Estimating Beta Inverse Weibull Distribution by the Method of Quantile Estimates

Faysal Satter, Mezbahur Rahman

Minnesota State University, Mankato, MN 56002, USA

Abstract: **The Beta Inverse Weibull distribution is one of the widely applied distribution. In reliability and biological studies, we can model failure rates by this distribution. In this paper, we extend the study of the Beta Inverse Weibull distribution (BIW). The unimodal BIW with three parameters are estimated by the Method of Quantile Estimates (MQE) and then compared with parameters estimated by the Maximum Likelihood Estimates (MLE). To estimate the model parameters, Newton-Raphson method is applied for MLE and Nelder-Mead Simplex method is applied for MQE and the table of simulation results is shown. Finally, practical use of the model is demonstrated by dint of an application to real data. For the real data computation, standard deviations of the estimated parameters are calculated by random permutation method.**

Key Words: **Nelder-Mead Simplex Method; Newton-Raphson root finding method; quantiles, bootstrap random permutation.**

Running Title: Method of Quantile Estimates in Beta Inverse Weibull distribution

I. INTRODUCTION

The Beta Inverse Weibull distribution (BIW) is derived from a generalized class of probability distributions discussed by Eugene et al. (2002). Let $F(x)$ be the cumulative distribution function (cdf) of a random variable X. The cdf for a generalized Beta class of distribution for the random variable X, defined by Eugene et al. (2002) as :

$$G(x) = I_{F(x)}(a,b), \quad a > 0 \text{ and } b > 0$$

where

$$I_{F(x)}(a,b) = \frac{B_{F(x)}(a,b)}{B(a,b)} \text{ and}$$

$$B_{F(x)}(a,b) = \int_0^{F(x)} t^{a-1}(1-t)^{b-1}dt \qquad (1)$$

and $B(a,b)=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

The BIW distribution is obtained by taking $F(x)$ in (1) as the cdf of Inverse Weibull distribution which is given by:

$$F(x) = e^{-x^{-\beta}} \qquad (2)$$

Thus by plugging (2) in (1), the cdf of BIW distribution is given as:

$$G(x) = \frac{1}{B(a,b)}\int_0^{e^{-x^{-\beta}}} t^{a-1}(1-t)^{b-1}dt, \; a > 0 \text{ and } b > 0 \qquad (3)$$

The probability density function is obtained by taking derivative of (3):

$$g(x) = \frac{1}{B(a,b)}\left(e^{-x^{-\beta}}\right)^a\left(1-e^{-x^{-\beta}}\right)^{b-1}\beta x^{-(\beta+1)} \qquad (4)$$

The random variables $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$, are defined as an ordered random sample from the BIW distribution (4). In this paper, the method of quantile estimates is implemented in estimating parameters in a three parameter BIW distribution. The parameters estimated by the method of quantile estimates are then compared with the parametes estimated by the method of maximum likelihood using simulation.

Beta Inverse Weibull distribution was also introduced and important properties were summarized by Khan (2010) and Hanooka et al. (2013).

II. ESTIMATION PROCEDURES

2.1 Method of Quantile Estimates (MQE)

The Method of Quantile Estimates (QE) uses the quantiles of the corresponding distribution. Percentiles or quantiles are used to estimate the parameters of the distribution. Schmid (1997) considered percentile estimators for the three-parameter Weibull distribution and Castillo and adi (1995) estimates the parameters of continuous distribution by using the quantiles. For historical background and for other details readers are referred to these two references. Quantile estimates (QE) in general can be summarized as follows:

Let the parameters to be estimated are $\theta = \{\theta_1, \theta_2, \ldots, \theta_r\}$ and $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics obtained from a random sample from $F(x;\theta)$, where, for fixed θ , $F(x;\theta)$ is assumed to be strictly increasing on the interior of its support. Also, consider a set of r distinct indices $I = \{i_1, i_2, \ldots, i_r\}$, , where $i_j \in \{1, 2, \ldots, n\}, j = \{1, 2, \ldots, r\}$. Then, one can write

$$F(x_{i:n};\theta) \cong p_{i:n}, i \in I$$

or, equivalently,

$$x_{i:n} = F^{-1}(p_{i:n}; \theta), i \in I \quad (5)$$

where

$$p_{i:n} = \frac{(i-a)}{n+b}$$

is an empirical distribution of $F(x_{i:n}; \theta)$ or suitable plotting positions, and a and b are constants. The values of a and b are chosen (either theoretically or based on simulation) so that the resulting estimators have certain desirable properties (e.g., the minimum root mean square error). Replacing the approximation by equality in (5), we get a set of r independent equations in r unknowns, $\theta_1, \theta_2, \dots, \theta_r$. An elemental estimate of θ can then be obtained by solving (5) for θ. Note that these elemental estimates are based on the percentile method.

The estimates obtained from (5) depend on r observations. A subset of size r observations is known as an elemental subset and the resultant estimate is known as an elemental estimate of θ. Thus, for a sample of size n, there are ${}_nC_r$ elemental estimates. For large ${}_nC_r$, the number of elemental subsets may be too large for the computations of all elemental estimates to be feasible. In such cases, instead of computing all possible elemental estimates, one may select a pre-specified number, N, of elemental subsets either systemically, based on some theoretical considerations, or at random.

For each of these subsets, an elemental estimate of θ is computed and is denoted as $\hat{\theta}_{j1}, \hat{\theta}_{j2}, \dots, \hat{\theta}_{jN}$, *j*=1,2,...,*r*. These elemental estimates can then be combined, using some suitable (preferably robust) functions, to obtain an overall final estimate of θ. A commonly used robust function is the median (MED). That is,

$$\hat{\theta} = median(\hat{\theta}_{j1}, \hat{\theta}_{j2}, \dots, \hat{\theta}_{jN}), j=1,2,\dots,r.$$

The estimates are unique even when the method of moments (MOM) and the MLE equations have multiple solutions or when the MOM and the MLE do not exist.

We can set up a system of nonlinear equations as:

$$\frac{i}{n+1} = \frac{1}{B(a,b)} \int_0^{e^{-x_i^{-\beta}}} t^{a-1}(1-t)^{b-1} dt$$

$$\frac{j}{n+1} = \frac{1}{B(a,b)} \int_0^{e^{-x_j^{-\beta}}} t^{a-1}(1-t)^{b-1} dt$$

$$\frac{k}{n+1} = \frac{1}{B(a,b)} \int_0^{e^{-x_k^{-\beta}}} t^{a-1}(1-t)^{b-1} dt \quad (6)$$

The x_i's in right side of the nonlinear system are coming from sampling. By solving the nonlinear system above, we can get an estimation of a, b and β.

Method used for optimizing the objective function:

Nelder-Mead Simplex Method

1. *Sort:* Evaluate (7) at the n+1 vertices and sort the vertices so that

$$f(x_1) \le f(x_2) \le \cdots \le f(x_{n+1}) \text{ holds.}$$

2. *Reflection:* Compute the reflection point x_r from

$$x_r = \bar{x} + \alpha(\bar{x} - x_{n+1})$$

Evaluate $f_r = f(x_r)$. If $f_1 \le f_r < f_n$, replace x_{n+1} with x_r.

3. *Expansion:* If $f_r < f_1$ then compute the expansion point x_e from

$$x_e = \bar{x} + \beta(x_r - \bar{x})$$

and evaluate $f_e = f(x_e)$. If $f_e < f_r$, replace x_{n+1} with x_e; otherwise replace x_{n+1} with x_r.

4. *Outside Contraction:* If $f_n \le f_r < f_{n+1}$, compute the outside contraction point $x_{oc} = \bar{x} + \gamma(x_r - \bar{x})$ and evaluate $f_{oc}=f(x_{oc})$.

If $f_{oc} < f_r$, replace x_{n+1} with x_{oc}; otherwise go to step 6.

5. *Inside Contraction:* If $f_r \ge f_{n+1}$, compute the inside contraction point x_{ic} from $x_{ic} = \bar{x} + \gamma(x_r - \bar{x})$ and evaluate $f_{ic}=f(x_{ic})$.

If $f_{ic} < f_{n+1}$, replace x_{n+1} with x_{ic}; otherwise, go to step 6.

6. *Shrink:* For $2 \le i \le n+1$, define $x_i = x_1 + \delta(x_i - x_1)$

Note: α, β, γ, and δ are respectively the reflection, expansion, contraction and shrink coefficient. Standard values of α,β,γ, and δ are 1, 2, $-\frac{1}{2}$ and $\frac{1}{2}$, respectively.

We can write the optimization function *T*(*x*) as:

$$T(x) = (f_i(x))^2 + (f_j(x))^2 + (f_k(x))^2 \quad (7)$$

where, $f_m(x) = \frac{1}{B(a,b)} \int_0^{e^{-x_m^{-\beta}}} t^{a-1}(1-t)^{b-1} dt - \frac{m}{n+1}$

for *m* = *I*, *j*, *k*.

2.2 Maximum Likelihood Estimates (MLE)

The likelihood function is the joint density function for all observations which can be written as:

$$g(x_1, x_2, \dots, x_n|a, b, \beta) = \prod_{i=1}^{n} g(x_i|a, b, \beta)$$

Log-likelihood is obtained by taking logarithm of the likelihood function,

$$L = \log g(x_1, x_2, \dots, x_n|a, b, \beta)$$

The log-likelihood function of the BIW distribution is:

$$L = n\, log\beta - nlog\, B(a,b) - a\sum_{i=1}^{n} x_i^{-\beta} + (b-1)\sum_{i=1}^{n}\log(1-e^{-x_i^{-\beta}}) + \sum_{i=1}^{n}\log(x_i^{-(\beta+1)}) \tag{8}$$

The first derivatives of log-likelihood function (8) with respect to a, b, and β are respectively,

$$L_a' = n\psi(a+b) - n\psi(a) - \sum_{i=1}^{n} x_i^{-\beta},$$

$$L_b' = n\psi(a+b) - n\psi(b) - \sum_{i=1}^{n}\log(1-e^{-x_i^{-\beta}}),$$

and

$$L_\beta' = \frac{n}{\beta} + a\sum_{i=1}^{n}(x_i^{-\beta}\log x_i) - (b-1)\sum_{i=1}^{n}\frac{e^{-x_i^{-\beta}}x_i^{-\beta}\log x_i}{1-e^{-x_i^{-\beta}}} - \sum_{i=1}^{n}\log x_i$$

The second derivatives of log-likelihood function (8) with respect to a, b, and β are respectively,

$$L_{aa}'' = n\psi_1(a+b) - n\psi_1(a),$$

$$L_{ab}'' = n\psi_1(a+b),$$

$$L_{a\beta}'' = \sum_{i=1}^{n}(x_i^{-\beta}\log x_i),$$

$$L_{bb}'' = n\psi_1(a+b) - n\psi_1(b),$$

$$L_{b\beta}'' = \sum_{i=1}^{n}\frac{e^{-x_i^{-\beta}}x_i^{-\beta}\log x_i}{e^{-x_i^{-\beta}}-1},$$

and

$$L_{\beta\beta}'' = -\frac{n}{\beta^2} + a\sum_{i=1}^{n}(x_i^{-\beta}\log x_i^2) + (b-1)\sum_{i=1}^{n}\log x_i^2\left[\frac{e^{-x_i^{-\beta}}}{x_i^{2\beta}(e^{-x_i^{-\beta}}-1)} - \frac{e^{-x_i^{-\beta}}x_i^{-\beta}}{e^{-x_i^{-\beta}}-1} - \frac{e^{-2x_i^{-\beta}}}{x_i^{2\beta}(e^{-x_i^{-\beta}}-1)^2}\right]$$

where $\psi_1(a) = \frac{\Gamma(a)\Gamma''(a) - (\Gamma'(a))^2}{(\Gamma(a))^2}$ and $\Gamma''(a)$ is the derivative of $\Gamma'(a)$ with respect to a. $\psi_1(a)$ is also known as the Tri-gamma function.

Maximum likelihood of the parameters can be obtained by setting these partial derivatives to zero and then solving the system of the nonlinear system.

Three parameters can be easily estimated by the Newton-Raphson method. The multivariate Newton-Raphson iteration is performed as

$$\begin{bmatrix}\hat{a}_L^{(l+1)}\\ \hat{b}_L^{(l+1)}\\ \hat{\beta}_L^{(l+1)}\end{bmatrix} = \begin{bmatrix}\hat{a}_L^{(l)}\\ \hat{b}_L^{(l)}\\ \hat{\beta}_L^{(l)}\end{bmatrix} - \begin{bmatrix}L_{aa}''^{(l)} & L_{ab}''^{(l)} & L_{a\beta}''^{(l)}\\ L_{ab}''^{(l)} & L_{bb}''^{(l)} & L_{b\beta}''^{(l)}\\ L_{a\beta}''^{(l)} & L_{b\beta}''^{(l)} & L_{\beta\beta}''^{(l)}\end{bmatrix}^{-1}\begin{bmatrix}L_a'^{(l)}\\ L_b'^{(l)}\\ L_\beta'^{(l)}\end{bmatrix} \tag{9}$$

where l is the index for the iteration.

III. SIMULATION RESULTS

One thousand samples are generated for two different parameter settings (a = 2, b = 2, β = 2) and (a = 3, b = 2, β =1) and for three different sample sizes n =100, n = 500 and n = 1000. A = 1,b = 1 and β = 1 are taken as initial guess. Means (MEAN), standard deviations (SD), biases (BIAS),mean of the absolute biases (MAB) and mean squared errors (MSE) are computed and displayed for MQE and MLE in Table 1 and Table 2 respectively.

MATLAB software is used in all computations and the codes are readily available.

Table 1: Simulation Results by the Method of Quantile Estimates

n=100						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	2	2.4571	2.2998	0.4571	1.2086	5.4981
b	2	2.3645	2.7169	0.3645	1.3254	7.5142
β	2	2.3076	0.8542	0.3076	0.6890	0.8243
n=500						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	2	1.9895	0.6519	-0.0105	0.4499	0.4250
b	2	1.9576	0.7715	0.0424	0.5363	0.5971
β	2	2.1308	0.3803	0.1308	0.3196	0.1618
n=1000						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	2	1.9393	0.3755	-0.0607	0.3012	0.1447
b	2	1.9049	0.4444	-0.0951	0.3629	0.2065
β	2	2.1038	0.2667	0.1038	0.2223	0.0819
n=100						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	3	3.9052	2.2597	0.9052	1.5681	5.9257
b	2	1.8214	1.3479	-0.1786	0.9509	1.8486
β	1	1.2620	0.4528	0.2620	0.3794	0.2737
n=500						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	3	2.9880	0.6778	-0.0120	0.5260	0.4595
b	2	1.7210	0.5439	-0.2790	0.5122	0.3737
β	1	1.1393	0.1990	0.1393	0.1920	0.0590
n=1000						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	3	2.9307	0.4687	-0.0693	0.3777	0.2245
b	2	1.7642	0.4004	-0.2358	0.3908	0.2159
β	1	1.0981	0.1393	0.0981	0.1347	0.0290

Table 2: Simulation Results by the Maximum Likelihood Estimates

n=100						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	2	2.6517	2.2190	0.6517	1.2782	5.3488
b	2	2.9365	3.0833	0.9365	1.6562	10.3839
β	2	2.1226	0.7858	0.1226	0.6253	0.6326
n=500						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	2	2.1239	0.6322	0.1239	0.4377	0.4150
b	2	2.1683	0.8103	0.1683	0.5390	0.6849
β	2	2.0100	0.3361	0.0100	0.2667	0.1131
n=1000						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	2	2.0564	0.3885	0.0564	0.2961	0.1541
b	2	2.0756	0.4787	0.0756	0.3605	0.2348
β	2	2.0067	0.2394	0.0067	0.1915	0.0564
n=100						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	3	3.6329	2.1257	0.6329	1.3762	4.9191
b	2	2.7744	2.4990	0.7744	1.4836	6.8448
β	1	1.0662	0.3933	0.0662	0.3113	0.1590
n=500						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	3	3.1250	0.7251	0.1250	0.5285	0.5414
b	2	2.1455	0.7622	0.1455	0.5356	0.6022
β	1	1.0112	0.1736	0.0112	0.1378	0.0303
n=1000						
Parameters	Values	MEAN	SD	BIAS	MAB	MSE
a	3	3.0623	0.4726	0.0623	0.3604	0.2272
b	2	2.0711	0.4807	0.0711	0.3598	0.2361
β	1	1.0046	0.1217	0.0046	0.0966	0.0148

IV. APPLICATION

The following data in Table 2 represents failure times of machine parts from manufacturer A and are taken from http://v8doc.sas.com/sashtml/stat/chap29/sect44.htm:

Table 3: Failure Times

620	470	260	89	388	242	103	100	39	460	284
1285	218	393	106	158	152	477	403	103	69	158
818	947	399	1274	32	12	134	660	548	381	203
871	193	531	317	85	1410	250	41	1101	32	421
32	343	376	1512	1792	47	95	76	515	72	1585
253	6	860	89	1055	537	101	385	176	11	565
164	16	1267	352	160	195	1279	356	751	500	803
560	151	24	689	1119	1733	2194	763	555	14	45
776	1									

The MQE estimates of the parameters are $\hat{a}$ = 0.9039, $\hat{b}$ = 5.4383, and $\hat{\beta}$ = 0.3486. Note that the data scaled by dividing the maximum value to avoid numerical computational complexities.

We get the standard deviation of the parameters using bootstrap random permutation. Standard deviation of $\hat{a}$ is 0.1179, Standard deviation of $\hat{b}$ is 0.0801, and Standard deviation of $\hat{\beta}$ is 0.0637.

In Table 4 we have displayed the computational results for the above data for MLE estimates.

Table 4: Estimates and the Variance-Covariance Matrix for MLE

			Variance-Covariance Matrix		
Parameter	Estimate	S.E	*a*	*b*	β
a	0.4492	0.2047	0.0419	0.2112	-0.0189
b	4.3545	1.2319	0.2112	1.5175	-0.0905
β	0.5046	0.0950	-0.0189	-0.0905	0.0090

V. CONCLUSION

The Beta Inverse Weibull distribution is introduced in this paper. Distribution parameters are estimated by the Method of Quantile Estimates and then compared with the estimates obtained by the Maximum Likelihood method. In terms of parameters *a*, *b* and β, the Method of Quantile Estimates perform better (have smaller BIAS and MSE) for sample sizes *n*=100 and *n*=500 and the Method of Maximum Likelihood Estimates perform better for sample size *n*=1000.

REFERENCES

[1]. Castillo, E. and Hadi, A. S. (1995). *A Method for Estimating Parameters and Quantiles of Distributions of Continuous Random Variables*. Computational Statistics & Data Analysis. 20, 421-439.

[2]. Cheng, R. C. H. and Amin, N. A. K. (1983). *Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin*. Journal of Royal Statistical Society, Series B, 45(3), 394-403.

[3]. Cheng, R. C. H. and Iles, T. C. (1987). *Corrected Maximum Likelihood in Non-Regular Problems*. Journal of Royal Statistical Society, Series B, 49(1), 95-101.

[4]. Choi, S. C. and Wette, R. (1969). *Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias*. Technometrics, 11(4), 683-690.

[5]. Dang, H. and Weerakkody, G. (2000). *Bounds for the Maximum Likelihood Estimates in Two-Parameter Gamma Distribution*. Journal of Mathematical Analysis and Applications, 245(1), 1-6.

[6]. Evans, M., Hastings, N. and Peacock, B. (2000). *Statistical Distributions, Third Edition*. John Wiley & Sons, Inc., New York.

[7]. Gusmao, F. R. S. de, Ortega, E. M. M and Cordeiro, G. M. (2011) *The Generalized inverse Weibull Distribution*". Springer-Verlag, Stat Papers (2011), 52:591-619, DOI 10.1007

[8]. Hanooka, S. Shahbaz, M. Q. Mohsin, and Kibria, B.M.G (2013). *A Note On Beta Inverse-Weibull Distribution*. Communications in Statistics - Theory and Methods, Volume 42, Number 2, Pages 320-335.

[9]. Harter, H. L. and Moore, A. H. (1965). *Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples*. Technometrics, 7(4), 639-643.

[10]. Jockovic, J. (2012). *Quantile Estimation for the Generalized Pareto Distribution with Application to Finance*. Yugoslav Journal of Operations Research, 22 (2012), Number 2, 297-311

[11]. Khan, M. S. (2010). *The Beta Inverse Weibull Distribution*. International Transactions in Mathematical Sciences and Computer, Volume 3, Number 1, Pages 113-119.

[12]. Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis, Second Edition*. John Wiley & Sons, Inc., New York.

[13]. Mathews, J. H and Fink, K. K (2004). *Numerical Methods Using Matlab*. 4th Edition, Prentice-Hall Inc., New Jersey, USA.

[14]. Pearson, L. M. and Rahman, M. (2008). *Quantile Estimates in Two-Parameter Gamma Distributions*. Journal of Statistical Research, 42(2) 141-149.

[15]. Rahman, M. and Pearson, L. M. (2003). *A Note on Estimating Parameters in Two-Parameter Pareto Distributions*. International Journal of Mathematical Education in Science and Technology, 34(2), 298-306.

[16]. Rahman, M. and Pearson, L. M. (2003). *A Note on Estimating Parameters in the Two-Parameter Weibull Distribution*. Journal of the Korean Data & Information Science Society, 14(4), 1091-1102.

[17]. Schmidt, U. (1997). *Percentile estimators for the three-parameter weibull distribution for use when all parameters are unknown*. Communications in Statistics — Theory and Methods. 26(3), 765-785.