

Original Research Article

Bayesian Inference Based on the Interval Grouped Data from the Weibull Model with Application

ABSTRACT

In this paper, based on the interval grouped data, Bayesian approach is used to obtain estimator for the Weibull scale parameter, and some lifetime such as the reliability and hazard functions. The estimation procedures have been developed and modified under squared error loss function (SELF) and general entropy loss function (GELF). The estimators are derived using the inverted gamma conjugate prior. Credible intervals and high posterior density (HPD) credible intervals are also obtained. Prediction for the future number of failures in the corresponding intervals is presented. Finally, real life example is applied to illustrate the performance of the estimation procedures.

Keywords: *Weibull distribution, Interval grouped data, Bayesian estimation, Loss functions, High posterior density estimator, Credible intervals, Prediction.*

1. INTRODUCTION

The Weibull distribution has been extensively used in life testing and reliability studies. Initially, it was proposed by Weibull (1951) for describing the fatigue failures from the wear out materials. Lieblein and Zelen (1956) have used the model for the ball bearing failures. Recently, the Weibull statistics have become the most popular to describe the breakdown events in solid dielectrics (Yazhou and Zhixin, 1995). Weibull distribution also provides a good model for lives of electrical and mechanical components or systems (Jiang and Kececioglu, 1992). Reliability and survival analysis based on the weibull distribution are applied and discussed by many authors (Lowe and Lewis, 1983; Lai and Murthy, 2003; Murthy et al. 2004; Govind et al., 2010).

The density function for the Weibull distribution with parameters α, β is given by

$$f(t) = \frac{\beta}{\sigma^\beta} t^{\beta-1} \exp\left(-\left(\frac{t}{\sigma}\right)^\beta\right) \quad ; \quad t > 0 ; \sigma > 0 ; \beta > 0$$

Using the transformation: $\theta = \sigma^\beta$, $\sigma = \theta^{\frac{1}{\beta}}$ implies

$$f(t) = \frac{\beta}{\theta} t^{\beta-1} \exp\left(-\frac{t^\beta}{\theta}\right) \quad ; \quad t > 0 ; \theta > 0 ; \beta > 0 \quad (1)$$

Hence, the hazard rate and the reliability functions are given respectively by

$$h(t) = \frac{\beta}{\theta} t^{\beta-1}, t > 0 \quad (2)$$

$$R(t) = \exp\left(-\frac{t^\beta}{\theta}\right), t > 0 \quad (3)$$

In the normal used conditions, the inspection units have long time for failure. Thus, having a complete failure data set is costly and time consuming. For this reason, censored life testing experiments are performed either by terminating the life test at a pre-assigned fixed time to have "Type I" censored data or after a pre-assigned number of failures to have "Type 2" censored data. In other situations when continuous monitoring is not feasible, the intermittent inspection life testing is the only probable procedure by testing the failure units periodically. In the reliability and survival data analysis, the data obtained from the intermittent inspection are often called interval grouped data. Interval grouped data are widely used in survival and reliability analysis (Gove and Fairweather, 1989; David and Tongdan, 2000; Jie, 2005; Shu-Jye, 2006)

Using different types of failure time's data, many researches are devoted for estimating the Weibull parameters. Nicholas et al. (1997) have proved the uniqueness of the maximum likelihood estimators obtained from a complete data. He also presented a regression method approach using both complete and censored data. Using singly censored data maximum likelihood and least square regression estimators are applied to the Weibull parameters by Montanari et al. (1997). Nadarajah (2005) has used progressively "Type II" censored data for the moments of the Weibull distribution. Kundu (2006) presents maximum likelihood and Bayesian inference procedures for the parameters when the data are mixture of "Type I" and "Type II" censoring schemes. Other referees of interest may be found in Berger and Sun (1993), Zhang and Tang (2007), Chen et al. (2007) and Govind et al. (2010).

In order to select a single value δ as the Bayesian estimator of θ , a loss function must be specified. A frequently used loss function is the squared error loss function (SELF) defined by $L_s(\theta; \delta) = (\theta - \delta)^2$ and the corresponding Bayesian estimator is the posterior mean (Berger, 1985). For mathematical tractability (Soliman and Elkahoul, 2005), the general entropy loss function (GELF) as a generalization of the entropy loss is defined by $L_g(\theta; \delta) = \left(\frac{\delta}{\theta}\right)^m - \theta \log\left(\frac{\delta}{\theta}\right) - 1$

Hence, the Bayesian estimator under GELF is obtained by $\theta_{Bg} = \left(E_{\tilde{\Pi}}(\theta^{-m})\right)^{-\frac{1}{m}}$, where $\tilde{\Pi}$ is the corresponding posterior.

2. LIKELIHOOD, PRIOR, AND POSTERIOR

Suppose that the lifetime T follow the Weibull distribution (θ, β) and assume the time scale line is divided into intervals by the constant inspection times $a_j, j = 1, 2, \dots, k$. Assuming $a_0 = 0$ and $a_{k+1} = \infty$, then the corresponding intervals are expressed as $I_1 = [0, a_1)$, $I_2 = [a_1, a_2)$, ..., $I_k = [a_{k-1}, a_k)$, $I_{k+1} = [a_k, \infty)$. If n units are put on the test for failure and the experimenter intermittently records the failure data which

consists of the number of failures f_j in the intervals $I_j, j = 1, 2, \dots, k+1$ then the probability of failure in the corresponding intervals are

$$P_j(\theta, \beta) = R(a_{j-1}) - R(a_j) = \exp\left(\frac{-a_{j-1}^\beta}{\theta}\right) - \exp\left(\frac{-a_j^\beta}{\theta}\right) \quad (4)$$

$$= \exp\left(\frac{-a_{j-1}^\beta}{\theta}\right) \left(1 - \exp\left(\frac{a_j^\beta - a_{j-1}^\beta}{\theta}\right)\right), j = 1, 2, \dots, k+1$$

Since $\sum_{j=1}^{k+1} f_j = n$, $\sum_{j=1}^{k+1} P_j(\theta) = 1$, the likelihood function given the interval grouped data is

$$L(\theta | \beta, G) \propto \exp\left(\frac{-\sum_{j=1}^{k+1} a_{j-1}^\beta f_j}{\theta}\right) \prod_{j=1}^k \left(1 - \exp\left(\frac{-(a_j^\beta - a_{j-1}^\beta)}{\theta}\right)\right)^{f_j} \quad (5)$$

A fundamental element in the Bayesian estimation is to specify a respective prior for the indexed parameter. In this research, we adopt the natural conjugate prior density for the Weibull scale parameter, namely the inverted gamma prior given by

$$\Pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} \exp\left(\frac{-b}{\theta}\right), a > 0, b > 0 \quad (6)$$

The mean and the variance of the above prior are

$$E(\theta) = \frac{b}{a-1}, b > 0, a > 1 \quad (7)$$

$$V(\theta) = b^2 \left(\frac{1}{(a-1)^2(a-2)} \right), b > 0, a > 2 \quad (8)$$

if $a = 0, b = 0$ then we get the non-informative prior.

Setting $S = \sum_{j=1}^{k+1} a_{j-1}^\beta f_j + b$ and combining (5) and (6), the posterior function of θ is

$$\Pi(\theta, \beta | G) = \frac{\theta^{-(a+1)} \exp\left(\frac{-S}{\theta}\right) \prod_{j=1}^k \left(1 - \exp\left(\frac{-(a_j^\beta - a_{j-1}^\beta)}{\theta}\right)\right)^{f_j}}{\int_0^\infty \theta^{-(a+1)} \exp\left(\frac{-S}{\theta}\right) \prod_{j=1}^k \left(1 - \exp\left(\frac{-(a_j^\beta - a_{j-1}^\beta)}{\theta}\right)\right)^{f_j} d\theta} \quad (9)$$

105 To obtain the integral in the denominator of (9), $\prod_{j=1}^k \left(1 - \exp\left(-\frac{(a_j^\beta - a_{j-1}^\beta)}{\theta}\right) \right)^{f_j}$ is simplified

106 by substituting the average $Z = \frac{1}{k} \sum_{j=1}^k (a_j^\beta - a_{j-1}^\beta)$ in both numerator and denominator and

107 hence, the posterior is

108

109

$$\Pi(\theta | \beta, G) = \frac{\sum_{j=0}^r (r-j) \theta^{-(a+1)} \exp\left(-\frac{(jZ+S)}{\theta}\right)}{\sum_{j=0}^r (r-j) \int_0^\infty \theta^{-(a+1)} \exp\left(-\frac{(jZ+S)}{\theta}\right) d\theta} \quad (10)$$

110 Where, $r = \sum_{j=1}^k f_j = n - f_{k+1}$

111 Using the transformations:

112

$$\theta = \frac{\beta t^{\beta-1}}{h(t)}, h(t) > 0, \quad \theta = \frac{t^{\beta-1}}{R(t)(\log(R(t)))^2}, 0 < R(t) < 1 \quad (11)$$

113

114 The posterior functions for both the hazard and the reliability functions are readily derived.

116 3. BAYESIAN ESTIMATION

117

118 In this section, based on the interval grouped data, the Bayesian and the HPD estimators for

119 the parameter θ are obtained and modified. Under SELF, the Bayesian estimator is given by

120 the mean of the posterior, defining:

121

$$\Psi(r, h, w) = \sum_{j=0}^r \binom{r}{j} (-1)^j h^{-w}, \text{ and taking the mean of } \Pi(\theta | \beta, G), \text{ the Bayesian}$$

122 estimator of θ under SELF is given from the following formula

123

124

$$\theta_{SB} = \frac{\Psi(r, jZ + S, a-1)}{(a-1)\Psi(r, jZ + S, a)} \quad (12)$$

125

126 and under GELF of shape parameter m , the Bayesian estimator

127

$$\theta_{GB} = \left(\frac{\Gamma(a+m)}{\Gamma(a)} \right)^{\frac{-1}{m}} \left(\frac{\Psi(r, jZ + S, a+m)}{\Psi(r, jZ + S, a)} \right)^{\frac{-1}{m}} \quad (13)$$

128

129 To base the estimation on the sufficient statistics for θ given the interval grouped data

130

$$S = \sum_{j=1}^{k+1} a_{j-1}^\beta f_j + b, \text{ the posterior can be approximated by}$$

$$\Pi(\theta | \beta, G) = \frac{\theta^{-(a+r+1)} \exp\left(-\frac{S}{\theta}\right)}{\int_0^{\infty} \theta^{-(a+r+1)} \exp\left(-\frac{S}{\theta}\right) d\theta} \quad (14)$$

hence, Bayesian estimator under SELF is

$$\theta_{SB} = \frac{S}{a+r-1} \quad (15)$$

and the corresponding Bayesian estimator under GELF is

$$\theta_{GB} = S \left(\frac{\Gamma(a+r+m)}{\Gamma(a+r)} \right)^{\frac{-1}{m}} \quad (16)$$

Mathematically, the estimators in (15), and (16) are sufficient for computing and modifications.

If it is accepted for some specific loss function, the estimation can be based on the maximum likelihood principle. In the Bayesian inference, this leads to the mode of the posterior density or the HPD estimator. Since the posterior defined in (14) is unimodal, the HPD estimator of θ denoted by θ_{HPD} , can be obtained by solving the equation

$$\exp\left(-\frac{S}{\theta}\right) \theta^{-(a+r+1)} \left(\frac{S}{\theta^2} - \frac{(a+r+1)}{\theta} \right) = 0$$

This implies, the HPD estimator is given by

$$\theta_{HPD} = \frac{S}{a+r+1} \quad (17)$$

using the transformations in (11) the Bayesian estimators for the hazard rate and the reliability functions under SELF at a given specified t are respectively

$$h_s(t) = \frac{\beta(a+r)t^{\beta-1}}{S}, \quad R_s(t) = \left(\frac{S}{S+t^\beta} \right)^{a+r}, \quad t > 0 \quad (18)$$

Also, the Bayesian estimators for the hazard rate, and the reliability functions under GELF at a given specified time t are respectively

$$h_G(t) = \beta t^{\beta-1} S^{\frac{r+m}{m}} \left(\frac{\Gamma(a+r-m)}{\Gamma(a+r)} \right)^{\frac{-1}{m}}, \quad R_G(t) = \left(\frac{S}{S-mt^{\beta-1}} \right)^{\frac{-1}{m}} \quad (19)$$

154
155
156

4. CREDIBLE INTERVALS

Another common Bayesian inference to obtain intervals $[C_1, C_2]$ for the unknown parameter θ is probably to lie. Based on the posterior distribution, the interval $[C_1, C_2]$ is said to be a $(1 - \alpha)\%$ credible interval for θ if

$$\int_{C_1}^{C_2} \Pi(\theta | \beta, G) d\theta = 1 - \alpha \quad (20)$$

In choosing a credible interval for θ , it is usually desirable to minimize its size subject to the condition (20) which requires

$$\Pi(C_1 | \beta, G) = \Pi(C_2 | \beta, G) \quad (21)$$

an interval $[C_1, C_2]$ which simultaneously satisfies (20), and (21) is called the $(1 - \alpha)\%$ HPD credible interval. Substituting for $\Pi(\theta | \beta, G)$ from (14) and using the integral

$$\int_h^\infty \theta^{-(a+r+1)} \exp\left(-\frac{S}{\theta}\right) d\theta = S^{-(a+r)} \int_0^{\frac{S}{h}} w^{(a+r)-1} \exp(-w) dw$$

Implies HPD credible intervals can be obtained by choosing C_1, C_2 that satisfied the following two equations, simultaneously

$$I_g\left(\frac{S}{C_1}, a + r\right) - I_g\left(\frac{S}{C_2}, a + r\right) = 1 - \alpha \quad (22)$$

$$\left(\frac{C_1}{C_2}\right)^{-(a+r+1)} = \exp\left(S\left(\frac{1}{C_1} - \frac{1}{C_2}\right)\right) \quad (23)$$

where, $I_g(x, y) = \frac{1}{\Gamma(y)} \int_0^x t^{y-1} \exp(-t) dt$ is the incomplete gamma function

Using the transformations in (11), credible intervals and HPD credible intervals for the hazard and the reliability functions at a pre-given time t can be derived.

5. PREDICTION

In the Bayesian frame work related to life testing experiments, prediction mainly concern with the future order failure time. In this research, we approach the prediction for the future number of failures in a pre-given interval.

Assume a new sample of size $\mathbf{N}$ is put on the test for failure, if the corresponding intervals $I_1 = [0, a_1)$, $I_2 = [a_1, a_2)$, ..., $I_k = [a_{k-1}, a_k)$, $I_{k+1} = [a_k, \infty)$ are still fixed and the number of failures $f_j, j = 1, 2, \dots, k+1$ in these intervals for the pre sample of size n are given, then the distribution for each of the new number of failures $f_j^*, j = 1, 2, \dots, k+1$ for a new sample of size N is binomial $(\mathbf{N}, P_j(\theta))$ where $P_j(\theta)$ is as defined in (4). This implies the predictive posterior of f_j^* which is given by

$$\Pi^*(f_j^* = x) = \frac{\binom{N}{x} \int_0^\infty (P_j(\theta))^x (1 - P_j(\theta))^{N-x} \theta^{-(a+r+1)} \exp\left(\frac{-S}{\theta}\right) d\theta}{\int_0^\infty \theta^{-(a+r+1)} \exp\left(\frac{-S}{\theta}\right) d\theta}, x = 0, 1, 2, \dots, m \quad (24)$$

In particular, the predictive posterior of f_1^*, f_{k+1}^* are given by substituting $P_1(\theta) = 1 - \exp\left(\frac{-a_1^\beta}{\theta}\right)$, $P_{k+1}(\theta) = \exp\left(\frac{-a_k^\beta}{\theta}\right)$ for $P_j(\theta)$ in (24), respectively.

This implies, the expected prediction value of f_1, f_{k+1} are

$$E_{\Pi^*}(f_1^*) = \frac{S^{a+r}}{\Gamma(a+r)} \sum_{x=0}^N \binom{N}{x} \frac{x a_1^x S^{a+r} \Gamma(a+r+x)}{(S + a_1^\beta (N-x))^{a+r+1}} \quad (25)$$

$$E_{\Pi^*}(f_{k+1}^*) = \frac{S^{a+r}}{\Gamma(a+r)} \sum_{x=0}^N \binom{N}{x} \frac{x a_{k+1}^x S^{a+r} \Gamma(a+r+x)}{(S + a_{k+1}^\beta (N-x))^{a+r+1}} \quad (26)$$

6. APPLICATION AND CONCLUSION

To illustrate the performance of the Bayesian inference, real life example is presented in this section. The data represent number of million revolutions before failure for 23 ball bearings dividing each by 17 (Lawless, 2003). The data are as follows:
1.05, 1.70, 1.94, 2.44, 2.48, 2.68, 2.85, 3.05, 3.06, 3.18, 3.27, 3.99, 4.04, 4.04, 4.05, 4.95, 5.48, 5.80, 6.18, 6.23, 7.52, 7.53, 10.20.

The above ungrouped data are fit to the Weibull distribution by the maximum likelihood estimation method using Minitab with 95% confidence interval. The corresponding maximum likelihood estimators using the complete ungrouped data for the scale and the shape parameters are $\hat{\alpha} = 4.81615$, $\hat{\beta} = 2.10185$ implies the value of θ , $\hat{\theta}_{mle} = 27.2228$. Then the data are grouped into intervals of fixed length = 1.5. Two cases for the informative priors are considered. In the first case assuming $\text{var}(\theta) = 10$ and in the second case $\text{var}(\theta) = 5$.

In both cases, the prior mean is assumed to satisfy $E_{\Pi}(\theta) = \hat{\theta}_{mle} = 27.2228$. Thus, the available prior information indicates **a= 76.108**, **b= 2044.65** for the first case and **a= 150.2161**, **b= 4062.08** for the second. For the general loss the value of m is fixed to be 2 in

both cases. The Bayesian estimators θ_{SB} , θ_{GB} and θ_{HPD} are computed assuming the end interval points **a_k = 4.5, 6**, and **7.5** using equations (15), (16), and (17) and their efficiencies is computed as the ratio of their values to $\hat{\theta}_{mle}$. 90% symmetric Credible Intervals are derived using (22) and (23). The predicted values for the number of failures are also computed using (25) and (26).

233
234

Table (1): Case 1 performance with a= 76.108, and b= 2044.65

a_j	θ_{SB} efficiency		θ_{GB} efficiency		θ_{HPD} efficiency		90% Credible Interval	Real and Predicted values of f_{k+1} and f^*_{k+1}	
4.5	24.5267	90.09%	24.1254	88.62%	23.9941	88.14%	(20.60,29.40)	8	7.35
6	24.6645	90.60%	24.2738	89.16%	24.1459	88.69%	(20.75,29.33)	5	4.42
7.5	26.3243	96.69%	25.9160	95.19%	25.7822	94.71%	(22.55,29.21)	3	2.85

235
236

Table (2): Case 2 performance with a= 150.2161, and b= 4062.08

a_j	θ_{SB} efficiency		θ_{GB} efficiency		θ_{HPD} efficiency		90% Credible Interval	Real and Predicted values of f_{k+1} and f^*_{k+1}	
4.5	25.7434	94.55%	25.5105	93.71%	25.4336	93.42%	(22.62,29.24)	8	7.76
6	26.2696	96.49%	25.5691	93.92%	25.9535	95.33%	(22.70,29.23)	5	4.71
7.5	27.2000	99.91%	26.4832	97.28%	26.2312	96.35%	(23.53,29.20)	3	2.96

237
238
239
240
241
242
243
244
245
246
247

Despite there is a considerable loss of information about the exact failure times. From the results in Tables (1) and (2), Bayesian inference based on the interval grouped data show significance efficiency as compared to the maximum likelihood estimation using the complete ungrouped data. This efficiency increases as the number of intervals increases. The available prior information clearly affects the accuracy of the Bayesian estimates. This depends on the variance of the indicated priors. The developed computations for both the posterior and the corresponding estimators and prediction gave a high precision, and thus, can be used for any further inference.

248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275

REFERENCES

- Berger J.O. 1985. Statistical decision theory and Bayesian analysis. Springer. New York.
- Berger J.O., Sun D. 1993. Bayesian analysis for the Poly-Weibull distribution. Journal of American association. 88, 69-92.
- Chen J., Hung K., Lam L. 2007. Bayesian single and double variable sampling plans for the Weibull distributions with censoring. European Journal of Operation Research. 177(2), 1062-1073.
- David W.C. and Tongdan J. 2000. Gamma distribution parameter estimation for field data with missing failure times. IIE Transaction. 32, 1161-1166.
- Gove J.H, Fairweather S.E. 1989. Maximum likelihood estimation of the Weibull function parameters using a general interactive optimizer and grouped data. Forest Ecology and Management. 28, 61-69.
- Govind S., Mudholkar, Kobby O., Asubont E. 2010. Data-transformation approach to life times data analysis: An Overview. Journal of Statistical Planning Inference. 140, 1-14.
- Jiang S., Kececioglu D.1992. Maximum likelihood estimators from censored data, for mixed Weibull distributions. IEEE Transaction on Reliability. 41(2), 152-161.
- Jie M. 2005. MLE of parameters of location-scale distribution for complete and partially grouped data. Journal of statistical planning and inference. 1, 1-18.
- Kundu D. 2006. On hyprid Censored Weibull distribution. Journal of Statistical Planning Inference. 6-43.
- Lai C.D., Murthy D.N.P. 2003. Amodified Weibull distribution. IEEE Transaction on Reliability. 52(1), 33-37.
- Lawless J.F. 2003. Statistical Models and Methods for Lifetime Data, second edition. Wiley, New York, NY.
- Lieblein J., Zelen M. 1956. Statistical investigation of Fatigue life of Deep Groove ball Bearing. J.of Res. Natl.Bur.Std. 57, 273-315.
- Lowe P.H., Lewis W. 1983. Reliability analysis based on Weibull distribution: An application to maintenance float factor. Int.J.Prod. Res. 21, 461-470.

276 Montanari G.C, Mazzanti G., Cacciari M., Fothergill J.C. 1997. Optimum Estimators for
 277 Weibull distribution of censored data. IEEE Transaction on Dielectrics and Electrical
 278 Insulation. 14(4).
 279 Murthy D.N.P., Bulmer M., Eocleston J.A. 2004. Weibull model selection for Reliability
 280 Modeling. Reliab.Eng.Sys.Saf. 68(3), 257-267.
 281 Nadarajah S. 2005. On the moments of the modified Weibull distribution for progressively
 282 Type II censored samples. IEEE Transaction Reliability. 45, 374-380.
 283 Nicholas R. Farnum, Paul P. 1997. Uniqueness of maximum likelihood estimators of the two
 284 parameter Weibull distribution. IEEE Transaction on Reliability. 46(4), 133-137.
 285 Shu-Jye W., Ying-Polin, Yi-Juchen, 2006. Planning step stress life testing with progressively
 286 type I group-censored Exponential data. Statistica Neerlandica. 60, 46-56.
 287 Soliman A.A. Elkahout G.R. 2005. Bayes estimation of logistic distribution based on
 288 progressively censored data using Burr XII model. IEEE Transaction on Reliability. 14,
 289 281-293.
 290 Weibull W. 1951. A statistical distribution function of wide applicability. Journal of Applied
 291 Mechanics.11, 293-297.
 292 Yazhou J., Zhixin J. 1995. Probability distribution of machining center failures. Reliability
 293 Engineering and System Safety. 50, 121-125.
 294 Zhang L.E, Tang L.C. 2007. A study of two estimation approaches for parameters of Weibull
 295 distribution based on WPP. Reliab.Eng.Sys.Saf. 92, 360-368.
 296
 297
 298