
  5. If $z^{'}$ is in the slice, the accept it as $z^{(t+1)}$.
  6. If not : make $z^{'}$ new end point of the region, and resample $z^{'}$.
  7. If resample point is not defined, **perform resampling until it is defined**.

Therefore the steps of sampling are as follows :

- Start with some initial choice of parameters $\mu_1^{(0)}$, $\mu_2^{(0)}$, $\sigma_1^{(0)}$, $\sigma_2^{(0)}$, $\alpha_0^{(0)}$, $\alpha_1^{(0)}$ and $\alpha_2^{(0)}$.
- Sample $\alpha_0^{(t+1)}$, $\alpha_1^{(t+1)}$, $\alpha_2^{(t+1)}$, $\mu_1^{(t+1)}$, $\mu_2^{(t+1)}$, $\sigma_1^{(t+1)}$, $\sigma_2^{(t+1)}$ from (5.3.1), (5.3.2), (5.3.3), (5.3.4), (5.3.5), (5.3.6) and (5.3.7) respectively using modified slice sampling.

5.4 Constructing credible intervals for Parameters

We have found the credible intervals for the parameters as described by Chen and Shao (1999) [11]. Let's assume $\underline{\theta}$ is a vector. To obtain the credible intervals of first variable θ_{1i}, we order $\{\theta_{1i}\}$, as $\theta_{1(1)} < \theta_{1(2)} < \cdots < \theta_{1(M)}$. Then 100(1 - γ)% credible interval of θ_1 become

$$(\theta_{1(j)}, \theta_{1(j+M-M\gamma)}), \qquad for \quad j = 1, \cdots, M\gamma$$

Therefore, 100(1 - γ)% credible interval for θ_1 becomes $(\theta_{1(j^*)}, \theta_{1(j^*+M-M\gamma)})$, where j^* is such that

$$\theta_{1(j^*+M-M\gamma)} - \theta_{1(j^*)} \leq \theta_{1(j+M-M\gamma)} - \theta_{1(j)}$$

for all $j = 1, \cdots, M\gamma$. Similarly, we can obtain the credible interval for other co-ordinates of θ.

In this context, we consider $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$ and $\gamma = 0.05$ to construct credible interval for all parameters.

5.5 Numerical Result

The numerical results are obtained by using the programming software R 3.2.3. The codes are run at IIT Guwahati computers with model : Intel(R) Core(TM) i5-6200U CPU 2.30 GHz. The codes will be available on request to the authors. Bayes estimates, mean square errors, credible intervals are calculated by two approaches for two sets of the parameters $\mu_1 = 0.3$, $\mu_2 = 0.4$, $\sigma_1 = 0.6$, $\sigma_2 = 0.7$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$ and $\mu_1 = 1.0$, $\mu_2 = 2.0$, $\sigma_1 = 0.4$, $\sigma_2 = 0.5$, $\alpha_0 = 0.5$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$. We use the following hyper parameters of prior as defined in Section-5.3 : $c_1 = 0.1, d_1 = 0.25, c_2 = 3, d_2 = 2, k_0 = 2, \theta_0 = 3, k_1 = 4, \theta_1 = 3, k_2 = 3, \theta_2 = 2, \mu_1^{'} = 0, \sigma_1^{'} = 1, \mu_2^{'} = 0, \sigma_2^{'} = 1$. However, we observe that the results does not vary much with different choice of hyper parameters.

Table-5.1, Table-5.2, Table-5.3 and Table-5.4 show Bayes estimates, 95% credible intervals for different parameter sets. We also calculate mean square errors and coverage probabilities for 95% credible intervals based on 200 different simulated samples from BB-BVPA. Table-5.1, Table-5.2 are results in Approach-1, whereas Table-5.3, Table-5.4 are th results obtained through Approach-2. In slice cum Gibbs sampling, we take the burn in period as 500. Bayes estimates are calculated based on 2000 and more iterations after burn-in period. We take bivariate sample size 450 and 1000 to get the estimates of the parameters. In both the slice sampling by step-out method, the width is taken 1. However, sampling procedure does not depend much on choice of width. Numerical result shows that estimates are very close to the original parameters. Mean square errors are on higher side for shape parameters as compared to other parameters. However, mean square errors decrease as sample size increases. Coverage Probabilities are more closer to 95% in Approach 2 than Approach 1. This shows the credible interval construction seems better fit for Approach 2 than Approach 1.

Slice-cum-Gibbs				
Gamma Prior				
n = 450				
Original Parameter Sets	$\mu_1 = 0.3$	$\mu_2 = 0.4$	$\sigma_1 = 0.6$	$\sigma_2 = 0.7$
Starting Value	0.1438	0.3462	0.7780	0.5386
Bayes Estimates	0.3001	0.4000	0.6984	0.6277
Mean Square Error	5.0340×10^{-7}	4.8804×10^{-7}	0.0397	0.0176
Credible Intervals	[0.2986, 0.3002]	[0.3998, 0.4014]	[0.2687, 0.5586]	[0.2960, 0.5923]
Coverage Probability	0.93	0.935	0.98	0.89
n = 450				
Original Parameter Sets		$\alpha_0 = 1.7$	$\alpha_1 = 1.2$	$\alpha_2 = 1.4$
Starting Value		0.4417	0.6138	0.4549
Bayes Estimates		1.2701	1.7544	1.4691
Mean Square Error		0.2921	0.6362	0.1078
Credible Intervals		[0.1496, 0.7516]	[0.3298, 1.0582]	[0.5259, 1.4262]
Coverage Probability		0.715	0.945	0.99
n = 1000				
Original Parameter Sets	$\mu_1 = 0.3$	$\mu_2 = 0.4$	$\sigma_1 = 0.6$	$\sigma_2 = 0.7$
Starting Value	0.1438	0.3462	0.7780	0.5386
Bayes Estimates	0.2999	0.4000	0.6463	0.6591
Mean Square Error	5.0900×10^{-8}	6.5169×10^{-8}	0.0182	0.0126
Credible Intervals	[0.2995, 0.3002]	[0.3994, 0.4002]	[0.3126, 0.5465]	[0.4007, 0.6287]
Coverage Probability	0.995	0.975	0.945	0.895
n = 1000				
Original Parameter Sets		$\alpha_0 = 1.7$	$\alpha_1 = 1.2$	$\alpha_2 = 1.4$
Starting Value		0.4417	0.6138	0.4549
Bayes Estimates		1.4754	1.4558	1.4248
Mean Square Error		0.0974	0.2059	0.0966
Credible Intervals		[0.8288, 1.4457]	[0.6454, 1.2945]	[0.5501, 1.1967]
Coverage Probability		0.79	0.945	0.96

Table 5.1: Approach 1 : The Bayes Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\mu_1 = 0.3$, $\mu_2 = 0.4$, $\sigma_1 = 0.6$, $\sigma_2 = 0.7$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$.

Slice-cum-Gibbs				
Gamma Prior				
n = 450				
Original Parameter Sets	$\mu_1 = 1.0$	$\mu_2 = 2.0$	$\sigma_1 = 0.4$	$\sigma_2 = 0.5$
Starting Value	0.4794	1.7308	0.7781	0.5386
Bayes Estimates	1.0000	1.1999	0.4493	0.5893
Mean Square Error	6.2330×10^{-7}	9.3982×10^{-7}	0.0110	0.0231
Credible Intervals	[0.9982, 1.0001]	[1.9986, 2.0011]	[0.1684, 0.3552]	[0.2401, 0.5190]
Coverage Probability	0.96	0.96	0.96	0.94
n = 450				
Original Parameter Sets		$\alpha_0 = 0.50$	$\alpha_1 = 0.70$	$\alpha_2 = 0.65$
Starting Value		0.4417	0.6138	0.4549
Bayes Estimates		0.6210	0.9100	0.8766
Mean Square Error		0.0625	0.0984	0.1009
Credible Intervals		[0.4630, 0.9564]	[0.2635, 0.7136]	[0.3051, 0.7579]
Coverage Probability		0.78	0.93	0.89
n = 1000				
Original Parameter Sets	$\mu_1 = 1.0$	$\mu_2 = 2.0$	$\sigma_1 = 0.4$	$\sigma_2 = 0.5$
Starting Value	0.4794	1.7308	0.7781	0.5386
Bayes Estimates	1.000	2.000	0.4175	0.5371
Mean Square Error	1.7234×10^{-7}	2.7793×10^{-7}	0.0053	0.0075
Credible Intervals	[0.9992, 1.0003]	[1.9994, 2.0006]	[0.2133, 0.3671]	[0.3366, 0.5236]
Coverage Probability	0.95	0.945	0.92	0.965
n = 1000				
Original Parameter Sets		$\alpha_0 = 0.50$	$\alpha_1 = 0.70$	$\alpha_2 = 0.65$
Starting Value		0.4417	0.6138	0.4549
Bayes Estimates		0.7198	0.7848	0.74445
Mean Square Error		0.0198	0.0335	0.0274
Credible Intervals		[0.6762, 1.0229]	[0.3201, 0.6573]	[0.2906, 0.6196]
Coverage Probability		0.895	0.915	0.925

Table 5.2: Approach 1 : The Bayes Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\mu_1 = 1.0$, $\mu_2 = 2.0$, $\sigma_1 = 0.4$, $\sigma_2 = 0.5$, $\alpha_0 = 0.5$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$.

Slice-cum-Gibbs				
Gamma Prior				
n = 450				
Original Parameter Sets	$\mu_1 = 1.0$	$\mu_2 = 2.0$	$\sigma_1 = 0.4$	$\sigma_2 = 0.5$
Starting Value	0.4794	1.7308	0.7781	0.5386
Bayes Estimates	1.0000	2.0000	0.4320	0.5822
Mean Square Error	6.6623×10^{-7}	1.2651×10^{-6}	0.0059	0.0155
Credible Intervals	[0.998, 1.0005]	[1.998, 2.0006]	[0.209, 0.414]	[0.293, 0.539]
Coverage Probability	0.93	0.935	0.99	0.96
n = 450				
Original Parameter Sets		$\alpha_0 = 0.50$	$\alpha_1 = 0.70$	$\alpha_2 = 0.65$
Starting Value		0.4417	0.6138	0.4549
Bayes Estimates		0.6034	0.9031	0.8882
Mean Square Error		0.0900	0.0973	0.1119
Credible Intervals		[0.1192, 0.3065]	[0.7859, 1.3529]	[0.3769, 0.9353]
Coverage Probability		0.775	0.915	0.86
n = 1000				
Original Parameter Sets	$\mu_1 = 1.0$	$\mu_2 = 2.0$	$\sigma_1 = 0.4$	$\sigma_2 = 0.5$
Starting Value	0.4794	1.7308	0.7781	0.5386
Bayes Estimates	1.0000	2.0000	0.4055	0.5443
Mean Square Error	1.4546×10^{-7}	2.7804×10^{-7}	0.0019	0.0053
Credible Intervals	[0.9992, 1.00008]	[1.9990, 2.0002]	[0.2256, 0.3636]	[0.3210, 0.5054]
Coverage Probability	0.945	0.955	0.98	0.975
n = 1000				
Original Parameter Sets		$\alpha_0 = 0.50$	$\alpha_1 = 0.70$	$\alpha_2 = 0.65$
Starting Value		0.4417	0.6138	0.4549
Bayes Estimates		0.7288	0.7611	0.7479
Mean Square Error		0.0207	0.0207	0.0264
Credible Intervals		[0.5037, 0.9138]	[0.3726, 0.7917]	[0.2776, 0.6572]
Coverage Probability		0.925	0.975	0.955

Table 5.3: Approach 2 : The Bayes Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\mu_1 = 1.0$, $\mu_2 = 2.0$, $\sigma_1 = 0.4$, $\sigma_2 = 0.5$, $\alpha_0 = 0.5$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$.

Slice-cum-Gibbs				
Gamma Prior				
n = 450				
Original Parameter Sets	$\mu_1 = 0.3$	$\mu_2 = 0.4$	$\sigma_1 = 0.6$	$\sigma_2 = 0.70$
Starting Value	0.1438	0.3462	0.7781	0.5386
Average Bayes Estimates	0.3001	0.40003	0.6376	0.6456
Mean Square Error	6.0022×10^{-7}	5.2204×10^{-7}	0.0099	0.0106
Credible Intervals	[0.2987, 0.3002]	[0.3986, 0.4003]	[0.3211, 0.6159]	[0.3301, 0.6358]
Coverage Probability	0.92	0.935	0.995	0.935
n = 450				
Original Parameter Sets		$\alpha_0 = 1.7$	$\alpha_1 = 1.2$	$\alpha_2 = 1.4$
Starting Value		0.4417	0.6138	0.4549
Average Bayes Estimates		1.3592	1.5300	1.4743
Mean Square Error		0.2360	0.2587	0.1252
Credible Intervals		[0.4086, 1.0383]	[0.5725, 1.6912]	[0.3620, 1.3647]
Coverage Probability		0.84	0.98	0.98
n = 1000				
Original Parameter Sets	$\mu_1 = 0.3$	$\mu_2 = 0.4$	$\sigma_1 = 0.6$	$\sigma_2 = 0.7$
Starting Value	0.1438	0.3462	0.7781	0.5386
Average Bayes Estimates	0.3000	0.3999	0.6229	0.6391
Mean Square Error	8.2786×10^{-8}	9.7774×10^{-8}	0.0054	0.0084
Credible Intervals	[0.2993, 0.3000]	[0.3996, 0.4003]	[0.3484, 0.5562]	[0.3801, 0.6119]
Coverage Probability	0.95	0.95	0.98	0.895
n = 1000				
Original Parameter Sets		$\alpha_0 = 1.7$	$\alpha_1 = 1.2$	$\alpha_2 = 1.4$
Starting Value		0.4417	0.6138	0.4549
Average Bayes Estimates		1.5937	1.3185	1.3127
Mean Square Error		0.0580	0.0620	0.0543
Credible Intervals		[1.0965, 1.8738]	[0.5557, 1.3337]	[0.5983, 1.3339]
Coverage Probability		0.975	0.99	0.955

Table 5.4: Approach 2 : The Bayes Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\mu_1 = 0.3$, $\mu_2 = 0.4$, $\sigma_1 = 0.6$, $\sigma_2 = 0.7$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$.

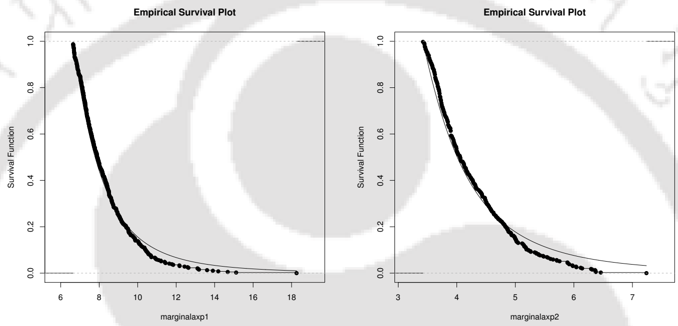
5.6 Case Study in real-life Data

We study one particular data set which is used by Paul, Dey and Kundu [43]. The data set is taken from *https://archive.ics.uci.edu/ml/machine-learning-databases*. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope. The data set contains related measurements. We extract a part of the data for the bivariate modeling. We consider only measurements related to the female population, where two of the variables are "Shucked weight" and "Viscera weight". They represent the weight of meat and the gut weight (after bleeding) of abalones, which are measured in grams. We use peak over threshold method on this data set.

From [15], we know that peak over threshold method on the random variable U (say) provides polynomial generalized Pareto distribution for any x_0 with $1 + log(G(x_0)) \in (0, 1)$ i.e. $P(U > tx_0|U > x_0) = t^{-\alpha}, \;\; t \geq 1$, where $G(\cdot)$ is the distribution function of U. We choose an appropriate t and x_0 on the marginals of the bivariate data, so that the data should behave more like Pareto distribution. The transformed data set does not have any singular component. Therefore, one possible assumption can be the absolute continuous Marshall Olkin bivariate Pareto.

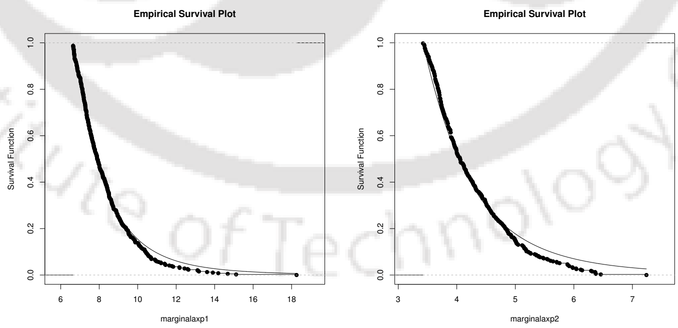
This data set are used to model the seven parameter BB-BVPA distribution. The EM estimates for BB-BVPA distribution produces the values as $\mu_1 = 6.644$, $\mu_2 = 3.423$, $\sigma_1 = 4.0375$, $\sigma_2 = 2.4821$, $\alpha_0 = 2.9622$, $\alpha_1 = 1.1400$ and $\alpha_2 = 1.4552$. Figure-5.1 and Figure-5.2 show that the empirical marginals coincide with the marginals calculated from the estimated parameters using Approach 1 and Approach 2 respectively.

We also verify our assumption by plotting empirical two dimensional density plot in Figure-5.3, which resembles closer to the surface of the absolute continuous Marshall-Olkin bivariate Pareto distribution.



(a) X_1 (b) X_2

Figure 5.1: Survival plots for two marginals of BB-BVPA of the transformed dataset through Approach 1



(a) X_1 (b) X_2

Figure 5.2: Survival plots for two marginals of BB-BVPA of the transformed dataset through Approach 2

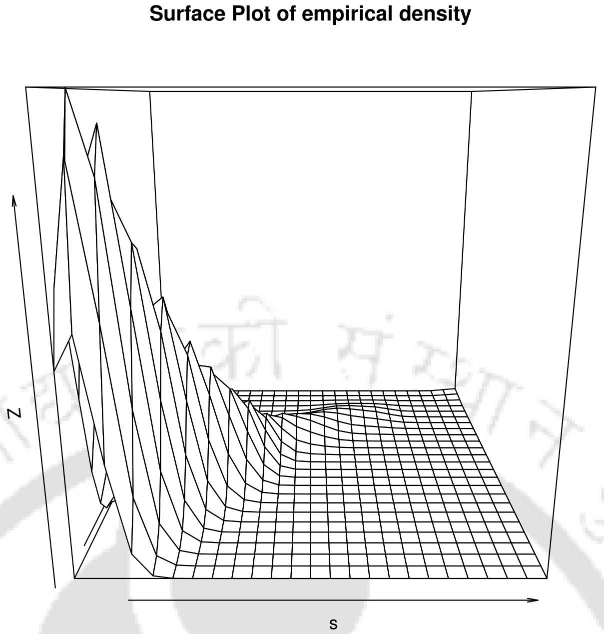


Figure 5.3: Two dimensional density plots for BB-BVPA of the transformed dataset

We observe the number of singular observations after transformation is zero. Therefore, this could be a reasonable data to work with. The sample size for the data set is 329. Note that even after location, scale transformation observed cardinalities of I_1 and I_2 are good representative of the actual ones. Bayesian estimates are calculated and provided in Table-5.5 and Table-5.6 using Approach 1 and Approach 2 respectively for the abalone data. We use the parametric bootstrap technique to generate different samples from the seven parameter BB-BVPA distribution using estimated parameters as original parameter and then find out the mean square error and the coverage probabilities for the constructed credible interval. Results are available in Table-5.8 and Table-5.7 using Approach 1 and Approach 2 respectively. Both the algorithm works quite well in estimating all parameters. However, we observe that the mean square error is on higher side for α_0 as compared to other parameters. It is also unable to capture the coverage probability of α_0 at desired level and provides a value significantly smaller than the desired 95% confidence level. However, we get the significant improvement in both MSEs and Coverage Probabilities if we increase the sample size. Therefore, it is safer to work with larger sample size to obtain a desired accuracy level, while applying these methods in real life data sets.

Slice-cum-Gibbs							
Gamma Prior							
n = 329							
Original Parameter Sets	μ_1	μ_2	σ_1	σ_2	α_0	α_1	α_2
Starting Value	3.1854	2.9623	0.7781	0.5386	0.4417	0.6138	0.4549
Bayes Estimates	6.6397	3.4208	4.0394	2.0505	3.1775	0.8194	0.8064

Table 5.5: The Bayes Estimates of the parameters of absolute continuous Marshall-Olkin bivariate Pareto distribution based on the abalone data set through Approach 1

Slice-cum-Gibbs							
Gamma Prior							
n = 329							
Original Parameter Sets	μ_1	μ_2	σ_1	σ_2	α_0	α_1	α_2
Starting Value	3.1854	2.9623	0.7781	0.5386	0.4417	0.6138	0.4549
Bayes Estimates	6.6394	3.4206	5.9161	3.0359	4.0606	1.2695	1.2749

Table 5.6: The Bayes Estimates of the parameters of absolute continuous Marshall-Olkin bivariate Pareto distribution based on the abalone data set through Approach 2

Slice-cum-Gibbs				
Gamma Prior				
n = 391				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	3.1832	2.9602	3.7781	1.5386
Average Bayes Estimates	6.6389	3.4201	7.1394	3.6475
Mean Square Error	1.7919×10^{-5}	3.8658×10^{-6}	3.2984	0.8284
Credible Intervals	[6.6293, 6.6429]	[3.4155, 3.4225]	[2.8745, 9.4807]	[1.5790, 5.0924]
Coverage Probability	0.985	0.97	1	1
n = 391				
Parameter Sets		α_0	α_1	α_2
Starting Value		2.4417	0.6137	0.4549
Average Bayes Estimates		4.9191	1.3711	1.2813
Mean Square Error		1.5919	0.0564	0.0389
Credible Intervals		[2.1636, 6.5914]	[0.1960, 2.3196]	[0.2220, 2.0860]
Coverage Probability		1	1	1
n = 1000				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	3.1832	2.9602	3.7781	1.5386
Average Bayes Estimates	6.6393	3.4206	6.5320	3.3487
Mean Square Error	3.1872×10^{-6}	7.6010×10^{-7}	1.1271	0.2915
Credible Intervals	[6.6376, 6.6429]	[3.4180, 3.4206]	[3.1986, 7.9750]	[1.6539, 4.1996]
Coverage Probability	0.94	0.95	1	0.995
n = 1000				
Parameter Sets		α_0	α_1	α_2
Starting Value		2.4417	0.6137	0.4549
Average Bayes Estimates		4.4982	1.3149	1.2895
Mean Square Error		0.5567	0.0587	0.0551
Credible Intervals		[2.6644, 5.3735]	[0.1893, 2.0423]	[0.1646, 1.8381]
Coverage Probability		0.995	1	1

Table 5.7: Results for Approach 2 : The Bayes Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\mu_1 = 6.6394$, $\mu_2 = 3.4206$, $\sigma_1 = 5.9161$, $\sigma_2 = 3.0359$, $\alpha_0 = 4.0606$, $\alpha_1 = 1.2695$, $\alpha_2 = 1.2749$.

Slice-cum-Gibbs				
Gamma Prior				
n = 391				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	3.1834	2.9604	2.7781	0.5386
Average Bayes Estimates	6.6395	3.4205	4.9587	2.2112
Mean Square Error	1.4547×10^{-5}	5.3310×10^{-6}	1.8326	0.2171
Credible Intervals	[6.6311, 6.6435]	[3.4172, 3.4229]	[1.8828, 4.5788]	[1.2090, 2.5855]
Coverage Probability	0.97	0.955	0.97	0.96
n = 391				
Parameter Sets		α_0	α_1	α_2
Starting Value		1.4417	0.6138	0.4549
Average Bayes Estimates		3.3195	1.1693	0.8664
Mean Square Error		0.5178	0.3975	0.0347
Credible Intervals		[2.3324, 3.9134]	[0.1719, 1.2620]	[0.1916, 1.3486]
Coverage Probability		0.865	0.99	1.00
n = 1000				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	3.1834	2.9604	2.7781	0.5386
Average Bayes Estimates	6.6397	3.4210	4.4543	2.1045
Mean Square Error	2.9711×10^{-06}	7.7316×10^{-07}	0.6800	0.0732
Credible Intervals	[6.6389, 6.6435]	[3.4186, 3.4209]	[2.3449, 4.3376]	[1.3212, 2.4482]
Coverage Probability	0.945	0.93	0.995	0.98
n = 1000				
Parameter Sets		α_0	α_1	α_2
Starting Value		1.4417	0.6138	0.4549
Average Bayes Estimates		3.2002	1.007	0.8554
Mean Square Error		0.1486	0.1599	0.047
Credible Intervals		[2.2964, 3.3024]	[0.2817, 1.1934]	[0.3384, 1.2530]
Coverage Probability		0.915	0.975	0.99

Table 5.8: Results for Approach 1 : The Bayes Estimates (BE), Mean Square Error (MSE) and credible intervals of absolute continuous Marshall-Olkin bivariate Pareto distribution with parameters $\mu_1 = 6.6397$, $\mu_2 = 3.4208$, $\sigma_1 = 4.0394$, $\sigma_2 = 2.0505$, $\alpha_0 = 3.1775$, $\alpha_1 = 0.8194$, $\alpha_2 = 0.8064$.

5.7 Conclusion

Bayes estimates of the parameters of the absolute continuous bivariate Pareto under square error loss are obtained using the slice cum Gibbs sampler approach. Two approaches proposed in this works quite well, even for moderately large sample size. However, coverage probabilities calculated based on the parametric bootstrap samples show that the proposed credible interval construction does not work very well for all the shape parameters. Although, we see its significant improvement for large sample, more research is needed to explore the best credible interval construction in small sample set up. The same study can be made using many other algorithms like importance sampling, HMC etc. This estimation procedure can be used in the Bayesian discrimination between two bivariate distributions using the Bayes factor. The Bayesian estimation of corresponding singular distribution in presence of location and scale parameters also a challenging work. We address this problem in Chapter-6.

Chapter 6

Bayesian analysis of singular Marshall-Olkin Bivariate Pareto Distribution with location and scale parameter

6.1 Introduction

In this chapter, our main aim is to compute the Bayesian estimates and the associated credible intervals of the unknown parameters of singular Marshall-Olkin Bivariate Pareto distribution (MOBVPA) with location and scale parameter under informative priors. We use the truncated normal distribution for the location parameters and the gamma distribution as prior for both scale and shape parameters. We assume all priors are mutually independent. Since, Bayes estimate can not be obtained in closed form, we propose to use the slice [38] cum Gibbs sampling [49] technique to generate samples from the posterior distribution of the parameters. However, the likelihood function of MOBVPA distribution is discontinuous in the case of seven parameters setup, which restrict direct slice sampling to generate samples from the posterior distribution. Therefore, we propose to use a trick, which allows direct slice sampling to use in this case. The generated samples are used to calculate the Bayes estimates and the credible intervals. Since convergence of the proposed algorithm depends on the choice of hyper-parameters and data, we also try to set the hyper-parameters through a modified empirical Bayes estimate. Therefore, the major focus of this work is not only the slice sampling implementation, but also the discussion of empirical Bayes and highlight few important related results. There are a few implementation related the issues, which make the work significantly different from the previous chapter.

The statistical literature contains many works related to the Bayesian estimation of bivariate exponential ([44], [27], [32]) or bivariate Weibull distribution ([23], [24]). However, the MOBVPA distribution model has not received much attention primarily due to the analytical intractability of the model in the presence of both location and scale parameters. Recently, Asimit et al. [3] used the EM algorithm to estimate the singular Marshall-Olkin bivariate Pareto distribution. Paul and Dey [40] also discussed the different variations of the algorithms and provided the best EM algorithms in this set up. However, the Bayesian analysis contains more information than the maximum likelihood estimation/EM estimates, which is better for further statistical analysis. The maximum likelihood estimators can be quite unstable, i.e., vary widely for small variations of the observations, at least for reduced sample sizes. Moreover, derivation of the confidence regions in both univariate and multivariate cases become easy with the knowledge of full expression of the posterior distribution.

The rest of the chapter is arranged in the following way. In Section-6.2, we briefly keep the Marshall-Olkin bivariate Pareto distribution formulation. Section-6.3 and Section-6.4 are kept in the Bayesian analysis of three and seven parameters Marshall-Olkin bivariate Pareto distribution respectively. We discuss about the construction of the credible interval in Section-6.5 and the empirical Bayesian inference in Section-6.6. The numerical analysis and the data analysis results are presented in Section-6.7 and Section-6.8 respectively. Finally, we conclude this chapter in Section-6.9.

6.2 Formulation of Marshall-Olkin bivariate Pareto distribution

The Marshall-Olkin bivariate Pareto distribution with location and scale parameters as described by Asimit et al. ([3], [2]) can be obtained in the following manner.

Let U_0, U_1 and U_2 are three independent univariate Pareto distributions $Pa(II)(0,1,\alpha_0)$, $Pa(II)(\mu_1,\sigma_1,\alpha_1)$ and $Pa(II)(\mu_2,\sigma_2,\alpha_2)$ respectively.

We define $X_1 = \min\{\sigma_1 U_0 + \mu_1, U_1\}$ and $X_2 = \min\{\sigma_2 U_0 + \mu_2, U_2\}$. We can show that (X_1, X_2) jointly follow the bivariate Pareto distribution of second kind with joint probability density function :

$$f(x_1,x_2) = \begin{cases} f_1(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x \end{cases}, \tag{6.2.1}$$

where

$$\begin{aligned} f_1(x_1,x_2) &= \frac{\alpha_1(\alpha_0+\alpha_2)}{\sigma_1\sigma_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2-1}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1-1} \\ f_2(x_1,x_2) &= \frac{\alpha_2(\alpha_0+\alpha_1)}{\sigma_1\sigma_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2-1}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1-1} \\ f_0(x) &= \alpha_0(1+x)^{-\alpha_0-\alpha_1-\alpha_2-1}. \end{aligned}$$

We denote this distribution as $MOBVPA(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$. We will call it seven parameters MOBVPA distribution. One can easily show that the marginal distribution of X_i is $PA(II)(\mu_i, \sigma_i, \alpha_0 + \alpha_i)$ where $i = 1, 2$.

If we choose $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$, then the joint PDF of (X_1, X_2) is

$$f(x_1, x_2) = \begin{cases} \alpha_1(\alpha_0 + \alpha_2)(1 + x_1)^{-\alpha_1-1}(1 + x_2)^{-\alpha_0-\alpha_2-1}, & \text{if } x_1 < x_2 \\ \alpha_2(\alpha_0 + \alpha_1)(1 + x_2)^{-\alpha_2-1}(1 + x_1)^{-\alpha_0-\alpha_1-1}, & \text{if } x_1 > x_2 \\ \alpha_0(1 + x)^{-\alpha_0-\alpha_1-\alpha_2-1}, & \text{if } x_1 = x_2 = \text{x} \end{cases} \tag{6.2.2}$$

We denote this distribution as $MOBVPA(\alpha_0, \alpha_1, \alpha_2)$. We will call it three parameter MOBVPA distribution.

6.3 Bayesian analysis of three parameters Marshall-Olkin bivariate Pareto distribution

Let us consider I $= \{(x_{11}, x_{21}), (x_{12}, x_{22}), \cdots, (x_{1n}, x_{2n})\}$ is a random sample of size n from three parameters MOBVPA distribution. We use the following notations;

$$I_0 = \{i \mid x_{1i} = x_{2i}\},\ I_1 = \{i \mid x_{1i} < x_{2i}\} \text{ and } I_2 = \{i \mid x_{1i} > x_{2i}\}$$

and

$$|I_0| = n_0,\ |I_1| = n_1,\ |I_2| = n_2, \text{ and } n = n_0 + n_1 + n_2.$$

The log-likelihood function based on the data I takes the form,

$$\begin{aligned} L(\alpha_0, \alpha_1, \alpha_2) = {} & n_1 \ln \alpha_1 + n_1 \ln(\alpha_0 + \alpha_2) - (\alpha_1 + \alpha_2 + 1) \sum_{i \in I_1}^{n} \ln(1 + x_{2i}) \\ & - (\alpha_1 + 1) \sum_{i \in I_1} \ln(1 + x_{1i}) + n_2 \ln \alpha_2 + n_2 \ln(\alpha_0 + \alpha_1) \\ & - (\alpha_0 + \alpha_1 + 1) \sum_{i \in I_2}^{n} \ln(1 + x_{1i}) - (\alpha_2 + 1) \sum_{i \in I_2}^{n} \ln(1 + x_{2i}) \\ & + n_0 \ln \alpha_0 - (\alpha_0 + \alpha_1 + \alpha_2 + 1) \sum_{i \in I_0}^{n} \ln(1 + x_{1i}). \end{aligned} \tag{6.3.1}$$

6.3.1 Prior Assumption

We assume that α_0, α_1, and α_2 are distributed according to the gamma distribution with shape parameters k_i and scale parameters θ_i, i.e.,

$$\alpha_0 \sim \Gamma(k_0, \theta_0) \equiv \text{Gamma}(k_0, \theta_0),$$

$$
\begin{aligned}
\alpha_1 &\sim \Gamma(k_1, \theta_1) \equiv \text{Gamma}(k_1, \theta_1), \\
\alpha_2 &\sim \Gamma(k_2, \theta_2) \equiv \text{Gamma}(k_2, \theta_2).
\end{aligned} \tag{6.3.2}
$$

The probability density function of the Gamma distribution is given by,

$$
f_\Gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}},
$$

where $\Gamma(k)$ is the gamma function evaluated at k. We can use some other prior like Gamma-Dirichlet Prior (for details see [44]), reference prior etc.

6.3.2 Bayes Estimates

In this section, we provide the Bayes estimates of the unknown parameters, namely α_0, α_1, and α_2 for the three parameters MOBVPA distribution using slice cum Gibbs Sampler method. In this work, we use step-out slice sampling as described by Neal [38]. We can provide the expression of full posterior when the posterior is constructed based on Gamma prior. The full posterior of $(\alpha_0, \alpha_1, \alpha_2)$ given the data I based on the gamma prior $\pi(\cdot)$ is,

$$
\begin{aligned}
\pi(\alpha_0, \alpha_1, \alpha_2 | I) \propto \quad & l(\alpha_0, \alpha_1, \alpha_2 | I) \pi(\alpha_0, \alpha_1, \alpha_2) \\
= & \alpha_0^{n_0} \alpha_1^{n_1} \alpha_2^{n_2} (\alpha_0 + \alpha_1)^{n_2} (\alpha_0 + \alpha_2)^{n_1} \\
& \prod_{i \in I_0} (1 + x_{1i})^{-(\alpha_0 + \alpha_1 + \alpha_2 - 1)} \prod_{i \in I_1} (1 + x_{1i})^{-\alpha_1 - 1} \\
& (1 + x_{2i})^{-\alpha_0 - \alpha_2 - 1} \prod_{i \in I_2} (1 + x_{1i})^{-\alpha_0 - \alpha_1 - 1} (1 + x_{2i})^{-\alpha_2 - 1} \\
& \times \alpha_0^{k_0 - 1} \alpha_1^{k_1 - 1} \alpha_2^{k_2 - 1} e^{-(\frac{\alpha_0}{\theta_0} + \frac{\alpha_1}{\theta_1} + \frac{\alpha_2}{\theta_2})} \\
= & \pi_1(\alpha_0, \alpha_1, \alpha_2 | I) \quad \text{(say)}.
\end{aligned} \tag{6.3.3}
$$

Therefore, if we want to compute the Bayes estimate of some function of α_0, α_1 and α_2, say $g(\alpha_0, \alpha_1, \alpha_2)$, the Bayes estimate of g, say $\hat{g}$ under the squared error loss function is the posterior mean of g, i.e.

$$
\hat{g} = \frac{\int_0^\infty \int_0^\infty \int_0^\infty g(\alpha_0, \alpha_1, \alpha_2) \pi_1(\alpha_0, \alpha_1, \alpha_2 | I) d\alpha_0 d\alpha_1 d\alpha_2}{\int_0^\infty \int_0^\infty \int_0^\infty \pi_1(\alpha_0, \alpha_1, \alpha_2 | I) d\alpha_0 d\alpha_1 d\alpha_2}. \tag{6.3.4}
$$

Since, the Bayes estimate of the unknown model parameters α_0, α_1 and α_2 can not be obtained in closed form using the integral (6.3.4), we use the slice cum Gibbs sampling technique to generates the sample from the corresponding posterior distribution and in turn use them to compute Bayes estimates. To obtain the Bayesian estimates through the slice cum Gibbs sampler, we need to use conditional distribution of each the parameter given the other parameters and the data. We provide the expressions of logarithm of those conditional distributions for gamma prior. We propose to perform the standard step-out method in slice sampling [38] on each of the conditional distributions based on gamma prior to generate the respective parameters. The algorithmic steps for the slice sampling via step out method can be provided as follows,

- Sample z and u uniformly from the area under the distribution, say $p(\cdot)$.
  1. Fix z, sample u uniformly from $[0, p(z)]$.
  2. Fix u, sample z uniformly from the slice through the region $\{z : p(z) > u\}$
- How to sample z from the slice.
  1. Start with the region of width w containing $z^{(t)}$.
  2. If end point is in slice, then extend region by w in that direction.
  3. Sample $z^{'}$ uniformly from the region.
  4. If $z^{'}$ is in the slice, the accept it as $z^{(t+1)}$.
  5. If not : make $z^{'}$ new end point of the region, and resample $z^{'}$.

Expression for the logarithm of conditional posterior distribution of α_0, α_1, and α_2 are,

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) + n_1\ln(\alpha_0+\alpha_2) \\ &- \alpha_0\sum_{i\in I_1}\ln(1+x_{2i}) + n_2\ln(\alpha_0+\alpha_1) \\ &- \alpha_0\sum_{i\in I_2}\ln(1+x_{1i}) + (k_0-1)\ln\alpha_0 - \frac{\alpha_0}{\theta_0}, \end{aligned} \tag{6.3.5}$$

$$\begin{aligned}\ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1\ln\alpha_1 - \alpha_1\sum_{i\in I_1}\ln(1+x_{1i}) + n_2\ln(\alpha_0+\alpha_1) \\ &- \alpha_1\sum_{i\in I_2}\ln(1+x_{1i}) + (k_1-1)\ln\alpha_1 - \frac{\alpha_1}{\theta_1} \end{aligned} \tag{6.3.6}$$

and

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \alpha_0, \alpha_1, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1\ln(\alpha_0+\alpha_2) - \alpha_2\sum_{i\in I_1}\ln(1+x_{2i}) + n_2\ln\alpha_2 \\ &- \alpha_2\sum_{i\in I_2}\ln(1+x_{2i}) + (k_2-1)\ln\alpha_2 - \frac{\alpha_2}{\theta_2}. \end{aligned} \tag{6.3.7}$$

6.4 Bayesian analysis of seven parameters Marshall-Olkin bivariate Pareto distribution

Let us assume that μ_1, μ_2, σ_1 and σ_2 are known. Now we divide our data set $I = \{(x_{11}, x_{21}), (x_{12}, x_{22}), \cdots, (x_{1n}, x_{2n})\}$ into three part as follows,

$$I_0 = \{i : \tfrac{x_{1i}-\mu_1}{\sigma_1} = \tfrac{x_{2i}-\mu_2}{\sigma_2}\},\; I_1 = \{i : \tfrac{x_{1i}-\mu_1}{\sigma_1} < \tfrac{x_{2i}-\mu_2}{\sigma_2}\},\; I_2 = \{i : \tfrac{x_{1i}-\mu_1}{\sigma_1} > \tfrac{x_{2i}-\mu_2}{\sigma_2}\}$$

and

$$|I_0| = n_0,\; |I_1| = n_1,\; |I_2| = n_2, \text{ and } n = n_0 + n_1 + n_2.$$

Usual log-Likelihood function of this MOBVPA can be written as

$$\begin{aligned}
L(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2) &= n_1 \ln \alpha_1 + n_1 \ln(\alpha_0 + \alpha_2) - n_1 \ln \sigma_1 - n_1 \ln \sigma_2 \\
&\quad -(\alpha_0 + \alpha_2 + 1) \sum_{i \in I_1} \ln\left(1 + \frac{x_{2i} - \mu_2}{\sigma_2}\right) - (\alpha_1 + 1) \sum_{i \in I_1} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) \\
&\quad -n_2 \ln \sigma_1 - n_2 \ln \sigma_2 + n_2 \ln \alpha_2 + n_2 \ln(\alpha_0 + \alpha_1) \\
&\quad -(\alpha_0 + \alpha_1 + 1) \sum_{i \in I_2} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right) - (\alpha_2 + 1) \sum_{i \in I_2} \ln\left(1 + \frac{x_{2i} - \mu_2}{\sigma_2}\right) \\
&\quad +n_0 \ln \alpha_0 - n_0 \ln \sigma_1 - (\alpha_0 + \alpha_1 + \alpha_2 + 1) \sum_{i \in I_0} \ln\left(1 + \frac{x_{1i} - \mu_1}{\sigma_1}\right). \qquad (6.4.1)
\end{aligned}$$

6.4.1 Prior Assumption

We assume that the shape parameters α_0, α_1 and α_2 are distributed according to the gamma distribution with shape parameters k_i and scale parameters θ_i, where $i = 0, 1, 2$, i.e.,

$$\begin{aligned}
\pi_0(\alpha_0) &= \text{Gamma}(k_0, \theta_0) = \frac{1}{\Gamma(k_0)\theta_0^{k_0}} \alpha_0^{k_0 - 1} e^{-\frac{\alpha_0}{\theta_0}}, \quad \alpha_0 > 0 \\
\pi_1(\alpha_1) &= \text{Gamma}(k_1, \theta_1) = \frac{1}{\Gamma(k_1)\theta_1^{k_1}} \alpha_1^{k_1 - 1} e^{-\frac{\alpha_1}{\theta_1}}, \quad \alpha_1 > 0 \\
\pi_2(\alpha_2) &= \text{Gamma}(k_2, \theta_2) = \frac{1}{\Gamma(k_2)\theta_2^{k_2}} \alpha_2^{k_2 - 1} e^{-\frac{\alpha_2}{\theta_2}}, \quad \alpha_2 > 0
\end{aligned}$$

where $k_i > 0$ and $\theta_i > 0$, $i = 0, 1, 2$. Here $\Gamma(k)$ is the gamma function evaluated at k. We also assume that the location parameters μ_1 and μ_2 are distributed according to the truncated normal distribution with mean $= \mu_i^{'}$ and variance $= \sigma_i^{'}$, where $i = 1, 2$,

$$\begin{aligned}
\pi_3(\mu_1) &= \frac{\phi(\mu_1; \mu_1^{'}, \sigma_1^{'})}{\Phi(\min\{x_{1i} : i = 1(1)n\}; \mu_1^{'}, \sigma_1^{'})}, \\
\pi_4(\mu_2) &= \frac{\phi(\mu_2; \mu_2^{'}, \sigma_2^{'})}{\Phi(\min\{x_{2i} : i = 1(1)n\}; \mu_2^{'}, \sigma_2^{'})}.
\end{aligned}$$

Further, we assume that the scale parameters σ_1 and σ_2 are distributed according to the gamma distribution with the hyper-parameters c_1, c_2, d_1 and d_2,

$$\pi_5(\sigma_1) = \text{Gamma}(c_1, d_1) = \frac{1}{\Gamma(c_1) d_1^{c_1}} \sigma_1^{c_1 - 1} e^{-\frac{\sigma_1}{d_1}}, \quad \sigma_1 > 0$$
$$\pi_6(\sigma_2) = \text{Gamma}(c_2, d_2) = \frac{1}{\Gamma(c_2) d_2^{c_2}} \sigma_2^{c_2 - 1} e^{-\frac{\sigma_2}{d_2}}, \quad \sigma_2 > 0$$

where $c_i > 0$ and $d_i > 0$, $i = 1, 2$.

6.4.2 Bayes Estimate

We know the Bayes estimate of an unknown parameter is the posterior mean of the corresponding parameter under the squared error loss function. But in this case it is not possible to calculate a Bayes estimate of unknown model parameters μ_1, μ_2, σ_1, σ_2, α_0, α_1 and α_2 in closed form. We propose to use the slice cum Gibbs sampling procedure to generate samples from the conditional posterior distribution. The full logarithm of the conditional posterior distributions of α_0, α_1, α_2 are given by,

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, x_1, x_2)) = {} & L(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_0)) - k_0 \ln(\theta_0) \\ & + (k_0 - 1) \ln \alpha_0 - \frac{\alpha_0}{\theta_0}\end{aligned}$$

$$\begin{aligned}\ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, x_1, x_2)) = {} & L(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_1)) - k_1 \ln(\theta_1) \\ & + (k_1 - 1) \ln \alpha_1 - \frac{\alpha_1}{\theta_1}\end{aligned}$$

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \alpha_0, \alpha_1, \mu_1, \mu_2, \sigma_1, \sigma_2, x_1, x_2)) = {} & L(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_2)) - k_2 \ln(\theta_2) \\ & + (k_2 - 1) \ln \alpha_2 - \frac{\alpha_2}{\theta_2}\end{aligned}$$

We already know that the log-likelihood function is discontinuous with respect to location and scale parameters. So, we use **marginal log-likelihood function** to formulate log-posterior. Hence the full logarithm of the conditional posterior distributions of μ_1, μ_2, σ_1, and σ_2 are given by,

$$\begin{aligned}\ln(\pi(\mu_1 \mid \alpha_0, \alpha_1, \alpha_2, \mu_2, \sigma_1, \sigma_2, x_1, x_2)) = {} & n \ln(\alpha_0 + \alpha_1) - n \ln(\sigma_1) - (\alpha_0 + \alpha_1 + 1) \\ & \sum_{i=1}^{n} \ln(1 + (x_{1i} - \mu_1)/\sigma_1) - \ln\Big(\Phi(\min\{x_{1i}\}; \mu_1^{'}, \sigma_1^{'}) \sqrt{2\pi} \sigma_1^{'}\Big) \\ & - 0.5 \frac{(m_1 - \mu_1^{'})^2}{(\sigma_1^{'})^2},\end{aligned}$$

$$\ln(\pi(\mu_2 \mid \alpha_0, \alpha_1, \alpha_2, \mu_1, \sigma_1, \sigma_2, x_1, x_2)) = n\ln(\alpha_0+\alpha_2) - n\ln(\sigma_2) - (\alpha_0+\alpha_2+1)$$
$$\sum_{i=1}^{n}\ln(1+(x_{2i}-\mu_2)/\sigma_2) - \ln\Big(\Phi(\min\{x_{2i}\};\mu_2^{'},\sigma_2^{'})\sqrt{2\pi}\sigma_2^{'}\Big)$$
$$-\,0.5\frac{(m_2-\mu_2^{'})^2}{(\sigma_2^{'})^2},$$

$$\ln(\pi(\sigma_1 \mid \alpha_0, \alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_2, x_1, x_2)) = n\ln(\alpha_0+\alpha_1) - n\ln(\sigma_1) - (\alpha_0+\alpha_1+1)$$
$$\sum_{i=1}^{n}(1+(x_{1i}-\mu_1)/\sigma_1) - \ln(\Gamma(c_1)) - c_1\ln(d_1)$$
$$+\,(c_1-1)\ln\sigma_1 - \frac{\sigma_1}{d_1}$$

and

$$\ln(\pi(\sigma_2 \mid \alpha_0, \alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, x_1, x_2)) = n\ln(\alpha_0+\alpha_2) - n\ln(\sigma_2) - (\alpha_0+\alpha_2+1)$$
$$\sum_{i=1}^{n}(1+(x_{2i}-\mu_2)/\sigma_2) - \ln(\Gamma(c_2)) - c_2\ln(d_2)$$
$$+\,(c_2-1)\ln\sigma_2 - \frac{\sigma_2}{d_2}.$$

6.5 Constructing Credible Intervals

We have found the credible intervals for parameters as described by Chen and Shao [11] in 1990. Let assume $\underline{\theta}$=$\{\theta_1, \theta_2, \cdots, \theta_p\}$ is a vector. To obtain the credible intervals of the first variable θ_1, we order $\{\theta_{1i}\}$, as $\theta_{1(1)} < \theta_{1(2)} < \cdots < \theta_{1(M)}$. Therefore 100(1 - γ)% credible interval for θ_1 becomes $(\theta_{1(j^*)}, \theta_{1(j^*+\lfloor M(1-\gamma)\rfloor)})$, where j^* is such that

$$\theta_{1(j^*+\lfloor M(1-\gamma)\rfloor)} - \theta_{1(j^*)} \leq \theta_{1(j+\lfloor M(1-\gamma)\rfloor)} - \theta_{1(j)}$$

for all $j = 1, 2, \cdots, M - \lfloor M(1-\gamma)\rfloor$, where $\lfloor x\rfloor$ is the greatest integer not greater than x. Similarly, we can obtain the credible interval for the other co-ordinates of θ. In this context, the parameter vector are $\theta = (\alpha_0, \alpha_1, \alpha_2)$ and $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$ for three and seven parameters cases respectively.

We have scope to construct such intervals when the full posterior is not known and intractable. We use an R package 'coda'[45] to obtain the credible intervals described above. Note that we construct the credible interval from empirical cdf of the sample of the parameters. We assume that the posterior distribution of the parameters are unimodal. In this case the credible interval and the highest posterior density (HPD) interval are same, as we calculate the interval using samples of the parameter.

6.6 Empirical Bayesian Inference

In empirical Bayesian inference [5], we sample from the posterior distribution of the parameter with a hyper-parameter set to its maximum likelihood estimate. Let us consider $\underline{x}$ as a sample generated from a population with pdf $f_{\theta,\lambda}(x)$, given θ and λ, where $(\theta, \lambda) \in \Theta \times \Lambda$. We treat λ as hyper-parameter and we denote $\pi(\theta|\lambda)$ as a prior distribution for θ. So the posterior distribution of θ given $\underline{x}$ and λ is,

$$\pi(\theta|\underline{x}, \lambda) = \frac{f_{\theta,\lambda}(\underline{x})\pi(\theta|\lambda)}{\pi(\underline{x}|\lambda)}, \tag{6.6.1}$$

where $\pi(\underline{x}|\lambda) = \int_\theta f_{\theta,\lambda}(\underline{x})\pi(\theta|\lambda)d\theta$ is a marginal distribution of $\underline{x}$ given λ.

The idea is to propose an estimate for λ based on the sample or data $\underline{x}$. This estimate is just a maximum likelihood estimate $\hat{\lambda}$ of λ given the data $\underline{x}$,

$$\hat{\lambda} = \text{Argmax} \quad \pi(\lambda|\underline{x}). \tag{6.6.2}$$

Then we generate θ from $\pi(\theta|\underline{x}, \hat{\lambda})$.

Note that in our case, likelihood function does not depend on hyper-parameter and we consider that $H(\theta, \lambda) = \nabla_\lambda \pi(\theta|\lambda)$. Now the empirical Bayesian algorithm based on the stochastic approximation (SA) algorithm and RMSprop [50] is given by,

Algorithm 6.1 Empirical Bayesian Algorithm

Initially, we start with $\theta_0 \in \Theta$, $\lambda_0 \in \Lambda$ and $H(\theta, \lambda)_0$, given (θ_t, λ_t),

1: Generate θ_{t+1} from posterior distribution of the parameters using Slice cum Gibbs sampling procedure with λ_t.
2: Calculate weighted average of the square of gradient $E[H^2(\theta, \lambda)]_t$ at time step t which depends on the previous average and the square of the current gradient :
$$E[H^2(\theta, \lambda)]_t = 0.9E[H^2(\theta, \lambda)]_{t-1} + 0.1H^2(\theta_{t+1}, \lambda_t).$$
and
3: $\lambda_{t+1} = \lambda_t + \frac{0.001}{\sqrt{E[H^2(\theta,\lambda)]_t + 10^{-8}}} H(\theta_{t+1}, \lambda_t)$

6.7 Numerical Analysis

The numerical results are obtained by using the programming software R 3.2.3. The codes are run at IIT Guwahati computers with model : Intel(R) Core(TM) i5-6200U CPU 2.30GHz. The codes will be available to the authors on request.

In the slice cum Gibbs sampling technique, we take burn in period as 500 and width of the slice as $w = 1$. It is observed that this technique works even for different width of slice namely, 0.5, 2, 5 etc. The Bayes estimates are calculated based on 2000 samples after the burn in period. We conduct the experiment based on two different sample sizes namely, $n = 450, 1000$. Average Bayes estimate (ABE), Mean squared error (MSE), Credible interval (CI), and Coverage Probability (CP) of the credible interval are calculated based on 200 different samples drawn from the original set of parameters. Note that all of the numerical results are based on the particular initial choice of the hyper-parameters. However, we observe that it also works for some other initial choice of hyper-parameters. We use an uniform random number from the interval (0, 1) as our initial choice for scale and shape parameters. For location parameters we also use a positive uniform random number as our initial choice which satisfies the condition $\mu_i < x_i$, $i = 1, 2$. It is observed that some other initial choice also worked.

Three Parameter MOBVPA: For the empirical Bayes inference, we use the following initial choice of hyper parameters : $k_0 = 2,\ \theta_0 = 3,\ k_1 = 4,\ \theta_1 = 3,\ k_2 = 3$ and $\theta_2 = 2$. We observe that it works for some other moderately small initial choice of hyper-parameters. However, the results can be obtained without implementing the empirical Bayes algorithm in case of three parameter MOBVPA distribution.

Here we take two different sets of parameters for the inference and observed the ABE, MSE, CI, and CP of credible interval, which are shown in Table-6.1, Table-6.2, Table-6.3 and Table-6.4 respectively. We observe that the slice-cum Gibbs sampler works for a sample size greater than 200 for three parameter MOBVPA distribution. Required sample size becomes more when original sample is drawn from higher values of the parameters.

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Average Bayesian Estimates	0.1014	0.2037	0.4037
Mean Squared Error	0.0001	0.0002	0.0005
Credible Interval	[0.0589 0.0938]	[0.1718 0.2176]	[0.3166 0.3886]
Coverage Probabilities	0.925	0.97	0.945

Table 6.1: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of three parameter MOBVPA with parameters $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 450$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Average Bayesian Estimates	0.1003	0.2015	0.4012
Mean Squared Error	$5.32e^{-05}$	$8.94e^{-05}$	0.0002
Credible Interval	[0.0674 0.0914]	[0.1761 0.2090]	[0.3504 0.4008]
Coverage Probabilities	0.95	0.94	0.93

Table 6.2: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of three parameter MOBVPA with parameters $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 1000$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Average Bayesian Estimates	4.0001	5.0589	9.9573
Mean Squared Error	0.1012	0.1565	0.2763
Credible Interval	[2.7511 3.9288]	[3.8364 5.1071]	[7.4483 9.4306]
Coverage Probabilities	0.96	0.925	0.975

Table 6.3: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of three parameter MOBVPA with parameters $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 450$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Average Bayesian Estimates	4.0054	5.0589	9.9536
Mean Squared Error	0.0515	0.0656	0.1531
Credible Interval	[3.1042 3.9184]	[4.3447 5.2663]	[8.1594 9.4974]
Coverage Probabilities	0.95	0.97	0.95

Table 6.4: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of three parameter MOBVPA with parameters $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 1000$

Seven Parameter MOBVPA: The empirical Bayes implementation is necessary for the seven parameter MOBVPA distribution, as one particular choice of hyper-parameter may often fail to provide the Bayes estimates for the different samples. However, the empirical Bayes may work for any generated samples for some specific set of choices. One such choice is : $\mu_1^{'} = 0.15,\ \sigma_1^{'} = 2,\ \mu_2^{'} = 0.20,\ \sigma_2^{'} = 2,\ c_1 = 3,\ d_1 = 2,\ c_2 = 3,\ d_2 = 2,\ k_0 = 0.10,\ \theta_0 = 0.25,\ k_1 = 0.10,\ \theta_1 = 0.25,\ k_2 = 0.10,\ \theta_2 = 0.25$. One of the pitfall of the procedure is such choices are made by trial and error method. However, all our computation and simulation works very well with this choice. There are multiple such choices are available.

We take two different set of parameters and calculate ABE, MSE, CI, and CP of credible interval, which are available in Table-6.5, Table-6.6, Table-6.7, Table-6.8, Table-6.9, Table-6.10, Table-6.11 and Table-6.12 respectively. It is observed that the cardinality of I_0, I_1 and I_2 do not reflect the actual value of n_0, n_1 and n_2 respectively, while calculating the value of the likelihood function due to location and scale transformation. We replace n_0, n_1 and n_2, the cardinality of cells I_0, I_1 and I_2 by $\tilde{n_0}$, $\tilde{n_1}$ and $\tilde{n_2}$ where $\tilde{n_i} = (n_0 + n_1 + n_2)\frac{\alpha_i}{(\alpha_0+\alpha_1+\alpha_2)}$ for $i = 0, 1, 2$ is an approximation made by taking the expectation of the multinomial distribution of cell cardinalities. This approximation can be obtained by using the distribution of unknown random cardinalities as the multinomial distribution with the parameters $(n_0 + n_1 + n_2)$ and $\frac{\alpha_i}{(\alpha_0+\alpha_1+\alpha_2)}$ for $i = 0, 1, 2$.

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	0.1438	0.3462	0.7781	0.5386
Average Bayesian Estimates	0.3001	0.4000	0.4986	0.5733
Mean Squared Error	$2.73e^{-07}$	$2.19e^{-07}$	0.0163	0.0242
Credible Interval	[0.2990 0.3001]	[0.3995 0.4006]	[0.2193 0.5072]	[0.2389 0.5911]
Coverage Probabilities	0.925	0.94	0.835	0.82

Table 6.5: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 0.30$, $\mu_2 = 0.40$, $\sigma_1 = 0.60$, $\sigma_2 = 0.70$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$ and sample size $n = 450$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4417	0.6138	0.45491
Average Bayesian Estimates	1.4156	1.0516	1.1853
Mean Squared Error	0.1076	0.0635	0.0924
Credible Interval	[0.8048 1.5705]	[0.3806 1.1071]	[0.4522 1.2110]
Coverage Probabilities	0.835	0.92	0.895

Table 6.6: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 0.30$, $\mu_2 = 0.40$, $\sigma_1 = 0.60$, $\sigma_2 = 0.70$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$ and sample size $n = 450$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	0.1438	0.3462	0.7781	0.5386
Average Bayesian Estimates	0.3000	0.4000	0.5373	0.6070
Mean Squared Error	$3.65e^{-08}$	$5.49e^{-08}$	0.0084	0.0150
Credible Interval	[0.2997 0.3002]	[0.3996 0.4001]	[0.3429 0.6113]	[0.3784 0.6733]
Coverage Probabilities	0.96	0.955	0.87	0.85

Table 6.7: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 0.30$, $\mu_2 = 0.40$, $\sigma_1 = 0.60$, $\sigma_2 = 0.70$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$ and sample size $n = 1000$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4417	0.6138	0.45491
Average Bayesian Estimates	1.4662	1.1806	1.2818
Mean Squared Error	0.0700	0.0427	0.0568
Credible Interval	[0.9673 1.6838]	[0.5975 1.2945]	[0.6437 1.3543]
Coverage Probabilities	0.89	0.945	0.95

Table 6.8: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 0.30$, $\mu_2 = 0.40$, $\sigma_1 = 0.60$, $\sigma_2 = 0.70$, $\alpha_0 = 1.7$, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$ and sample size $n = 1000$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	0.4794	1.7308	0.7781	0.5386
Average Bayesian Estimates	0.9999	2.0000	0.3986	0.4960
Mean Squared Error	$2.80e^{-07}$	$5.80e^{-07}$	0.0047	0.0063
Credible Interval	[0.9991 1.0006]	[1.9982 2.0001]	[0.2066 0.3824]	[0.2375 0.4728]
Coverage Probabilities	0.96	0.955	0.92	0.935

Table 6.9: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 0.40$, $\sigma_2 = 0.50$, $\alpha_0 = 0.80$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$ and sample size $n = 450$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4417	0.6138	0.45491
Average Bayesian Estimates	0.7466	0.7206	0.6718
Mean Squared Error	0.0088	0.0176	0.0157
Credible Interval	[0.5518 0.8145]	[0.3588 0.7285]	[0.2749 0.6917]
Coverage Probabilities	0.885	0.98	0.96

Table 6.10: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 0.40$, $\sigma_2 = 0.50$, $\alpha_0 = 0.80$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$ and sample size $n = 450$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	0.4794	1.7308	0.7781	0.5386
Average Bayesian Estimates	1.0000	2.0000	0.4018	0.5023
Mean Squared Error	$6049e^{-08}$	$1.01e^{-07}$	0.0020	0. 0038
Credible Interval	[0.9999 1.0006]	[1.9994 2.0002]	[0.2350 0.4492]	[0.3096 0.5370]
Coverage Probabilities	0.95	0.97	0.925	0.95

Table 6.11: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 0.40$, $\sigma_2 = 0.50$, $\alpha_0 = 0.80$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$ and sample size $n = 1000$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4417	0.6138	0.45491
Average Bayesian Estimates	0.7595	0.7329	0.6861
Mean Squared Error	0.0051	0.0095	0.0102
Credible Interval	[0.5965 0.8339]	[0.3662 0.7390]	[0.3419 0.6814]
Coverage Probabilities	0.92	0.965	0.955

Table 6.12: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA with parameters $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 0.40$, $\sigma_2 = 0.50$, $\alpha_0 = 0.80$, $\alpha_1 = 0.70$, $\alpha_2 = 0.65$ and sample size $n = 1000$

6.8 Data Analysis

We analyze one real life data set from the R package 'evd'[16]. This data set contains the indemnity payments (Loss) and allocated loss adjustment expense (ALAE) relating to 1500 general liability claims from insurance companies. Paul anad Dey [40] also used this data set to model MOBVPA distribution through the EM algorithm. From [15], we know that peak over threshold method on random variable U (say) provides generalized Pareto distribution (GPD) for any x_0 with $1 + log(G(x_0)) \in (0, 1)$ i.e. $P(U > tx_0|U > x_0) = t^{-\alpha}, \quad t \geq 1$, where G is an univariate Extreme value distribution (EVD) function. We choose an appropriate t and x_0 on the marginals of the bivariate data, so that the data should behave more like Pareto distribution. After needful transformation the size of data becomes $n = 468$.

Three Parameter MOBVPA: The estimated parameters of the above data set through the EM algorithm ([40]) for MOBVPA distribution are $\mu_1 = 1.000$, $\mu_2 = 0.4000$, $\sigma_1 = 2.5594$, $\sigma_2 = 0.7176$, $\alpha_0 = 0.7557$, $\alpha_1 = 0.8162$, $\alpha_2 = 0.9792$. Since, the direct real life data which will follow this three parameter MOBVPA distribution is not available, we modify the data with location and scale transformation. This transformation will affect cardinalities of I_0, I_1 and I_2 and consequently the value of the likelihood function of MOBVPA distribution. Therefore, we modify the algorithm by making a suitable approximation of the number of observations in each of I_0, I_1 and I_2, while calculating the value of the likelihood function. Here, we replace n_0, n_1 and n_2, the cardinality of cells I_0, I_1 and I_2 by $\tilde{n_0}$, $\tilde{n_1}$ and $\tilde{n_2}$, where $\tilde{n_i} = (n_0 + n_1 + n_2)\frac{\alpha_i}{(\alpha_0+\alpha_1+\alpha_2)}$ for $i = 0, 1, 2$. The Bayesian estimates are calculated and provided in Table-6.13 and simulated data analysis results are available in Table-6.14, where sample size is $n = 468$.

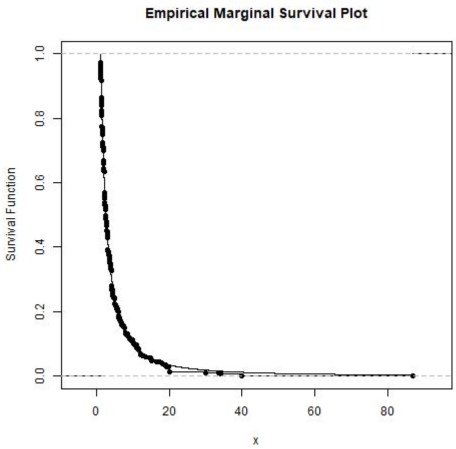
Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Bayesian Estimates	0.7296	0.8630	1.0189

Table 6.13: Bayesian Estimates (BE) of the parameters of three parameter MOBVPA based on data

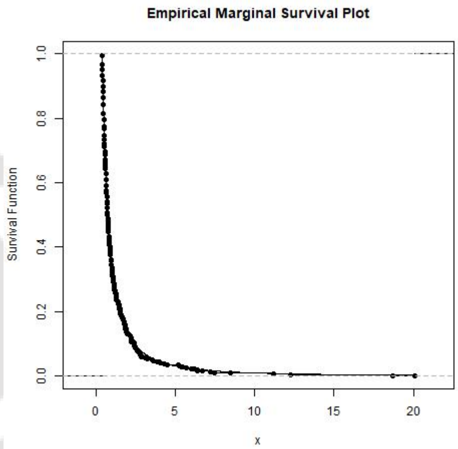
Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781
Average Bayesian Estimates	0.7355	0.8763	1.0325
Mean Squared Error	0.0027	0.0041	0.0044
Credible Interval	[0.4914 0.6639]	[0.5789 0.7827]	[0.7606 0.9670]
Coverage Probabilities	0.95	0.95	0.945

Table 6.14: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of three parameter MOBVPA based on Simulated data with parameters $\alpha_0 = 0.7296$, $\alpha_1 = 0.8630$, $\alpha_2 = 1.0189$ and sample size $n = 468$

Seven Parameter MOBVPA: It is observed that this data fits to seven parameter MOBVPA distribution. In Figure-6.1, we verify the distributional assumption by plotting the marginal empirical survival functions with the marginals of the distribution using these Bayes estimates of the parameters. Two dimensional empirical density plot in Figure-6.2 also shows that this seven parameter MOBVPA distribution can be an alternative model for the transformed data set. The Bayes estimates are shown in Table-6.15, whereas Table-6.16 and Table-6.17 show the ABE, MSE, CI and CP of the credible interval of this data based on simulated data of size $n = 468$.



(a) For X_1



(b) For X_2

Figure 6.1: Survival plots for two marginals of MOBVPA based on Bayesian analysis results of the transformed dataset

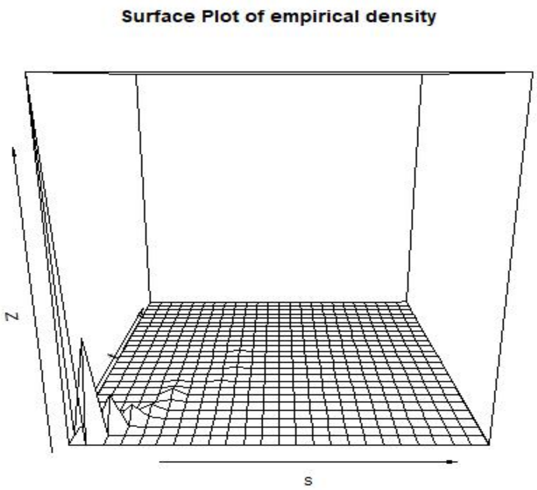


Figure 6.2: Two dimensional density plots for MOBVPA of the transformed dataset

Slice-cum-Gibbs							
Gamma Prior							
Parameter Sets	μ_1	μ_2	σ_1	σ_2	α_0	α_1	α_2
Starting Value	0.4794	0.3462	0.7781	0.5386	0.4417	0.6138	0.45491
Bayesian Estimates	0.9963	0.3990	2.9178	0.7766	0.7928993	0.8794	0.9998

Table 6.15: Bayesian Estimates (BE) of the parameters of Seven parameter MOBVPA based on data

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	μ_1	μ_2	σ_1	σ_2
Starting Value	0.4794	0.3462	0.7781	0.5386
Average Bayesian Estimates	0.9961	0.3991	2.7722	0.7550
Mean Squared Error	$1.25e^{-05}$	$9.64e^{-07}$	0.2470	0.0192
Credible Interval	[0.9878 0.9965]	[0.3977 0.3997]	[1.3532 2.6263]	[0.3782 0.7202]
Coverage Probabilities	0.96	0.94	0.88	0.86

Table 6.16: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA based on Simulated data with parameters $\mu_1 = 0.9963$, $\mu_2 = 0.3990$, $\sigma_1 = 2.9178$, $\sigma_2 = 0.7766$, $\alpha_0 = 0.7929$, $\alpha_1 = 0.8794$, $\alpha_2 = 0.9998$ and sample size $n = 468$

Slice-cum-Gibbs			
Gamma Prior			
Parameter Sets	α_0	α_1	α_2
Starting Value	0.4417	0.6138	0.45491
Average Bayesian Estimates	0.7270	0.8548	0.9993
Mean Squared Error	0.0120	0.0258	0.0304
Credible Interval	[0.4942 0.7937]	[0.3818 0.8988]	[0.6133 1.0853]
Coverage Probabilities	0.90	0.965	0.97

Table 6.17: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of seven parameter MOBVPA based on simulated data with parameters $\mu_1 = 0.9963$, $\mu_2 = 0.3990$, $\sigma_1 = 2.9178$, $\sigma_2 = 0.7766$, $\alpha_0 = 0.7929$, $\alpha_1 = 0.8794$, $\alpha_2 = 0.9998$ and sample size $n = 468$

6.9 Conclusion

We implement the empirical Bayes estimate in Bayesian analysis of the parameters of singular Marshall-Olkin bivariate Pareto distribution through slice cum Gibbs sampler technique. We observe for any particular choice of starting value of hyper-parameter, the algorithm works for any data set. We also construct credible intervals based on the generated samples and find the corresponding coverage probabilities. We observe that the results related to the coverage probabilities for location parameters are quite satisfactory, which was a matter for great concern in frequentist set up. Based on MOBVPA distribution, we can formulate the Geometric MOBVPA distribution. Due to the presence of extra parameter, it is more flexible in practical application. We study the distribution in the next chapter.

Chapter 7

Parameter Estimation of four parameters Geometrics Marshall-Olkin bivariate Pareto Distribution

7.1 Introduction

In this chapter, we study the EM algorithm and the Bayesian analysis of the four parameters Geometric Marshall-Olkin bivariate Pareto distribution (G-MOBVPA). Due to the presence of one extra parameter, this Geometric MOBVPA distribution is more flexible for the bivariate data, whose some of the components are equal and marginals are heavy-tail in nature. This distribution can be used to model the data related to finance, insurance, environmental sciences and internet network. In reliability, it can be used to model a series of the systems, which have two systems say 1 and 2, each having N number of independent and identical components attached in the series. Here N is a random variable. If $X_1, X_2, \cdots$ denotes the lifetime of the components of system 1, and $Y_1, Y_2, \cdots$ denotes the lifetime of the components of the system 2, then the lifetime of the two systems becomes (X, Y), where $X = \min\{X_1, X_2, \cdots, X_N\}$ and $Y = \min\{Y_1, Y_2, \cdots, Y_N\}$. Similar situation may arise in a two component competing risk model. In Insurance, such situation may arise when the number of insurance claims is random and two types of losses are attached to each claim.

The maximum likelihood estimates may be computationally expensive in this case too, as it requires to solve a system of four non-linear equations. Newton-Raphson or Gauss-Newton type iterative procedure is needed to solve this system of equations. Moreover, the choice of initial guesses and the convergence of the iterative algorithm are important issues. To resolve these issues, we treat the problem as the missing value problem and construct an EM algorithm. We also use the Bayesian approach through the slice cum Gibbs sampler. Usual step-out procedure works quite well and easy to implement. Since, it is a very flexible model and easy to implement, it gives the practitioner a choice of an alternative bivariate Pareto model, which may provide a better fit than the existing Marshall-Olkin bivariate Pareto distribution.

The idea came from [20] and [6], where the authors introduced the bivariate exponential geometric and bivariate gamma geometric distributions respectively, along the same line. A series of papers can also be found on the statistical inferences on different distributions in the work of Kundu et al. [[22], [31] and [25]]. However, there is no paper available on statistical inference on the Geometric Marshall-Olkin bivariate Pareto distribution. A few recent paper of Asimit et al [[3], [4], [2]] and [40] discussed the statistical inference of the singular bivariate Pareto with location scale parameter and its applications. However, this chapter is mainly dedicated to the four parameters bivariate Pareto distribution without location and scale parameters. We also explore the Bayesian analysis under the informative prior. Both frequentist and Bayesian confidence intervals are provided with an illustrative real-life data example. This work also provides a research opening of much more flexible model with location and scale.

Te rest of this chapter is arranged as follows. In section-7.2, we show the formulation of the geometric Marshall-Olkin bivariate Pareto distribution. Section-7.3 discusses the maximum likelihood estimation through the EM algorithm. The Bayesian analysis is discussed in Section-7.4. Numerical results are shown in Section-7.5. Data analysis is shown for all the methods in section-7.6. Finally, we conclude this chapter in Section-7.7.

7.2 Formulation of Geometric Marshall-Olkin bivariate Pareto Distribution

7.2.1 Type II Univariate Pareto distribution

A random variable X is said to be distributed according to the Pareto distribution type II, i.e., $X \sim PA(II)(\mu, \sigma, \alpha)$ if it has the survival function

$$S_X(x) = P(X > x) = \left(1 + \frac{x-\mu}{\sigma}\right)^{-\alpha}$$

and the probability density function

$$f(x) = \frac{\alpha}{\sigma}\left(1 + \frac{x-\mu}{\sigma}\right)^{-\alpha-1}$$

with $x > \mu \in \mathbb{R}$, $\sigma > 0$ and $\alpha > 0$.

Let us consider $\{X_n : n = 1, 2, \cdots\}$ is a sequence of iid non-negative random variables with a common cdf $F(\cdot)$ and N is a Geometric random variable with the parameter $0 < \theta \leq 1$, i.e., $N \sim GM(\theta)$. Also N is independent with X_n's. Now the mass function for N is,

$$P(N = n) = \theta(1-\theta)^{n-1}; \quad n = 1, 2, \cdots \tag{7.2.1}$$

Let us defined a new random variable $Y = \min\{X_n : n = 1, 2, \cdots, N\}$. Then the survival function of Y is given by,

$$\begin{aligned}\bar{G}(y) &= P(Y \geq y) \\ &= \sum_{n=1}^{\infty} P(Y \geq y | N = n) P(N = n) \\ &= \sum_{n=1}^{\infty} \bar{F}(y)^n \theta(1-\theta)^{n-1} \\ &= \frac{\theta \bar{F}(y)}{1-(1-\theta)\bar{F}(y)},\end{aligned} \tag{7.2.2}$$

where $\bar{F}(\cdot)$ is the survival function of X_i, $i = 1, 2, \cdots N$. So the corresponding probability density function (pdf) of Y becomes,

$$g(y) = \frac{\theta f(y)}{\{1-(1-\theta)\bar{F}(y)\}^2},$$

where $f(\cdot)$ is the pdf of X_i, $i = 1, 2, \cdots$. So the pdf of the univariate Pareto Geometric distribution with location parameter 0 and scale parameter 1, takes the following form,

$$g(y; \theta, \alpha) = \frac{\theta\alpha(1+y)^{-\alpha-1}}{\{1-(1-\theta)(1+y)^{-\alpha}\}^2} \tag{7.2.3}$$

We will denote it $UPG(II)(\theta, 0, 1, \alpha)$ or simply we denote $UPG(II)(\theta, \alpha)$.

7.2.2 Marshall Olkin Bivariate Pareto distribution

A bivariate random variable (X_1, X_2) is said to be distributed according to the Marshall-Olkin bivariate Pareto distribution i.e., $(X_1, X_2) \sim MOBVPA(\alpha_0, \alpha_1, \alpha_2)$ if it has the cumulative survival function

$$\begin{aligned}S(x_1, x_2) &= (1+z)^{-\alpha_0}\Big(1+x_1\Big)^{-\alpha_1}\Big(1+x_2\Big)^{-\alpha_2} \\ &= \begin{cases} S_1(x_1, x_2), & \text{if } x_1 < x_2 \\ S_2(x_1, x_2), & \text{if } x_1 > x_2 \\ S_0(x), & \text{if } x_1 = x_2 = x, \end{cases}\end{aligned}$$

where

$$S_1(x_1, x_2) = \Big(1+x_2\Big)^{-\alpha_0-\alpha_2}\Big(1+x_1\Big)^{-\alpha_1}$$

$$S_2(x_1,x_2) = \Big(1+x_2\Big)^{-\alpha_2}\Big(1+x_1\Big)^{-\alpha_0-\alpha_1}$$
$$S_0(x) = (1+x)^{-\alpha_0-\alpha_1-\alpha_2}.$$

So its joint pdf can be written as

$$f(x_1,x_2) \quad = \quad \begin{cases} f_1(x_1,x_2), & \text{if } x_1 < x_2 \\ f_2(x_1,x_2), & \text{if } x_1 > x_2 \\ f_0(x), & \text{if } x_1 = x_2 = x, \end{cases}$$

where

$$f_1(x_1,x_2) = \alpha_1(\alpha_0+\alpha_2)\Big(1+x_2\Big)^{-\alpha_0-\alpha_2-1}\Big(1+x_1\Big)^{-\alpha_1-1}$$
$$f_2(x_1,x_2) = \alpha_2(\alpha_0+\alpha_1)\Big(1+x_2\Big)^{-\alpha_2-1}\Big(1+x_1\Big)^{-\alpha_0-\alpha_1-1}$$
$$f_0(x) = \alpha_0(1+x)^{-\alpha_0-\alpha_1-\alpha_2-1}.$$

7.2.3 Geometric bivariate Pareto distribution

We follow the similar procedure of univariate geometric distribution to construct the Geometric bivariate Pareto distribution. Suppose $\{(X_{1i}, X_{2i}) : i = 1, 2, \cdots\}$ is a sequence of iid bivariate random variables with same cdf $F(\cdot,\cdot)$ and pdf $f(\cdot,\cdot)$. N is an univariate discrete Geometric random variable with $0 < \theta \leq 1$ and independent with (X_{1N}, X_{2N}). We define (Y_1, Y_2), another bivariate random variable as,

$$Y_1 = \min\{X_{11}, X_{12}, \cdots, X_{1N}\}$$
$$Y_2 = \min\{X_{21}, X_{22}, \cdots, X_{2N}\}.$$

Now the joint cumulative survival function can be obtained as,

$$\bar{G}(y_1,y_2) = \frac{\theta S(y_1,y_2)}{1-(1-\theta)S(y_1,y_2)}. \tag{7.2.4}$$

So its joint density function can be presented as,

$$g(y_1,y_2) = \frac{\theta\{1-(1-\theta)S(y_1,y_2)\}\frac{\partial^2 S(y_1,y_2)}{\partial y_1 \partial y_2} + 2\theta(1-\theta)\frac{\partial S(y_1,y_2)}{\partial y_1}\frac{\partial S(y_1,y_2)}{\partial y_2}}{\{1-(1-\theta)S(y_1,y_2)\}^3}. \tag{7.2.5}$$

If we use the survival function of MOBVPA distribution for distribution of $\{(X_1, X_2)\}$, then the joint distribution of (Y_1, Y_2) is called the Geometric Marshall-Olkin bivariate Pareto (G-MOBVPA) distribution. Therefore, the joint survival function of (Y_1, Y_2) can be written as,

$$\bar{G}(y_1,y_2) = \begin{cases} \bar{G}_1(y_1,y_2), & \text{if } y_1 < y_2 \\ \bar{G}_2(y_1,y_2), & \text{if } y_1 > y_2 \\ \bar{G}_0(y), & \text{if } y_1 = y_2 = y, \end{cases}$$

where

$$\bar{G}_1(y_1,y_2) = \frac{\theta\big(1+y_1\big)^{-\alpha_1}\big(1+y_2\big)^{-\alpha_0-\alpha_2}}{1-(1-\theta)\big(1+y_1\big)^{-\alpha_1}\big(1+y_2\big)^{-\alpha_0-\alpha_2}},$$
$$\bar{G}_2(y_1,y_2) = \frac{\theta\big(1+y_1\big)^{-\alpha_0-\alpha_1}\big(1+y_2\big)^{-\alpha_2}}{1-(1-\theta)\theta\big(1+y_1\big)^{-\alpha_0-\alpha_1}\big(1+y_2\big)^{-\alpha_2}},$$
$$\bar{G}_0(y) = \frac{\theta(1+y)^{-\alpha_0-\alpha_1-\alpha_2}}{1-(1-\theta)(1+y)^{-\alpha_0-\alpha_1-\alpha_2}}.$$

Hence the joint pdf is,

$$g(y_1,y_2) = \begin{cases} g_1(y_1,y_2), & \text{if } y_1 < y_2 \\ g_2(y_1,y_2), & \text{if } y_1 > y_2 \\ g_0(y), & \text{if } y_1 = y_2 = y, \end{cases} \tag{7.2.6}$$

where

$$g_1(y_1,y_2) = \frac{\theta\alpha_1(\alpha_0+\alpha_2)\big(1+y_1\big)^{-\alpha_1-1}\big(1+y_2\big)^{-\alpha_0-\alpha_2-1}\{1+(1-\theta)S_1(y_1,y_2)\}}{\sigma_1\sigma_2\{1-(1-\theta)S_1(y_1,y_2)\}^3}$$
$$g_2(y_1,y_2) = \frac{\theta(\alpha_0+\alpha_1)\alpha_2\big(1+y_1\big)^{-\alpha_0-\alpha_1-1}\big(1+y_2\big)^{-\alpha_2-1}\{1+(1-\theta)S_2(y_1,y_2)\}}{\{1-(1-\theta)S_2(y_1,y_2)\}^3}$$
$$g_0(y) = \frac{\theta\alpha_0(1+y)^{-\alpha_0-\alpha_1-\alpha_2-1}}{\{1-(1-\theta)(1+y)^{-\alpha_0-\alpha_1-\alpha_2}\}^2}.$$

We denote this distribution as $G-MOBVPA(\theta,\alpha_0,\alpha_1,\alpha_2)$. Rest of the chapter, We will call it four parameters G-MOBVPA distribution. Surface and contour plots of the pdf of this distribution are shown in Figure-7.1 and Figure-7.2. The following two different sets of parameters provide two figures. $\xi_1 : \theta = 0.20, \alpha_0 = 0.10, \alpha_1 = 0.20, \alpha_2 = 0.40$; $\xi_2 : \theta = 0.80, \alpha_0 = 4, \alpha_1 = 5, \alpha_2 = 10$;

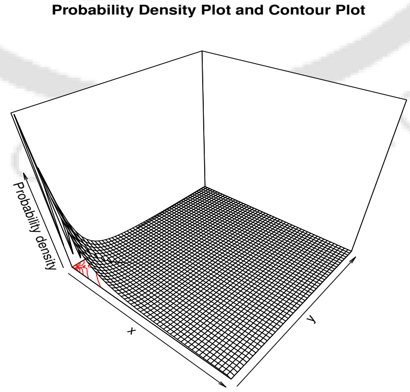


Figure 7.1: Probability density plot and Contour Plot of three parameter G-MOBVPA for the parameter values $\theta = 0.20$, $\alpha_0 = 0.10$, $\alpha_1 = 0.20$ and $\alpha_2 = 0.40$

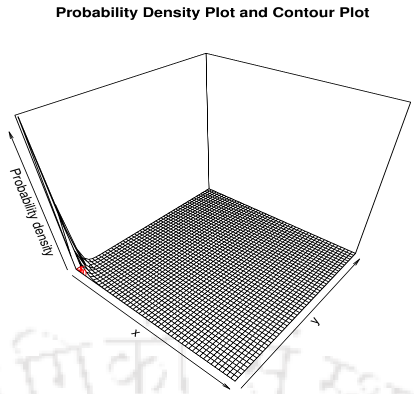


Figure 7.2: Probability density plot and Contour Plot of three parameter G-MOBVPA for the parameter values $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$ and $\alpha_2 = 10$

It is clear that the bivariate random variable (Y_1, Y_2) given $N = n$ follow the three parameters MOBVPA distribution with the parameters $n\alpha_0$, $n\alpha_1$ and $n\alpha_2$ i.e. $(Y_1, Y_2|N = n) \sim MOBVPA(n\alpha_0, n\alpha_1, n\alpha_2)$. Now the joint survival function of $(Y_1, Y_2) \sim G{-}MOBVPA(\theta, \alpha_0, \alpha_1, \alpha_2)$ and $N \sim GM(\theta)$ is given as follows,

$$P(Y_1 > y_1, Y_2 > y_2, N = n) = P(Y_1 > y_1, Y_2 > y_2|N = n)P(N = n) \tag{7.2.7}$$

$$= \begin{cases} \theta(1-\theta)^{n-1}(1+y_1)^{-n\alpha_1}(1+y_2)^{-n(\alpha_0+\alpha_2)} & \text{if } y_1 < y_2 \\ \theta(1-\theta)^{n-1}(1+y_1)^{-n(\alpha_0+\alpha_1)}(1+y_2)^{-n\alpha_2} & \text{if } y_1 > y_2 \\ \theta(1-\theta)^{n-1}(1+y)^{-n(\alpha_0+\alpha_1+\alpha_2)} & \text{if } y_1 = y_2 = y. \end{cases} \tag{7.2.8}$$

So the joint pdf of (Y_1, Y_2, N) is

$$g_{Y_1,Y_2,N}(y_1, y_2) = \begin{cases} \theta(1-\theta)^{n-1}n^2\alpha_1(\alpha_0+\alpha_2)(1+y_1)^{-n\alpha_1-1}(1+y_2)^{-n(\alpha_0+\alpha_2)-1} & \text{if } y_1 < y_2 \\ \theta(1-\theta)^{n-1}n^2(\alpha_0+\alpha_1)\alpha_2(1+y_1)^{-n(\alpha_0+\alpha_1)-1}(1+y_2)^{-n\alpha_2-1} & \text{if } y_1 > y_2 \\ \theta(1-\theta)^{n-1}n\alpha_0(1+y)^{-n(\alpha_0+\alpha_1+\alpha_2)-1} & \text{if } y_1 = y_2 = y. \end{cases}$$

Now we would like to find the conditional distribution of N given $Y_1 = y_1$ and $Y_2 = y_2$. The conditional mass function of N given $Y_1 = y_1$ and $Y_2 = y_2$ becomes,

$$g_N(n|y_1, y_2) = \begin{cases} \frac{(1-a)^3}{1+a}n^2a^{n-1} & \text{if } y_1 < y_2 \\ \frac{(1-b)^3}{1+b}n^2b^{n-1} & \text{if } y_1 > y_2 \\ (1-c)^2nc^{n-1} & \text{if } y_1 = y_2, \end{cases}$$

where

$$a = (1-\theta)(1+y_1)^{-\alpha_1}(1+y_2)^{-\alpha_0-\alpha_2}$$

$$b = (1-\theta)\big(1+y_1\big)^{-\alpha_0-\alpha_1}\big(1+y_2\big)^{-\alpha_2}$$
$$c = (1-\theta)(1+y_1)^{-\alpha_0-\alpha_1-\alpha_2}.$$

Then the conditional mean of N given $Y_1 = y_1$, and $Y_2 = y_2$ can be written as,

$$E[N|y_1,y_2] = \begin{cases} \frac{a^2+4a+1}{1-a^2} & \text{if } y_1 < y_2 \\ \frac{b^2+4b+1}{1-b^2} & \text{if } y_1 > y_2 \\ \frac{1+c}{1-c} & \text{if } y_1 = y_2. \end{cases}$$

Note: Suppose $N \sim GM(1-c)$, i.e., $P(N=n) = (1-c)(1-\overline{1-c})^{n-1}$, $\quad n = 1,2,\cdots$. Then $E[N] = \frac{1}{1-c}$, $Var(N) = \frac{c}{(1-c)^2}$, $E[N^2] = \frac{1+c}{(1-c)^2}$ and $E[N^3] = \frac{c^2+4c+1}{(1-c)^3}$ (using probability generating function).

7.3 EM Estimation

In this section, we discuss the maximum likelihood estimates using the EM algorithm for unknown parameters of the four parameters G-MOBVPA distribution. Let us consider $I = \{(y_{11}, y_{21}), (y_{12}, y_{22}), \cdots, (y_{1m}, y_{2m})\}$ is sample of size m from the four parameters G-MOBVPA distribution. Now we use the following notation:

$$I_0 = \{i : y_{1i} = y_{2i}\},\ I_1 = \{i : y_{1i} < y_{2i}\},\ I_2 = \{i : y_{1i} > y_{2i}\}$$

Also $m_0 = |I_0|$, $m_1 = |I_1|$, $m_2 = |I_2|$ and $m_0 + m_1 + m_2 = m$. Usual log-likelihood function of this G-MOBVPA can be written as

$$\begin{aligned}
&L(\theta,\alpha_0,\alpha_1,\alpha_2) \\
&= m\ln(\theta) + m_1\ln\alpha_1 + m_1\ln(\alpha_0+\alpha_2) - (\alpha_0+\alpha_2+1)\sum_{i\in I_1}\ln(1+y_{2i}) \\
&- (\alpha_1+1)\sum_{i\in I_1}\ln(1+y_{1i}) + \sum_{i\in I_1}\ln\Big(1+(1-\theta)(1+y_{1i})^{-\alpha_1}(1+y_{2i})^{-\alpha_0-\alpha_2}\Big) \\
&- 3\sum_{i\in I_1}\ln\Big(1-(1-\theta)(1+y_{1i})^{-\alpha_1}(1+y_{2i})^{-\alpha_0-\alpha_2}\Big) + m_2\ln\alpha_2 + m_2\ln(\alpha_0+\alpha_1) \\
&- (\alpha_0+\alpha_1+1)\sum_{i\in I_2}\ln(1+y_{1i}) - (\alpha_2+1)\sum_{i\in I_2}\ln(1+y_{2i}) \qquad (7.3.1) \\
&+ \sum_{i\in I_2}\ln\Big(1+(1-\theta)(1+y_{1i})^{-\alpha_0-\alpha_1}(1+y_{2i})^{-\alpha_2}\Big) \\
&- 3\sum_{i\in I_2}\ln\Big(1-(1-\theta)(1+y_{1i})^{-\alpha_0-\alpha_1}(1+y_{2i})^{-\alpha_2}\Big) + m_0\ln\alpha_0 \\
&- (\alpha_0+\alpha_1+\alpha_2+1)\sum_{i\in I_0}\ln(1+y_{1i}) - 2\sum_{i\in I_0}\ln\Big(1-(1-\theta)(1+y_{1i})^{-\alpha_0-\alpha_1-\alpha_2}\Big). \qquad (7.3.2)
\end{aligned}$$

It is clear that MLEs of the unknown parameters cannot be found is closed form. It can be obtained only by solving four non-linear equations simultaneously, which is computationally too expensive. To avoid this problem, we treat this as a missing value problem and use the EM algorithm.

Now, let us see how we can formulate this as a missing value problem. Let us assume that we observed n and consider the following three independent random variables,

$$\{U_0|N=n\} \sim PA(II)(0,1,n\alpha_0)$$

$$\{U_1|N=n\} \sim PA(II)(0,1,n\alpha_1)$$

$$\{U_2|N=n\} \sim PA(II)(0,1,n\alpha_2)$$

Also it is well known that

$$\{Y_1|N=n\} = \min\{\{U_0|N=n\},\{U_1|N=n\}\}$$

$$\{Y_2|N=n\} = \min\{\{U_0|N=n\},\{U_2|N=n\}\}$$

The complete log likelihood function of the random variable (Y_1, Y_2, N) based on the complete data $I = \{(y_{11}, y_{21}, n_1), (y_{12}, y_{22}, n_2), \cdots, (y_{1m}, y_{2m}, n_m)\}$ is

$$\begin{aligned}
&L^*(\theta,\alpha_0,\alpha_1,\alpha_2)\\
&= (\sum_{i\in I_0} n_i + 2\sum_{i\in I_1} n_i + 2\sum_{i\in I_2} n_i) + m_1\ln\alpha_1 + m_1\ln(\alpha_0+\alpha_2) + m_2\ln\alpha_2 + m_2\ln(\alpha_0+\alpha_1)\\
&- \sum_{i\in I_1}(n_i(\alpha_0+\alpha_2)+1)\ln(1+y_{2i}) - (n_i\alpha_1+1)\sum_{i\in I_1}\ln(1+y_{1i})\\
&- \sum_{i\in I_2}(n_i(\alpha_0+\alpha_1)+1)\ln(1+y_{1i}) - \sum_{i\in I_2}(n_i\alpha_2+1)\ln(1+y_{2i}) + m_0\ln\alpha_0\\
&- \sum_{i\in I_0}(n_i(\alpha_0+\alpha_1+\alpha_2)+1)\ln(1+x_{1i}) + m\ln\left(\frac{\theta}{1-\theta}\right)\\
&+ \ln(1-\theta)(\sum_{i\in I_0} n_i + \sum_{i\in I_1} n_i + \sum_{i\in I_2} n_i).
\end{aligned} \tag{7.3.3}$$

Now we introduce two new random variables as missing variables for the EM algorithm $(\Delta_1; \Delta_2)$ as

$$\Delta_1 = \begin{cases} 0 & \text{if } Y_1 = U_0 \\ 1 & \text{if } Y_1 = U_1 \end{cases}$$

and

$$\Delta_2 = \begin{cases} 0 & \text{if } Y_2 = U_0 \\ 2 & \text{if } Y_2 = U_2. \end{cases}$$

To calculate the E-step, we need to calculate the conditional distribution of Δ_1 and Δ_2. Using the definition of Y_1, Y_2, Δ_1 and Δ_2 we have,

$\triangleright$ For group I_0, both Δ_1 and Δ_2 are known,

$$\Delta_1 = \Delta_2 = 0$$

$\triangleright$ For group I_1, Δ_1 is known, Δ_2 is unknown,

$$\Delta_1 = 1, \Delta_2 = 0 \text{ or } 2$$

Therefore, we need to find out $u_1 = P(\Delta_2 = 0|I_1)$and $u_2 = P(\Delta_2 = 2|I_1)$

$\triangleright$ For group I_2, Δ_2 is known, Δ_1 is unknown,

$$\Delta_1 = 0 \text{ or } 1, \Delta_2 = 2$$

Moreover, we need $w_1 = P(\Delta_1 = 0|I_2)$ and $w_2 = P(\Delta_1 = 1|I_2)$

Since, each posterior probability corresponds to one of the ordering from Table-7.1, we calculate u_1, u_2, w_1, w_2 using the probability of appropriate ordering. We have the following expressions for u_1, u_2, w_1, w_2

$$u_1 = \frac{P(U_1 < U_0 < U_2)}{P(U_1 < U_0 < U_2) + P(U_1 < U_2 < U_0)}$$

$$u_2 = \frac{P(U_1 < U_2 < U_0)}{P(U_1 < U_0 < U_2) + P(U_1 < U_2 < U_0)}$$

$$w_1 = \frac{P(U_2 < U_0 < U_1)}{P(U_2 < U_0 < U_1) + P(U_2 < U_1 < U_0)}$$

$$w_2 = \frac{P(U_2 < U_1 < U_0)}{P(U_2 < U_0 < U_1) + P(U_2 < U_1 < U_0)}$$

Ordering	(X_1, X_2)	Group
$U_0 < U_1 < U_2$	(U_0, U_0)	I_0
$U_0 < U_2 < U_1$	(U_0, U_0)	I_0
$U_1 < U_0 < U_2$	(U_1, U_0)	I_1
$U_1 < U_2 < U_0$	(U_1, U_2)	I_1
$U_2 < U_0 < U_1$	(U_0, U_2)	I_2
$U_2 < U_1 < U_0$	(U_1, U_2)	I_2

Table 7.1: Groups and corresponding orderings of hidden random variables U_0, U_1 and U_2

Using this above result, we evaluate the other probabilities to get values of u_1, u_2, w_1, w_2 as : $u_1 = \frac{\alpha_0}{\alpha_0+\alpha_2}$, $u_2 = \frac{\alpha_2}{\alpha_0+\alpha_2}$, $w_1 = \frac{\alpha_0}{\alpha_0+\alpha_1}$ and $w_2 = \frac{\alpha_1}{\alpha_0+\alpha_1}$.

Now we consider the pseudo likelihood contribution of an observation from I_0, I_1 and I_2 with conditioning on $N = n$. Hence the pseudo likelihood contribution of any observation from I_0 becomes,

$$n_i\alpha_0(1+y_{1i})^{-n_i(\alpha_0+\alpha_1+\alpha_2)-1}$$

Similarly the pseudo likelihood contributions of any observation from I_1 and I_2 becomes,

$$n_i\alpha_1(1+y_{1i})^{-n_i\alpha_1-1}n_i\alpha_0(1+y_{2i})^{-n_i(\alpha_0+\alpha_2)-1} \text{ with the probability } u_1,$$
$$n_i\alpha_1(1+y_{1i})^{-n_i\alpha_1-1}n_i\alpha_2(1+y_{2i})^{-n_i(\alpha_0+\alpha_2)-1} \text{ with the probability } u_2$$

and

$$n_i\alpha_2(1+y_{2i})^{-n_i\alpha_2-1}n_i\alpha_0(1+y_{1i})^{-n_i(\alpha_0+\alpha_1)-1} \text{ with the probability } w_1,$$
$$n_i\alpha_2(1+y_{2i})^{-n_i\alpha_2-1}n_i\alpha_1(1+y_{1i})^{-n_i(\alpha_0+\alpha_1)-1} \text{ with the probability } w_2$$

respectively. Here n is missing, so we replace n by $E[N|y_1, y_2]$ to compute pseudo log-likelihood function in the EM algorithm. Let us consider $E[N|y_{1i}, y_{2i}] = a_i$. The pseudo log-likelihood function is obtained by multiplying all log-likelihood contributions by its posterior probabilities of hidden random variables. So the pseudo log-likelihood function is,

$$\begin{aligned}
Q &= L_{pseudo}(\theta, \alpha_0, \alpha_1, \alpha_2)\\
&= \Big(\sum_{i\in I_0}\ln(a_i) + 2\sum_{i\in I_1}\ln(a_i) + 2\sum_{i\in I_2}\ln(a_i)\Big) + (m_0 + u_1 m_1 + w_1 m_2)\ln(\alpha_0)\\
&+ (m_1 + w_2 m_2)\ln(\alpha_1) + (u_2 m_1 + m_2)\ln(\alpha_2) - \Big(\sum_{i\in I_0}\ln(1+y_{1i}) + \sum_{i\in I_1\cup I_2}\ln(1+y_{1i})\\
&+ \sum_{i\in I_1\cup I_2}\ln(1+y_{2i})\Big) - \alpha_0\Big(\sum_{i\in I_0} a_i\ln(1+y_{1i}) + \sum_{i\in I_1} a_i\ln(1+y_{2i}) + \sum_{i\in I_2} a_i\ln(1+y_{1i})\Big)\\
&- \alpha_1\Big(\sum_{i\in I_0} a_i\ln(1+y_{1i}) + \sum_{i\in I_1} a_i\ln(1+y_{1i}) + \sum_{i\in I_2} a_i\ln(1+y_{1i})\Big) - \alpha_2\Big(\sum_{i\in I_0} a_i\ln(1+y_{1i}) \quad (7.3.4)\\
&+ \sum_{i\in I_1} a_i\ln(1+y_{2i}) + \sum_{i\in I_2} a_i\ln(1+y_{2i})\Big) + m\ln\Big(\frac{\theta}{1-\theta}\Big) + \ln(1-\theta)\Big(\sum_{i\in I_0} a_i + \sum_{i\in I_1} a_i \quad (7.3.5)\\
&+ \sum_{i\in I_2} a_i\Big) \quad (7.3.6)
\end{aligned}$$

The M-step of the EM algorithm involves the maximization of the above equation 8.3.1 w.r.t. unknown parameters θ, α_0, α_1, and α_2. Hence the estimator of the unknown parameters at $t+1$ step are,

$$\hat{\alpha}_0^{(t+1)} = \frac{m_0 + u_1^{(t)} m_1 + w_1^{(t)} m_2}{\sum_{i\in I_0} a_i^{(t)}\ln(1+y_{1i}) + \sum_{i\in I_1} a_i^{(t)}\ln(1+y_{2i}) + \sum_{i\in I_2} a_i^{(t)}\ln(1+y_{1i})}. \quad (7.3.7)$$

$$\hat{\alpha}_1^{(t+1)} = \frac{m_1 + w_2^{(t)} m_2}{\sum_{i\in I_0} a_i^{(t)}\ln(1+y_{1i}) + \sum_{i\in I_1} a_i^{(t)}\ln(1+y_{1i}) + \sum_{i\in I_2} a_i^{(t)}\ln(1+y_{1i})} \quad (7.3.8)$$

$$\hat{\alpha}_2^{(t+1)} = \frac{m_2 + u_2^{(t)} m_1}{\sum_{i\in I_0} a_i^{(t)}\ln(1+y_{1i}) + \sum_{i\in I_1} a_i^{(t)}\ln(1+y_{2i}) + \sum_{i\in I_2} a_i^{(t)}\ln(1+y_{2i})} \quad (7.3.9)$$

and

$$\hat{\theta}^{(t+1)} = \frac{m}{\sum_{i \in I_0} a_i^{(t)} + \sum_{i \in I_1} a_i^{(t)} + \sum_{i \in I_2} a_i^{(t)}} \tag{7.3.10}$$

7.4 Bayesian Estimation

In this section, we compute the average Bayesian estimates (ABE) of the unknown parameters and also the associated mean squared error (MSE), credible intervals (CI) and coverage probability (CP) of CIs based on the proper priors with fixed hyper-parameters. Mainly, we use the slice cum Gibbs sampling technique for sampling from the posterior distribution.

7.4.1 Prior Assumption

We assume that α_0, α_1 and α_2 are distributed according to the gamma distribution with shape parameters k_i and scale parameters θ_i, where $i = 0, 1, 2$, i.e.,

$$\begin{aligned} \alpha_0 &\sim \Gamma(k_0, \theta_0) \equiv \text{Gamma}(k_0, \theta_0), \\ \alpha_1 &\sim \Gamma(k_1, \theta_1) \equiv \text{Gamma}(k_1, \theta_1), \\ \alpha_2 &\sim \Gamma(k_2, \theta_2) \equiv \text{Gamma}(k_2, \theta_2). \end{aligned} \tag{7.4.1}$$

The probability density function of the Gamma distribution is given by,

$$f_\Gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}},$$

where $\Gamma(k)$ is the gamma function evaluated at k.

Further, we assume that the geometric parameter θ is distributed according to Beta distribution first kind with the parameters a and b, i.e.,

$$\pi(\theta) = \frac{1}{\beta(a, b)} \theta^{a-1} (1-\theta)^{b-1}, \quad \theta \in [0, 1]. \tag{7.4.2}$$

7.4.2 Posterior Distribution

We know the Bayes estimate of an unknown parameter under the squared error loss function is the posterior mean of the corresponding parameter. But in this case, it is not easy to calculate the Bayes estimate of the unknown model parameters θ, α_0, α_1 and α_2 in closed form. We propose to use the slice cum Gibbs sampling procedure to generate samples from the conditional posterior distribution. The full logarithm of the conditional posterior distributions of θ, α_0, α_1 and α_2 are given as follows,

$$\ln(\pi(\theta \mid \alpha_0, \alpha_1, \alpha_2, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\beta(a, b)) + (a-1)\ln\theta$$

$$+ (b-1)\ln(1-\theta),$$

$$\ln(\pi(\alpha_0 \mid \theta, \alpha_1, \alpha_2, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_0)) - k_0 \ln(\theta_0) + (k_0 - 1)\ln\alpha_0 - \frac{\alpha_0}{\theta_0},$$

$$\ln(\pi(\alpha_1 \mid \theta, \alpha_0, \alpha_2, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_1)) - k_1 \ln(\theta_1) + (k_1 - 1)\ln\alpha_1 - \frac{\alpha_1}{\theta_1}$$

and

$$\ln(\pi(\alpha_2 \mid \theta, \alpha_0, \alpha_1, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_2)) - k_2 \ln(\theta_2) + (k_2 - 1)\ln\alpha_2 - \frac{\alpha_2}{\theta_2}$$

respectively.

7.5 Numerical Analysis

We use the software R 3.5.0 to calculate all the estimates. The codes are run at IIT Guwahati computers with model : Intel(R) Core(TM) i5-6200U CPU 2.30GHz. The codes will be available on request to the authors. Here, we take two different sets of parameters with two different sample size as $n = 450$ and $n = 1000$ in both of the estimation technique.

7.5.1 EM estimates

For the EM estimate, we calculate average estimates (AE), mean square error (MSE) and 95% parametric bootstrap confidence interval (CI) for the parameters based on 1000 replications. We also observe the average number of iterations for each of the cases. We use different initial choice of the parameters and it is observed that all results are near about same with other initial choices. We have used the stopping criterion as the relative difference of log-likelihood and pseudo log-likelihood of each of the iteration with stopping tolerance limit as 0.00001. All the EM algorithm results are shown in Table-7.2, Table-7.3, Table-7.4 and Table-7.5 respectively.

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.2061	0.1030	0.2056	0.4106
Mean Squared Error	0.0010	0.0003	0.0008	0.0031
Confidence Interval	[0.1512 0.2751]	[0.0725 0.1394]	[0.1552 0.2653]	[0.3084 0.5292]
Average Number of Iteration			309	

Table 7.2: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter Geometric MOBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 450$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.2030	0.1017	0.2027	0.4064
Mean Squared Error	0.0004	0.0001	0.0004	0.0013
Confidence Interval	[0.1657 0.2463]	[0.0822 0.1245]	[0.1670 0.2421]	[0.3417 0.4813]
Average Number of Iteration			320	

Table 7.3: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter Geometric MOBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 1000$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.7892	3.9544	4.9544	9.9321
Mean Squared Error	0.0103	0.1960	0.2616	0.8451
Confidence Interval	[0.6415 0.9792]	[3.1198 4.8381]	[4.0186 6.0154]	[8.2129 11.7032]
Average Number of Iteration			163	

Table 7.4: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter Geometric MOBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 450$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.7788	3.9348	4.9157	9.8307
Mean Squared Error	0.0073	0.1137	0.1648	0.5389
Confidence Interval	[0.6459, 0.9154]	[3.2995, 4.5493]	[4.1326, 5.6472]	[8.3958, 11.0804]
Average Number of Iteration			139	

Table 7.5: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter Geometric MOBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 1000$

7.5.2 For Bayesian Estimation:

Here, we use different sets of parameters to show the performance of the proposed method. We calculate average Bayesian estimates (ABE), mean square errors (MSE), credible intervals (CI) and coverage probabilities (CP) of the parameters. Coverage probability increases near equal to the theoretical confidence level as sample size increases. All the Bayesian results are shown in Table-7.6, Table-7.7, Table-7.8 and Table-7.9 respectively. The convergence of the results does not depend on the choice of hyper-parameters and the initial choice of the starting value.

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.2099	0.1052	0.2062	0.4168
Mean Squared Error	0.0011	0.0004	0.0008	0.0035
Credible Interval	[0.0992 0.1815]	[0.0449 0.0900]	[0.0840 0.1612]	[0.2134 0.3807]
Coverage Probabilities	0.94	0.95	0.95	0.96

Table 7.6: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter Geometric MOBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.10$, $\alpha_1 = 0.20$, $\alpha_2 = 0.40$ and sample size $n = 450$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.2040	0.1022	0.2038	0.4080
Mean Squared Error	0.0004	0.0001	0.0004	0.0011
Credible Interval	[0.1454 0.2089]	[0.0629 0.0978]	[0.1290 0.1898]	[0.2626 0.3766]
Coverage Probabilities	0.96	0.95	0.96	0.95

Table 7.7: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter Geometric MOBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.10$, $\alpha_1 = 0.20$, $\alpha_2 = 0.40$ and sample size $n = 1000$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.7220	3.6768	4.6113	9.1104
Mean Squared Error	0.0108	0.2436	0.3201	1.2677
Credible Interval	[0.8380 0.9999]	[2.3196 3.5563]	[2.6072 4.0985]	[6.2088 8.6390]
Coverage Probabilities	0.85	0.92	0.94	0.87

Table 7.8: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter Geometric MOBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 450$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.7665	3.8943	4.8000	9.5275
Mean Squared Error	0.0039	0.0729	0.1307	0.4857
Credible Interval	[0.5696, 0.7497]	[2.9846, 3.9414]	[3.7911, 4.8773]	[7.3480, 9.1772]
Coverage Probabilities	0.92	0.95	0.96	0.92

Table 7.9: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter Geometric MOBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 1000$

7.6 Data Analysis

We analyze a data set of the indemnity payments (Loss) and allocated loss adjustment expense (ALAE) relating to 1500 general liability claims from insurance companies are available in the R package evd ([54]). From [15], we know that peak over threshold method on random variable U (say) provides polynomial generalized Pareto distribution for any x_0 with $1 + log(G(x_0)) \in (0, 1)$ i.e. $P(U > tx_0|U > x_0) = t^{-\alpha}, \quad t \geq 1$, where $G(\cdot)$ is the distribution function of U. We choose an appropriate t and x_0 on the marginals of the bivariate data, so that the data should behave more like Pareto distribution.

We assume the data follow near equal to singular Marshall Olkin bivariate Pareto and try to verify our assumption. We fit the empirical survival functions with the marginals of this bivariate Pareto, whose parameters can be obtained from the EM algorithm that we have developed. Figure-7.4 shows a good fit for both the marginals, where the empirical survival function and the survival function based on estimated parameters are plotted together. We check the same with the estimates of Bayesian analysis in Figure-7.5. The data analysis is performed based on sample size 1500.

The empirical two dimensional density plot in Figure-7.3 verifies that Geometric Marshall-Olkin bivariate Pareto can be an alternative model for the transformed data set. Parameter estimates through the EM algorithm of this bivariate distribution based on the transformed data set of size $n = 468$ are $\theta = 0.6878$, $\alpha_0 = 0.0234$, $\alpha_1 = 1.2745$ and $\alpha_2 = 1.4087$. We use the parametric bootstrap technique to find the mean square errors, confidence intervals and average number of iterations. All results are available in Table-7.10. Parameter estimation found in the Bayesian estimation is closer to the estimates in the EM algorithm. They are $\theta = 0.6917$, $\alpha_0 = 0.0294$, $\alpha_1 = 1.2694$ and $\alpha_2 = 1.4046$. Results obtained through parametric bootstrap taking the original parameters as estimates through Bayesian approach are kept in Table-7.11.

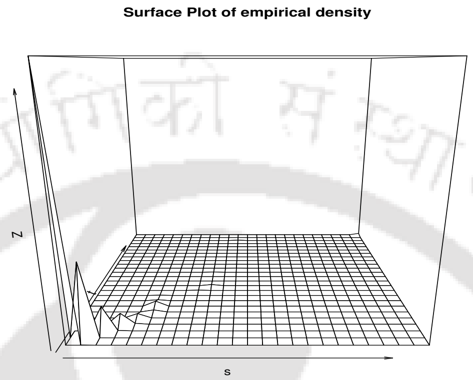
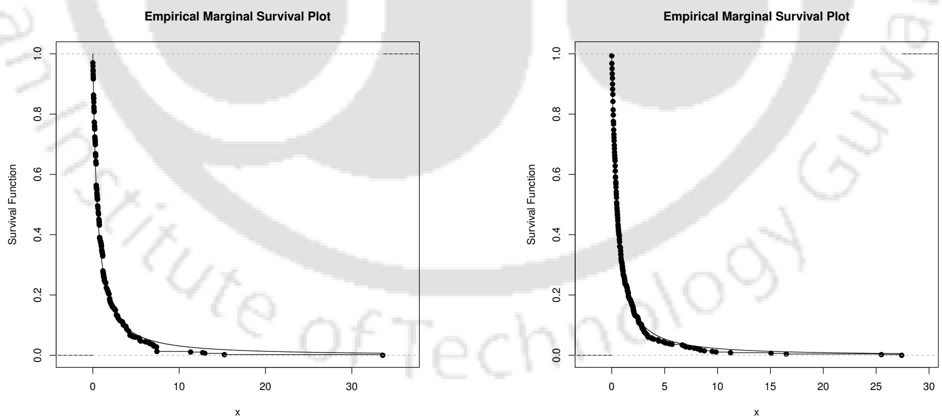


Figure 7.3: Two dimensional density plots for three parameter G-MOBVPA of the transformed dataset

7.6.1 Data Analysis : EM algorithm results



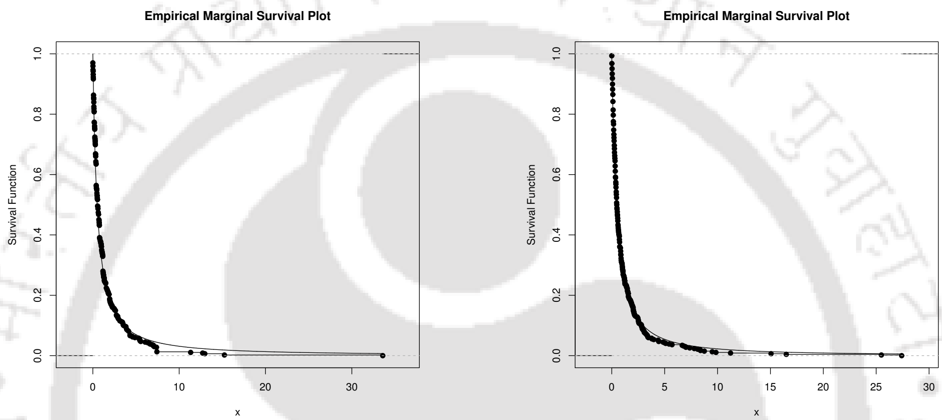
(a) For X_1 (b) For X_2

Figure 7.4: Survival plots for two marginals of three parameter G-MOBVPA through EM algorithm of the transformed dataset

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Mean Squared Error	0.0047	0.0001	0.0087	0.0117
Confidence Interval	[0.5686, 0.8420]	[0.0055, 0.0480]	[1.1106, 1.4656]	[1.2267, 1.6538]
Average Number of Iteration			143	

Table 7.10: Average Estimates (AE), Mean Squared Errors (MSE) and Confidence Intervals (CI) of four parameter Geometric MOBVPA with parameters $\theta = 0.6878$, $\alpha_0 = 0.0234$, $\alpha_1 = 1.2745$, $\alpha_2 = 1.4087$ and sample size $n = 468$

7.6.2 Data Analysis : Bayesian Estimation



(a) For X_1 (b) For X_2

Figure 7.5: Survival plots for two marginals three parameter G-MOBVPA through Bayesian estimation of the transformed dataset

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.7082	0.0336	1.2730	1.4142
Mean Squared Error	0.0051	0.0002	0.0087	0.0109
Credible Interval	[0.4854, 0.7072]	[0.0002, 0.0216]	[0.9132, 1.2177]	[1.0137, 1.3525]
Coverage Probabilities	0.92	0.96	0.96	0.95

Table 7.11: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter Geometric MOBVPA with parameters $\theta = 0.6917$, $\alpha_0 = 0.0294$, $\alpha_1 = 1.2694$, $\alpha_2 = 1.4046$ and sample size $n = 468$

7.7 Conclusion

We observe that it is a more flexible model than three parameter Marshall-Olkin bivariate Pareto distribution. It is easy to implement the EM algorithm and the Bayesian analysis for estimation of the parameters. Therefore, modeling the data through G-MOBVPA may have more appeal for the practitioner. However, statistical inference for this model is much more difficult with location and scale due to its discontinuous nature of likelihood with respect to the parameters. An absolute continuous version of this above distribution can be an interesting model to explore. We address the parameter estimation procedure of the distribution in Chapter-8.

Chapter 8

Parameter Estimation of Four Parameters Block-Basu Bivariate Pareto Geometric Distribution

8.1 Introduction

Geometric absolute continuous Marshall-Olkin bivariate Pareto (GBBVPA) is more flexible model than the absolute continuous Marshall-Olkin bivariate Pareto. In this chapter, we explore the statistical inference of the GBBVPA distribution through the EM algorithm. We also study the Bayesian analysis in this setup through the slice cum Gibbs sampler approach. Like the other cases, finding initial guesses can be an integral issue in computing maximum likelihood estimates. To resolve the issue, we construct an the EM algorithm. We also explore the Bayesian approach through the slice cum Gibbs sampler. Usual step-out procedure works quite well and easy to implement. Since, it is a very flexible model when components are not equal, it gives the practitioner a choice of an alternative bivariate Pareto model, which may provide a better fit than existing the Marshall-Olkin bivariate Pareto distribution.

Kozubowski and Panorska (2008) [20] and Barreto-Souza (2012) [6], introduced the geometric bivariate exponential and the geometric bivariate gamma distributions, respectively, along the same line. A series of papers can also be found on statistical inferences on different distributions in the work of Kundu et al. [[22], [31] and [25]]. A few recent paper of Asimit et al [[3], [4], [2]] and [40] discussed statistical inference of the singular bivariate Pareto with location scale parameter and its applications. Recently Paul, Dey and Gupta [42] also studied the singular four parameters Geometric Marshall Olkin bivariate Pareto distribution. However, there is no paper available in statistical inference on the absolute continuous Geometric Marshall-Olkin bivariate Pareto distribution. We also explore the Bayesian analysis under the informative prior. Both frequentist and Bayesian confidence intervals are provided along with an illustrative real-life data example.

The rest of the chapter is arranged as follows. In section-8.2, we show the formulation of the Marshall-Olkin bivariate Pareto distribution. Section-8.3 discusses the maximum likelihood estimation through the EM algorithm. Bayesian analysis is discussed in Section-8.4. Numerical results are shown in Section-8.5. Data analysis is shown for all methods in section-8.6. Finally, we conclude this chapter in Section-8.7.

8.2 Formulation of Block-Basu bivariate Pareto Geometric Distribution

8.2.1 Brief of singular Geometric bivariate Pareto Distribution

A bivariate random variable (X_1, X_2) is said to be distributed according to the Marshall-Olkin bivariate Pareto distribution, i.e., $(X_1, X_2) \sim MOBVPA(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$ if it has the cumulative survival function

$$
\begin{aligned}
S(x_1, x_2) &= (1+z)^{-\alpha_0}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2} \\
&= \begin{cases} S_1(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ S_2(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ S_0(x), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x, \end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
S_1(x_1, x_2) &= \Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1} \\
S_2(x_1, x_2) &= \Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1} \\
S_0(x) &= (1+x)^{-\alpha_0-\alpha_1-\alpha_2}.
\end{aligned}
$$

So it's joint pdf can be written as

$$
f(x_1, x_2) = \begin{cases} f_1(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1, x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2} = x, \end{cases} \tag{8.2.1}
$$

where

$$
\begin{aligned}
f_1(x_1, x_2) &= \frac{\alpha_1(\alpha_0+\alpha_2)}{\sigma_1\sigma_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_0-\alpha_2-1}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_1-1} \\
f_2(x_1, x_2) &= \frac{\alpha_2(\alpha_0+\alpha_1)}{\sigma_1\sigma_2}\Big(1+\frac{x_2-\mu_2}{\sigma_2}\Big)^{-\alpha_2-1}\Big(1+\frac{x_1-\mu_1}{\sigma_1}\Big)^{-\alpha_0-\alpha_1-1} \\
f_0(x) &= \alpha_0(1+x)^{-\alpha_0-\alpha_1-\alpha_2-1}.
\end{aligned}
$$

Suppose $\{(X_{1i}, X_{2i}) : i = 1, 2, \cdots\}$ is a sequence of iid bivariate random variables with same the cdf $F(\cdot,\cdot)$ and the pdf $f(\cdot,\cdot)$. N is an univariate random variable independent with (X_{1n}, X_{2n})'s follows the Geometric distribution with $0 < \theta \leq 1$. We also consider that (Y_1, Y_2) is an another bivariate random variable defined as,

$$Y_1 = \min\{X_{11}, X_{12}, \cdots, X_{1N}\}$$
$$Y_2 = \min\{X_{21}, X_{22}, \cdots, X_{2N}\}.$$

The joint cumulative survival function can be obtained as,

$$\bar{G}(y_1, y_2) = \frac{\theta S(y_1, y_2)}{1 - (1-\theta) S(y_1, y_2)}. \tag{8.2.2}$$

So it's a joint density function that can be presented as,

$$g(y_1, y_2) = \frac{\theta\{1 - (1-\theta)S(y_1, y_2)\}\frac{\partial^2 S(y_1,y_2)}{\partial y_1 \partial y_2} + 2\theta(1-\theta)\frac{\partial S(y_1,y_2)}{\partial y_1}\frac{\partial S(y_1,y_2)}{\partial y_2}}{\{1 - (1-\theta)S(y_1, y_2)\}^3}. \tag{8.2.3}$$

Now we use the survival function of the MOBVPA distribution in equation-8.2.3, then the joint distribution of (Y_1, Y_2) is called the Geometric Marshall-Olkin bivariate Pareto (G-MOBVPA) distribution. Therefore, the joint survival function of (Y_1, Y_2) can be written as,

$$\bar{G}(y_1, y_2) = \begin{cases} \bar{G}_1(y_1, y_2), & \text{if } \frac{y_1 - \mu_1}{\sigma_1} < \frac{y_2 - \mu_2}{\sigma_2} \\ \bar{G}_2(y_1, y_2), & \text{if } \frac{y_1 - \mu_1}{\sigma_1} > \frac{y_2 - \mu_2}{\sigma_2} \\ \bar{G}_0(y), & \text{if } \frac{y_1 - \mu_1}{\sigma_1} = \frac{y_2 - \mu_2}{\sigma_2}, \end{cases} \tag{8.2.4}$$

where

$$\bar{G}_1(y_1, y_2) = \frac{\theta\left(1 + \frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_1}\left(1 + \frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_2}}{1 - (1-\theta)\left(1 + \frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_1}\left(1 + \frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_2}}$$
$$\bar{G}_2(y_1, y_2) = \frac{\theta\left(1 + \frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1}\left(1 + \frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_2}}{1 - (1-\theta)\theta\left(1 + \frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1}\left(1 + \frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_2}}$$
$$\bar{G}_0(y) = \frac{\theta(1+y)^{-\alpha_0-\alpha_1-\alpha_2}}{1 - (1-\theta)(1+y)^{-\alpha_0-\alpha_1-\alpha_2}}.$$

Hence the joint pdf is,

$$g(y_1, y_2) = \begin{cases} g_1(y_1, y_2), & \text{if } \frac{y_1 - \mu_1}{\sigma_1} < \frac{y_2 - \mu_2}{\sigma_2} \\ g_2(y_1, y_2), & \text{if } \frac{y_1 - \mu_1}{\sigma_1} > \frac{y_2 - \mu_2}{\sigma_2} \\ g_0(y), & \text{if } \frac{y_1 - \mu_1}{\sigma_1} = \frac{y_2 - \mu_2}{\sigma_2} = y, \end{cases} \tag{8.2.5}$$

where

$$g_1(y_1, y_2) = \frac{\theta\alpha_1(\alpha_0 + \alpha_2)\left(1 + \frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_1-1}\left(1 + \frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_0-\alpha_2-1}\{1 + (1-\theta)S_1(y_1, y_2)\}}{\sigma_1\sigma_2\{1 - (1-\theta)S_1(y_1, y_2)\}^3}$$
$$g_2(y_1, y_2) = \frac{\theta(\alpha_0 + \alpha_1)\alpha_2\left(1 + \frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_0-\alpha_1-1}\left(1 + \frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_2-1}\{1 + (1-\theta)S_2(y_1, y_2)\}}{\sigma_1\sigma_2\{1 - (1-\theta)S_2(y_1, y_2)\}^3}$$

$$g_0(y) = \frac{\theta\alpha_0(1+y)^{-\alpha_0-\alpha_1-\alpha_2-1}}{\{1-(1-\theta)(1+y)^{-\alpha_0-\alpha_1-\alpha_2}\}^2}.$$

We denote this distribution as $G - MOBVPA(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2, \alpha_0, \alpha_1, \alpha_2)$. From the Lebesgue decomposition theorem, the joint pdf of (Y_1, Y_2) can be written as,

$$g(y_1, y_2) = \frac{\alpha_1+\alpha_2}{\alpha_0+\alpha_1+\alpha_2} g_{ac}(y_1, y_2) + \frac{\alpha_0}{\alpha_0+\alpha_1+\alpha_2} g_s(y), \tag{8.2.6}$$

where $g_{ac}(y_1, y_2)$ and $g_s(y_1, y_2)$ are the absolute continuous part and the singular part of the G-MOBVPA distribution. Here we are interested in the absolute continuous part only.

8.2.2 Absolute continuous Geometric Marshall-Olkin bivariate Pareto distribution (G-BBBVPA)

In this work, we are interested in parameter estimation of absolute continuous part only. We consider the case when $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$ and $\sigma_2 = 1$. We call the distribution as the four parameters Geometric Block-Basu bivariate Pareto distribution and denote this as $G - BBBVPA(\theta, \alpha_0, \alpha_1, \alpha_2)$. Then the joint density function of (Y_1, Y_2) becomes,

$$g(y_1, y_2) = \begin{cases} \frac{p\theta\alpha_1(\alpha_0+\alpha_2)(1+y_1)^{-\alpha_1-1}(1+y_2)^{-\alpha_0-\alpha_2-1}\{1+(1-\theta)(1+y_1)^{-\alpha_1}(1+y_2)^{-\alpha_0-\alpha_2}\}}{\{1-(1-\theta)(1+y_1)^{-\alpha_1}(1+y_2)^{-\alpha_0-\alpha_2}\}^3}, & \text{if } y_1 < y_2 \\ \frac{p\theta(\alpha_0+\alpha_1)\alpha_2(1+y_1)^{-\alpha_0-\alpha_1-1}(1+y_2)^{-\alpha_2-1}\{1+(1-\theta)(1+y_1)^{-\alpha_0-\alpha_1}(1+y_2)^{-\alpha_2}\}}{\{1-(1-\theta)(1+y_1)^{-\alpha_0-\alpha_1}(1+y_2)^{-\alpha_2}\}^3}, & \text{if } y_1 > y_2 \end{cases} \tag{8.2.7}$$

where $p = \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}$. The probability density plot and corresponding contour plot of different parameter sets are provided in Figure-8.1 and Figure-8.2 respectively.

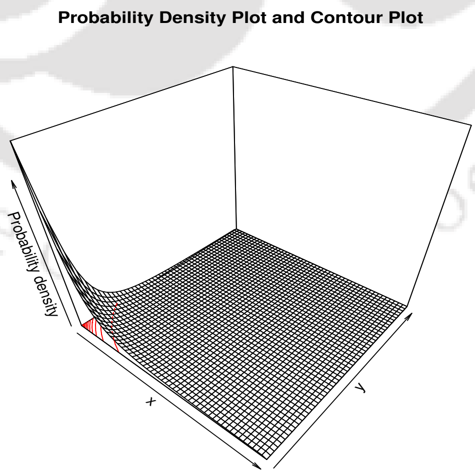


Figure 8.1: Probability density plot and Contour Plot of three parameter G-BBBVPA for the parameter values $\theta = 0.20$, $\alpha_0 = 0.10$, $\alpha_1 = 0.20$ and $\alpha_2 = 0.40$

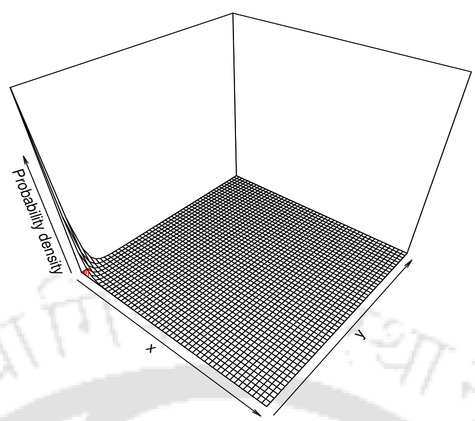


Figure 8.2: Probability density plot and Contour Plot of three parameter G-BBBVPA for the parameter values $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$ and $\alpha_2 = 10$

8.2.3 Marginal distributions

The marginal distributions are easy to obtain from the above bivariate distribution which is given as,

$$\bar{G}(y_1) = p[\frac{\theta(1+\frac{y_1-\mu_1}{\sigma_1})^{-\alpha_0-\alpha_1}}{1-(1-\theta)(1+\frac{y_1-\mu_1}{\sigma_1})^{-\alpha_0-\alpha_1}} - (1-\frac{1}{p})\frac{\theta(1+\frac{y_1-\mu_1}{\sigma_1})^{-(\alpha_0+\alpha_1+\alpha_2)}}{1-(1-\theta)(1+\frac{y_1-\mu_1}{\sigma_1})^{-(\alpha_0+\alpha_1+\alpha_2)}}],$$
$$\text{if } y_1 > \mu_1 \qquad (8.2.8)$$

$$\bar{G}(y_2) = p[\frac{\theta(1+\frac{y_2-\mu_2}{\sigma_2})^{-\alpha_0-\alpha_2}}{1-(1-\theta)(1+\frac{y_2-\mu_2}{\sigma_2})^{-\alpha_0-\alpha_2}} - (1-\frac{1}{p})\frac{\theta(1+\frac{y_2-\mu_2}{\sigma_2})^{-(\alpha_0+\alpha_1+\alpha_2)}}{1-(1-\theta)(1+\frac{y_2-\mu_2}{\sigma_2})^{-(\alpha_0+\alpha_1+\alpha_2)}}],$$
$$\text{if } y_2 > \mu_2 \qquad (8.2.9)$$

where $p = \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}$.

8.3 Parameter estimation through EM algorithm

Let us consider, $I = \{(y_{11}, y_{21}), (y_{12}, y_{22}), \cdots, (y_{1m}, y_{2m})\}$ is a sample of size m from the four parameters G-BBBVPA distribution. Let us also assume that μ_1, μ_2, σ_1 and σ_2 are known. Now we use the following notation:

$$I_1 = \{i : y_{1i} < y_{2i}\},\ I_2 = \{i : y_{1i} > y_{2i}\},$$

also $m_1 = |I_1|$, $m_2 = |I_2|$, $m = m_1 + m_2$.

We assume this data corresponds to a fictitious singular distribution, where cardinality of singular observation is m_0. We form the usual EM in singular set up first. Suppose we observe not only (Y_1, Y_2) but also the corresponding N value. Hence the complete data would be of the form,

$$\{(y_{11}, y_{21}, n_1), (y_{12}, y_{22}, n_2) \cdots, (y_{1m^*}, y_{2m^*}, n_{m^*})\},$$

here $m^* = m_0 + m_1 + m_2$.

We can imagine the following three independent hidden random variables corresponding to (Y_1, Y_2),

$$\{U_0|N = n\} \sim PA(II)(0, 1, n\alpha_0),$$

$$\{U_1|N = n\} \sim PA(II)(0, 1, n\alpha_1),$$

$$\{U_2|N = n\} \sim PA(II)(0, 1, n\alpha_2).$$

Also it is well known that

$$\{Y_1|N = n\} = \min\{\{U_0|N = n\}, \{U_1|N = n\}\},$$

$$\{Y_2|N = n\} = \min\{\{U_0|N = n\}, \{U_2|N = n\}\}.$$

Pseudo log-likelihood function described in ([42]) can be obtained as :

$$\begin{aligned}
Q &= L_{pseudo}(\theta, \alpha_0, \alpha_1, \alpha_2) \\
&= \Big(\sum_{i\in I_0} \ln(a_i) + 2\sum_{i\in I_1} \ln(a_i) + 2\sum_{i\in I_2} \ln(a_i)\Big) + (m_0 + u_1 m_1 + w_1 m_2)\ln(\alpha_0) \\
&+ (m_1 + w_2 m_2)\ln(\alpha_1) + (u_2 m_1 + m_2)\ln(\alpha_2) - \Big(\sum_{i\in I_0} \ln(1 + y_{1i}) + \sum_{i\in I_1\cup I_2} \ln(1 + y_{1i}) \\
&+ \sum_{i\in I_1\cup I_2} \ln(1 + y_{2i})\Big) - \alpha_0\Big(\sum_{i\in I_0} a_i \ln(1 + y_{1i}) + \sum_{i\in I_1} a_i \ln(1 + y_{2i}) + \sum_{i\in I_2} a_i \ln(1 + y_{1i})\Big) \\
&- \alpha_1\Big(\sum_{i\in I_0} a_i \ln(1 + y_{1i}) + \sum_{i\in I_1} a_i \ln(1 + y_{1i}) + \sum_{i\in I_2} a_i \ln(1 + y_{1i})\Big) - \alpha_2\Big(\sum_{i\in I_0} a_i \ln(1 + y_{1i}) \\
&+ \sum_{i\in I_1} a_i \ln(1 + y_{2i}) + \sum_{i\in I_2} a_i \ln(1 + y_{2i})\Big) + m^* \ln\Big(\frac{\theta}{1-\theta}\Big) + \ln(1-\theta)\Big(\sum_{i\in I_0} a_i + \sum_{i\in I_1} a_i \\
&+ \sum_{i\in I_2} a_i\Big).
\end{aligned} \tag{8.3.1}$$

The EM updates for the parameters θ, α_0, α_1 and α_2 are given as follows,

$$\hat{\alpha}_0^{(t+1)} = \frac{m_0 + u_1^{(t)} m_1 + w_1^{(t)} m_2}{\sum_{i\in I_0} a_i^{(t)} \ln(1 + y_{1i}) + \sum_{i\in I_1} a_i^{(t)} \ln(1 + y_{2i}) + \sum_{i\in I_2} a_i^{(t)} \ln(1 + y_{1i})}, \tag{8.3.2}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{m_1 + w_2^{(t)} m_2}{\sum_{i\in I_0} a_i^{(t)} \ln(1+y_{1i}) + \sum_{i\in I_1} a_i^{(t)} \ln(1+y_{1i}) + \sum_{i\in I_2} a_i^{(t)} \ln(1+y_{1i})}, \tag{8.3.3}$$

$$\hat{\alpha}_2^{(t+1)} = \frac{m_2 + u_2^{(t)} m_1}{\sum_{i\in I_0} a_i^{(t)} \ln(1+y_{1i}) + \sum_{i\in I_1} a_i^{(t)} \ln(1+y_{2i}) + \sum_{i\in I_2} a_i^{(t)} \ln(1+y_{2i})} \tag{8.3.4}$$

and

$$\hat{\theta}^{(t+1)} = \frac{m^*}{\sum_{i\in I_0} a_i^{(t)} + \sum_{i\in I_1} a_i^{(t)} + \sum_{i\in I_2} a_i^{(t)}}, \tag{8.3.5}$$

where $a_i^{(t)} = E[N|y_{1i}, y_{2i}]$ is the conditional mean of N given $Y_1 = y_1$ and $Y_2 = y_2$ at the step t and $u_1^{(t)}, u_2^{(t)}, w_1^{(t)}$ and $w_2^{(t)}$ are posterior probabilities at time step t.

Important Issues and Suggested Solutions :

We do not observe m_0 and each of the observation U_0 falling in I_0. We also do not know a_i when the observations are in I_0. One of the straightforward solution is to replace all the unknown quantities by its estimate $\tilde{m}_0$, $\tilde{b}_0$, $\tilde{a}_{0i}$. So we replace m_0 and each of the observation U_0 falling in I_0 by its estimates $\tilde{m}_0 = (m_1+m_2)\frac{\alpha_0}{\alpha_1+\alpha_2}$ and $\tilde{b}_0 = E[U_0|U_0 < \min\{U_1, U_2\}] = \frac{1}{n(\alpha_0+\alpha_1+\alpha_2)-1}$, where n falling in I_0 should be replaced by $\tilde{a}_{0i}$. **This method is valid when** $\tilde{a}_{0i}(\alpha_0+\alpha_1+\alpha_2) > 1$. To make this a valid method for any range of the parameters, we estimate $\log(1+U_0)$ by $\tilde{b}_0^* = E[\log(1+U_0)|\log(1+U_0) < \min\{\log(1+U_1), \log(1+U_2)\}] = \frac{1}{n(\alpha_0+\alpha_1+\alpha_2)}$ instead of estimating U_0. Hence the modified EM estimates for the parameters θ, α_0, α_1 and α_2 of the four parameters G-BBBVPA distribution will look like,

$$\tilde{a}_{0i}^{(t+1)} = \frac{1 + (1-\hat{\theta}^{(t)}) e^{-1/\tilde{a}_{0i}^{(t)}}}{1 - (1-\hat{\theta}^{(t)}) e^{-1/\tilde{a}_{0i}^{(t)}}}, \tag{8.3.6}$$

$$\tilde{b}_0^{*(t+1)} = \frac{1}{\tilde{a}_{0i}^{(t+1)} \left(\hat{\alpha}_0^{(t)} + \hat{\alpha}_1^{(t)} + \hat{\alpha}_2^{(t)}\right)}, \tag{8.3.7}$$

$$\hat{\alpha}_0^{(t+1)} = \frac{\tilde{m}_0 + u_1^{(t)} m_1 + w_1^{(t)} m_2}{\tilde{m}_0 \tilde{a}_{0i}^{(t+1)} \tilde{b}_0^{*(t+1)} + \sum_{i\in I_1} a_i^{(t)} \ln(1+y_{2i}) + \sum_{i\in I_2} a_i^{(t)} \ln(1+y_{1i})}, \tag{8.3.8}$$

$$\hat{\alpha}_1^{(t+1)} = \frac{m_1 + w_2^{(t)} m_2}{\tilde{m}_0 \tilde{a}_{0i}^{(t+1)} \tilde{b}_0^{*(t+1)} + \sum_{i\in I_1} a_i^{(t)} \ln(1+y_{1i}) + \sum_{i\in I_2} a_i^{(t)} \ln(1+y_{1i})}, \tag{8.3.9}$$

$$\hat{\alpha}_2^{(t+1)} = \frac{m_2 + u_2^{(t)} m_1}{\tilde{m}_0 \tilde{a}_{0i}^{(t+1)} \tilde{b}_0^{*(t+1)} + \sum_{i\in I_1} a_i^{(t)} \ln(1+y_{2i}) + \sum_{i\in I_2} a_i^{(t)} \ln(1+y_{2i})} \tag{8.3.10}$$

and

$$\hat{\theta}^{(t+1)} = \frac{\tilde{m}_0 + m_1 + m_2}{\tilde{m}_0 \tilde{a}_{0i}^{(t+1)} \tilde{b}_0^{*(t+1)} + \sum_{i\in I_1} a_i^{(t)} + \sum_{i\in I_2} a_i^{(t)}}. \tag{8.3.11}$$

Therefore the algorithmic steps of final working version can be given as :

Algorithm 8.1 Final modified EM procedure for four parameter Geometric absolute continuous Marshall Olkin bivariate Pareto distribution

1: Set initial α_0, α_1, α_2 and $\theta_{(i+1)}$ and a_i.
2: **while** $\Delta Q/Q < tol$ **do**
3: Compute $\tilde{a}_{0i}^{(t+1)}$ using $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$
4: Compute $\tilde{b}_{0i}^{*(t+1)}$ using $\tilde{a}_{0i}^{(t+1)}$, $\alpha_0^{(i)}$, $\alpha_1^{(i)}$ and $\alpha_2^{(i)}$.
5: Compute $u_1^{(i)}$, $u_2^{(i)}$, $w_1^{(i)}$, $w_2^{(i)}$, $\tilde{m}_0$ from $\alpha_0^{(i)}$, $\alpha_1^{(i)}$, $\alpha_2^{(i)}$ and $\tilde{a}_{0i}^{(t+1)}$.
6: Update $\alpha_0^{(i+1)}$, $\alpha_1^{(i+1)}$, $\alpha_2^{(i+1)}$ and $\theta^{(i+1)}$ using Equation (8.3.8), (8.3.9), (8.3.10) and (8.3.11).
7: Calculate Q for the new iterate.
8: **end while**

8.4 Bayesian Estimation

We use the slice cum Gibbs sampler technique to calculate the Bayes estimate. Usual step out is easy to implement in case of informative prior, which makes the Bayesian procedure also appealing for the practitioners. At each Gibbs sampling step we plan to use the slice sampling to generate the sample from the posterior.

8.4.1 Prior Assumption

We assume that α_0, α_1, and α_2 are distributed according to the gamma distribution with the shape parameters k_i and scale parameters θ_i, i.e.,

$$\begin{aligned}
\alpha_0 &\sim \Gamma(k_0, \theta_0) \equiv \text{Gamma}(k_0, \theta_0) \\
\alpha_1 &\sim \Gamma(k_1, \theta_1) \equiv \text{Gamma}(k_1, \theta_1) \\
\alpha_2 &\sim \Gamma(k_2, \theta_2) \equiv \text{Gamma}(k_2, \theta_2).
\end{aligned} \tag{8.4.1}$$

The probability density function of the Gamma Distribution is given by,

$$f_\Gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}},$$

where $\Gamma(k)$ is the gamma function evaluated at k.

Further, we assume that the geometric parameter θ is distributed according to the Beta distribution first kind with the parameters a and b

$$\pi(\theta) = \frac{1}{\beta(a,b)} \theta^{a-1}(1-\theta)^{b-1}, \quad \theta \in [0,1]. \tag{8.4.2}$$

8.4.2 Posterior Distribution

We know the Bayes estimate of an unknown parameter under the squared error loss function is the posterior mean of the corresponding parameter. But in this case, it is not easy to calculate Bayes estimate of unknown model parameters θ, α_0, α_1 and α_2 in closed form. We propose to use slice cum Gibbs sampling procedure to generates sample from the conditional posterior distribution. The full logarithm of conditional posterior distributions of θ, α_0, α_1 and α_2 are given as,

$$\begin{aligned}\ln(\pi(\theta \mid \alpha_0, \alpha_1, \alpha_2, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\beta(a, b)) + (a-1)\ln\theta \\ + (b-1)\ln(1-\theta),\end{aligned}$$

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \theta, \alpha_1, \alpha_2, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_0)) - k_0\ln(\theta_0) \\ + (k_0-1)\ln\alpha_0 - \frac{\alpha_0}{\theta_0},\end{aligned}$$

$$\begin{aligned}\ln(\pi(\alpha_1 \mid \theta, \alpha_0, \alpha_2, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_1)) - k_1\ln(\theta_1) \\ + (k_1-1)\ln\alpha_1 - \frac{\alpha_1}{\theta_1}\end{aligned}$$

and

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \theta, \alpha_0, \alpha_1, I)) = L(\theta, \alpha_0, \alpha_1, \alpha_2) - \ln(\Gamma(k_2)) - k_2\ln(\theta_2) \\ + (k_2-1)\ln\alpha_2 - \frac{\alpha_2}{\theta_2}\end{aligned}$$

respectively.

8.5 Numerical Analysis

We use the software R 3.5.0 to calculate all estimates. The codes are run at IIT Guwahati computers with model : Intel(R) Core(TM) i5-6200U CPU 2.30GHz. The codes will be available on request to the authors. Here we take two different sets of parameters with two different sample size as $n = 450$ and $n = 1000$ in both of the estimation technique.

EM Estimates: For the EM estimates, we generate sample from this distribution with different sample sizes ($n = 300, 450, 1000$) for two different parameter sets to calculate average estimates (AE), mean squared error (MSE) and 95% parametric bootstrap confidence interval (CI) of the parameters based on 1000 replications. We have also observed the average iteration number for each of the cases. We use different initial choice of the parameters for different parameter sets, but it remains same for different sample size in the EM algorithm. It is observed that all results are near about same with other initial choice within its usual range. We have used the stopping criterion as the relative difference of log-likelihood and pseudo log-likelihood of each of the iteration with stopping tolerance limit as 0.00001. All MLE results are shown in Table-8.1, Table-8.2, Table-8.3, Table-8.4, Table-8.5 and Table-8.6 respectively. The simulation provides average estimates closer to the original parameter with low MSEs which indicates that the method works quite well even for moderate sample size.

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.2253	0.0657	0.2325	0.4655
Mean Squared Error	0.0025	0.0032	0.0037	0.0125
Confidence Interval	[0.1417 0.3135]	[0.0002 0.1588]	[0.1433 0.3352]	[0.2958 0.6522]
Average Number of Iteration			363	

Table 8.1: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBPA with parameters $\theta = 0.20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 300$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.2230	0.0636	0.2319	0.4630
Mean Squared Error	0.0017	0.0027	0.0028	0.0096
Confidence Interval	[0.1624, 0.2925]	[0.0003, 0.1380]	[0.1525, 0.3198]	[0.3292, 0.6119]
Average Number of Iteration			349	

Table 8.2: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBPA with parameters $\theta = 0.20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 450$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.2205	0.0632	0.2302	0.4593
Mean Squared Error	0.0009	0.0020	0.0016	0.0057
Confidence Interval	[0.1845, 0.2634]	[0.0130, 0.1178]	[0.1820, 0.2820]	[0.3767, 0.5531]
Average Number of Iteration			303	

Table 8.3: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.4$ and sample size $n = 1000$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.8249	3.9399	5.1484	10.2878
Mean Squared Error	0.0089	1.6446	0.9181	2.0799
Confidence Interval	[0.6508, 0.9998]	[1.3485, 6.3114]	[3.4932, 7.3207]	[7.7573, 13.1848]
Average Number of Iteration			330	

Table 8.4: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 300$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.8236	3.9427	5.1358	10.2333
Mean Squared Error	0.0069	1.0905	0.6411	1.4749
Confidence Interval	[0.6780, 0.9894]	[1.8354, 5.8152]	[3.7078, 6.7657]	[7.9997, 12.6352]
Average Number of Iteration			302	

Table 8.5: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 450$

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.8212	3.9539	5.1086	10.2066
Mean Squared Error	0.0031	0.4898	0.2825	0.6507
Confidence Interval	[0.7231, 0.9285]	[2.6038, 5.3579]	[4.1389, 6.1406]	[8.7452, 11.7135]
Average Number of Iteration			266	

Table 8.6: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 1000$

For Bayesian Estimation: We compute the average Bayesian estimates (ABE) of unknown parameters and also the associated mean squared error (MSE), credible intervals (CI) and coverage probability (CP) of CIs based on the proper priors with fixed hyper-parameters. Although, the codes are done based on one particular set of hyper-parameters, it may work for any set of hyper-parameters. To step out method of the slice sampling technique, we choose our width as one. However, we cross check that the algorithm works for both larger and smaller choices of width. The confidence intervals are constructed directly using R package 'coda'. All Bayesian results are shown in Table-8.7, Table-8.8, Table-8.9 and Table-8.10 respectively. Results indicate that the method used in this case works really well, even for moderate sample size. We use the following hyper-parameters $a = 0.70$, $b = 0.75$, $k_0 = 0.70$, $\theta_0 = 0.75$, $k_1 = 0.70$, $\theta_1 = 0.75$, $k_2 = 0.70$ and $\theta_2 = 0.75$ for gamma priors. We have observed in a simulation experiment that the method works for multiple different choices of hyper-parameters indicating that the algorithm is independent of the choice of the hyper-parameters.

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.2279	0.0668	0.2364	0.4670
Mean Squared Error	0.0017	0.0020	0.0025	0.0084
Credible Interval	[0.1897 0.3265]	[0.0001 0.1607]	[0.1908 0.3614]	[0.3281 0.5755]
Coverage Probabilities	0.89	0.755	0.90	0.85

Table 8.7: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter G-BBBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.10$, $\alpha_1 = 0.20$, $\alpha_2 = 0.40$ and sample size $n = 450$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.2217	0.0604	0.2335	0.4634
Mean Squared Error	0.0009	0.0022	0.0017	0.0059
Credible Interval	[0.1567 0.0.2243]	[0.0001 0.0879]	[0.1164 0.1979]	[0.3680 0.5112]
Coverage Probabilities	0.90	0.80	0.84	0.85

Table 8.8: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter G-BBBVPA with parameters $\theta = 0.20$, $\alpha_0 = 0.10$, $\alpha_1 = 0.20$, $\alpha_2 = 0.40$ and sample size $n = 1000$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.7333	4.0535	4.4503	8.7605
Mean Squared Error	0.0093	0.9245	0.7420	2.2692
Credible Interval	[0.7219 0.9996]	[3.2085 7.2971]	[3.0594 5.8117]	[6.6314 11.0700]
Coverage Probabilities	0.83	0.84	0.78	0.73

Table 8.9: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter G-BBBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 450$

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.7844	4.0068	4.7777	9.5861
Mean Squared Error	0.0027	0.4754	0.2992	0.7041
Credible Interval	[0.6967 0.9068]	[3.7685 6.3895]	[3.1651 4.7518]	[7.6523 10.3458]
Coverage Probabilities	0.92	0.955	0.91	0.905

Table 8.10: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter G-BBBVPA with parameters $\theta = 0.80$, $\alpha_0 = 4$, $\alpha_1 = 5$, $\alpha_2 = 10$ and sample size $n = 1000$

8.6 Data Analysis

We use the same data set which is used by Paul, Dey and Kundu (2018)[41] for the data analysis of the three parameters BB-BVPA distribution. The data set is taken from `https://archive.ics.uci.edu/ml/machine-learning-databases`. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope. The data set contains related measurements. We extract a part of the data for bivariate modeling. We consider only measurements related to the female population, where two of the variables are "Shucked weight" and "Viscera weight". They represent the weight of meat and the gut weight (after bleeding) of abalones, which are measured in grams. We use peak over threshold method on this data set. It is observed that the transform data set does not have any singular component. So instead of modeling with BBBVPA distribution, we can plan to choose more generalized/flexible Geometric BBBVPA distribution, as one of the possible distributional assumption. We fit the empirical survival functions with the marginals of this distribution, whose parameters are obtained from the both of proposed EM algorithm and Bayesian estimation. For the both marginals it has a good fit, which are shown in Figure-8.4 and Figure-8.5 respectively, for the EM algorithms and the Bayesian estimation respectively. We also verify our assumption by plotting empirical two dimensional density plot of this data, which is shown in Figure-8.3.

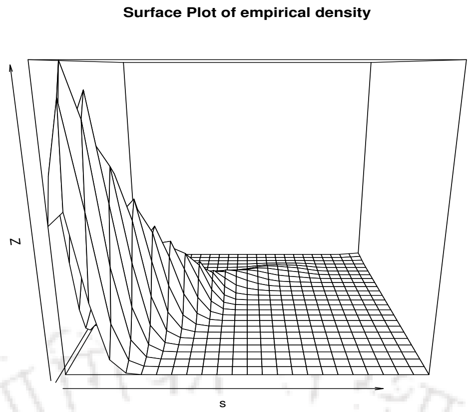
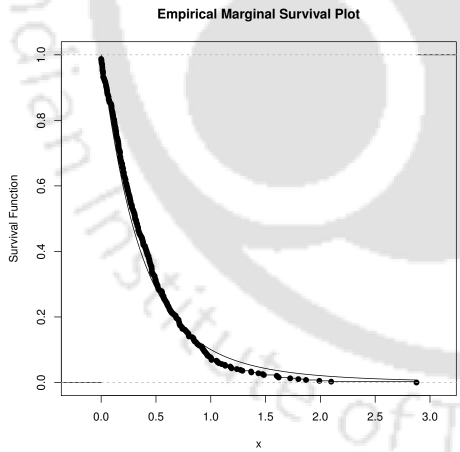
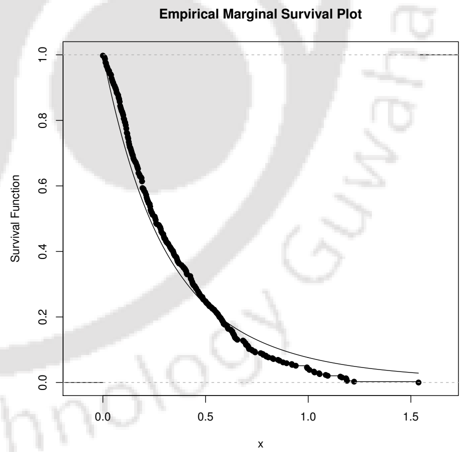


Figure 8.3: Two dimensional density plots for three parameter G-BBBVPA of the transformed dataset

For EM algorithm: The EM estimates of the parameters of four parameters G-BBBVPA based on transforming data are $\theta = 0.9999$, $\alpha_0 = 2.9845$, $\alpha_1 = 1.1265$ and $\alpha_2 = 1.4379$. Average estimates, mean square errors, confidence intervals and average number of iterations based on parametric bootstrap are available in Table-8.11.



(a) For X_1



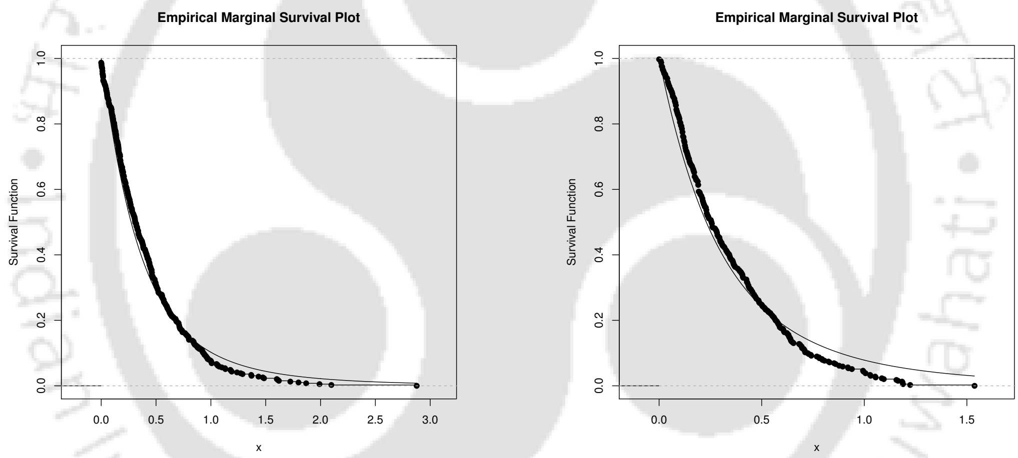
(b) For X_2

Figure 8.4: Survival plots for two marginals of three parameter G-BBBVPA through EM algorithm of the transformed dataset

EM Algorithm				
Parameter Sets	θ	α_0	α_1	α_2
Average Estimates	0.9616	2.9449	1.0989	1.4047
Mean Squared Error	0.0059	0.2520	0.0955	0.1455
Confidence Interval	[0.8196, 1.0000]	[2.0047, 3.9773]	[0.4981, 1.6922]	[0.6672, 2.1279]
Average Number of Iteration			1057	

Table 8.11: Average Estimates (AE), Mean Square Errors (MSE) and Confidence Intervals (CI) of four parameter G-BBBVPA with parameters $\theta = 0.9999$, $\alpha_0 = 2.9845$, $\alpha_1 = 1.1265$, $\alpha_2 = 1.4379$ and sample size $n = 391$

For Bayesian Estimation: The Bayesian estimates of the parameters based on this data are $\theta = 0.9877$, $\alpha_0 = 3.0056$, $\alpha_1 = 1.0733$ and $\alpha_2 = 1.3681$. We also calculate average Bayesian estimates, mean squared error, credible intervals and coverage probability of the credible intervals using the simulation technique which is available in Table-8.12.



(a) For X_1 (b) For X_2

Figure 8.5: Survival plots for two marginals of three parameter G-BBBVPA through Bayesian Estimation of the transformed dataset

Slice-cum-Gibbs				
Gamma Prior				
Parameter Sets	θ	α_0	α_1	α_2
Starting Value	0.4794	0.8654	0.7781	0.5386
Average Estimates	0.9202	2.8991	1.0237	1.2916
Mean Squared Error	0.0060	0.2123	0.0773	0.1194
Credible Interval	[0.8593, 1.0000]	[2.4865, 4.5771]	[0.1876, 1.4484]	[0.3136, 1.8329]
Coverage Probabilities	0.98	0.94	0.904	0.925

Table 8.12: Average Bayesian Estimates (ABE), Mean Square Errors (MSE), Credible Intervals (CI) and Coverage Probabilities (CP) of the parameters of four parameter G-BBBVPA with parameters $\theta = 0.9877$, $\alpha_0 = 3.0056$, $\alpha_1 = 1.0733$, $\alpha_2 = 1.3681$ and sample size $n = 391$

8.7 Conclusion

We observe that it is a more flexible model than the three parameter absolute continuous Marshall-Olkin bivariate Pareto distribution. We have proposed different modifications while implementing the EM algorithm. Bayesian analysis through the slice cum Gibbs sampler is used for estimation of the parameters too. Modeling the data through G-MOBVPA may have more appeal for the practitioner when there is no equality in components of the model. However, the statistical inference for this model is much more difficult with location and scale due to its discontinuous nature of the likelihood with respect to the parameters. A different alternative version of this Geometric distribution can also be proposed in the same direction. The work is in progress.

Chapter 9

Conclusion and Future Works

9.1 Conclusion

This thesis work mainly contains information related to the issues involved in parameter estimation through the EM algorithm of certain types of the Marshall Olkin bivariate Pareto distribution. It also covers the Bayesian estimation through slice sampling methods. In most of the cases, we see that finding confidence interval is a difficult task. Even the proposed procedures require more modification and separate attention to explore in greater details. Since, we have been making parametric modeling of bivariate data, data sets created in this case require theoretical justification of the distribution used in every case. However, such theoretical structure is not available or may not lead to the desired distributions. However, since the approach resembles very much like that of Peak over the threshold method of extreme value theory, we can guess that the transformed data should behave near to some bivariate Pareto type distribution. The empirical plots and the goodness of fit tests validate the same in each of the chapter. Therefore, the distributional assumptions made in this thesis is not surprising. In fact, it opens the door of few important problems in extreme value theory. The theoretical exploration on rate of the convergence of the constructed Markov chain and their mixing properties used in different the Bayesian analysis can be the another dimension of the future work. Some more aspects of the thesis are provided in the following future directions.

9.2 General Future Works

We point out some generic future works. However, each of the direction requires separate attention to various problems.

1. Statistical Inference of Bivariate Transmuted distributions. (*Work under progress*)

2. Statistical Inference of Higher dimensional distributions under censoring.

3. Discrimination of Higher dimensional distribution.

4. Role of Bayes factor and Bayesian discrimination.

5. Modeling multivariate data with the Copula.

Bibliography

[1] Barry C. Arnold. A note on multivariate distributions with specified marginals (Ref: V62 p30-44). *Journal of the American Statistical Association*, 62:1460–1461, 1967.

[2] Alexandru V Asimit, Edward Furman, and Raluca Vernic. On a multivariate Pareto distribution. *Insurance: Mathematics and Economics*, 46(2):308–316, 2010.

[3] Alexandru V Asimit, Edward Furman, and Raluca Vernic. Statistical inference for a new class of multivariate Pareto distributions. *Communications in Statistics-Simulation and Computation*, 45(2):456–471, 2016.

[4] Alexandru V Asimit, Raluca Vernic, and Riċardas Zitikis. Evaluating risk measures and capital allocations based on multi-losses driven by a heavy-tailed background risk: The multivariate Pareto-II model. *Risks*, 1(1):14–33, 2013.

[5] Yves F Atchadé. A computational framework for empirical Bayes inference. *Statistics and computing*, 21(4):463–473, 2011.

[6] Wagner Barreto-Souza. Bivariate Gamma-geometric law and its induced Lévy process. *Journal of Multivariate Analysis*, 109:130–145, 2012.

[7] James O Berger and José M Bernardo. On the development of the reference prior method. *Bayesian statistics*, 4:35–60, 1992.

[8] Jose M Bernardo. Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 113–147, 1979.

[9] Henry W Block and AP Basu. A continuous, bivariate exponential extension. *Journal of the American Statistical Association*, 69(348):1031–1037, 1974.

[10] George Casella and Edward I George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[11] Ming-Hui Chen and Qi-Man Shao. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 1999.

[12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[13] Arabin Kumar Dey and Debasis Kundu. Discriminating between the bivariate generalized exponential and bivariate Weibull distributions. *Chilean Journal of Statistics (ChJS)*, 3(1), 2012.

[14] Bradley Efron. *The jackknife, the bootstrap, and other resampling plans*, volume 38. Siam, 1982.

[15] Michael Falk and Armelle Guillou. Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis*, 99(4):715–734, 2008.

[16] Alec Stephenson Function fbvpot by Chris and Maintainer Alec Stephenson. Package evd. 2018.

[17] Rameshwar D Gupta and Debasis Kundu. Theory & methods: Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41(2):173–188, 1999.

[18] David D Hanagal. A multivariate Pareto distribution. *Communications in Statistics-Theory and Methods*, 25(7):1471–1488, 1996.

[19] Dimitris Karlis. ML estimation for multivariate shock models via an EM algorithm. *Annals of the Institute of Statistical Mathematics*, 55(4):817–830, 2003.

[20] Tomasz J Kozubowski and Anna K Panorska. A mixed bivariate distribution connected with geometric maxima of exponential variables. *Communications in StatisticsTheory and Methods*, 37(18):2903–2923, 2008.

[21] Debasis Kundu. On sarhan-balakrishnan bivariate distribution. *Journal of Statistics Applications & Probability*, 1(3):163, 2012.

[22] Debasis Kundu. Multivariate geometric skew-normal distribution. *Statistics*, 51(6):1377–1397, 2017.

[23] Debasis Kundu and Arabin Kumar Dey. Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm. *Computational Statistics & Data Analysis*, 53(4):956–965, 2009.

[24] Debasis Kundu and Arjun K Gupta. Bayes estimation for the Marshall–Olkin bivariate Weibull distribution. *Computational Statistics & Data Analysis*, 57(1):271–281, 2013.

[25] Debasis Kundu and Arjun K Gupta. On bivariate Weibull-geometric distribution. *Journal of Multivariate Analysis*, 123:19–29, 2014.

[26] Debasis Kundu and Rameshwar D Gupta. Generalized exponential distribution: Bayesian estimations. *Computational Statistics & Data Analysis*, 52(4):1873–1883, 2008.

[27] Debasis Kundu and Rameshwar D Gupta. Bivariate generalized exponential distribution. *Journal of Multivariate Analysis*, 100(4):581–593, 2009.

[28] Debasis Kundu and Rameshwar D Gupta. A class of absolutely continuous bivariate distributions. *Statistical Methodology*, 7(4):464–477, 2010.

[29] Debasis Kundu and Rameshwar D Gupta. Absolute continuous bivariate generalized exponential distribution. *AStA Advances in Statistical Analysis*, 95(2):169–185, 2011.

[30] Debasis Kundu, Ankush Kumar, and Arjun K Gupta. Absolute continuous multivariate generalized exponential distribution. *Sankhya B*, 77(2):175–206, 2015.

[31] Debasis Kundu and Vahid Nekoukhou. Univariate and bivariate geometric discrete generalized exponential distributions. *Journal of Statistical Theory and Practice*, pages 1–20, 2018.

[32] Debasis Kundu and Biswabrata Pradhan. Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring. *Communications in StatisticsTheory and Methods*, 38(12):2030–2041, 2009.

[33] Dennis V Lindley. Approximate bayesian methods. *Trabajos de estadística y de investigación operativa*, 31(1):223–245, 1980.

[34] Albert W Marshall. Copulas, marginals, and joint distributions. *Lecture Notes-Monograph Series*, pages 213–222, 1996.

[35] Albert W Marshall and Ingram Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44, 1967.

[36] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[37] Seyed Mohsen Mirhosseini, Mohammad Amini, D Kundu, and Ali Dolati. On a new absolutely continuous bivariate generalized exponential distribution. *Statistical Methods & Applications*, 24(1):61–83, 2015.

[38] Radford M Neal. Slice sampling. *Annals of statistics*, pages 705–741, 2003.

[39] Roger B Nelsen. *An introduction to copulas.* Springer Science & Business Media, 2007.

[40] Biplab Paul and Arabin Dey. Some variations of EM algorithms for Marshall-Olkin bivariate Pareto distribution with location and scale. *Journal of Statistical Theory and Practice*, 2018.

[41] Biplab Paul, Arabin Dey, and Debasis Kundu. Bayesian analysis of three parameter absolute continuous Marshall-Olkin bivariate Pareto distribution. *Communications in Statistics - Case Studies and Data Analysis*, 2018.

[42] Biplab Paul, Arabin Kumar Dey, and Arjun K Gupta. Parameter Estimation of four parameter Geometric Marshall-Olkin bivariate Pareto Distribution. *(under review)*, 2018.

[43] Biplab Paul, Arabin Kumar Dey, and Debasis Kundu. Em algorithm for absolute continuous bivariate Pareto distribution. *(under review)*, 2018.

[44] Edsel A Pena and Arjun K Gupta. Bayes estimation for the Marshall-Olkin exponential distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 379–389, 1990.

[45] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Package coda. 2016.

[46] Biswabrata Pradhan and Debasis Kundu. Bayes estimation for the Block and Basu bivariate and multivariate Weibull distributions. *Journal of Statistical Computation and Simulation*, 86(1):170–182, 2016.

[47] Pál Rakonczai and András Zempléni. Bivariate generalized Pareto distribution in practice: models and estimation. *Environmetrics*, 23(3):219–227, 2012.

[48] Miroslav M Ristić and Debasis Kundu. Marshall-Olkin generalized exponential distribution. *Metron*, 73(3):317–333, 2015.

[49] Christian Robert and George Casella. *Monte Carlo statistical methods.* Springer Science & Business Media, 2013.

[50] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[51] PG Sankaran and Debasis Kundu. A bivariate Pareto model. *Statistics*, 48(2):241–255, 2014.

[52] Ammar M Sarhan and N Balakrishnan. A new class of bivariate distributions and its mixture. *Journal of Multivariate Analysis*, 98(7):1508–1527, 2007.

[53] Richard L Smith. Multivariate threshold methods. In *Extreme Value Theory and Applications*, pages 225–248. Springer, 1994.

[54] Alec G Stephenson. evd: Extreme value distributions. *R News*, 2(2):31–32, 2002.

[55] Hsiaw-Chan Yeh. Two multivariate Pareto distributions and their related inferences. *Bulletin of the Institute of Mathematics, Academia Sinica.*, 28(2):71–86, 2000.

[56] Hsiaw-Chan Yeh. Some properties and characterizations for generalized multivariate Pareto distributions. *Journal of Multivariate Analysis*, 88(1):47–60, 2004.

List of Accepted or Communicated Papers

Based on the work in this thesis, the following research articles have been accepted or communicated.

1. Biplab Paul, Arabin Kumar Dey, Some variations of EM algorithms for estimating parameters of singular Bivariate Marshall-Olkin Pareto distribution, *Journal of Statistical Theory and Practice* (*accepted, Doi number : 10.1007/s42519-018-0009-4*)

2. Biplab Paul, Arabin Kumar Dey, Debasis Kundu, An EM algorithm for absolutely continuous Marshall-Olkin bivariate Pareto distribution with location and scale. (*under review*)

3. Biplab Paul, Arabin Kumar Dey, Debasis Kundu, Bayesian analysis of three parameter absolute continuous Marshall-Olkin bivariate Pareto distribution, *Communications in Statistics Case Studies and Data Analysis* (*accepted, Doi number: 10.1080/23737484.2018.1482479*)

4. Biplab Paul, Arabin Kumar Dey, Bayesian estimation of Marshal-Olkin bivariate Pareto Distribution with location and scale parameters. (*under review*)

5. Biplab Paul, Arabin Kumar Dey, Sanku Dey, Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters. (*under review*)

6. Biplab Paul, Arabin Kumar Dey, Arjun K. Gupta, Parameter Estimation of Four Parameter Geometric Marshall-Olkin Bivariate Pareto Distribution. (*under review*)

7. Biplab Paul, Arabin Kumar Dey, Arjun K. Gupta, Debasis Kundu, Parameter Estimation of Four Parameter Geometric Block-Basu Bivariate Pareto Distribution. (*under review*)