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

Biplab Paul, Arabin Kumar Dey and Debasis Kundu

Department of Mathematics, IIT Guwahati

Department of Mathematics, IIT Guwahati

Department of Mathematics and Statistics, IIT Kanpur

February 17, 2018

Abstract

This paper provides Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution. We consider only three parameter for this Marshall-Olkin bivariate Pareto distribution. We take two types of prior - reference prior and gamma prior for our analysis. Bayesian estimate of the parameters are calculated based on slice cum gibbs sampler and Lindley approximation. Credible intervals are also provided for all methods and all prior distributions. A real-life data analysis is shown for illustrative purpose.

Keywords: Bivariate Pareto distribution; Absolute continuous bivariate distribution; Slice sampling; Lindley Approximation.

1 Introduction

In this paper we consider the Bayesian analysis of absolute continuous version of 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). Finding efficient estimation technique to estimate the parameters of BB-BVPA was a major challenge in last few decades. Parameter estimation by EM algorithm for absolute continuous BVPA is also available in a recent work by Dey and Kundu (2017). There is no work in Bayesian set up for BB-BVPA. There are very less work using reference prior as posterior becomes extremely complicated while dealing with such complicated prior. We make an easy formulation for the same through Slice cum Gibbs sampler method which is not available in the literature. The formulation makes sense as Slice is also a type of Gibbs sampler. In this paper we restrict ourselves only up to three parameter BBBVPA. BBBVPA can be very useful in modelling data related to finance, climate, network-security etc. This is one of the higher dimensional distribution which is heavy-tail in nature. A variety of bivariate (multivariate) extensions of bivariate Pareto distribution also have been studied in the literature. These include the distributions described in the following works : Sankaran and Kundu (2014), Yeh (2000), Yeh (2004), Asimit et al. (2010), Asimit et al. (2016).

Bayesian analysis contains more information than maximum likelihood estimation, which is better for further statistical analysis. For instance, not only the mean, mode, or median 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 in Bayesian set up with such a complicated distribution has its own advantages. In this paper 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 [Lindley (1980)] (2) Slice cum Gibbs Sampler [Neal (2003), Casella and George (1992)]. However we can use other Monte Carlo methods for the same. In this paper we made slight modification in calculation of the Lindley approximation. We use EM algorithms instead of MLE. We also calculate credible intervals for the parameters.

Rest of the paper is organized as follows. Section 2 is kept for Bayesian analysis of BB-BVPA. Section 3 deals with the construction of credible interval. Numerical results are discussed in section 4. Data analysis is shown in section 5. We conclude the paper in section 6.

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

BB-BVPA is introduced and parameter estimation through EM algorithm is available at Dey and Kundu (2017). In this section we discuss Bayesian estimate of three parameter absolute continuous bivariate Pareto through Lindley and Slice cum Gibbs sampler technique. Lindley approximation is done based on EM estimates instead of direct MLE.

Three parameter MOBVPA bivariate survival function as we derived in previous section,

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

The above expression can be decomposed as follows,

$$S_{X_1,X_2}(x_1,x_2) = pS_{ac}(x_1,x_2) + (1-p)S_s(x_1,x_2) \tag{2}$$

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,

$$S_s(x_1,x_2) = (1+z)^{-(\alpha_0+\alpha_1+\alpha_2)} \tag{3}$$

when $x_1 = x_2 = z$.

$$\frac{\partial^2}{\partial x_1 \partial x_2} S_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2) & \text{if } x_1 < x_2 \\ f_2(x_1,x_2) & \text{if } x_1 > x_2 \end{cases} \tag{4}$$

Therefore,

$$f_1(x_1,x_2) = \alpha_1(\alpha_0+\alpha_2)(1+x_1)^{-(\alpha_1+1)}(1+x_2)^{-(\alpha_0+\alpha_2+1)} \tag{5}$$

$$f_2(x_1,x_2) = \alpha_2(\alpha_0+\alpha_1)(1+x_1)^{-(\alpha_0+\alpha_1+1)}(1+x_2)^{-(\alpha_2+1)} \tag{6}$$

and,

$$\iint_{x_1<x_2} f_1(x_1,x_2)dx_1dx_2 + \iint_{x_1>x_2} f_2(x_1,x_2)dx_1dx_2 = p \tag{7}$$

Consider,

$$\iint_{x_1<x_2} f_1(x_1,x_2)dx_1dx_2 \tag{8}$$

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

$$= \frac{\alpha_1}{\alpha_0+\alpha_1+\alpha_2} \tag{9}$$

Similarly,

$$\iint_{x_1>x_2} f_2(x_1,x_2)dx_1dx_2 = \frac{\alpha_2}{\alpha_0+\alpha_1+\alpha_2} \tag{10}$$

Therefore, from (7), we get,

$$\frac{\alpha_1+\alpha_2}{\alpha_0+\alpha_1+\alpha_2} = p \tag{11}$$

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

$$f_{BVPAC}(x_1, x_2) = \begin{cases} cf_1(x_1, x_2) & \text{if } x_1 < x_2 \\ cf_2(x_1, x_2) & \text{if } x_1 > x_2 \end{cases} \tag{12}$$

where c is a normalizing constant and

$$c = \frac{1}{p} = \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_1 + \alpha_2} \tag{13}$$

2.1 Likelihood Function

The likelihood function corresponding to this pdf is given by,

$$\begin{aligned} l(x_1, x_2; \alpha_0, \alpha_1, \alpha_2) &= \prod_i^n (cf_1(x_1, x_2))^{\mathrm{I}_{(x_1,x_2)\in I_1}} \prod_i^n (cf_2(x_1, x_2))^{\mathrm{I}_{(x_1,x_2)\in I_2}} \\ &= c^{n_1+n_2} \alpha_1^{n_1} (\alpha_0 + \alpha_2)^{n_1} \alpha_2^{n_2} (\alpha_0 + \alpha_1)^{n_2} \\ & \quad \prod_{i \in I_1} (1 + x_{1i})^{-(\alpha_1+1)} (1 + x_{2i})^{-(\alpha_0+\alpha_2+1)} \\ &\times \prod_{i \in I_2} (1 + x_{1i})^{-(\alpha_0+\alpha_1+1)} (1 + x_{2i})^{-(\alpha_2+1)} \end{aligned} \tag{14}$$

where $I_1 = \{(x_1, x_2) \mid x_1 < x_2\}$ and $I_2 = \{(x_1, x_2) \mid x_1 > x_2\}$ and $|I_1| = n_1$, $|I_2| = n_2$, also $n_1 + n_2 = n$.

Therefore log-likelihood function takes the form,

$$\begin{aligned} L(x_1, x_2; \alpha_0, \alpha_1, \alpha_2) &= (n_1 + n_2) \ln(\alpha_0 + \alpha_1 + \alpha_2) - (n_1 + n_2) \ln(\alpha_1 + \alpha_2) \\ &+ n_1 \ln \alpha_1 - (\alpha_1 + 1) \sum_{i \in I_1} \ln(1 + x_{1i}) + n_1 \ln(\alpha_0 + \alpha_2) \\ &- (\alpha_0 + \alpha_2 + 1) \sum_{i \in I_1} \ln(1 + x_{2i}) + n_2 \ln(\alpha_0 + \alpha_1) \\ &- (\alpha_0 + \alpha_1 + 1) \sum_{i \in I_2} \ln(1 + x_{1i}) + n_2 \ln(\alpha_2) \\ &- (\alpha_2 + 1) \sum_{i \in I_2} \ln(1 + x_{2i}) \end{aligned} \tag{15}$$

2.2 Prior Assumption

2.2.1 Gamma Prior

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, i.e.,

$$\begin{aligned} \alpha_0 &\sim \Gamma(k_0, \theta_0) \equiv \text{Gamma}(k_0, \theta_0) \\ \alpha_1 &\sim \Gamma(k_1, \theta_1) \equiv \text{Gamma}(k_1, \theta_1) \\ \alpha_2 &\sim \Gamma(k_2, \theta_2) \equiv \text{Gamma}(k_2, \theta_2) \end{aligned} \tag{16}$$

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

$$f_\Gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \tag{17}$$

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

2.2.2 Reference Prior Assumption

We have taken reference prior as prior for each parameter conditional on the others. We calculate the expression using Bernardo's reference Prior [Berger et al. (1992), Bernardo (1979)] 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 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

$$\pi(\theta_2|\theta_1) = |I_{22}(\theta)|^{\frac{1}{2}} c(\theta_1) \tag{18}$$

, 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

$$E(\log \frac{\pi(\theta|x)}{\pi(\theta)}) = \frac{p}{2} \log n - \frac{p}{2} \log(2\pi e) + \int \pi(\theta) \log \frac{|I(\theta)|^{\frac{1}{2}}}{\pi(\theta)} d\theta + O(n^{-\frac{1}{2}}). \tag{19}$$

Similarly,

$$E(\log \frac{\pi(\theta_2|\theta_1, x)}{\pi(\theta_2|\theta_1)}) = \frac{p_2}{2} \log n - \frac{p_2}{2} \log(2\pi e) + \int \pi(\theta) \log \frac{|I_{22}(\theta)|^{\frac{1}{2}}}{\pi(\theta_2|\theta_1)} d\theta + O(n^{-\frac{1}{2}}). \tag{20}$$

From (19) and (20), we find

$$E(\log \frac{\pi(\theta_1|x)}{\pi(\theta_1)}) = \frac{p_1}{2} \log n - \frac{p_1}{2} \log 2\pi e + \int \pi(\theta) \log \frac{|I_{11.2}(\theta)|^{\frac{1}{2}}}{\pi(\theta_1)} d\theta + O(n^{-\frac{1}{2}}) \tag{21}$$

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 (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 integral above, we choose $\pi(\theta_1) = \psi(\theta_1)$. Note that $I^{-1}_{11.2}(\theta) = I_{11}(\theta)$ where

$$I^{-1}(\theta) = \begin{pmatrix} I_{11}(\theta) & -I_{11}(\theta)I^{-1}_{12}(\theta)I_{22}(\theta) \\ -I_{22}(\theta)I^{-1}_{21}(\theta)I_{11}(\theta) & I_{22}(\theta) \end{pmatrix}$$

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 expession of 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 :

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

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

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

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

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,

$$\begin{aligned}
\pi(\alpha_0,\alpha_1,\alpha_2|D_2) \propto \;& l(\alpha_0,\alpha_1,\alpha_2|D_2)\pi(\alpha_0,\alpha_1,\alpha_2|a,b,a_0,a_1,a_2) \\
=&\left(\frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_1+\alpha_2}\right)^n \alpha_1^{n_1}\alpha_2^{n_2}(\alpha_0+\alpha_1)^{n_2}(\alpha_0+\alpha_2)^{n_1} \\
&\prod_{i\in I_1}(1+x_{1i})^{-\alpha_1-1}(1+x_{2i})^{-\alpha_0-\alpha_2-1} \\
&\prod_{i\in I_2}(1+x_{1i})^{-\alpha_0-\alpha_1-1}(1+x_{2i})^{-\alpha_2-1} \\
&\times \alpha_0^{k_0-1}\alpha_1^{k_1-1}\alpha_2^{k_2-1} \\
&e^{-(\frac{\alpha_0}{\theta_0}+\frac{\alpha_1}{\theta_1}+\frac{\alpha_2}{\theta_2})} \\
=&\pi_1(\alpha_0,\alpha_1,\alpha_2|D_2) \quad \text{(say)}
\end{aligned} \tag{22}$$

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.

$$\hat{g} = \frac{\int_0^\infty \int_0^\infty \int_0^\infty g(\alpha_0, \alpha_1, \alpha_2)\pi_1(\alpha_0, \alpha_1, \alpha_2|D_2)d\alpha_0 d\alpha_1 d\alpha_2}{\int_0^\infty \int_0^\infty \int_0^\infty \pi_1(\alpha_0, \alpha_1, \alpha_2|D_2)d\alpha_0 d\alpha_1 d\alpha_2} \tag{23}$$

2.3.1 Lindley Approximation in BB-BVPA

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

$$\hat{g}_B = \frac{\int_{(\alpha_0,\alpha_1,\alpha_2)} g(\alpha_0, \alpha_1, \alpha_2)e^{[L(\alpha_0,\alpha_1,\alpha_2)+\rho(\alpha_0,\alpha_1,\alpha_2)]}d(\alpha_0, \alpha_1, \alpha_2)}{\int_{(\alpha_0,\alpha_1,\alpha_2)} e^{[L(\alpha_0,\alpha_1,\alpha_2)+\rho(\alpha_0,\alpha_1,\alpha_2)]}d(\alpha_0, \alpha_1, \alpha_2)} \tag{24}$$

where $L(\alpha_0, \alpha_1, \alpha_2)$ is log-likelihood function and $\rho(\alpha_0, \alpha_1, \alpha_2)$ is logarithm of 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 (24), 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.

$$\begin{aligned}b_i =& \rho_0\sigma_{i0} + \rho_1\sigma_{i1} + \rho_2\sigma_{i2}, \quad i = 0, 1, 2 \\ b_3 =& g_{01}\sigma_{01} + g_{02}\sigma_{02} + g_{12}\sigma_{12} \\ b_4 =& \frac{1}{2}(g_{00}\sigma_{00} + g_{11}\sigma_{11} + g_{22}\sigma_{22}) \\ A =& \sigma_{00}L_{000} + 2\sigma_{01}L_{010} + 2\sigma_{02}L_{020} + 2\sigma_{12}L_{120} + \sigma_{11}L_{110} + \sigma_{22}L_{220} \\ B =& \sigma_{00}L_{001} + 2\sigma_{01}L_{011} + 2\sigma_{02}L_{021} + 2\sigma_{12}L_{121} + \sigma_{11}L_{111} + \sigma_{22}L_{221} \\ C =& \sigma_{00}L_{002} + 2\sigma_{01}L_{012} + 2\sigma_{02}L_{022} + 2\sigma_{12}L_{122} + \sigma_{11}L_{113} + \sigma_{22}L_{222}\end{aligned}$$

Also

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

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

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

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 already mentioned the log-likelihood function for BB-BVPA 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 BBBVPA or absolute continuous bivariate Pareto distribution are $\hat{\alpha}_0, \hat{\alpha}_1$ and $\hat{\alpha}_2$. Now $\rho = \log \pi(\alpha_0, \alpha_1, \alpha_2)$ then $\rho_0 = \frac{k_0-1}{\alpha_0} - \frac{1}{\theta_0}$,

$\rho_1 = \frac{k_1-1}{\alpha_1} - \frac{1}{\theta_1}$, $\rho_2 = \frac{k_2-1}{\alpha_2} - \frac{1}{\theta_2}$. 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,

$$\hat{\alpha}_{0B} = \hat{\alpha}_0 + b_0 + \frac{1}{2}[A\sigma_{00} + B\sigma_{10} + C\sigma_{20}] \tag{25}$$

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

$$\hat{\alpha}_{1B} = \hat{\alpha}_1 + b_1 + \frac{1}{2}[A\sigma_{01} + B\sigma_{11} + C\sigma_{21}] \tag{26}$$

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

$$\hat{\alpha}_{2B} = \hat{\alpha}_2 + b_2 + \frac{1}{2}[A\sigma_{02} + B\sigma_{12} + C\sigma_{22}] \tag{27}$$

Remark : In this calculation we replace MLEs of parameters by parameters obtained through EM algorithm (Dey and Paul (2017)).

2.3.2 The Full Conditional distributions

To obtain the Bayesian estimates through slice cum gibbs sampler we need conditional distribution of each parameter given the other parameters and the data. We provide the expressions for logarithm of those conditional distributions for both reference and gamma prior. We propose to perform standard stepout method in slice sampling (Neal (2003)) on each 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 reference prior as reference prior is not defined in certain interval for some set of parameter values. As a result 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 nor 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 different other techniques like importance sampling, HMC etc taking Gamma as a prior. However we avoid the calculation of full posterior using reference prior as it is much easier to express conditional posterior distribution of each parameter given the other parameters and the data.

2.3.3 The full log-conditional distributions in Gamma prior

The log full conditional distributions of α_0, α_1, and α_2 are given by,

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) + n_1\ln(\alpha_0+\alpha_2) \\ &- \alpha_0 \sum_{i \in I_1} \ln(1+x_{2i}) + n_2 \ln(\alpha_0+\alpha_1) \\ &- \alpha_0 \sum_{i \in I_2} \ln(1+x_{1i}) + (k_0-1)\ln\alpha_0 - \frac{\alpha_0}{\theta_0} \qquad (28)\end{aligned}$$

$$\begin{aligned}\ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1\ln\alpha_1 - \alpha_1 \sum_{i \in I_1} \ln(1+x_{1i}) + n_2 \ln(\alpha_0+\alpha_1) \\ &- \alpha_1 \sum_{i \in I_2} \ln(1+x_{1i}) + (k_1-1)\ln\alpha_1 - \frac{\alpha_1}{\theta_1} \qquad (29)\end{aligned}$$

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \alpha_0, \alpha_1, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1\ln(\alpha_0+\alpha_2) - \alpha_2 \sum_{i \in I_1} \ln(1+x_{2i}) + n_2 \ln\alpha_2 \\ &- \alpha_2 \sum_{i \in I_2} \ln(1+x_{2i}) + (k_2-1)\ln\alpha_2 - \frac{\alpha_2}{\theta_2} \qquad (30)\end{aligned}$$

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

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

$$\begin{aligned}\ln(\pi(\alpha_0 \mid \alpha_1, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) + n_1\ln(\alpha_0+\alpha_2) \\ &- \alpha_0 \sum_{i\in I_1} \ln(1+x_{2i}) + n_2 \ln(\alpha_0+\alpha_1) \\ &- \alpha_0 \sum_{i\in I_2} \ln(1+x_{1i}) + \log(R_1); \end{aligned} \tag{31}$$

$$\begin{aligned}\ln(\pi(\alpha_1 \mid \alpha_0, \alpha_2, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1 \ln\alpha_1 - \alpha_1 \sum_{i\in I_1} \ln(1+x_{1i}) + n_2 \ln(\alpha_0+\alpha_1) \\ &- \alpha_1 \sum_{i\in I_2} \ln(1+x_{1i}) + \log(R_2); \end{aligned} \tag{32}$$

$$\begin{aligned}\ln(\pi(\alpha_2 \mid \alpha_0, \alpha_1, x_1, x_2)) &= (n_1+n_2)\ln(\alpha_0+\alpha_1+\alpha_2) - (n_1+n_2)\ln(\alpha_1+\alpha_2) \\ &+ n_1 \ln(\alpha_0+\alpha_2) - \alpha_2 \sum_{i\in I_1} \ln(1+x_{2i}) + n_2 \ln\alpha_2 \\ &- \alpha_2 \sum_{i\in I_2} \ln(1+x_{2i}) + \log(R_3); \end{aligned} \tag{33}$$

3 Constructing credible Intervals for $\underline{\theta}$

We find the credible intervals for parameters as described by Chen and Shao (1999). Let assume $\underline{\theta}$ is vector. To obtain 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 full posterior is not known and tractable. In this paper we calculate the Bayesian confidence interval for both gamma prior and reference prior. We skip working with full expression of posterior under reference prior as it is not tractable. We use R package coda to obtain the credible intervals described above.

4 Numerical Results

The numerical results are obtained by using package 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 authors.

We use the following hyper parameters of prior as gamma : $k_0 = 2,\ \theta_0 = 3,\ k_1 = 4,\ \theta_1 = 3,\ k_2 = 3,\ \theta_2 = 2$. 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 does not depend much on choice of hyper parameters. Just to verify the fact we have shown some results taking gamma as prior in Table-6. Table-4 and Table-5 show results obtained by different methods, e.g. Lindley and Slice sampling etc for absolute continuous Marshall-Olkin bivariate Pareto distribution with two different parameter sets. In slice cum gibbs sampling we take burn in period as 500. Bayesian estimates are calculated based on 2000 iterations after burn-in period. We made further investigation on sample size needed for all the methods to work. In case of BB-BVPA, slice cum gibbs with gamma as prior demands 250 or more sample 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 case of reference prior. 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. Mean square error is calculated based on 200 different samples drawn from 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 posterior based on reference prior as posterior in this case becomes a discontinuous function within specific ranges of parameter values. We can easily see Figure-1. In this Figure-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.

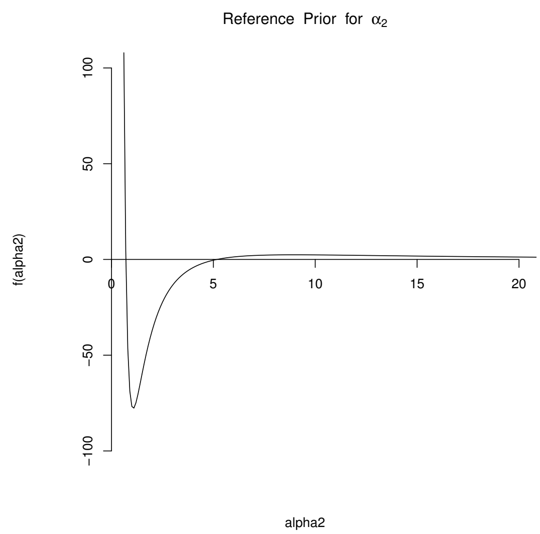


Figure 1: Square of the prior with respect to α_2 when α_1 and α_0 is given

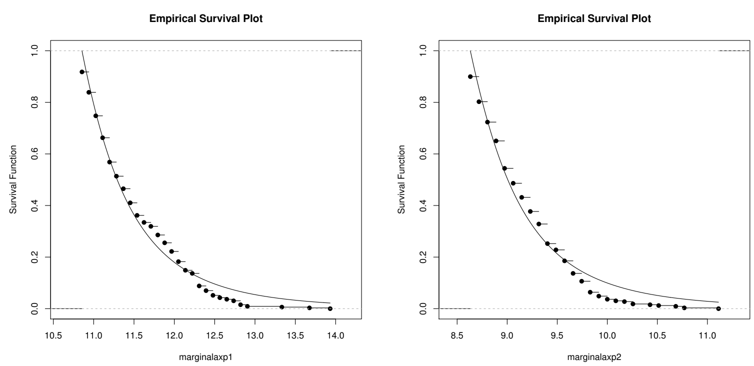
5 Data Analysis

We study one particular data sets which is used by Dey and Kundu (2017). 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 female population where one of the variable is Length as Longest shell measurement and other variable is Diameter which is perpendicular to length. We use peak over threshold method on this data set.

From Falk and Guillou (2008), we know that peak over threshold method on random variable U 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 appropriate t and x_0 so that data should behave more like Pareto distribution. The transformed data set does not have any singular component. Therefore one possible assumption can be absolute continuous Marshall Olkin bivariate Pareto.

These data set are used to model seven parameter BB-BVPA. EM estimates for BB-BVPA produces the values as $\mu_1 = 10.855$, $\mu_2 = 8.632$, $\sigma_1 = 2.124$, $\sigma_2 = 1.7110$, $\alpha_0 = 3.124$, $\alpha_1 = 1.743$, $\alpha_2 = 1.602$. Figure-2 shows that the empirical marginals coincide with the marginals calculated from the estimated parameters.

We also verify our assumption by plotting empirical two dimensional density plot in Figure-3 which resembles closer to the surface of Marshall-Olkin bivariate Pareto distribution.



(a) ξ_1

(b) ξ_2

Figure 2: Survival plots for two marginals of the transformed dataset

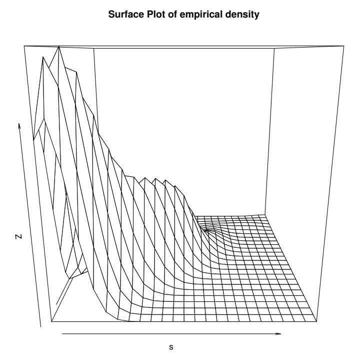


Figure 3: Two dimensional density plots for the transformed dataset

Direct real life data which will model three parameter BB-BVPA is not available. Therefore we modify the data with location and scale transformation. We observe number of singular observations after transformation is zero. Therefore this could be a reasonable data to work with. 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 for the actual ones. Bayesian estimates are calculated and provided in Table-1 under different methods and different choice of priors based on this data set. We use parametric bootstrap technique to generate different samples from three parameter BB-BVPA and then find out the mean square error and coverage probabilities for the constructed credible interval. Results are available in Table-2 and Table-3. For gamma prior, we observe that the mean square error is high for α_0 as compared to α_1 and α_2 and coverage probabilities are much lower than the desired 95% confidence level. In this case study we observe that we can not use 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, 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 coverage probabilities closer to 95% using 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 1: 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 2: 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 3: 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 $k_0 = 2,\ \theta_0 = 3,\ k_1 = 4,\ \theta_1 = 3,\ k_2 = 3,\ \theta_2 = 2$

6 Conclusion

Bayesian estimates of the parameters of 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. Algorithm does not work always for small sample size with reference as a prior. Coverage Probability for the credible interval is coming little low for gamma prior. More work in needed that direction. The same study can be made using many other algorithms like importance sampling, HMC etc. Bayesian estimation in case of seven parameter bivariate Pareto with location and scale as parameters is a challenging problem. The work is on progress.

References

Arnold, B. C. (1967). A note on multivariate distributions with specified marginals. *Journal of the American Statistical Association*, 62(320):1460–1461.

Asimit, A. V., Furman, E., and Vernic, R. (2010). On a multivariate pareto distribution. *Insurance: Mathematics and Economics*, 46(2):308–316. 2

Asimit, A. V., Furman, E., and Vernic, R. (2016). Statistical inference for a new class of multivariate pareto distributions. *Communications in Statistics-Simulation and Computation*, 45(2):456–471. 2

Berger, J. O., Bernardo, J. M., et al. (1992). On the development of reference priors. *Bayesian statistics*, 4(4):35–60. 6

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 113–147. 6

Block, H. W. and Basu, A. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association*, 69(348):1031–1037. 2

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174. 3

Chen, M.-H. and Shao, Q.-M. (1999). Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92. 16

Christian, P. R. and Casella, G. (1999). Monte carlo statistical methods.

Dey, A. K. and Kundu, D. (2017). Em algorithm for absolute continuous bivariate pareto distribution. *https://fr.arxiv.org/pdf/1608.02199*. 2, 3, 18

Dey, A. K. and Paul, B. (2017). Some variations of em algorithms for marshall-olkin bivariate pareto distribution with location and scale. *arXiv preprint arXiv:1707.09974*. 13

Falk, M. and Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized pareto distributions. *Journal of Multivariate Analysis*, 99(4):715–734. 19

Hanagal, D. D. (1996). A multivariate pareto distribution. *Communications in Statistics-Theory and Methods*, 25(7):1471–1488.

Khosravi, M., Kundu, D., and Jamalizadeh, A. (2015). On bivariate and a mixture of bivariate birnbaum–saunders distributions. *Statistical Methodology*, 23:1–17.

Kundu, D. (2012). On sarhan-balakrishnan bivariate distribution. *Journal of Statistics Applications & Probability*, 1(3):163.

Kundu, D. and Gupta, R. D. (2010). A class of absolute continuous bivariate distributions. *Statistical Methodology*, 7(4):464–477.

Kundu, D. and Gupta, R. D. (2011). Absolute continuous bivariate generalized exponential distribution. *AStA Advances in Statistical Analysis*, 95(2):169–185.

Kundu, D., Kumar, A., and Gupta, A. K. (2015). Absolute continuous multivariate generalized exponential distribution. *Sankhya B*, 77(2):175–206.

Lindley, D. V. (1980). Approximate bayesian methods. *Trabajos de estadística y de investigación operativa*, 31(1):223–245. 3

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Mirhosseini, S. M., Amini, M., Kundu, D., and Dolati, A. (2015). On a new absolute continuous bivariate generalized exponential distribution. *Statistical Methods & Applications*, 24(1):61–83.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, pages 705–741. 3, 13

Rakonczai, P. and Zempléni, A. (2012). Bivariate generalized pareto distribution in practice: models and estimation. *Environmetrics*, 23(3):219–227.

Ristić, M. M. and Kundu, D. (2015). Marshall-olkin generalized exponential distribution. *Metron*, 73(3):317–333.

Sankaran, P. and Kundu, D. (2014). A bivariate pareto model. *Statistics*, 48(2):241–255. 2

Sarhan, A. M. and Balakrishnan, N. (2007). A new class of bivariate distributions and its mixture. *Journal of Multivariate Analysis*, 98(7):1508–1527.

Yeh, H. C. (2000). Two multivariate pareto distributions and their related inferences. *Bulletin of the Institute of Mathematics, Academia Sinica.*, 28(2):71–86. 2

Yeh, H.-C. (2004). Some properties and characterizations for generalized multivariate pareto distributions. *Journal of Multivariate Analysis*, 88(1):47–60. 2

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: 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 5: 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 6: 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 $k_0 = 0.2,\ \theta_0 = 0.3,\ k_1 = 0.4,\ \theta_1 = 0.3,\ k_2 = 0.3,\ \theta_2 = 0.2$.