Bayes Estimation for the Block and Basu Bivariate and Multivariate Weibull Distributions

Biswabrata Pradhan [*]& Debasis Kundu [†]

Abstract

Block and Basu bivariate exponential distribution is one of the most popular absolute continuous bivariate distributions. Recently, Kundu and Gupta ('A class of absolute continuous bivariate distributions', *Statistical Methodology*, 2010) introduced Block and Basu bivariate Weibull distribution, which is a generalization of the Block and Basu bivariate exponential distribution, and provided the maximum likelihood estimators using EM algorithm. In this paper we consider the Bayesian inference of the unknown parameters of the Block and Basu bivariate Weibull distribution. The Bayes estimators are obtained with respect to the squared error loss function, and the prior distributions allow for prior dependence among the unknown parameters. Prior independence also can be obtained as a special case. It is observed that the Bayes estimators of the unknown parameters cannot be obtained in explicit forms. We propose to use the importance sampling technique to compute the Bayes estimates and also to construct the associated highest posterior density credible intervals. The analysis of two data sets have been performed for illustrative purposes. The performances of the proposed estimators are quite satisfactory. Finally we generalize the results for the multivariate case.

Key Words and Phrases Bivariate exponential model; maximum likelihood estimators; Prior distribution; Posterior analysis; Credible intervals.

[*]SQC & OR Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, Pin 700108, India

[†]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. Corresponding author, e-mail: kundu@iitk.ac.in

1 Introduction

Exponential distribution is the most widely used distribution in analyzing the life time data. The main reason about the overwhelming popularity of the exponential distribution is mainly due to its analytical tractability. One of the major problems about the exponential distribution is that it can have only constant hazard function, which often may not be very practical. Due to this reason several other extensions of the exponential distributions have been suggested, for example gamma, Weibull, generalized exponential distributions may have non-constant hazard functions also. A number of bivariate/ multivariate extensions of the exponential distribution can be found in the literature. These include, Gumbel (1960), Freund (1961), Henrich and Jensen (1995), Marshall and Olkin (1967), Downtown (1970), Block and Basu (1974) etc. Several books and books chapters have been devoted exclusively for exponential distribution, see for example Balakrishnan and Basu (1995), Johnson, Kotz and Balakrishnan (1995) etc. For different bivariate exponential distributions, the readers are referred to Kotz, Balakrishnan and Johnson (2000).

The most popular bivariate exponential distribution is the Marshall-Olkin bivariate exponential (MOBE) distribution. Marshall and Olkin (1967) obtained the MOBE distribution as a shock model, from three independent exponential distributions. It has a nice physical interpretation also. The marginals of the MOBE distribution are exponential distributions, and it is a singular distribution. Therefore, if there are ties in the data, the MOBE distribution can be used quite effectively to model such data sets.

Block and Basu (1974) obtained the Block-Basu bivariate exponential (BBBE) distribution from the MOBE distribution, by removing the singular part, and only retaining the absolute continuous part. Unlike the MOBE distribution, BBBE distribution is an absolute continuous distribution. Among the different bivariate exponential distributions, the BBBE distribution has received the maximum attention, because of its flexibility, and analytical tractability. It has been used quite extensively for the analysis of bivariate data sets, particularly when there are no ties in the data set, even though it is known that the marginals are not exponential distributions unlike the MOBE distribution.

Since the marginals of the MOBE distribution are exponential distributions, Marshall and Olkin (1967) in the same paper proposed a more general model along the same line when the marginals are Weibull distribution with the same shape parameter. We call it as the Marshall-Olkin bivariate Weibull (MOBW) model. Recently, Kundu and Gupta (2010) considered a Block and Basu bivariate Weibull (BBBW) distribution which can be obtained along the same line as the BBBE model. The BBBW distribution has been obtained from the MOBW distribution by removing the singular part, and retaining only the absolute continuous part. Clearly, it is a more flexible model than the BBBE model because of the presence of the shape parameter. Although extensive work has been done on BBBE model, but not much attention has been paid on the BBBW model, mainly due to analytical intractability. It may be mentioned that computing the maximum likelihood estimators (MLEs) of the unknown parameters of the BBBW model, one needs to solve a four dimensional optimization problem. Recently, Kundu and Gupta (2010) developed an EM algorithm to compute the maximum likelihood estimators (MLEs) of the unknown parameters, which can be obtained by solving one non-linear equation at each 'E' step of the EM algorithm.

The main purpose of this paper is to develop the Bayesian inference of the unknown parameters of the BBBW model. The Bayes estimators have some advantages in this case. In case of BBBW model, although EM algorithm works quite well when the MLEs exist, MLEs may not exist always in this case. For example, if we have the data $\{(x_{1i}, x_{2i}); i = 1, \ldots, n\}$ and $x_{1i} \leq x_{2i}$ for all $i = 1, \ldots, n$, then MLEs do not exist, but the Bayes estimates will always exist. Although it has been observed that the EM algorithm usually converges, the convergence of EM algorithm could not be established, because of the complicated nature of the model. Further, the convergence of the EM algorithm may depend of the initial guesses. To compute the Bayes estimates we do not need any initial guesses. Moreover, the method we propose to compute the Bayes estimates is guaranteed to converge because of the strong law of large numbers. Finally, it should be mentioned that the confidence intervals based on the MLEs are obtained using the asymptotic results of the MLEs. Constructions of exact confidence intervals may not be very simple. On the other hand the credible intervals obtained using the Bayes estimates have finite sample properties, and they can be obtained even for small sample size also.

In this paper, we want to compute the Bayes estimates of the unknown parameters and the associated credible intervals under proper priors. Achcar and Santander (1993) considered a fairly flexible prior on the scale parameters of the BBBE model, and proposed to use the approximate Bayes estimates with respect to the squared error loss function. They have used a very flexible Dirichlet-Gamma conjugate prior on the scale parameters, similarly as it was used by Pena and Gupta (1990) for the MOBE model parameters. Depending on the hyper-parameter, the Dirichlet-Gamma prior allows the stochastic dependence and independence among the associated model parameters. Moreover, Jeffry's non-informative prior also can be obtained as a limiting case of the proposed Dirichlet-Gamma prior.

In this paper, we have taken the same Dirichlet-Gamma prior on the scale parameters, and on the common shape parameter we do not assume any specific prior. It is only assumed that the prior defined on the shape parameter is independent on the prior defined on the scale parameters, and the support of the prior is on $(0, \infty)$, whose probability density function (PDF) is log-concave. The assumption of the log-concave prior mainly on the shape parameter is not very uncommon in the statistical literature, see for example Berger and Sun (1993) or Kundu (2008). Moreover, many common distributions like normal, log-normal, gamma and Weibull distribution have log-concave PDFs.

The joint posterior distribution of the unknown parameters can be easily obtained based on the above prior distribution. In this paper we have restricted our attention on the squared error loss function, although any other loss functions can be easily incorporated. Unfortunately, although expected, the Bayes estimates of the unknown parameters cannot be obtained in closed form. We propose to use the importance sampling technique to generate samples from the joint posterior distribution function, and use them to compute the Bayes estimates and also to construct highest posterior density (HPD) credible intervals. For illustrative purposes we have analyzed two data sets; one real data set and one simulated data set, and it is observed that the performance of the Bayes estimators are quite satisfactory.

We then consider the multivariate generalization of the BBBW model along the same line and name it as the Block and Basu multivariate Weibull (BBMW) model. As expected, the p-variate BBMW model has total p+2 unknown parameters. Based on the p+1 variate Gamma-Dirichlet prior on the scale parameter and log-concave prior on the shape parameter we develop Bayesian inference of the unknown parameters based on importance sampling procedure.

Rest of the paper is organized as follows. In Section 2, we briefly discuss the BBBW model. Prior assumptions and the available data are presented in Section 3. Bayesian inference of the unknown parameters are discussed in Section 4. In Section 5 we present the analysis of two data sets for illustrative purposes. The Bayesian inference of the unknown parameters when the data are obtained from a p-variate BBMW model, are discussed in Section 6. Finally we conclude the paper in Section 7.

2 Block and Basu Bivariate Weibull Model

In this section we provide a brief description of the BBBW model. We will be using the following notations for the rest of the paper. If X is an exponential random variable with parameter $\lambda > 0$, then the PDF of X is given by

$$f_{EX}(x;\lambda) = \lambda e^{-\lambda x}; \quad x > 0 \tag{1}$$

and it will be denoted by Exp(λ). Similarly, if X has an univariate Weibull distribution with the shape and scale parameter as α and λ respectively, then for $x > 0$, the PDF of X is defined as

$$f_{WE}(x;\alpha,\lambda) = \alpha\lambda x^{\alpha-1} e^{-\lambda x^{\alpha}}; \quad x > 0, \tag{2}$$

and it will be denoted by WE(α, λ). If X follows ($\sim$) WE(α, λ), the hazard function (HF) and survival function (SF) will be denoted by $h_{WE}(x;\alpha,\lambda)$ and $S_{WE}(x;\alpha,\lambda)$ respectively.

First we will define MOBW distribution. Suppose $U_0 \sim$ WE(α, λ_0), $U_1 \sim$ WE(α, λ_1) and $U_2 \sim$ WE(α, λ_2) and they are independently distributed. Consider the random variables $X_1 = \min\{U_0, U_1\}$ and $X_2 = \min\{U_0, U_2\}$. The bivariate random vector (X_1, X_2) is said to have Marshall-Olkin bivariate Weibull distribution. The joint survival function of (X_1, X_2) can be written for $z = \max\{x_1, x_2\}$, as;

$$\begin{aligned} S_{MO}(x_1,x_2) &= P(X_1 > x_1, X_2 > x_2) = P(U_0 > z, U_1 > x_1, U_2 > x_2) \\ &= S_{WE}(x_1;\alpha,\lambda_1) S_{WE}(x_2;\alpha,\lambda_2) S_{WE}(z;\alpha,\lambda_0) \\ &= \begin{cases} S_{WE}(x_1,\alpha,\lambda_1) S_{WE}(x_2;\alpha,\lambda_0+\lambda_2) & \text{if} \quad 0 < x_1 < x_2 < \infty \\ S_{WE}(x_1,\alpha,\lambda_0+\lambda_1) S_{WE}(x_2;\alpha,\lambda_2) & \text{if} \quad 0 < x_2 < x_1 < \infty \\ S_{WE}(x;\alpha,\lambda_0+\lambda_1+\lambda_2) & \text{if} \quad 0 < x_1 = x_2 = x < \infty. \end{cases} \end{aligned} \tag{3}$$

It may be noted that in (3) when $\alpha = 1$, it becomes the celebrated MOBE model. It can be easily observed that that joint survival function of (X_1, X_2) can be written as follows:

$$S_{MO}(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_0 + \lambda_1 + \lambda_2} S_a(x_1, x_2) + \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} S_s(x_1, x_2), \qquad (4)$$

where $S_a(\cdot, \cdot)$ is the absolute continuous part, and $S_s(\cdot, \cdot)$ is the singular part.

The BBBW distribution has been obtained from the MOBW model, by removing the singular part, and retaining only the absolute continuous part. Therefore, the joint PDF of BBBW can be written as follows;

$$f_{BB}(y_1, y_2) = \begin{cases} cf_1(y_1, y_2) = cf_{WE}(y_1; \alpha, \lambda_1) f_{WE}(y_2; \alpha, \lambda_0 + \lambda_2) & \text{if} \quad 0 < y_1 < y_2 \\ cf_2(y_1, y_2) = cf_{WE}(y_1; \alpha, \lambda_0 + \lambda_1) f_{WE}(y_2; \alpha, \lambda_2) & \text{if} \quad 0 < y_2 < y_1, \end{cases} \qquad (5)$$

here c is the normalizing constant and $c = \dfrac{\lambda_0 + \lambda_1 + \lambda_2}{\lambda_1 + \lambda_2}$. From now on if a random vector has the joint PDF (5), then it will be denoted by BBBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$.

Several properties of the BBBW have been established by Kundu and Gupta (2010). It is observed that under the restriction $\lambda_1 = \lambda_2$, BBBW has a total positivity of order two property. The maximum likelihood estimators can be obtained by solving a four dimensional optimization process. To avoid that Kundu and Gupta (2010) suggested to use EM algorithm to estimate the unknown parameters, which can be obtained very conveniently by solving a one dimensional optimization problem at each 'E' step.

3 Priors & Data

3.1 Prior Assumptions

Following the approach of Pena and Gupta (1999), Achcar and Santander (1993) assumed the following priors on the scale parameters λ_0, λ_1 and λ_2. It is assumed that $\lambda = \lambda_0 + \lambda_1 + \lambda_2$ has a Gamma(a, b) prior, say $\pi_0(\cdot|a, b)$. It is assumed that the PDF of Gamma(a, b) for $\lambda > 0$,

$a > 0$ and $b > 0$ is

$$\pi_0(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}, \tag{6}$$

and 0 otherwise.

To bring the dependence between λ_0, λ_1 and λ_2, Achcar and Santander (1993), see also Pena and Gupta (1999) or Kundu and Gupta (2013), made the following assumptions; given λ, $\left(\frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}\right)$ has a Dirichlet prior, say $\pi_1(\cdot|a_0,a_1,a_2)$, *i.e.*

$$\pi_1\left(\frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}|\lambda, a_0, a_1, a_2\right) = \frac{\Gamma(a_0+a_1+a_2)}{\Gamma(a_0)\Gamma(a_1)\Gamma(a_2)}\left(\frac{\lambda_0}{\lambda}\right)^{a_0-1}\left(\frac{\lambda_1}{\lambda}\right)^{a_1-1}\left(\frac{\lambda_2}{\lambda}\right)^{a_2-1}, \tag{7}$$

for $\lambda_0 > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$, where $\lambda_0 = \lambda - \lambda_1 - \lambda_2$. Here all the hyper parameters a, b, a_0, a_1, a_2 are all greater than 0. After simplification, the joint prior of λ_0, λ_1 and λ_2 becomes

$$\begin{aligned}\pi_1(\lambda_0,\lambda_1,\lambda_2|a,b,a_0,a_1,a_2) &= \frac{\Gamma(a_0+a_1+a_2)}{\Gamma(a)} \times (b\lambda)^{a-a_0-a_1-a_2} \times \frac{b^{a_0}}{\Gamma(a_0)}\lambda_0^{a_0-1}e^{-b\lambda_0} \\ &\quad \times \frac{b^{a_1}}{\Gamma(a_1)}\lambda_1^{a_1-1}e^{-b\lambda_1} \times \frac{b^{a_2}}{\Gamma(a_2)}\lambda_2^{a_2-1}e^{-b\lambda_2}. \end{aligned} \tag{8}$$

Using the notation $\bar{a} = a_0 + a_1 + a_2$, (8) can be written as

$$\pi_1(\lambda_0,\lambda_1,\lambda_2|a,b,a_0,a_1,a_2) = \frac{\Gamma(\bar{a})}{\Gamma(a)}(b\lambda)^{a-\bar{a}} \times \prod_{i=0}^{2}\frac{b^{a_i}}{\Gamma(a_i)}\lambda_i^{a_i-1}e^{-b\lambda_i}. \tag{9}$$

The prior as defined in (9) is also known as the Gamma-Dirichlet prior, with parameters a, b, a_0, a_1, a_2, and from now on we will denote this as $\text{GD}(a,b,a_0,a_1,a_2)$. It is clear from (9) that if we take the hyper parameter $a = \bar{a}$, then λ_0, λ_1 and λ_2 will be independent, otherwise they will be dependent. Since the conditional distribution of $(\lambda_i/\lambda, \lambda_j/\lambda)$ $(i \neq j)$ given λ, is Dirichlet, therefore, the correlation between λ_i/λ and λ_j/λ $(i \neq j)$, given λ is negative. However, as the following result shows that the correlation between λ_i and λ_j $(i \neq j)$ can be positively correlated also. The following result presents the distributional properties of the proposed prior distribution. The proof is routine, hence omitted. It is immediate that the above mentioned prior is a very flexible prior.

Result: If $(\lambda_0, \lambda_1, \lambda_2) \sim \text{GD}(a, b, a_0, a_1, a_2)$, then for i, j = 0,1,2,

(a) $\text{E}(\lambda) = a/b, \quad \text{Var}(\lambda) = a/b^2$.

(b) $\text{E}(\lambda_i) = (a/\bar{a})(a_i/b)$, $\text{Var}(\lambda_i) = (a_i/b^2)(a/\bar{a})\{(a_i+1)(a+1)/(\bar{a}+1) - a_i a/\bar{a}\}$

(c) $\text{Cov}(\lambda_i, \lambda_j) = (a_i a_j/b^2)(a/\bar{a})\{(a+1)/(\bar{a}+1) - a/\bar{a}\}, (i \neq j)$.

At this moment we do not assume any specific form of the prior on α. It is assumed that the prior on α has a non-negative support on $(0, \infty)$, and the probability density function of the prior of α, $\pi_2(\alpha)$, is log-concave, and it is independent of $\pi_1(\lambda_0, \lambda_1, \lambda_2)$. From now on the joint prior of α, λ_0, λ_1 and λ_2 will be denoted by $\pi(\alpha, \lambda_0, \lambda_1, \lambda_2)$, and

$$\pi(\alpha, \lambda_0, \lambda_1, \lambda_2) = \pi_1(\lambda_0, \lambda_1, \lambda_2)\pi_2(\alpha). \tag{10}$$

The assumption of log-concave prior is quite common in the Bayesian analysis, see for example Berger and Sun (1993) and Berger (1995). Several well known distributions have log-concave PDF, for example, normal and log-normal distribution have always log-concave PDFs. Weibull and gamma distributions have log-concave PDFs when the corresponding shape parameter is greater than one.

3.2 Data

In this subsection we mention different kinds of data available to us for analysis purposes.

Bivariate Data: It is assumed that we have a bivariate sample of size n, from a BBBW distribution, and it is as follows:

$$\mathcal{D}_2 = \{(x_{11}, x_{21}), \cdots, (x_{1n}, x_{2n})\}. \tag{11}$$

We will be using the following notations:

$$I_1 = \{i; x_{1i} < x_{2i}\} \quad I_2 = \{i; x_{1i} > x_{2i}\}, \quad I = \{1, \cdots, n\},$$

and $|I_1| = n_1$, and $|I_2| = n_2$, here $|I_j|$, for $j = 1$ and 2 denote the number of elements in the set I_j.

Multivariate Data: It is assumed that we have a p-variate sample of size n from BBMW distribution, and the data are of the following form;

$$\mathcal{D}_p = \{(x_{11}, \cdots, x_{p1}), \cdots, (x_{1n}, \cdots, x_{pn})\}. \tag{12}$$

In the multivariate case we will be using the following notations;

$$J_{i_1 i_2 \cdots i_p} = \{i; x_{i_1 i} < \cdots < x_{i_p i}\},$$

for all possible $n!$ permutations $(i_1 i_2 \cdots i_p)$ of $(1, \cdots, n)$. Moreover, the number of elements in $J_{i_1 i_2 \cdots i_p}$ will be denoted by $n_{i_1 i_2 \cdots i_p}$.

4 Bayes Estimates and Credible Intervals

In this section we provide the Bayes estimates of the unknown parameters and the associated credible intervals when the data are obtained from BBBW model. We consider two cases separately. First we consider the case when the shape parameter α is known, and then we consider the more important case when the shape parameter α is not known. In both the cases it is observed that the Bayes estimators cannot be obtained in explicit form. We propose to use the importance sampling technique to compute the Bayes estimates and also to construct the associated credible intervals. In computing the Bayes estimates we have mainly assumed the squared error loss function, although any other loss function can be easily incorporated.

Now based on the observations (11), the joint likelihood of the observed data can be written as

$$l(\mathcal{D}_2|\alpha, \lambda_0, \lambda_1, \lambda_2) = c^n \times \prod_{i \in I_1} f_1(y_{1i}, y_{2i}) \times \prod_{i \in I_2} f_2(y_{1i}, y_{2i}), \tag{13}$$

with the usual convention that if I_j is empty, then $\prod_{i \in I_j} f_j(y_{1i}, y_{2i}) = 1$, for $j = 1, 2$. Now we consider the two cases separately.

4.1 Shape Parameter α is Known

In case of bivariate Weibull model or in the competing risks set up, the assumption of known shape parameter is quite common in the statistical literature, see for example Kundu and Gupta (2013) or Mukhopadhyay and Basu (1997). In this case based on the prior $\pi_1(\cdot)$ on $(\lambda_0, \lambda_1, \lambda_2)$, we obtain the posterior density function of $(\lambda_0, \lambda_1, \lambda_2)$ given $\mathcal{D}_2$ as

$$
\begin{aligned}
l(\lambda_0, \lambda_1, \lambda_2 | \alpha, \mathcal{D}_2) &\propto l(\mathcal{D}_2 | \lambda_0, \lambda_1, \lambda_2, \alpha) \pi_1(\lambda_0, \lambda_1, \lambda_2 | a, b, a_0, a_1, a_2) \\
&\propto \left\{ \prod_{i \in I} x_{1i}^{\alpha-1} x_{2i}^{\alpha-1} \right\} \alpha^{2n} \times c^n \lambda_1^{n_1} (\lambda_0 + \lambda_2)^{n_1} \\
&\quad \times e^{-\lambda_1 \sum_{i \in I_1} x_{1i}^{\alpha}} e^{-(\lambda_0 + \lambda_2) \sum_{i \in I_1} x_{2i}^{\alpha}} \\
&\quad \times \lambda_2^{n_2} (\lambda_0 + \lambda_1)^{n_2} e^{-(\lambda_0 + \lambda_1) \sum_{i \in I_2} x_{1i}^{\alpha}} e^{-\lambda_2 \sum_{i \in I_2} x_{2i}^{\alpha}} \\
&\quad \times \frac{\Gamma(\bar{a})}{\Gamma(a)} (b\lambda)^{a - \bar{a}} \times \prod_{i=0}^{2} \frac{b^{a_i}}{\Gamma(a_i)} \lambda_i^{a_i - 1} e^{-b\lambda_i}. \qquad (14)
\end{aligned}
$$

Therefore, based on (14), the joint posterior density function of $(\lambda_0, \lambda_1, \lambda_2)$ can be written as

$$
\begin{aligned}
l(\lambda_0, \lambda_1, \lambda_2 | \alpha, \mathcal{D}_2) &\propto \lambda^{n + a - \bar{a}} \times \left(1 + \frac{\lambda_2}{\lambda_0}\right)^{n_1} \times \left(1 + \frac{\lambda_1}{\lambda_0}\right)^{n_2} \times \frac{1}{(\lambda_1 + \lambda_2)^n} \\
&\quad \times \text{Gamma}(\lambda_0; a_0 + n, T_0(\alpha) + b) \times \text{Gamma}(\lambda_1; a_1 + n_1, T_1(\alpha) + b) \\
&\quad \times \text{Gamma}(\lambda_2; a_2 + n_2, T_2(\alpha) + b), \qquad (15)
\end{aligned}
$$

where

$$
T_0(\alpha) = \sum_{i \in I_2} y_{1i}^{\alpha} + \sum_{i \in I_1} y_{2i}^{\alpha}, \quad T_1(\alpha) = \sum_{i \in I} y_{1i}^{\alpha}, \quad T_2(\alpha) = \sum_{i \in I} y_{2i}^{\alpha}.
$$

Therefore, if we want to compute the Bayes estimate of some function of λ_0, λ_1 and λ_2, say $\theta = \theta(\lambda_0, \lambda_1, \lambda_2)$, the Bayes estimate of θ, say $\widehat{\theta}_B$, under squared error loss function is the posterior mean of θ, *i.e.*

$$\widehat{\theta}_B = \frac{\int_0^\infty \int_0^\infty \int_0^\infty \theta(\lambda_0, \lambda_1, \lambda_2) l_N(\lambda_0, \lambda_1, \lambda_2 | \mathcal{D}_2) d\lambda_0 d\lambda_1 d\lambda_2}{\int_0^\infty \int_0^\infty \int_0^\infty l_N(\lambda_0, \lambda_1, \lambda_2 | \mathcal{D}_2) d\lambda_0 d\lambda_1 d\lambda_2}. \tag{16}$$

Here $l_N(\lambda_0, \lambda_1, \lambda_2)$ is the right hand side of (15), and it differs with $l(\lambda_0, \lambda_1, \lambda_2)$ only with the proportionality constant. In general the Bayes estimate of θ cannot be obtained explicitly. We need some numerical technique to compute the Bayes estimate of θ. Alternatively, we may use Lindley's approximation technique to compute the approximate Bayes estimate of θ under squared error loss function. Unfortunately, using Lindlay's method although it is possible to compute the approximate Bayes estimate, it is not possible to compute the credible interval of the unknown parameter. Due to this reason we propose to use the importance sampling technique to compute the Bayes estimate and also to construct the highest posterior density credible interval of θ.

We use the following algorithm to compute the Bayes estimate of θ.

Step 1: Generate for $i = 1, \cdots, N$,

$$\lambda_{0i} \sim \text{Gamma}(a_0 + n, T_0(\alpha) + b) \tag{17}$$

$$\lambda_{1i} \sim \text{Gamma}(a_1 + n_1, T_1(\alpha) + b) \tag{18}$$

$$\lambda_{2i} \sim \text{Gamma}(a_2 + n_2, T_2(\alpha) + b). \tag{19}$$

Step 2: Compute

$$\begin{aligned} h(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}) &= (\lambda_{0i} + \lambda_{1i} + \lambda_{2i})^{n+a-\bar{a}} \times \left(1 + \frac{\lambda_{2i}}{\lambda_{0i}}\right)^{n_1} \times \left(1 + \frac{\lambda_{1i}}{\lambda_{0i}}\right)^{n_2} \times \frac{1}{(\lambda_{1i} + \lambda_{2i})^n} \\ \theta_i &= \theta_i(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}) \end{aligned}$$

Step 3: A simulation consistent Bayes estimate of $\theta(\lambda_0, \lambda_1, \lambda_2)$ can be obtained as

$$\widehat{\theta}_B = \sum_{i=1}^{N} \frac{\theta_i h(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})}{\sum_{k=1}^{N} h(\lambda_{0k}, \lambda_{1k}, \lambda_{2k})}. \tag{20}$$

Note that although, $h(\cdot)$ is an unbounded function, if λ_0, λ_1 and λ_2 follow (17), (18) and (19) respectively, and they are independently distributed, then $E(h^k(\lambda_0, \lambda_1, \lambda_2)) < \infty$ for $k = 1, \cdots$. This forms the basis of the above importance sampling procedure.

Step 4: Construction of credible intervals:

The same method can be used to compute a HPD credible interval of θ, any function of λ_0, λ_1 and λ_2. Suppose for $0 < p < 1$, θ_p is $P[\theta \leq \theta_p | \mathcal{D}_2] = p$. Now consider the following function

$$g(\lambda_0, \lambda_1, \lambda_2) = \begin{cases} 1 & \text{if} \quad \theta \leq \theta_p \\ 0 & \text{if} \quad \theta > \theta_p. \end{cases} \tag{21}$$

Clearly, $E(g(\lambda_0, \lambda_1, \lambda_2)|\mathcal{D}_2) = p$. Therefore, a consistent estimator of θ_p under the squared error loss function can be obtained from the generated sample $\{(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}); i = 1, \cdots N\}$, as follows. Let

$$w_i = \frac{h(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})}{\sum_{j=1}^{N} h(\lambda_{0j}, \lambda_{1j}, \lambda_{2j})}. \tag{22}$$

Rearrange, $\{(\theta_1, w_1), \cdots, (\theta_N, w_N)\}$ as $\{(\theta_{(1)}, w_{(1)}), \cdots, (\theta_{(N)}, w_{(N)})\}$, where $\theta_{(1)} < \cdots < \theta_{(N)}$, and $w_{(i)}$'s are not ordered, they are just associated with $\theta_{(i)}$. Then a consistent Bayes estimator of θ_p is $\widehat{\theta}_p = \theta_{(N_p)}$, where N_p is the integer satisfying

$$\sum_{i=1}^{N_p} w_{(i)} \leq p < \sum_{i=1}^{N_p+1} w_{(i)}. \tag{23}$$

Now using the above procedure, a 100(1-γ)% credible interval of θ can be obtained as $(\widehat{\theta}_\delta, \widehat{\theta}_{\delta+1-\gamma})$, for $\delta = w_{(1)}, w_{(1)} + w_{(2)}, \cdots, \sum_{i=1}^{N_\gamma} w_{(i)}$. Therefore, a 100(1-$\gamma$)% HPD credible interval of θ becomes $(\widehat{\theta}_{\delta^*}, \widehat{\theta}_{\delta^*+1-\gamma})$, where δ^* satisfies

$$\widehat{\theta}_{\delta^*+1-\gamma} - \widehat{\theta}_{\delta^*} \leq \widehat{\theta}_{\delta+1-\gamma} - \widehat{\theta}_\delta, \quad \text{for all} \;\; \delta.$$

4.2 Shape Parameter α is Unknown

Now we consider the more important case when the shape parameter α is unknown. In this case the joint posterior density of λ_0, λ_1, λ_2 and α based on the priors (10), can be written as

$$l(\lambda_0, \lambda_1, \lambda_2, \alpha|\mathcal{D}_2) = l(\lambda_0, \lambda_1, \lambda_2|\alpha, \mathcal{D}_2) \times l(\alpha|\mathcal{D}_2), \tag{24}$$

where

$$\begin{aligned} l(\lambda_0, \lambda_1, \lambda_2|\alpha, \mathcal{D}_2) \quad \propto \quad & h(\lambda_0, \lambda_1, \lambda_2) \times \text{Gamma}(\lambda_0; a_0 + n, T_0(\alpha) + b) \\ & \times \text{Gamma}(\lambda_1; a_1 + n_1, T_1(\alpha) + b) \\ & \times \text{Gamma}(\lambda_2; a_2 + n_2, T_2(\alpha) + b) \end{aligned} \tag{25}$$

and

$$l(\alpha|\mathcal{D}_2) \propto \frac{\pi_2(\alpha) \times \left\{\prod_{i \in I} x_{1i}^{\alpha-1} x_{2i}^{\alpha-1}\right\} \times \alpha^{2n}}{(T_0(\alpha) + b)^{a_0+n} \times (T_1(\alpha) + b)^{a_1+n_1} \times (T_2(\alpha) + b)^{a_2+n_2}}. \tag{26}$$

It is immediate that if we want to compute the Bayes estimate of $\theta = \theta(\lambda_0, \lambda_1, \lambda_2, \alpha)$ under squared error loss function it cannot be obtained in general even if we know $\pi_2(\alpha)$ explicitly. We propose to use the importance sampling technique to compute the Bayes estimate of $\theta = \theta(\lambda_0, \lambda_1, \lambda_2, \alpha)$, and for that we need the following result.

Lemma 4.1: If the PDF of $\pi_2(\alpha)$ is log-concave then $l(\alpha|\mathcal{D}_2)$ is log-concave.

Proof: It can be obtained following the same line of proof of Lemma 2 of Kundu (2008), and therefore it is avoided.

From Lemma 4.1, it is clear that the generation from $l(\alpha|\mathcal{D}_2)$ can be easily performed using the general result of Devroye (1984), or using a similar method suggested by Kundu (2008), a very good approximation of $l(\alpha|\mathcal{D}_2)$ can be obtained using a two-parameter gamma distribution. Therefore, we suggest the following algorithm to compute simulation consistent Bayes estimate of θ and also to construct associated credible interval:

Step 1: Generate for $i = 1, \cdots, N$,

$$\alpha_i \sim l(\alpha|\mathcal{D}_2) \tag{27}$$

$$\lambda_{0i}|\alpha_i \sim \text{Gamma}(a_0 + n, T_0(\alpha_i) + b) \tag{28}$$

$$\lambda_{1i}|\alpha_i \sim \text{Gamma}(a_1 + n_1, T_1(\alpha_i) + b) \tag{29}$$

$$\lambda_{2i}|\alpha_i \sim \text{Gamma}(a_2 + n_2, T_2(\alpha_i) + b). \tag{30}$$

Step 2, Step 3 and Step 4 are exactly same as in Section 4.1, where

$$\theta_i = \theta(\alpha_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}).$$

5 Data Analysis

In this section we analyze two data sets for illustrative purpose. The first data set is a real life data and the second one is a simulated data set.

Example 1: This data set has been obtained from (Johnson and Wichern, 1999, pp. 374). It represents the bone mineral density (BMD) measured in g/cm^2 for 24 children after one year of birth. These bivariate data represent the BMD for Dominant radius and Radius bones and they are as follows;

Data set: (1.027, 1.051), (0.857, 0.817), (0.875, 0.880), (0.873, 0.698), (0.811, 0.813), (0.640, 0.734), (0.947, 0.865), (0.886, 0.806), (0.991, 0.923), (0.977, 0.925), (0.825, 0.826), (0.851, 0.765), (0.770, 0.730), (0.912, 0.875), (0.905, 0.826), (0.756, 0.727), (0.765, 0.764), (0.932, 0.914), (0.843, 0.782), (0.879, 0.906), (0.673, 0.537), (0.949, 0.900), (0.463, 0.937), (0.776, 0.743).

We consider the Bayes estimation of the parameters α, λ_0, λ_1 and λ_2. We assume that α has also Gamma prior with pdf Gamma(c, d). Since we do not have any knowledge on hyperparameter values, we obtain the Bayes estimates under non-informative priors, where we take $a = b = a_0 = a_1 = a_2 = c = d = 0$. Note that this prior setting leads to non-informative priors. The posterior density of α becomes

$$l(\alpha|\mathcal{D}_2) \propto \frac{\left\{\prod_{i\in I} x_{1i}^{\alpha-1} x_{2i}^{\alpha-1}\right\} \times \alpha^{2n-1}}{(T_0(\alpha))^n \times (T_1(\alpha))^{n_1} \times (T_2(\alpha))^{n_2}}. \tag{31}$$

As described in Section 4.2, we obtain the estimates by importance sampling technique. To apply the importance sampling technique, first we need to generate observation from $l(\alpha|\mathcal{D}_2)$. One can generate observation from $l(\alpha|\mathcal{D}_2)$ using the method of Devroye(1984), but we consider a more simpler method proposed by Kundu(2008). Using the idea of Kundu(2008), we approximate (31) by Gamma(k_1, k_2). Here the approximate values of k_1 and k_2 becomes 77.28 and 8.88, respectively. We provide the generated samples along with the actual posterior density function of α in Figure 1. They match very well.

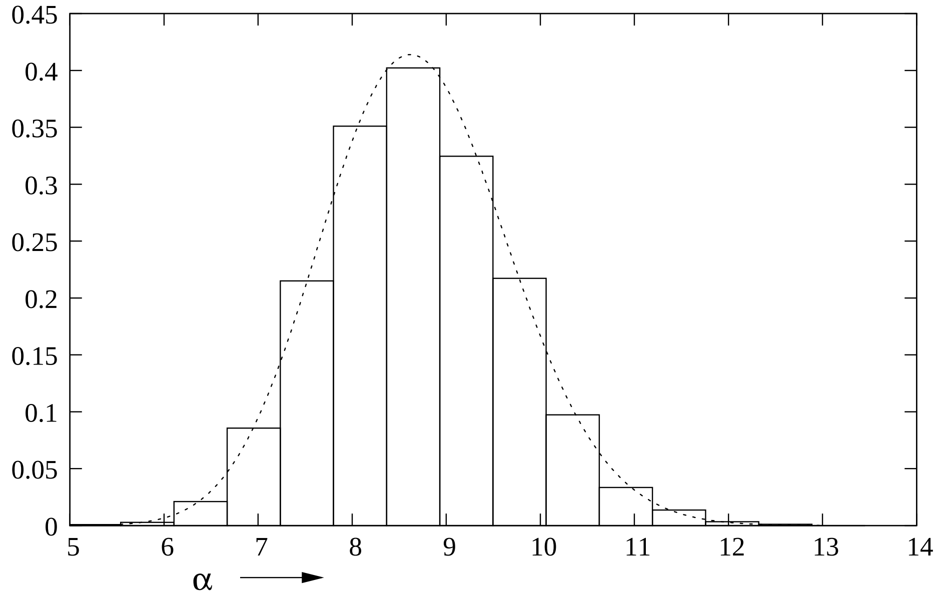

Figure 1: The histogram of the generated samples and the posterior density function of α for BMD data set are provided.

Based on 10000 importance samples, the Bayes estimates of α, λ_0, λ_1 and λ_2 are 8.3125, 3.8092, 0.6339 and 1.4589, respectively. The 95% HPD credible intervals for α, λ_0, λ_1 and λ_2 are (7.1880, 10.0870), (2.1652, 4.6996), (0.2960, 1.2774) and (0.8574, 2.4229), respectively.

Some of the points are quite clear from the Bayes estimates of the unknown parameters. For example, since α is significantly different from 1, therefore clearly BBBE model cannot be used to analyze this data set. Since $\lambda_1 < \lambda_2$, the expected BMD for Dominant radius bones will be more than the expected BMD of the Radius bones for the children of this population. Moreover, since we have the estimates of all the unknown parameters of the parametric model, we have the complete inference of the corresponding population characteristics.

Example 2: In this example, we analyze a simulated data set from BB Weibull distribution with $\alpha = 2$, $\lambda_0 = 1.2$, $\lambda_1 = 0.5$ and $\lambda_2 = 1.5$. The generated data set for $n = 30$ is as follows. (0.973, 0.275), (0.437, 0.547), (0.198, 0.616), (1.197, 0.596), (0.676, 0.546), (0.756, 0.372), (0.740, 0.727), (0.120, 0.252), (0.324, 0.419), (0.918, 1.003) (0.448, 0.335), (1.110, 0.504), (0.474, 0.373), (0.946, 1.021), (0.652, 0.512) (0.911, 0.663), (0.376, 0.314), (0.911, 0.616), (1.190, 0.873), (0.773, 0.771), (1.228, 0.771), (0.706, 0.409), (0.075, 0.606), (1.560, 0.917), (0.954, 0.692) (0.997, 0.556), (1.008, 0.726), (0.975, 0.684), (0.154, 0.911), (0.667, 0.581).

We obtain Bayes estimates of the parameters under non-informative priors as in example 1. First we obtain Bayes estimates of λ_0, λ_1 and λ_2 assuming α is known. The estimates of λ_0, λ_1 and λ_2 are 1.5307, 0.4900 and 1.2866, respectively. The corresponding 95% HPD credible intervals are (0.6466, 2.0789), (0.1657, 0.8376) and (0.7578, 2.1098). Next we obtain the Bayes estimates of α, λ_0, λ_1 and λ_2. First we have generated α as suggested in the previous example. In this case the generated samples along with the actual posterior density function of α are provided in Figure 2. The match is excellent.

The estimates of α, λ_0, λ_1 and λ_2 are 2.3516, 1.3808, 0.5614 and 1.6094, respectively. The corresponding 95% HPD credible intervals are (1.9554, 2.9172), (0.7160, 2.0909), (0.1464, 1.0046) and (0.8934, 2.8719).

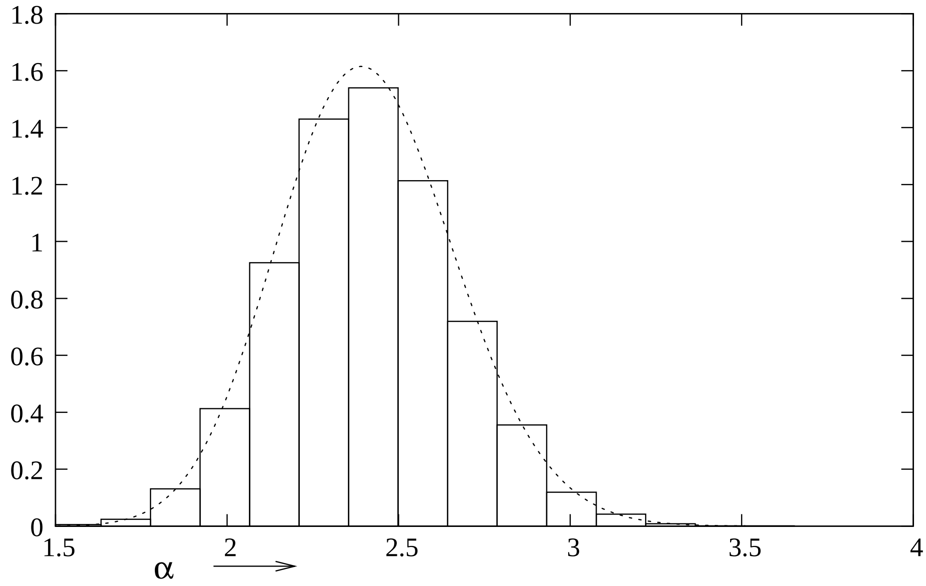

Figure 2: The histogram of the generated samples and the posterior density function of α for simulated data set are provided.

6 Block and Basu Multivariate Weibull Distribution

In this subsection first we will define Block and Basu multivariate Weibull (BBMW) distribution along the same line as the BBBW distribution and then consider the Bayesian inference of the unknown parameters under similar set of prior assumptions as in Section 3.

Definition: A p-variate random vector $(Y_1, \cdots, Y_p)$ is said to have a BBMW distribution with parameters $\lambda_0, \lambda_1, \cdots, \lambda_p$ and α, if the joint PDF of $(Y_1, \cdots, Y_p)$ is of the form;

$$f(y_1, \cdots, y_p) = cf_{WE}(y_{i_1}; \alpha, \lambda_{i_1}) \times \cdots \times f_{WE}(y_{i_{p-1}}; \alpha, \lambda_{i_{p-1}}) \times f_{WE}(y_{i_p}; \alpha, \lambda_0 + \lambda_{i_p}), \quad (32)$$

here c is the normalizing constant, and $(i_1, \cdots, i_p)$ is a permutation of $\{1, \cdots, p\}$, such that $y_{i_1} < \cdots < y_{i_p}$.

The normalizing constant c can be obtained from the identity

$$\int_{\mathbf{R}^p} f(y_1, \cdots, y_p) dy_1 \cdots dy_p = 1. \quad (33)$$

It can be easily seen by simple integration that

$$c^{-1} = \sum_{\mathcal{P}} \frac{\lambda_{i_1}}{\lambda_{i_1} + \cdots + \lambda_{i_p} + \lambda_0} \times \cdots \times \frac{\lambda_{i_{p-1}}}{\lambda_{i_{p-1}} + \lambda_{i_p} + \lambda_0}. \tag{34}$$

Note that when $p = 2$, $c = \dfrac{\lambda_0 + \lambda_1 + \lambda_2}{\lambda_1 + \lambda_2}$ as obtained in Section 2.

Following exactly the same approach as in Section 3, we can define the $(p+1)$ variate Gamma-Dirichlet prior on $\lambda_0, \lambda_1, \cdots, \lambda_p$ as follows

$$\pi_1(\lambda_0, \lambda_1, \cdots, \lambda_p | a, b, a_0, a_1 \cdots, a_p) = \frac{\Gamma(\bar{a})}{\Gamma(a)} (b\lambda)^{a-\bar{a}} \times \prod_{i=0}^{p} \frac{b^{a_i}}{\Gamma(a_i)} \lambda_i^{a_i - 1} e^{-b\lambda_i}, \tag{35}$$

here $\bar{a} = a_0 + a_1 + \cdots + a_p$, and $\lambda = \lambda_0 + \lambda_1 + \cdots + \lambda_p$. Similarly, we consider $\pi_2(\alpha)$, as defined in Section 3, as the prior of α, and it is independent of $\pi_1(\lambda_0, \lambda_1, \cdots, \lambda_p | a, b, a_0, a_1 \cdots, a_p)$.

For further development we use the following notations:

$$J_{i_1 \cdots i_p} = \{i : x_{i_1 i} < x_{i_2 i} < \cdots < x_{i_p i}\},$$

$$J_k = \{i : x_{i_1 i} < x_{i_2 i} < \cdots < x_{i_{p-1} i} < x_{ki}\}; \qquad k = 1, \cdots, p,$$

here for any k, $\{i_1, i_2, \cdots, i_{p-1}\}$ is a permutation of $(1, \cdots, k-1, k+1, \cdots, p)$. The number of elements in J_k is denoted by m_k, and let us denote $n_k = n - m_k$, for $k = 1, \cdots, p$. Now based on the observations (12) the joint likelihood function can be written as

$$\begin{aligned} l(\mathcal{D}_p | \alpha, \lambda_0, \cdots, \lambda_p) &= c^n \prod_{i_1 \cdots i_p} \prod_{i \in J_{i_1 \cdots i_p}} f_{WE}(x_{i_1 i}; \alpha, \lambda_{i_1}) \times \cdots \times f_{WE}(x_{i_{p-1} i}; \alpha, \lambda_{i_{p-1}}) \times \\ & \qquad f_{WE}(x_{i_p}; \alpha, \lambda_0 + \lambda_{i_p}). \end{aligned} \tag{36}$$

If the shape parameter α is known, the posterior PDF of $\lambda_0, \cdots, \lambda_p$ can be written as

$$\begin{aligned} l(\lambda_0, \cdots, \lambda_p | \alpha, \mathcal{D}_p) &= c^n(\lambda_0, \cdots, \lambda_p) \times \lambda^{a-\bar{a}} \times \prod_{i=1}^{p} \left(1 + \frac{\lambda_i}{\lambda_0}\right)^{n_i} \times \\ & \quad \text{Gamma}(\lambda_0; a_0 + n, T_0(\alpha) + b) \times \text{Gamma}(\lambda_1; a_1 + n_1, T_1(\alpha) + b) \times \\ & \quad \cdots \times \text{Gamma}(\lambda_1; a_p + n_p, T_p(\alpha) + b), \end{aligned} \tag{37}$$

here

$$T_0(\alpha) = \sum_{i \in J_1} x_{1i}^{\alpha} + \cdots + \sum_{i \in J_p} x_{pi}^{\alpha}, \quad T_1(\alpha) = \sum_{i=1}^{n} x_{1i}^{\alpha}, \cdots, T_p(\alpha) = \sum_{i=p}^{n} x_{pi}^{\alpha}.$$

Therefore, it is possible to use similar importance sampling procedure as in Section 4.1, to compute Bayes estimate of any function of $\lambda_0, \cdots, \lambda_p$, and also to compute associated credible interval by taking

$$h(\lambda_0, \cdots, \lambda_p) = c^n(\lambda_0, \cdots, \lambda_p) \times \lambda^{a-\bar{a}} \times \prod_{i=1}^{p} \left(1 + \frac{\lambda_i}{\lambda_0}\right)^{n_i}.$$

Similarly, when α is unknown the posterior PDF of α, $\lambda_0, \cdots, \lambda_p$ can be written as

$$l(\lambda_0, \cdots, \lambda_p, \alpha | \mathcal{D}_p) = l(\alpha|\mathcal{D}_p) \times l(\lambda_0, \cdots, \lambda_p | \alpha, \mathcal{D}_p),$$

where $l(\lambda_0, \cdots, \lambda_p | \alpha, \mathcal{D}_p)$ is same as defined (37), and

$$l(\alpha|\mathcal{D}_p) \propto \frac{\pi_2(\alpha) \times \prod_{i \in I} \left\{x_{1i}^{\alpha-1} \cdots x_{pi}^{\alpha-1}\right\} \times \alpha^{2n}}{(T_0(\alpha)+b)^{a_0+n} \times (T_1(\alpha)+b)^{a_1+n_1} \times \cdots \times (T_p(\alpha)+b)^{a_p+n_p}}. \tag{38}$$

Again, following the similar approach as in Lemma 2 of Kundu (2008), it can be proved that $l(\alpha|\mathcal{D}_p)$ is log-concave. Therefore, as in Section 4.2, the Bayes estimate of any function of $\alpha, \lambda_0, \cdots, \lambda_p$ and the associated credible interval can be constructed.

7 Conclusions

In this work, we have proposed Bayes estimation of the unknown parameters of Block and Basu bivariate Weibull distribution based on complete data. We have considered fairly flexible priors. A priori the scale parameters can be dependent or independent. The prior on the shape parameter also can be quite flexible. Based on these priors we obtain the joint posterior distribution. It is observed that the Bayes estimates cannot be obtained in closed form. Although, Lindley's approximation may be used to compute the approximate Bayes estimates of the unknown parameters. It is not pursued here. Instead, we have obtained the Bayes estimates under squared error loss function and also the associated credible intervals by importance sampling procedure. The implementation of the Bayesian inference is quite straightforward, and it can be easily used in practice.

References

[1] Achcar, J.A. and Santander, L.A.M. (1993), "Use of approximate Bayesian methods for the Block and Basu bivariate exponential distribution", *Journal of Italian Statistical Society*, vol. 3, 233 - 250.

[2] Balakrishnan, N. and Basu, A.P. (1995), *Exponential Distribution: Theory, Methods and Applications*, John Wiley, New York.

[3] Balakrishnan, N. and Lai, C.D. (2009), *Continuous Bivariate Distribution*, 2-nd edition, Springer, New York.

[4] Berger, J.O. and Sun, D. (1993), "Bayesian analysis for Poly-Weibull distribution", *Journal of the American Statistical Association*, vol. 88, 1412 - 1418.

[5] Block, H. and Basu, A. P. (1974), "A continuous bivariate exponential extension", *Journal of the American Statistical Association*, vol. 69, 1031–1037.

[6] Gumbel, E.J. (1960), "Bivariate exponential distribution", *Journal of the American Statistical Association*, vol. 55, 698 - 707.

[7] Devroye, L. (1984), "A simple algorithm for generating random variables with log-concave density", *Computing*, vol. 33, 247 - 257.

[8] Downton, F. (1970), "Bivariate exponential distributions in reliability theory", *Journal of the Royal Statistical Society*, B, vol. 32, 508 - 417.

[9] Freund, J.E. (1961), "A bivariate extension of the exponential distribution", *Journal of the American Statistical Association*, vol. 56, 971 - 977.

[10] Henrich, G. and Jensen, U. (1995), "Parameter estimation for a bivariate lifetime distribution in reliability with multivariate extension", *Metrika*, vol. 42, 49 - 65.

[11] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995), *Continuous Univariate Distribution*, Vol. 1, John Wiley and Sons, New York.

[12] Johnson, R. A. and Wichern, D. W. *Applied Multivariate Analysis*, fourth ed., Prentice Hall, New Jersey, 1999.

[13] Kundu, D. (2008), "Bayesian inference and life testing plan for Weibull distribution in presence of progressive censoring", *Technometrics*, vol. 50, 144 - 154.

[14] Kundu, D. and Gupta, R. D. (2010), "A class of absolutely continuous bivariate distributions", *Statistical Methodology* vol. 7, 464-477.

[15] Kundu, D. and Gupta, A. (2013), "Bayes estimation for the Marshall-Olkin bivariate Weibull distribution", *Computational Statistics and Data Analysis*, vol. 57, 271 - 281.

[16] Marshall, A.W. and Olkin, I. (1967), "A multivariate exponential distribution", *Journal of the American Statistical Association*, vol. 62, 30–44.

[17] Mukhopadhyay, C. and Basu, A.P. (1997), "Bayesian analysis of incomplete time and cause of failure time", *Journal of Statistical Planning and Inference*, vol. 59, 79 - 100.

[18] Pena, E. A. and Gupta, A. K. (1990), "Bayes estimation for the Marshall-Olkin exponential distribution", *Journal of the Royal Statistical Society*, Ser. B, vol. 52, 379 - 389.