

Bayesian Analysis of the Marshall-Olkin Bivariate Weibull Distribution *

ZHANG Hui XU Ancha*

(College of Mathematics and Information Science, Wenzhou University, Wenzhou, 325035, China)

Abstract: Kundu and Gupta^[1] proposed to use the importance sampling method to compute the Bayesian estimation of the unknown parameters of the Marshall-Olkin bivariate Weibull distribution. However, we find that the performance of the importance sampling method becomes worse as the sample size gets larger. In this paper, we introduce latent variables to simplify the likelihood function, and use MCMC algorithm to estimate the unknown parameters. Numerical simulations are carried out to assess the performance of the proposed method by comparing with the maximum likelihood estimation, and we find that the Bayesian estimates perform better even for the case of small sample size. A real data is also analyzed for illustrative purpose.

Keywords: Marshall-Olkin bivariate Weibull distribution; Bayesian estimation; MCMC algorithm

2010 Mathematics Subject Classification: 62N05

§1. Introduction

Weibull distribution plays a crucial role in reliability theory and life testing experiments. It reduces to exponential distribution when the shape parameter is one. A univariate Weibull distribution with the shape parameter α and the scale parameter λ has the following probability density function (PDF),

$$f_W(x|\alpha, \lambda) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}, \quad x, \alpha, \lambda > 0, \quad (1)$$

denoted as $W(\alpha, \lambda)$. Thus, the cumulative distribution function (CDF) and survival function (SF) are

$$F_W(x|\alpha, \lambda) = 1 - e^{-\lambda x^\alpha} \quad \text{and} \quad S_W(x|\alpha, \lambda) = e^{-\lambda x^\alpha}, \quad (2)$$

*The research was supported by National Natural Science Foundation of China (11201345), China Postdoctoral Science Foundation (2015M572598) and Natural Science Foundation of Zhejiang Province (LY15G010006).

*Corresponding author, E-mail: xuancha@wzu.edu.cn.

Received September 10, 2015. Revised December 31, 2015.

respectively. Marshall and Olkin^[2] proposed the Marshall-Olkin bivariate Weibull (MOB-W) distribution and it can be described as follows: W_0 , W_1 and W_2 are supposed to three independent random variables, and

$$W_0 \sim W(\alpha, \lambda_0), \quad W_1 \sim W(\alpha, \lambda_1), \quad W_2 \sim W(\alpha, \lambda_2),$$

where ‘ \sim ’ means follows in the distribution, and α is the common shape parameter, λ_0 , λ_1 , λ_2 are the corresponding scale parameters. Let

$$X_1 = \min(W_0, W_1) \quad \text{and} \quad X_2 = \min(W_0, W_2). \quad (3)$$

Then the bivariate random vector (X_1, X_2) follows a MOBW distribution with parameters α , λ_0 , λ_1 , λ_2 , denoted as MOBW($\alpha, \lambda_0, \lambda_1, \lambda_2$). Thus, the joint SF of (X_1, X_2) has the following form

$$\begin{aligned} S_{X_1, X_2}(x_1, x_2) &= P(X_1 > x_1, X_2 > x_2) = P(W_1 > x_1, W_0 > x_1, W_2 > x_2, W_0 > x_2) \\ &= P(W_1 > x_1, W_2 > x_2, W_0 > \max(x_1, x_2)) \\ &= S_W(x_1 | \alpha, \lambda_1) S_W(x_2 | \alpha, \lambda_2) S_W(\max(x_1, x_2) | \alpha, \lambda_0) \\ &= \begin{cases} S_W(x_1 | \alpha, \lambda_1) S_W(x_2 | \alpha, \lambda_0 + \lambda_2), & \text{if } 0 < x_1 < x_2 < \infty; \\ S_W(x_1 | \alpha, \lambda_0 + \lambda_1) S_W(x_2 | \alpha, \lambda_2), & \text{if } 0 < x_2 < x_1 < \infty; \\ S_W(x | \alpha, \lambda_0 + \lambda_1 + \lambda_2), & \text{if } 0 < x_1 = x_2 = x < \infty. \end{cases} \end{aligned} \quad (4)$$

Obviously, when $\alpha = 1$, it reduces to the Marshall-Olkin bivariate exponential (MOBE) distribution with parameters λ_0 , λ_1 , λ_2 . For more details about MOBE distribution, see [3] and the references therein. From (4), we can write the joint PDF of (X_1, X_2) as

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} f_1(x_1, x_2), & \text{if } 0 < x_1 < x_2 < \infty; \\ f_2(x_1, x_2), & \text{if } 0 < x_2 < x_1 < \infty; \\ f_0(x), & \text{if } 0 < x_1 = x_2 = x < \infty. \end{cases} \quad (5)$$

where

$$\begin{aligned} f_1(x_1, x_2) &= f_W(x_1 | \alpha, \lambda_1) f_W(x_2 | \alpha, \lambda_0 + \lambda_2), \\ f_2(x_1, x_2) &= f_W(x_1 | \alpha, \lambda_0 + \lambda_1) f_W(x_2 | \alpha, \lambda_2), \\ f_0(x) &= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_W(x | \alpha, \lambda_0 + \lambda_1 + \lambda_2). \end{aligned}$$

The MOBW distribution has both an absolute continuous part and a singular part. See [4], [5] for more detailed discussion.

Assume that we have a bivariate sample of size n from $\text{MOBW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$. The observed data is as follows:

$$D = \{(x_{11}, x_{21}), (x_{12}, x_{22}), \dots, (x_{1n}, x_{2n})\}. \quad (6)$$

According to the relationship of x_{1i} and x_{2i} ($i = 1, 2, \dots, n$), the random sample of the MOBW distribution can be divided into three parts, and they are as follows: $D_1 = \{(x_{1i}, x_{2i}) | x_{1i} < x_{2i}\}$, $D_2 = \{(x_{1i}, x_{2i}) | x_{1i} > x_{2i}\}$ and $D_0 = \{(x_{1i}, x_{2i}) | x_{1i} = x_{2i} = x_i\}$. Obviously, $D = D_1 \cup D_2 \cup D_0$. Let $I_1 = \{i | x_{1i} < x_{2i}\}$, $I_2 = \{i | x_{1i} > x_{2i}\}$, $I_0 = \{i | x_{1i} = x_{2i} = x_i\}$, $I = \{1, 2, \dots, n\}$, $I = I_1 \cup I_2 \cup I_0$ and $|I_1| = n_1$, $|I_2| = n_2$, $|I_0| = n_0$, where $|I_m|$ ($m = 0, 1, 2$) denotes the number of elements in the set I_m . Thus, the likelihood function based on the observed data D is

$$L(D | \alpha, \lambda_0, \lambda_1, \lambda_2) = \prod_{i \in I_1} f_1(x_{1i}, x_{2i}) \prod_{i \in I_2} f_2(x_{1i}, x_{2i}) \prod_{i \in I_0} f_0(x_i). \quad (7)$$

It should be mentioned that if $n_m = 0$, for some $m = 0, 1, 2$, then the maximum likelihood estimation (MLE) does not exist. Then we assume $n_m > 0$ for $m = 0, 1$ and 2 . Kundu and Dey^[5] have discussed the computation of the MLE of the unknown parameters using EM algorithm, but the convergence speed of the algorithm is slow, and it is highly dependent on the initial value. By assigning a Gamma-Dirichlet distribution as the prior of $\lambda = \lambda_0 + \lambda_1 + \lambda_2$, λ_1 and λ_2 , and a log-concave prior for α , Kundu and Gupta^[1] used the importance sampling method to obtain the Bayesian estimators of α , λ_0 , λ_1 and λ_2 based on the likelihood function (7). However, Neal^[6] indicated that the efficiency of the importance sampling method depends on the choice of the proposed density function. If the proposed density function is close to the target function, then the approximation is reasonably accurate even with thousands of random samples from the proposed density function. Otherwise, there needs much more random samples. Thus, the computational efficiency will be encountered. Besides, we show by numerical study that the method of [1] behaves worse as the sample size becomes larger, which will be discussed in Section 2. Then a data-augmented method is used to simplify the likelihood function (7), and MCMC is implemented easily, these details are introduced in Section 3. We compare the proposed Bayesian method with EM algorithm via simulations in Section 4. Section 5 is devoted to a real data analysis. Finally, some concluding remarks are made in Section 6.

§2. The Method of [1]

Unlike [1], we assumed that the parameters are independent, and that the prior of all

the parameters are Gamma distributions, that is,

$$\begin{aligned}\pi_0(\lambda_0 | a_0, b_0) &= \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_0^{a_0-1} e^{-b_0 \lambda_0}, \\ \pi_1(\lambda_1 | a_1, b_1) &= \frac{b_1^{a_1}}{\Gamma(a_1)} \lambda_1^{a_1-1} e^{-b_1 \lambda_1}, \\ \pi_2(\lambda_2 | a_2, b_2) &= \frac{b_2^{a_2}}{\Gamma(a_2)} \lambda_2^{a_2-1} e^{-b_2 \lambda_2}, \\ \pi(\alpha | a, b) &= \frac{b^a}{\Gamma(a)} \alpha^{a-1} e^{-b\alpha}.\end{aligned}\quad (8)$$

Based on (7) and (8), the joint posterior density function of $(\alpha, \lambda_0, \lambda_1, \lambda_2)$ can be written as

$$\begin{aligned}L(\alpha, \lambda_0, \lambda_1, \lambda_2 | D) &= \pi_0(\lambda_0 | a_0, b_0) \pi_1(\lambda_1 | a_1, b_1) \pi_2(\lambda_2 | a_2, b_2) \pi(\alpha | a, b) L(D | \alpha, \lambda_0, \lambda_1, \lambda_2) \\ &\propto \prod_{i \in I_1} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_2} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_0} x_i^{\alpha-1} (\lambda_0 + \lambda_2)^{n_1} (\lambda_0 + \lambda_1)^{n_2} e^{-b\alpha} \\ &\quad \times \alpha^{n_0+2n_1+2n_2+a-1} \text{Gamma}(\lambda_0; a_0 + n_0, T_0(\alpha) + b_0) \\ &\quad \times \text{Gamma}(\lambda_1; a_1 + n_1, T_1(\alpha) + b_1) \\ &\quad \times \text{Gamma}(\lambda_2; a_2 + n_2, T_2(\alpha) + b_2),\end{aligned}\quad (9)$$

where $T_0(\alpha) = \sum_{i \in I_2} x_{1i}^\alpha + \sum_{i \in I_1} x_{2i}^\alpha + \sum_{i \in I_0} x_i^\alpha$, $T_1(\alpha) = \sum_{i \in I_1 \cup I_2} x_{1i}^\alpha + \sum_{i \in I_0} x_i^\alpha$, $T_2(\alpha) = \sum_{i \in I_1 \cup I_2} x_{2i}^\alpha + \sum_{i \in I_0} x_i^\alpha$.

Then, the marginal posterior of α is

$$l(\alpha | D) = \frac{\alpha^{n_0+2n_1+2n_2+a-1} e^{-b\alpha} \prod_{i \in I_1} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_2} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_0} x_i^{\alpha-1}}{(T_0(\alpha) + b_0)^{a_0+n_0} (T_1(\alpha) + b_1)^{a_1+n_1} (T_2(\alpha) + b_2)^{a_2+n_2}}. \quad (10)$$

Denote that $\theta = (\alpha, \lambda_0, \lambda_1, \lambda_2)$. Then, the Bayesian estimator of θ under the squared error loss function is

$$\hat{\theta}_B = \frac{\int_0^\infty \int_0^\infty \int_0^\infty \theta L(\alpha, \lambda_0, \lambda_1, \lambda_2 | D) d\lambda_0 d\lambda_1 d\lambda_2}{\int_0^\infty \int_0^\infty \int_0^\infty L(\alpha, \lambda_0, \lambda_1, \lambda_2 | D) d\lambda_0 d\lambda_1 d\lambda_2}. \quad (11)$$

Kundu and Gupta^[1] proposed the following importance sampling procedure to obtain $\hat{\theta}_B$:

1. Generate α_1 from the log-concave density $l(\alpha | D)$ as given in (10), using the method proposed by [7].

2. Generate

$$\begin{aligned}\lambda_{01} | \alpha, D &\sim \text{Gamma}(\lambda_0; a_0 + n_0, T_0(\alpha) + b_0), \\ \lambda_{11} | \alpha, D &\sim \text{Gamma}(\lambda_1; a_1 + n_1, T_1(\alpha) + b_1), \\ \lambda_{21} | \alpha, D &\sim \text{Gamma}(\lambda_2; a_2 + n_2, T_2(\alpha) + b_2).\end{aligned}$$

3. Repeat steps 1 and 2 to obtain $\{(\alpha_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}); i = 1, 2, \dots, N\}$.

4. A consistent estimator of (11) can be obtained as

$$\frac{\sum_{i=1}^N \theta_i h(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})}{\sum_{i=1}^N h(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})},$$

where $\theta_i = (\alpha_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, and $h(\lambda_0, \lambda_1, \lambda_2) = (\lambda_0 + \lambda_2)^{n_1} (\lambda_0 + \lambda_1)^{n_2}$.

Suppose for $0 < p < 1$, θ_p satisfies $P(\theta \leq \theta_p | D) = p$. Let

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

Rearrange $\{(\theta_1, w_1), (\theta_2, w_2), \dots, (\theta_N, w_N)\}$ as $\{(\theta_{(1)}, w_{(1)}), (\theta_{(2)}, w_{(2)}), \dots, (\theta_{(N)}, w_{(N)})\}$, here $\theta_{(1)} < \theta_{(2)} < \dots < \theta_{(N)}$, and $w_{(i)}$ s are associated with $\theta_{(i)}$, which are not ordered. Then a consistent Bayesian estimator of $\hat{\theta}_p = \theta_{N_p}$, where N_p is the integer satisfying

$$\sum_{i=1}^{N_p} w_{(i)} \leq p < \sum_{i=1}^{N_p+1} w_{(i)}. \quad (12)$$

Then using the above procedure, a $100(1 - \gamma)\%$ credible interval of θ can be obtained.

In the process of data simulation, we take parameters $(\alpha, \lambda_0, \lambda_1, \lambda_2) = (2, 1, 1, 1)$, and assign $a = b = 0.001$, $a_0 = b_0 = 1$, $a_1 = b_1 = 1$, $a_2 = b_2 = 1$. We get the relative bias (RB), mean squared error (MSE), 95% coverage probability of the parameters (CP) over 10,000 replications to explore the effectiveness of the method. In addition, we also compute the probabilities of $X_1 < X_2$, $X_1 > X_2$, $X_1 = X_2$ in the simulations, that is, $p_1 = P(X_1 < X_2) = \lambda_1/(\lambda_0 + \lambda_1 + \lambda_2)$, $p_2 = P(X_1 > X_2) = \lambda_2/(\lambda_0 + \lambda_1 + \lambda_2)$, $p_0 = P(X_1 = X_2) = \lambda_0/(\lambda_0 + \lambda_1 + \lambda_2)$. p_0, p_1 and p_2 are usually considered in the stress-strength models. See for example [8]. The results are shown in the Table 1. From Table 1, we can see that the importance sampling performs well when the sample size is small, i.e., $n = 15$. However, as the sample size get larger, the estimates of the parameters, especially the CPs of the parameters are much farther to the nominal level. This is because the efficiency of

the importance sampling method depends on the choice of the proposed density function. Kundu and Gupta^[1] used the gamma density function as the proposed density function (see step 2 of the importance sampling procedure), and the ratio between the proposed density function and the target density function is proportional to $h(\lambda_0, \lambda_1, \lambda_2)$, which is a function of sample size and the parameters. Thus, large sample size will make the proposed density function be far away from target density function. We show all the weights w_i s based on a simulated sample in Figure 1. From Figure 1, we see that the weight w_i based on a certain posterior sample of the parameters is extremely big, which makes the other weights negligible. Thus, the credible interval estimation will fail according to (12).

Table 1 The RB(%), MSE and 95% CP of the parameters based on 10,000 replications based on the method of [1] when $(\alpha, \lambda_0, \lambda_1, \lambda_2) = (2, 1, 1, 1)$

n		α	λ_0	λ_1	λ_2	p_0	p_1	p_2
15	RB(%)	1.10	8.71	10.97	16.30	1.34	0.64	0.43
	CP	0.959	0.877	0.931	0.934	0.952	0.952	0.945
	MSE	0.122	0.805	0.309	0.271	0.016	0.013	0.013
25	RB(%)	2.44	3.20	4.66	5.39	0.63	0.29	1.15
	CP	0.951	0.810	0.922	0.893	0.947	0.950	0.947
	MSE	0.066	0.500	0.146	0.224	0.010	0.008	0.009
50	RB(%)	4.88	5.39	0.70	3.43	2.04	2.91	2.44
	CP	0.922	0.595	0.850	0.807	0.955	0.947	0.943
	MSE	0.039	0.274	0.067	0.216	0.006	0.005	0.006
100	RB(%)	5.91	15.41	6.11	10.69	4.47	6.07	4.19
	CP	0.860	0.219	0.661	0.606	0.958	0.933	0.940
	MSE	0.028	0.116	0.038	0.234	0.004	0.004	0.004

§3. Main Results

3.1 Latent Variables

From (3), there is a missing data (w_{0i}, w_{1i}, w_{2i}) observed from (W_0, W_1, W_2) for each (x_{1i}, x_{2i}) from $\text{MOBW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$. For $(x_{1i}, x_{2i}) \in D_1$, that is, $x_{1i} < x_{2i}$, there are two cases of order for w_{0i} , w_{1i} and w_{2i} : $w_{1i} < w_{0i} < w_{2i}$ and $w_{1i} < w_{2i} < w_{0i}$. Thus, we introduce y_i , which is defined as

$$y_i = \begin{cases} 0, & \text{if } w_{1i} < w_{0i} < w_{2i}; \\ 1, & \text{if } w_{1i} < w_{2i} < w_{0i}. \end{cases}$$

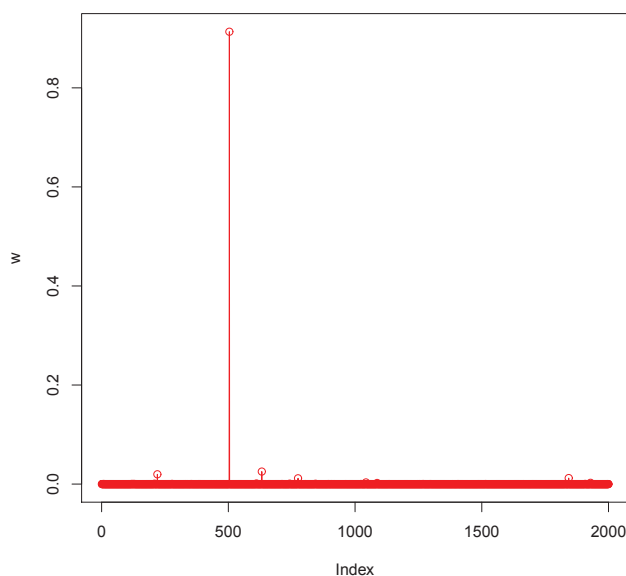


Figure 1 Trace plot of the weight based on the importance sampling

If $y_i = 0$, $w_{1i} = x_{1i}$, $w_{0i} = x_{2i}$ and $w_{2i} > x_{2i}$; otherwise, $w_{1i} = x_{1i}$, $w_{2i} = x_{2i}$ and $w_{0i} > x_{2i}$. Thus, the likelihood function of the complete data (x_{1i}, x_{2i}, y_i) is

$$g_{1i} = [f_W(x_{1i} | \alpha, \lambda_1) f_W(x_{2i} | \alpha, \lambda_0) S_W(x_{2i} | \alpha, \lambda_2)]^{1-y_i} \\ \times [f_W(x_{1i} | \alpha, \lambda_1) f_W(x_{2i} | \alpha, \lambda_2) S_W(x_{2i} | \alpha, \lambda_0)]^{y_i},$$

Hence, the likelihood function of the augmented data $\{(x_{1i}, x_{2i}, y_i), i \in I_1\}$ is $G_1 = \prod_{i \in I_1} g_{1i}$. Similarly, for $(x_{1i}, x_{2i}) \in D_2$, a latent variable z_i is defined as

$$z_i = \begin{cases} 0, & \text{if } w_{2i} < w_{0i} < w_{1i}; \\ 1, & \text{if } w_{2i} < w_{1i} < w_{0i}. \end{cases}$$

The likelihood function of (x_{1i}, x_{2i}, z_i) is

$$g_{2i} = [f_W(x_{2i} | \alpha, \lambda_2) f_W(x_{1i} | \alpha, \lambda_0) S_W(x_{1i} | \alpha, \lambda_1)]^{1-z_i} \\ \times [f_W(x_{2i} | \alpha, \lambda_2) f_W(x_{1i} | \alpha, \lambda_1) S_W(x_{1i} | \alpha, \lambda_0)]^{z_i},$$

Then we could know the likelihood function of the augmented data $\{(x_{1i}, x_{2i}, z_i), i \in I_2\}$ is $G_2 = \prod_{i \in I_2} g_{2i}$. If $(x_{1i}, x_{2i}) \in D_0$, then from (5), the likelihood function of (x_{1i}, x_{2i}) is

$$g_{0i} = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_W(x_i | \alpha, \lambda_0 + \lambda_1 + \lambda_2).$$

Therefore, the likelihood function of the observed data D_0 is

$$G_0 = \prod_{i \in I_0} g_{0i} = \prod_{i \in I_0} \lambda_0 \alpha \exp\{-(\lambda_0 + \lambda_1 + \lambda_2)x_i^\alpha\}.$$

Thus, we can obtain the likelihood function of the complete data $\mathcal{D} = \{(x_{1i}, x_{2i}, y_i), i \in I_1; (x_{1i}, x_{2i}, z_i), i \in I_2; (x_{1i}, x_{2i}), i \in I_0\}$ is

$$\begin{aligned} L(\mathcal{D} | \alpha, \lambda_0, \lambda_1, \lambda_2) &= G_0 \cdot G_1 \cdot G_2 = \prod_{i \in I_0} g_{0i} \prod_{i \in I_1} g_{1i} \prod_{i \in I_2} g_{2i} \\ &= \lambda_0^{n_0 + \sum_{i \in I_1} (1-y_i) + \sum_{i \in I_2} (1-z_i)} \lambda_1^{n_1 + \sum_{i \in I_2} z_i} \lambda_2^{n_2 + \sum_{i \in I_1} y_i} \alpha^{n_0 + 2(n_1 + n_2)} \\ &\quad \times \prod_{i \in I_1} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_2} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_0} x_i^{\alpha-1} \\ &\quad \times \exp\{-\lambda_0 \cdot T_0(\alpha) - \lambda_1 \cdot T_1(\alpha) - \lambda_2 \cdot T_2(\alpha)\}. \end{aligned} \quad (13)$$

3.2 Gibbs Sampling

Based on (13) and (8), we have the joint posterior density function of $(\alpha, \lambda_0, \lambda_1, \lambda_2)$ is

$$\begin{aligned} L(\alpha, \lambda_0, \lambda_1, \lambda_2 | \mathcal{D}) &= \pi_0(\lambda_0 | a_0, b_0) \pi_1(\lambda_1 | a_1, b_1) \pi_2(\lambda_2 | a_2, b_2) \pi(\alpha | a, b) L(\mathcal{D} | \alpha, \lambda_0, \lambda_1, \lambda_2) \\ &\propto \lambda_0^{n_0 + \sum_{i \in I_1} (1-y_i) + \sum_{i \in I_2} (1-z_i) + a_0 - 1} \lambda_1^{n_1 + \sum_{i \in I_2} z_i + a_1 - 1} \lambda_2^{n_2 + \sum_{i \in I_1} y_i + a_2 - 1} \\ &\quad \times \alpha^{n_0 + 2(n_1 + n_2) + a - 1} \\ &\quad \times \prod_{i \in I_1} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_2} (x_{1i} x_{2i})^{\alpha-1} \prod_{i \in I_0} x_i^{\alpha-1} \exp\{-b\alpha\} \\ &\quad \times \exp\{-\lambda_0(T_0(\alpha) + b_0) - \lambda_1(T_1(\alpha) + b_1) - \lambda_2(T_2(\alpha) + b_2)\}. \end{aligned} \quad (14)$$

It is observed that the Bayesian estimators cannot be obtained easily in explicit forms in general. We propose to use the Gibbs sampling to compute the Bayesian estimators. Before using the Gibbs sampling to do the corresponding calculation, we have to prove that $l(\alpha | D)$ is log-concave, where $l(\alpha | D)$ denotes that the posterior density function of α , when $\lambda_0, \lambda_1, \lambda_2$ is known in (14). See the proof in the Appendix.

Let $Y = \sum_{i \in I_1} y_i$, $Z = \sum_{i \in I_2} z_i$. Then the full conditional posterior densities of Y , Z , and θ are as follows.

1. The full conditional posterior densities of Y and Z , given $\alpha, \lambda_0, \lambda_1$ and λ_2 , are Binomial distributions. That is

$$Y \sim \text{Binom}\left(n_1, \frac{\lambda_2}{\lambda_0 + \lambda_2}\right), \quad Z \sim \text{Binom}\left(n_2, \frac{\lambda_1}{\lambda_0 + \lambda_1}\right).$$

2. The full conditional posterior densities of λ_0 , λ_1 and λ_2 are gamma distributions. That is,

$$\lambda_0 | \theta_{-\lambda_0} \sim \text{Gamma}(n - Y - Z + a_0, T_0(\alpha) + b_0),$$

$$\lambda_1 | \theta_{-\lambda_1} \sim \text{Gamma}(n_1 + Z + a_1, T_1(\alpha) + b_1),$$

$$\lambda_2 | \theta_{-\lambda_2} \sim \text{Gamma}(n_2 + Y + a_2, T_2(\alpha) + b_2),$$

where $\theta_{-\eta}$ denotes θ excluding η .

3. The full conditional posterior density of α is log-concave. Thus, the adaptive rejection sampling ([7]) can be used to generate α .

Then the Gibbs sampling can be implemented based on the above full conditional posterior densities, and we find that the convergence is attained by only 500 iterations in the simulations.

§4. Numerical Simulations

In this section, in order to verify the performance of the proposed Bayesian method, simulations are carried out to compare the Bayesian method with the MLEs obtained by EM algorithm ([5]).

In the process of simulation, we assume that the prior distributions of λ_0 , λ_1 and λ_2 are Gamma(1, 1). And the prior distribution of α is Gamma(α ; 0.001, 0.001), which is same as [1]. $(\alpha, \lambda_0, \lambda_1, \lambda_2) = (0.8, 1, 1.2, 2)$, $(1, 1, 1, 1)$ and $(2, 1.2, 1, 0.8)$. The RB, MSE, CP of the parameters have been computed over 10,000 replications. In the simulations, the Gibbs sampling algorithm converges after 500 iterations. Thus, we run the Gibbs sampling 2,000 iterations, discard the initial 500 burn-in iterations. The thinning interval is 3. Thus, 500 posterior samples are used to calculate the estimates of the parameters. The results are listed in Tables 2–4. From Tables 2–4, we see that

1. Both Bayesian method and maximum likelihood method perform better when the sample size increases.
2. When the sample size is small or moderate, both the RBs and the MSEs of the parameters based on the Bayesian method are smaller than these based on EM algorithm.
3. The CPs based on Bayesian method are much more close to the nominal level 0.95 even in the case of small sample size. While the CPs based on the EM algorithm are not so satisfying when the sample size $n = 15$ or 25.

4. When p_0, p_1 and p_2 are of interest, both Bayesian method and EM algorithm perform well even for the case of small sample size.

Table 2 The RB(%), MSE and 95% CP of the parameters based on 10,000 replications when $(\alpha, \lambda_0, \lambda_1, \lambda_2) = (0.8, 1, 1.2, 2)$

n			α	λ_0	λ_1	λ_2	p_0	p_1	p_2
15	Bayes	RB(%)	3.96	10.91	3.10	3.56	11.58	1.46	6.67
		CP	0.956	0.974	0.967	0.951	0.971	0.971	0.954
		MSE	0.018	0.175	0.201	0.361	0.008	0.007	0.010
	MLE	RB(%)	7.23	6.94	14.02	15.22	1.76	0.72	0.45
		CP	0.954	0.867	0.921	0.916	0.814	0.868	0.878
		MSE	0.024	0.344	0.487	1.031	0.016	0.013	0.014
25	Bayes	RB(%)	2.33	7.25	2.33	1.50	7.36	0.87	4.20
		CP	0.954	0.956	0.958	0.953	0.957	0.958	0.954
		MSE	0.010	0.120	0.136	0.256	0.006	0.005	0.006
	MLE	RB(%)	4.59	5.14	7.75	9.58	0.98	0.12	0.56
		CP	0.950	0.896	0.932	0.920	0.863	0.892	0.898
		MSE	0.013	0.194	0.237	0.494	0.009	0.008	0.008
50	Bayes	RB(%)	1.30	3.49	1.45	0.66	3.73	0.61	2.23
		CP	0.951	0.952	0.950	0.950	0.949	0.954	0.950
		MSE	0.005	0.061	0.075	0.144	0.003	0.003	0.003
	MLE	RB(%)	2.13	2.15	4.05	4.26	0.63	0.37	0.09
		CP	0.954	0.910	0.940	0.907	0.893	0.919	0.908
		MSE	0.005	0.083	0.103	0.201	0.004	0.004	0.004
100	Bayes	RB(%)	0.67	2.06	0.84	0.35	2.11	0.36	1.27
		CP	0.946	0.950	0.952	0.947	0.948	0.952	0.950
		MSE	0.002	0.030	0.036	0.074	0.002	0.002	0.002
	MLE	RB(%)	1.09	1.13	1.93	1.94	0.22	0.22	0.02
		CP	0.951	0.918	0.947	0.906	0.914	0.926	0.916
		MSE	0.003	0.038	0.046	0.088	0.002	0.002	0.002

§5. Data Analysis

A data set from [9] is reanalyzed for illustration. The data is from the UEFA Champion's League, and includes two variables: X_1 represents the time in minutes of the first kick goal scored by any team and X_2 represents the first goal of any type scored by the home team. Meintanis^[9] analyzed this data by using MOBE distribution. Kundu and Dey^[5], Kundu and Gupta^[1] suggested MOBW distribution to fit this data. As Kundu and Gupta^[1] suggested, all the data points have been divided by 100 so that the shape

Table 3 The RB(%), MSE and 95% CP of the parameters based on 10,000 replications when $(\alpha, \lambda_0, \lambda_1, \lambda_2) = (1, 1, 1, 1)$

n			α	λ_0	λ_1	λ_2	p_0	p_1	p_2
15	Bayes	RB(%)	6.58	7.40	7.35	6.83	1.90	0.73	1.17
		CP	0.948	0.955	0.957	0.957	0.962	0.962	0.958
		MSE	0.035	0.149	0.166	0.166	0.010	0.009	0.009
	MLE	RB(%)	7.90	10.62	11.52	10.73	1.52	0.57	0.96
		CP	0.948	0.911	0.922	0.909	0.873	0.883	0.881
		MSE	0.040	0.271	0.314	0.294	0.019	0.015	0.015
25	Bayes	RB(%)	3.83	4.09	5.06	5.08	0.66	0.34	0.32
		CP	0.944	0.949	0.956	0.953	0.955	0.954	0.952
		MSE	0.019	0.087	0.104	0.104	0.007	0.006	0.006
	MLE	RB(%)	4.68	5.57	6.68	6.82	0.34	0.24	0.10
		CP	0.951	0.920	0.935	0.920	0.904	0.909	0.905
		MSE	0.020	0.133	0.155	0.153	0.011	0.009	0.009
50	Bayes	RB(%)	1.95	2.58	2.65	2.33	0.80	0.25	0.56
		CP	0.946	0.950	0.947	0.949	0.948	0.950	0.948
		MSE	0.009	0.044	0.052	0.052	0.004	0.003	0.003
	MLE	RB(%)	2.18	3.09	3.07	2.70	0.66	0.16	0.50
		CP	0.950	0.934	0.946	0.929	0.925	0.928	0.927
		MSE	0.009	0.056	0.064	0.063	0.005	0.004	0.004
100	Bayes	RB(%)	0.79	0.93	1.17	0.95	0.31	0.04	0.26
		CP	0.948	0.952	0.949	0.949	0.952	0.950	0.951
		MSE	0.004	0.021	0.025	0.025	0.002	0.002	0.002
	MLE	RB(%)	1.12	1.40	1.54	1.19	0.27	0.04	0.32
		CP	0.950	0.936	0.952	0.935	0.939	0.943	0.935
		MSE	0.004	0.026	0.029	0.030	0.003	0.002	0.002

and scale parameters are of the same order. Then we use the proposed method to analyze the data set. 2,000 iterations are run, which is shown in Figure 2. And the corresponding Bayesian posterior means and 95% credible intervals of the parameters are listed in Table 5. We also list the estimates based on the EM algorithm ([5]) and Bayesian method ([1]), which are denoted as KD09 and KG13 in Table 5.

From the Table 5, we see that the point estimates based on the EM algorithm and the proposed Bayesian method are very close to each other. However, the lengths of interval estimates based on the proposed Bayesian method are much shorter. The estimates based on KG13 are much different, and the lengths of interval estimates are the shortest. However, as we have indicated in Section 2, the interval estimates of [1] are not reliable

Table 4 The RB(%), MSE and 95% CP of the parameters based on 10,000 replications when $(\alpha, \lambda_0, \lambda_1, \lambda_2) = (2, 1.2, 1, 0.8)$

n			α	λ_0	λ_1	λ_2	p_0	p_1	p_2
15	Bayes	RB(%)	6.71	3.79	8.64	13.17	2.32	0.57	4.18
		CP	0.941	0.959	0.960	0.969	0.962	0.960	0.973
		MSE	0.154	0.165	0.178	0.143	0.011	0.009	0.008
	MLE	RB(%)	6.64	11.01	12.86	12.51	0.65	0.55	0.29
		CP	0.964	0.934	0.924	0.906	0.896	0.886	0.875
		MSE	0.139	0.331	0.344	0.251	0.020	0.015	0.014
25	Bayes	RB(%)	3.86	2.67	5.29	8.07	1.20	0.51	2.44
		CP	0.942	0.950	0.953	0.954	0.954	0.953	0.950
		MSE	0.079	0.105	0.108	0.088	0.007	0.006	0.006
	MLE	RB(%)	4.55	6.55	6.74	5.74	1.09	0.47	1.05
		CP	0.957	0.943	0.938	0.919	0.921	0.920	0.900
		MSE	0.080	0.149	0.158	0.123	0.011	0.009	0.008
50	Bayes	RB(%)	2.10	1.58	2.87	3.64	0.34	0.21	0.77
		CP	0.945	0.951	0.951	0.947	0.951	0.952	0.948
		MSE	0.036	0.052	0.054	0.044	0.004	0.003	0.003
	MLE	RB(%)	2.26	3.39	3.07	1.83	0.90	0.13	1.18
		CP	0.953	0.942	0.950	0.930	0.932	0.933	0.924
		MSE	0.036	0.067	0.066	0.052	0.006	0.004	0.004
100	Bayes	RB(%)	1.05	0.70	1.58	2.31	0.39	0.11	0.73
		CP	0.944	0.950	0.950	0.950	0.948	0.948	0.949
		MSE	0.017	0.026	0.027	0.022	0.002	0.002	0.002
	MLE	RB(%)	1.12	1.40	1.33	1.12	0.31	0.14	0.29
		CP	0.948	0.946	0.952	0.937	0.942	0.943	0.937
		MSE	0.017	0.030	0.031	0.025	0.003	0.002	0.002

Table 5 Parameter estimates based on three methods

method		α	λ_0	λ_1	λ_2
KD09	MLE	1.695	2.692	1.219	2.805
	95% confidence interval	(1.328, 2.062)	(1.500, 3.885)	(0.270, 2.141)	(1.202, 4.448)
KG13	Posterior mean	1.705	2.075	0.957	3.037
	95% credible interval	(1.328, 1.892)	(1.485, 2.357)	(0.505, 1.482)	(2.190, 3.557)
Ours	Posterior mean	1.670	2.577	1.209	2.625
	95% credible interval	(1.386, 1.999)	(1.666, 3.749)	(0.530, 2.061)	(1.573, 4.023)

when sample size is moderate or large.

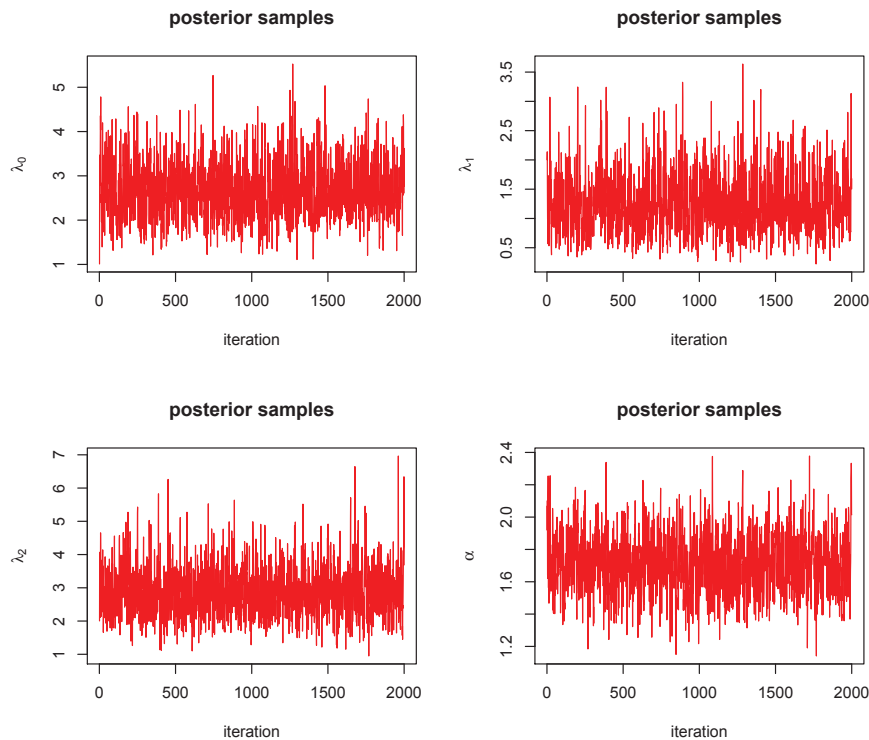


Figure 2 Trace plot of the posterior samples of the parameters

§6. Conclusions

In this paper, due to inefficiency of importance sampling method by [1], we have proposed a Bayesian method to estimate the parameters of MOBW distribution. We use Gamma prior of the parameters, and introduce latent variables to simplify the likelihood function. Then a MCMC procedure is given to obtain the Bayesian estimates. The proposed method is compared with EM algorithm via simulations. We find that the Bayesian estimates perform much better when the sample size is small or moderate.

Appendix

The conditional posterior density function of α is log-concave.

Proof

$$\frac{\partial^2 l(\alpha | D)}{\partial \alpha^2} = -\lambda_0 \left(\sum_{i \in I_2} x_{1i}^\alpha (\ln x_{1i})^2 + \sum_{i \in I_1} x_{2i}^\alpha (\ln x_{2i})^2 + \sum_{i \in I_0} x_i^\alpha (\ln x_i)^2 \right)$$

$$\begin{aligned}
& -\lambda_1 \left(\sum_{i \in I_1 \cup I_2} x_{1i}^\alpha (\ln x_{1i})^2 + \sum_{i \in I_0} x_i^\alpha (\ln x_i)^2 \right) \\
& -\lambda_2 \left(\sum_{i \in I_1 \cup I_2} x_{2i}^\alpha (\ln x_{2i})^2 + \sum_{i \in I_0} x_i^\alpha (\ln x_i)^2 \right) < 0.
\end{aligned}$$

Thus, the result holds. \square

References

- [1] Kundu D, Gupta A K. Bayes estimation for the Marshall-Olkin bivariate Weibull distribution [J]. *Comput. Statist. Data Anal.*, 2013, **57**(1): 271–281.
- [2] Marshall A W, Olkin I. A multivariate exponential distribution [J]. *J. Amer. Statist. Assoc.*, 1967, **62**(317): 30–44.
- [3] Guan Q, Tang Y C, Xu A C. Objective Bayesian analysis for bivariate Marshall-Olkin exponential distribution [J]. *Comput. Statist. Data Anal.*, 2013, **64**: 299–313.
- [4] Bemis B M, Bain L J, Higgins J J. Estimation and hypothesis testing for the parameters of a bivariate exponential distribution [J]. *J. Amer. Statist. Assoc.*, 1972, **67**(340): 927–929.
- [5] Kundu D, Dey A K. Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm [J]. *Comput. Statist. Data Anal.*, 2009, **53**(4): 956–965.
- [6] Neal R M. Annealed importance sampling [J]. *Stat. Comput.*, 2001, **11**(2): 125–139.
- [7] Gilks W R, Wild P. Adaptive rejection sampling for Gibbs sampling [J]. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 1992, **41**(2): 337–348.
- [8] Kotz S, Lumelskii Y, Pensky M. *The Stress-Strength Model and Its Generalizations: Theory and Applications* [M]. New Jersey: World Scientific, 2003.
- [9] Meintanis S G. Test of fit for Marshall - Olkin distributions with applications [J]. *J. Statist. Plann. Inference*, 2007, **137**(12): 3954–3963.
- [10] Dinse, G E. Nonparametric estimation for partially-complete time and type of failure data [J]. *Biometrics*, 1982, **38**(2): 417–431.

Marshall-Olkin威布尔分布的贝叶斯分析

张 慧 徐安察

(温州大学数学与信息科学学院, 温州, 325035)

摘 要: Kundu与Gupta^[1]提出用重要抽样法来计算Marshall-Olkin两元威布尔分布参数的贝叶斯估计, 然而我们发现在样本量变大的情况下, 重要抽样算法的参数估计效果却不理想. 在这篇文章中, 我们引入潜在变量来简化似然函数, 并且提出利用MCMC算法实现对该模型未知参数的估计. 为了评价我们提出方法的有效性, 我们将提出的贝叶斯方法与极大似然估计数据模拟结果作对比, 可以发现: 即使在样本量很小的情况下, 提出的贝叶斯方法的参数估计效果更理想. 实例分析也说明了这一点.

关键词: Marshall-Olkin威布尔分布; 贝叶斯估计; MCMC算法

中图分类号: O212.8