Pak. J. Statist.
2017 Vol. 33(4), 249-267

BAYESIAN ANALYSIS OF RESTRICTED RANDOM CENSORSHIP MODEL WITH BURR TYPE XII DISTRIBUTION

Muhammad Yameen Danish[1§] and **Muhammad Aslam**[2]
[1] Department of Statistics, Government College Fateh Jang
Attock, Pakistan. Email: yameendanish42sb@yahoo.com
[2] Department of Statistics, Riphah International University
Islamabad, Pakistan. Email: aslamsdqu@yahoo.com
[§] Corresponding Author

ABSTRACT

The article provides the Bayesian analysis of restricted random censorship model with Burr type XII distribution. It is seen that the joint continuous conjugate prior of the proposed model parameters cannot be derived; we consider in this paper the commonly used gamma prior distributions of these parameters. It is further seen that the explicit expressions of Bayes estimates are not possible; we suggest two different methods of Bayesian computation, namely, Lindley's method and Markov chain Gibbs sampling. The classical method of maximum likelihood estimation is considered for comparison purposes. Numerical experiments are performed to check the properties of Maximum likelihood estimators and Bayes estimators. A real life data example is analyzed to illustrate the proposed methodology. The proposed model is validated using posterior predictive simulation in order to ascertain its appropriateness.

KEYWORDS

Log-concave function, Lindley's approximation, Gibbs sampling, Predictive Simulation.

1. INTRODUCTION

Censoring is a necessary part of most of the life-testing experiments. It happens when the exact survival times until some event of interest are known for only a portion of subjects in the study and known to exceed certain values in the remainder. In life-testing experiments the units on test are dropped out or lost from test so that the event of interest may not always be happened for all the subjects in the study. It is not possible for the researcher to wait for the last outcome as it is necessary to report the results from the study as soon as possible. Several censoring schemes are used in survival analysis to reduce the experimental time and cost. The most popular among these are right censoring schemes because of their crucial importance in failure time studies. The unique feature of type I and II right censoring schemes is that these are completely under the control of an investigator. The third type of right censoring is random censoring where censoring time is not fixed but depends on other factors which are modeled by an independent random variable known as censoring time variable. Consider a clinical trial in which the patients with colorectal cancer enter simultaneously after their tumors have been removed by surgery and we want to observe their survival times but censoring occurs in one of the following forms: loss to follow-up (e.g. the patient may decide to move elsewhere), drop out (e.g. due to bad side effects or refusal to participate), death from other diseases or termination of the study. Clearly, all these random factors are beyond the control of an investigator and are modeled by a censoring time variable.

In survival analysis, a number of distributions are encountered in practice. The most important among these are exponential, Rayleigh, Weibull, gamma and lognormal. In Danish and Aslam (2014), the authors proposed a similar model for Weibull distribution that also includes the exponential and Rayleigh distributions as special cases. In the present paper, we take into consideration the Burr type XII distribution originally developed by Irving Burr (1942) as the outgrowth of his research into methods for fitting distribution functions rather than probability density functions to frequency data. The motivation behind this distribution is that much of the region covered by the gamma and lognormal distributions in the skewness-kurtosis plane is also covered by the Burr type XII distribution; see Rodriguez (1977) for detail. It has nice closed-form expressions for its distribution function and hazard function which are not possible for gamma and lognormal distributions. It is one of the general parametric families that covers a variety of curve shapes and provides a wide range of values of skewness and kurtosis.

Suppose $X_1,...,X_n$ are independent and identically distributed (i.i.d.) random variables with distribution function $F_X(t)$ and density function $f_X(t)$. Consider another sequence, independent of $\{X_i\}$, $T_1,...,T_n$ of i.i.d. random variables with distribution function $F_T(t)$ and density function $f_T(t)$. In the terminology of life-testing experiments, X_i ’s are the true life times of n subjects censored by T_i ’s from the right such that one observes i.i.d. random pairs $(Y_1,D_1),...,(Y_n,D_n)$, where $Y_i=\min(X_i,T_i)$ and $D_i=\mathrm{I}(X_i\le T_i)$, the indicator of non-censored observation, for $1\le i\le n$. Thus the observed $Y_i's$ constitute a random sample from the distribution function $F_Y(t)$, where $1-F_Y(t)=(1-F_X(t))(1-F_T(t))$. This is the general random censorship model studied by Kaplan and Meier (1958), Efron (1967), Breslow and Crowley (1974). Under this model, Kaplan and Meier introduced their historic product limit estimator given by

$$S(y)=\prod_{i:Y_i\le t}\left(\frac{n-R_i}{n-R_i+1}\right)^{d_i}. \tag{1}$$

Now it is simple to show that

$$f_{Y,D}(y,d)=\left\{f_X(y)(1-F_T(y))\right\}^d\left\{f_T(y)(1-F_X(y))\right\}^{1-d} \quad ;y\ge 0,d=0,1. \tag{2}$$

In this paper, it is assumed that the survival time variable X and the censoring time variable T follow a Burr type XII distribution with same shape parameter λ but different shape parameters θ and β. The probability density function and distribution function of the Burr type XII distribution are

$$f_X(x;\theta,\lambda)=\frac{\theta\lambda x^{\lambda-1}}{\left(1+x^{\lambda}\right)^{\theta+1}}, \tag{3}$$

and

$$F_X(x;\theta,\lambda)=1-\left(1+x^{\lambda}\right)^{-\theta}, \tag{4}$$

respectively. Using expressions (3) and (4) in (2), the joint density function of Y and D can be written as

$$g(y,d;\theta,\lambda,\beta)=\lambda y^{\lambda-1}\left(1+y^{\lambda}\right)^{-\theta-\beta-1}\theta^{d}\beta^{1-d};\quad y>0,\ d=0,1. \tag{5}$$

The density function of observed random variable Y is

$$f_Y(y;\theta,\lambda,\beta)=\lambda(\theta+\beta)y^{\lambda-1}\left(1+y^{\lambda}\right)^{-\theta-\beta-1}, \tag{6}$$

The remainder of the paper is organized as follows. In the next section we derive maximum likelihood (ML) estimators with some of their asymptotic properties. Section 3 contains the prior distributions, Bayes estimates based on Lindley's approximation and MCMC Gibbs sampling. Numerical simulation experiments are carried out in Section 4. Numerical illustration based on a real data set is provided in Section 5. Model validation using posterior predictive simulation is discussed briefly in Section 6. This section also gives numerical results of model validation based on real data example and finally Section 7 presents the concluding remarks.

2. MAXIMUM LIKELIHOOD ESTIMATION

This section provides the ML estimators $\hat{\theta}$, $\hat{\lambda}$ and $\hat{\beta}$ of the parameters θ, λ and β; and the ML estimators $\hat{h}(x_0)$ and $\hat{R}(x_0)$ of the hazard function $h(x_0)$ and the reliability function $R(x_0)$ at mission time x_0. For an observed sample $(y_1,d_1),\ldots,(y_n,d_n)=(\mathrm{y},\mathrm{d})$ from (5), the likelihood function is

$$l(\theta,\lambda,\beta|\mathrm{y},\mathrm{d})=\lambda^{n}\left(\prod_{i=1}^{n}\frac{y_i^{\lambda-1}}{1+y_i^{\lambda}}\right)\prod_{i=1}^{n}\left(1+y_i^{\lambda}\right)^{-\theta-\beta}\theta^{\sum_{i=1}^{n}d_i}\beta^{n-\sum_{i=1}^{n}d_i}. \tag{7}$$

The log-likelihood function $w=\sum_{i=1}^{n}d_i$ is

$$L(\theta,\lambda,\beta|\mathrm{y},\mathrm{d})=n\ln\lambda+w\ln\theta+(n-w)\ln\beta$$
$$-(\theta+\beta)\sum_{i=1}^{n}\ln\left(1+y_i^{\lambda}\right)+\sum_{i=1}^{n}\ln\left(\frac{y_i^{\lambda-1}}{1+y_i^{\lambda}}\right). \tag{8}$$

The likelihood equations are obtained from (8) as

$$\frac{w}{\theta}-\sum_{i=1}^{n}\ln\left(1+y_i^{\lambda}\right)=0,$$

$$\frac{n}{\lambda}-\left(\theta+\beta\right)\sum_{i=1}^{n}\frac{y_i^{\lambda}\ln y_i}{1+y_i^{\lambda}}+\sum_{i=1}^{n}\frac{\ln y_i}{1+y_i^{\lambda}}=0$$

$$\frac{1}{\beta}\left(n-w\right)-\sum_{i=1}^{n}\ln\left(1+y_i^{\lambda}\right)=0.$$

Solving these equations simultaneously, we have

$$\hat{\theta}\left(\lambda\right)=\frac{w}{\sum_{i=1}^{n}\ln\left(1+y_i^{\lambda}\right)},\ \hat{\beta}\left(\lambda\right)=\frac{n-w}{\sum_{i=1}^{n}\ln\left(1+y_i^{\lambda}\right)} \tag{9}$$

and

$$\lambda=\delta(\lambda), \tag{10}$$

where $\delta\left(\lambda\right)=\left[\frac{\sum_{i=1}^{n}\frac{y_i^{\lambda}\ln y_i}{1+y_i^{\lambda}}}{\sum_{i=1}^{n}\ln\left(1+y_i^{\lambda}\right)}-\frac{1}{n}\sum_{i=1}^{n}\frac{\ln y_i}{1+y_i^{\lambda}}\right]^{-1}$.

Any iterative principle can be used to solve (10). Start with an initial value of λ, say $\lambda^{(0)}$, and obtain $\lambda^{(1)}$ from $\lambda^{(1)}=\delta\left(\lambda^{(0)}\right)$, $\lambda^{(2)}$ from $\lambda^{(2)}=\delta\left(\lambda^{(1)}\right)$ and so on $\lambda^{(m)}$ from $\lambda^{(m)}=\delta\left(\lambda^{(m-1)}\right)$. Stop the process when $\left|\lambda^{(m)}-\lambda^{(m-1)}\right|<\varepsilon$ is satisfied, where ε is a pre-assigned tolerance limit. Now the ML estimators of θ, β, $h\left(x_0\right)$ and $R\left(x_0\right)$ are

$$\hat{\theta}=\frac{w}{\sum_{i=1}^{n}\ln\left(1+y_i^{\hat{\lambda}}\right)},\ \hat{\beta}=\frac{n-w}{\sum_{i=1}^{n}\ln\left(1+y_i^{\hat{\lambda}}\right)},\ \hat{h}\left(x_0\right)=\frac{\hat{\theta}\hat{\lambda}x_0^{\hat{\lambda}-1}}{1+x_0^{\hat{\lambda}}}\ \text{and}\ \hat{R}\left(x_0\right)=\left(1+x_0^{\hat{\lambda}}\right)^{-\hat{\theta}}.$$

Theorem 1:

The observed information matrix $O\left(\theta\right)=\left(\left(\frac{\partial^2 L\left(\theta;y,d\right)}{\partial\theta_r\partial\theta_s}\right)_{\theta=\hat{\theta}}\right)$ evaluated at $\hat{\theta}$ is negative-definite, where L is the log-likelihood function, $\theta=\left(\theta,\lambda,\beta\right)$ and $\hat{\theta}$ is solution to the likelihood equations.

Proof:

The determinant of the matrix $\mathrm{O}(\theta,\lambda,\beta)$ is

$$\left|\mathrm{O}(\theta)\right| = -\frac{w(n-w)}{\theta^2\beta^2}\left\{\frac{n}{\lambda^2}+(\theta+\beta+1)\sum_{i=1}^{n}\frac{y_i^{\lambda}(\ln y_i)^2}{\left(1+y_i^{\lambda}\right)^2}\right\}$$

$$+\left\{\frac{w}{\theta^2}+\frac{(n-w)}{\beta^2}\right\}\left(\sum_{i=1}^{n}\frac{y_i^{\lambda}\ln y_i}{1+y_i^{\lambda}}\right)^2 .$$

Substituting the ML estimates and simplifying, we have

$$\left|\mathrm{O}\left(\hat\theta,\hat\lambda,\hat\beta\right)\right| = -\frac{n\left(\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)\right)^4}{\hat\lambda^2(n-w)w}\left(\frac{1}{n}+\frac{1}{\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)}\right)\sum_{i=1}^{n}\frac{y_i^{\hat\lambda}\left(\ln y_i^{\hat\lambda}\right)^2}{\left(1+y_i^{\hat\lambda}\right)^2}$$

$$-\frac{n\left(\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)\right)^2}{\hat\lambda^2(n-w)w}\left(\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)-\sum_{i=1}^{n}\frac{y_i^{\hat\lambda}\ln y_i^{\hat\lambda}}{1+y_i^{\hat\lambda}}\right)\left(\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)+\sum_{i=1}^{n}\frac{y_i^{\hat\lambda}\ln y_i^{\hat\lambda}}{1+y_i^{\hat\lambda}}\right).$$

Since

$$\left(\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)-\sum_{i=1}^{n}\frac{y_i^{\hat\lambda}\ln y_i^{\hat\lambda}}{1+y_i^{\hat\lambda}}\right)\geq 0 \text{ for } y_i^{\hat\lambda}\geq 1, \qquad (11)$$

so $\left|\mathrm{O}\left(\hat\theta,\hat\lambda,\hat\beta\right)\right|<0$ for $y_i^{\hat\lambda}\geq 1$.

To prove that the expression (11) also holds for $y_i^{\hat\lambda}<1$, suppose that

$$\varphi(y)=\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)-\sum_{i=1}^{n}\frac{y_i^{\hat\lambda}\ln y_i^{\hat\lambda}}{1+y_i^{\hat\lambda}}.$$

Since the derivative of $\varphi(y)$ i.e. $\varphi'(y)=-\lambda\sum_{i=1}^{n}\frac{y_i^{\hat\lambda-1}\ln y_i^{\hat\lambda}}{\left(1+y_i^{\hat\lambda}\right)^2}>0$ for $0<y_i^{\hat\lambda}<1$, it follows that the function $\varphi(y)=\sum_{i=1}^{n}\ln\left(1+y_i^{\hat\lambda}\right)-\sum_{i=1}^{n}\frac{y_i^{\hat\lambda}\ln y_i^{\hat\lambda}}{1+y_i^{\hat\lambda}}>0$. Therefore, $\left|\mathrm{O}\left(\hat\theta,\hat\lambda,\hat\beta\right)\right|<0$ for $0<y_i^{\hat\lambda}<1$. This completes the proof.

Thus the likelihood function has a relative maximum at $\left(\hat{\theta},\hat{\lambda},\hat{\beta}\right)$ with probability tending to unity. Now we state the asymptotic normality result based on the Theorem 1. It can be stated as

$$\left[\left(\hat{\theta}-\theta\right),\left(\hat{\lambda}-\lambda\right),\left(\hat{\beta}-\beta\right)\right]\rightarrow N_3\left(0,\, \mathrm{I}^{-1}\left(\theta,\lambda,\beta\right)\right).$$

The Fisher information matrix $\mathrm{I}\left(\theta,\lambda,\beta\right)$ is derived in Appendix. The approximate $100(1-\alpha)\%$ confidence interval for θ is $\left[\hat{\theta}\ \pm\ Z_{\alpha/2}\left(\hat{V}_{11}\right)^{-\frac{1}{2}}\right]$, where $\hat{V}_{11}$ is (1,1)th element of the inverse of Fisher information matrix evaluated at ML estimate of θ and Z_α is the αth quantile of standard normal distribution. We can similarly obtain the confidence intervals for λ, β, $h\left(x_0\right)$ and $R\left(x_0\right)$.

3. BAYESIAN ESTIMATION

This section presents the Bayes estimates of parameters, the hazard and reliability functions. The model under consideration has three parameters; and the joint conjugate prior of these parameters is not possible. We assume that the parameters have gamma priors given by

$$\left.\begin{aligned}\pi_1\left(\theta\right)&=\frac{b_1^{a_1}}{\Gamma\left(a_1\right)}\theta^{a_1-1}e^{-b_1\theta}; && a_1,b_1,\theta>0\\ \pi_2\left(\lambda\right)&=\frac{b_2^{a_2}}{\Gamma\left(a_2\right)}\lambda^{a_2-1}e^{-b_2\lambda}; && a_2,b_2,\lambda>0\\ \pi_3\left(\beta\right)&=\frac{b_3^{a_3}}{\Gamma\left(a_3\right)}\beta^{a_3-1}e^{-b_3\beta}; && a_3,b_3,\beta>0\end{aligned}\right\} \tag{12}$$

Several authors have used the gamma priors for the scale and shape parameters of lifetime distributions in their Bayesian analysis; see Kundu and Pradhan (2009), Berger and Sun (1993), Joarder et al. (2011), among others. Actually the gamma density is quite flexible; many non-informative priors are special cases of it. For example as the hyper-parameters in the gamma density approach zero, it becomes inversely proportional to its argument. This density is often used as non-informative gamma prior for the parameters in the range 0 to ∞.

We can write the joint prior as

$$\pi\left(\theta,\lambda,\beta\right)\propto\theta^{a_1-1}e^{-b_1\theta}\lambda^{a_2-1}e^{-b_2\lambda}\beta^{a_3-1}e^{-b_3\beta}. \tag{13}$$

Now combining the likelihood function in (7) and the joint prior in (13), the joint posterior distribution of θ, λ and β given data is

$$\pi\left(\theta,\lambda,\beta\middle|y,d\right)\propto\theta^{a_1+w-1}e^{-\theta\left(b_1+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)}\beta^{a_3+n-w-1}e^{-\beta\left(b_3+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)}$$
$$\times\lambda^{a_2+n-1}e^{-\lambda\left(b_2-\sum_{i=1}^{n}\ln y_i\right)}\prod_{i=1}^{n}\left(1+y_i^\lambda\right)^{-1}. \tag{14}$$

Thus the Bayes estimate of any function of parameters, say $U(\theta,\lambda,\beta)$, with respect to squared error loss function is

$$E\left(U\left(\theta,\lambda,\beta\middle|\mathrm{y},\mathrm{d}\right)\right)=\frac{\int_0^\infty\int_0^\infty\int_0^\infty U(\theta,\lambda,\beta)\,\theta^{a_1+w-1}e^{-\theta\left(b_1+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)}\beta^{a_3+n-w-1}e^{-\beta\left(b_3+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)}}{\int_0^\infty\int_0^\infty\int_0^\infty\theta^{a_1+w-1}e^{-\theta\left(b_1+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)}\beta^{a_3+n-w-1}e^{-\beta\left(b_3+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)}}$$
$$\times\frac{\lambda^{a_2+n-1}e^{-\lambda\left(b_2-\sum_{i=1}^{n}\ln y_i\right)}\prod_{i=1}^{n}\left(1+y_i^\lambda\right)^{-1}d\theta d\lambda d\beta}{\lambda^{a_2+n-1}e^{-\lambda\left(b_2-\sum_{i=1}^{n}\ln y_i\right)}\prod_{i=1}^{n}\left(1+y_i^\lambda\right)^{-1}d\theta d\lambda d\beta}. \tag{15}$$

It is clear from (15) that we cannot valuate it in explicit form. Two different methods namely (a) Gibbs sampling and (b) Lindley's method are considered to evaluate it.

3.1 Gibbs Sampling

In order to obtain the Bayes estimates using Gibbs sampling it is necessary to be able to draw samples of the quantities involved from the posterior distribution. We need the following results for this purpose.

Theorem 2

(a) The conditional posterior distribution of θ given λ and data y,d is

$$\pi_1\left(\theta\middle|\lambda,\mathrm{y},\mathrm{d}\right)=\mathrm{gamma}\left(\theta;a_1+w,\ b_1+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right). \tag{16}$$

(b) The marginal posterior distribution of λ given data y,d is

$$\pi_2\left(\lambda\middle|\mathrm{y},\mathrm{d}\right)=\frac{\lambda^{a_2+n-1}e^{-\lambda\left(b_2-\sum_{i=1}^{n}\ln y_i\right)}\prod_{i=1}^{n}\left(1+y_i^\lambda\right)^{-1}}{\left(b_1+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)^{a_1+w}\left(b_3+\sum_{i=1}^{n}\ln\left(1+y_i^\lambda\right)\right)^{a_3+n-w}}. \tag{17}$$

(c) The conditional posterior distribution of β given λ and data y,d is

$$\pi_3\left(\beta|\lambda, \mathrm{y}, \mathrm{d}\right) = \text{gamma}\left(\beta; a_3 + n - w,\ b_3 + \sum_{i=1}^{n} \ln\left(1 + y_i^{\lambda}\right)\right). \tag{18}$$

(d) The density function in part (b) is log-concave.

Proof:

Parts (a), (b) and (c) are trivial, and part (d) is proved below. Taking the derivative of $\ln \pi_2\left(\lambda|\mathrm{y}, \mathrm{d}\right)$, we have

$$\frac{d^2 \ln \pi_2\left(\lambda|\mathrm{y}, \mathrm{d}\right)}{d\lambda^2} = -\frac{\left(a_2 + n - 1\right)}{\lambda^2} - \sum_{i=1}^{n} \frac{y_i^{\lambda}\left(\ln y_i\right)^2}{\left(1 + y_i^{\lambda}\right)^2}$$

$$-\left(\frac{a_1 + w}{b_1 + \sum_{i=1}^{n} \ln\left(1 + y_i^{\lambda}\right)} + \frac{a_3 + n - w}{b_3 + \sum_{i=1}^{n} \ln\left(1 + y_i^{\lambda}\right)}\right)\left\{\sum_{i=1}^{n} \frac{y_i^{\lambda}\left(\ln y_i\right)^2}{\left(1 + y_i^{\lambda}\right)^2} - \left(\sum_{i=1}^{n} \frac{y_i^{\lambda} \ln y_i}{1 + y_i^{\lambda}}\right)^2\right\}.$$

The right hand side of above expression is negative by virtue of Theorem 1. Therefore, the density given in (17) is log-concave. The following procedure can be used to sample the posterior distribution in (14):

Step 1. Generate λ_1 from the log-concave density $\pi_2\left(\lambda|\mathrm{y}, \mathrm{d}\right)$.

Step 2. Generate θ_1 from $\text{gamma}\left(\theta;\ a_1 + w,\ b_1 + \sum_{i=1}^{n} \ln\left(1 + y_i^{\lambda_1}\right)\right)$.

Step 3. Generate β_1 from $\text{gamma}\left(\beta;\ a_3 + n - w,\ b_3 + \sum_{i=1}^{n} \ln\left(1 + y_i^{\lambda_1}\right)\right)$.

Step 4. Repeat Steps 1–3 M times to get $\left(\lambda_1, \theta_1, \beta_1\right), \ldots, \left(\lambda_M, \theta_M, \beta_M\right)$.

Step 5. Obtain the approximate value of posterior expectation in (15) from

$$\frac{1}{M} \sum_{j=1}^{M} U\left(\theta_j, \lambda_j, \beta_j\right)$$

To obtain the highest posterior density interval for θ, arrange the posterior sample $\theta_1, \ldots, \theta_{M-N}$ as $\theta_{(1)} < \ldots < \theta_{(M-N)}$. Construct all 100 $\left(1 - \alpha\right)$% credible intervals for θ as $\left(\theta_{(1)},\ \theta_{\left(\left[M(1-\alpha)\right]\right)}\right), \ldots, \left(\theta_{\left([M\alpha]\right)},\ \theta_{(M)}\right)$. Here [w] represents the greatest integer value less than or equal to *w*. The highest posterior density interval for θ is the interval that has the shortest length. Similarly, the highest posterior density intervals for λ, β, $h\left(x_0\right)$ and $R\left(x_0\right)$ can be obtained.

3.2 Lindley's Method

Lindley (1980) introduced a procedure to evaluate the ratio of two integrals such as in (15). The procedure is explained in appendix. The Bayes estimates with respect to squared error loss function (SELF) using the Lindley's method are

$$\hat{\theta}_{BL} = \hat{\theta} + \rho_1\sigma_{11} + \rho_2\sigma_{21} + \rho_3\sigma_{31} + \frac{1}{2}\left(A_1\sigma_{11} + A_2\sigma_{21} + A_3\sigma_{31}\right), \tag{19}$$

$$\hat{\lambda}_{BL} = \hat{\lambda} + \rho_1\sigma_{12} + \rho_2\sigma_{22} + \rho_3\sigma_{32} + \frac{1}{2}\left(A_1\sigma_{12} + A_2\sigma_{22} + A_3\sigma_{32}\right), \tag{20}$$

$$\hat{\beta}_{LB} = \hat{\beta} + \rho_1\sigma_{13} + \rho_2\sigma_{23} + \rho_3\sigma_{33} + \frac{1}{2}\left(A_1\sigma_{13} + A_2\sigma_{23} + A_3\sigma_{33}\right), \tag{21}$$

$$\begin{aligned}\hat{h}_{BL}\left(x_0\right) = \hat{U} &+ \left(\rho_1\sigma_{11} + \rho_2\sigma_{12} + \rho_3\sigma_{13}\right)\hat{U}_1 + \left(\rho_1\sigma_{21} + \rho_2\sigma_{22} + \rho_3\sigma_{23}\right)\hat{U}_2 + \sigma_{12}\hat{U}_{12} \\ &+ \frac{1}{2}\left\{\left(A_1\sigma_{11} + A\sigma_{12} + A_3\sigma_{13}\right)\hat{U}_1 + \left(A_1\sigma_{21} + A_2\sigma_{22} + A_3\sigma_{23}\right)\hat{U}_2 + \sigma_{22}\hat{U}_{22}\right\},\end{aligned} \tag{22}$$

$$\begin{aligned}\hat{R}_{BL}\left(x_0\right) = \hat{U} &+ \left(\rho_1\sigma_{11} + \rho_2\sigma_{12} + \rho_3\sigma_{13}\right)\hat{U}_1 + \left(\rho_1\sigma_{21} + \rho_2\sigma_{22} + \rho_3\sigma_{23}\right)\hat{U}_2 + \sigma_{12}\hat{U}_{12} \\ &+ \frac{1}{2}\left\{\begin{aligned}&\left(A_1\sigma_{11} + A_2\sigma_{12} + A_3\sigma_{13}\right)\hat{U}_1 \\ &\qquad + \left(A_1\sigma_{21} + A_2\sigma_{22} + A_3\sigma_{23}\right)\hat{U}_2 + \sigma_{11}\hat{U}_{11} + \sigma_{22}\hat{U}_{22}\end{aligned}\right\},\end{aligned} \tag{23}$$

For closed-form expressions of $\rho_1, \rho_2, \rho_3, A_1, A_2, A_3$, see appendix.

3.3 Non-Informative Bayesian Analysis

Previous section presents the Bayes estimates assuming gamma prior distributions of the parameters involved. The gamma prior is quite flexible and often the non-informative priors (uniform and Jeffreys) are special cases of it. For example, when the hyper-parameters (*a*, *b*) in a gamma density of θ approach zero, we have $\pi\left(\theta\right) \propto 1/\theta$. This prior density is often used as the non-informative prior for scale and shape parameters having range $(0, \infty)$. Thus in case of no prior information about the unknown parameters, the independent gamma priors with hyper-parameters equal to zero can be used to perform the non-informative Bayesian analysis.

4. SIMULATION STUDY

In this section we perform a simulation to observe the behavior of the Bayes estimators and to compare with the ML estimators of θ, λ, β, the hazard function $h\left(x_0\right)$ and the reliability function $R\left(x_0\right)$ at the mission time $x_0 = 0.80$ under SELF. We consider sample sizes: $n = 20, 40, 60$; censoring rates: p = 0.50, 0.80; different sets of parameter values and different priors as provided in Table 1.

Table 1
The Values of Parameters and Hyper-Parameters Used in Simulation Study

θ	λ	β	$h(x_0)$	$R(x_0)$		a_1	b_1	a_2	b_2	a_3	b_3	Notation
2	1.5	2	1.5641	0.3398	0	0	0	0	0	0		NGP
2	1.5	2	1.5641	0.3398	4	2	3	2	4	2		IGP
2	1.5	0.5	1.5641	0.3398	0	0	0	0	0	0		NGP
2	1.5	0.5	1.5641	0.3398	4	2	3	2	2	4		IGP

Table 2
The Average ML and Bayes Estimates and the Corresponding MSEs (in Parenthesis) when $p = 0.50$

$U(\theta,\lambda,\beta)$	n	ML	Bayes (MCMC)		Bayes (Lindley)	
			NGP	IGP	NGP	IGP
θ	20	2.3094 (1.0318)	2.2428 (0.5970)	2.0713 (0.1263)	2.1199 (0.1966)	1.9588 (0.0402)
	40	2.1125 (0.1884)	2.1189 (0.2047)	2.0705 (0.1025)	2.0735 (0.1155)	2.0112 (0.0312)
	60	2.0724 (0.1138)	2.0671 (0.1138)	2.0434 (0.0750)	2.0349 (0.0576)	2.0171 (0.0319)
λ	20	1.6006 (0.0985)	1.5850 (0.0880)	1.5518 (0.0565)	1.5403 (0.0355)	1.5032 (0.0214)
	40	1.5417 (0.0356)	1.5413 (0.0376)	1.5295 (0.0298)	1.5266 (0.0231)	1.5115 (0.0167)
	60	1.5285 (0.0232)	1.5252 (0.0229)	1.5210 (0.0202)	1.5132 (0.0126)	1.5088 (0.0109)
β	20	2.3094 (1.0318)	2.2430 (0.5983)	2.0710 (0.1263)	2.1199 (0.1966)	1.9588 (0.0402)
	40	2.1125 (0.1884)	2.1192 (0.2057)	2.0676 (0.1026)	2.0735 (0.1155)	2.0112 (0.0312)
	60	2.0722 (0.1138)	2.0698 (0.1142)	2.0459 (0.0753)	2.0349 (0.0576)	2.0171 (0.0319)
$h(x_0)$	20	1.9489 (1.5131)	1.8819 (0.7921)	1.6736 (0.1630)	1.7061 (0.2298)	1.5374 (0.0530)
	40	1.6985 (0.2198)	1.7131 (0.2436)	1.6540 (0.1212)	1.6526 (0.1288)	1.5884 (0.0416)
	60	1.6515 (0.1268)	1.6498 (0.1284)	1.6231 (0.0849)	1.6061 (0.0601)	1.5880 (0.0355)
$R(x_0)$	20	0.3173 (0.0094)	0.3451 (0.0076)	0.3508 (0.0033)	0.3412 (0.0038)	0.3546 (0.0009)
	40	0.3299 (0.0041)	0.3406 (0.0040)	0.3424 (0.0025)	0.3412 (0.0027)	0.3469 (0.0011)
	60	0.3335 (0.0028)	0.3417 (0.0027)	0.3431 (0.0020)	0.3415 (0.0015)	0.3432 (0.0010)

Table 3
The Average ML and Bayes Estimates and the Corresponding MSEs (in Parenthesis) when *p* = 0.80

$U(\theta,\lambda,\beta)$	n	ML	Bayes (MCMC)		Bayes (Lindley)	
			NGP	IGP	NGP	IGP
θ	20	2.1740 (0.3723)	2.1393 (0.3397)	2.0734 (0.1476)	2.0778 (0.1344)	2.0267 (0.0492)
	40	2.0804 (0.1347)	2.0575 (0.1212)	2.0447 (0.0893)	2.0458 (0.0793)	2.0264 (0.0465)
	60	2.0489 (0.0798)	2.0443 (0.0779)	2.0326 (0.0655)	2.0285 (0.0464)	2.0217 (0.0354)
λ	20	1.5911 (0.0887)	1.5742 (0.0883)	1.5522 (0.0666)	1.5387 (0.0347)	1.5225 (0.0259)
	40	1.5401 (0.0348)	1.5322 (0.0375)	1.5305 (0.0340)	1.5226 (0.0215)	1.5163 (0.0182)
	60	1.5241 (0.0216)	1.5217 (0.0219)	1.5175 (0.0206)	1.5129 (0.0142)	1.5107 (0.0129)
β	20	0.5445 (0.0233)	0.5345 (0.0212)	0.5124 (0.0059)	0.5237 (0.0087)	0.5010 (0.0010)
	40	0.5201 (0.0084)	0.5144 (0.0076)	0.5090 (0.0044)	0.5132 (0.0050)	0.5045 (0.0016)
	60	0.5122 (0.0050)	0.5111 (0.0049)	0.5074 (0.0033)	0.5077 (0.0029)	0.5046 (0.0017)
$h(x_0)$	20	1.7841 (0.4251)	1.7438 (0.4188)	1.6603 (0.1684)	1.6548 (0.1350)	1.6022 (0.0543)
	40	1.6615 (0.1358)	1.6359 (0.1262)	1.6227 (0.0949)	1.6170 (0.0777)	1.5972 (0.0480)
	60	1.6229 (0.0785)	1.6169 (0.0777)	1.6042 (0.0627)	1.5960 (0.0461)	1.5892 (0.0364)
$R(x_0)$	20	0.3285 (0.0074)	0.3452 (0.0064)	0.3464 (0.0039)	0.3410 (0.0034)	0.3449 (0.0017)
	40	0.3336 (0.0036)	0.3432 (0.0033)	0.3433 (0.0026)	0.3410 (0.0022)	0.3426 (0.0015)
	60	0.3360 (0.0023)	0.3409 (0.0022)	0.3414 (0.0018)	0.3402 (0.0014)	0.3408 (0.0011)

It may be noted that NGP represents the non-informative gamma priors when all the hyper-parameters in (12) are zero and the IGP represents informative gamma priors with priors' means equal to the corresponding parameter values. Purposely, the two extreme cases are considered. One prior indicates that we are completely uninformative about the unknown parameters and the other indicates that we know the priors' means. For a particular case, 1000 random samples are simulated from the model in (5) and for each sample we compute the ML estimate and the corresponding 95% confidence interval estimates, the Bayes estimate and the corresponding 95% Bayesian credible interval estimates based on 20,000 Markov chain Monte Carlo (MCMC) samples taking every 10^{th} draw from each of the two independent chains. The average ML and Bayes estimates, mean square errors, average lengths of 95% confidence/credible intervals and coverage percentages are obtained from these computations. The results are reported in Tables 2–4. It is observed that the biases, the MSEs and the confidence/credible interval lengths decrease reasonably with increasing sample sizes. However, it is seen that the biases, MSEs and confidence/credible interval lengths decrease at relatively higher rate for small to medium sample sizes as compared with medium to large sample sizes.

Table 4
The Average Lengths of 95% Confidence/Credible Intervals and the Associated Coverage Percentages (in Parenthesis)

$U(\theta,\lambda,\beta)$	n	*ML*		Bayesian $(p=0.50)$		Bayesian $(p=0.80)$	
		$p=0.50$	$p=0.80$	NGP	IGP	NGP	IGP
θ	20	3.3407 (98)	2.2284 (96)	3.0391 (99)	2.2484 (100)	2.1499 (97)	1.8398 (99)
	40	2.0157 (99)	1.4751 (96)	1.9871 (99)	1.7340 (100)	1.4459 (96)	1.3490 (98)
	60	1.5960 (99)	1.1801 (97)	1.5718 (99)	1.4409 (99)	1.1702 (97)	1.1156 (97)
λ	20	1.0217 (94)	1.0541 (95)	1.0091 (94)	0.9157 (96)	1.0439 (94)	0.9847 (95)
	40	0.6912 (95)	0.7199 (95)	0.6900 (94)	0.6574 (95)	0.7170 (95)	0.7004 (95)
	60	0.5587 (94)	0.5816 (95)	0.5570 (94)	0.5406 (94)	0.5805 (95)	0.5709 (95)
β	20	3.3407 (98)	1.1091 (100)	3.0395 (99)	2.2492 (100)	1.0345 (100)	0.8095 (100)
	40	2.0157 (99)	0.7350 (100)	1.9872 (99)	1.7324 (100)	0.7073 (100)	0.6260 (100)
	60	1.5960 (99)	0.5871 (100)	1.5729 (99)	1.4422 (99)	0.5750 (100)	0.5292 (100)
$h(x_0)$	20	3.9886 (97)	2.8206 (98)	3.0618 (98)	2.1738 (100)	2.1269 (97)	1.7975 (99)
	40	2.2733 (99)	1.8176 (99)	1.9159 (98)	1.6519 (99)	1.3843 (96)	1.2961 (97)
	60	1.7782 (99)	1.4413 (99)	1.4914 (98)	1.3622 (99)	1.1128 (96)	1.0626 (97)
$R(x_0)$	20	0.3897 (97)	0.2255 (98)	0.4341 (99)	0.3722 (100)	0.3443 (97)	0.3129 (99)
	40	0.2871 (98)	0.2071 (98)	0.3173 (99)	0.2894 (100)	0.2490 (96)	0.2355 (98)
	60	0.2369 (98)	0.1706 (99)	0.2619 (99)	0.2455 (100)	0.2046 (97)	0.1969 (98)

It is further observed that the Bayes estimates under NGP based MCMC Gibbs sampling perform slightly better than the ML estimates for small sample sizes and their performance is very similar for large sample sizes. The Bayes estimates under NGP based on the Lindley's approximation perform relatively better than both the ML estimates and the Bayes estimates under NGP based on MCMC sampling. A similar behavior is observed under the informative priors. However, the Bayes estimates under IGP perform quite better than both the ML estimates and the Bayes estimates under NGP. Table 4 compares the ML estimators and the Bayes estimates in terms of average confidence/credible interval lengths and the corresponding 95% coverage percentages. This table further clarifies the previous results. It is interesting to note in Table 4 that coverage percentages of the estimators are larger than the nominal value of 95% in general.

5. ANALYSIS OF REAL LIFE DATA

This section is devoted to the description of the proposed methods with the help of real data analysis. McIllmurray and Turkie (1987) reported the survival times (in months) of 24 patients with Dukes' C colorectal cancer as 3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 12+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+ and 42. We perform the non-informative Bayesian analysis explained in Section 3.3 for this data set. We obtain the ML estimates, 95% ML confidence interval estimates, the Bayes estimates, 95% Bayesian credible interval estimates under SELF, linear exponential loss function (LELF) and general entropy loss function (GELF) using the Gibbs sampling and the Lindley's method.

Table 5
The ML and Bayes Estimates and the Associated *p*-values of the K-S Test

$U(\theta,\lambda,\beta)$	ML		Bayes (Gibbs Sampling)				Bayes SELF
			SELF		LELF	GELF	
	Point	Interval	Point	Interval			Lindley
θ	0.293	(0.138, 0.561)	0.296	(0.135, 0.527)	0.302	0.289	0.304
λ	2.822	(1.772, 4.401)	2.876	(1.820, 4.282)	2.871	2.623	2.910
β	0.293	(0.137, 0.560)	0.297	(0.137, 0.524)	0.302	0.290	0.305
$h(x_0)$	0.392	(0.123, 1.058)	0.387	(0.198, 0.645)	0.393	0.378	0.395
$R(x_0)$	0.849	(0.761, 0.920)	0.845	(0.736, 0.932)	0.846	0.847	0.850
p-value	0.3154		0.3174		0.3471	0.3275	0.3384

The detail about different loss functions can be seen in Danish and Aslam (2013). The hazard and reliability functions are computed at the mission time $x_0 = 0.9$. It is seen that the Bayesian credible interval lengths are much smaller than the corresponding ML confidence interval lengths in case of θ, λ, β and $h(x_0)$ and larger in case of reliability function $R(x_0)$. To check the goodness-of-fit of the proposed model, we compute the

D statistics of Kolomogrov and Smirnov (K-S) test of fit and the associated *p*-values. The high *p*-values of this test in Table 5 indicate the plausibility of the proposed model to analyze this data. The Bayes estimates under LELF and GELF using Gibbs sampling and the Bayes estimates under SELF using the Lindley's approximation perform slightly better than both the Bayes estimates under SELF using Gibbs sampling and the ML estimates in terms of K-S test. It is clear that all the methods provide good fit to the data at hand with slightly better result for the Bayesian method under LELF corresponding to loss function parameter value 1.5. Figure 1 shows the survival function of Burr type XII distribution fitted to the K-M survival curve of the data using different methods of estimation. It is clear that the fitted survival functions of the Burr type XII distribution provide a good summary of the K-S survival curve with relatively better fit for the Bayes estimators under LELF.

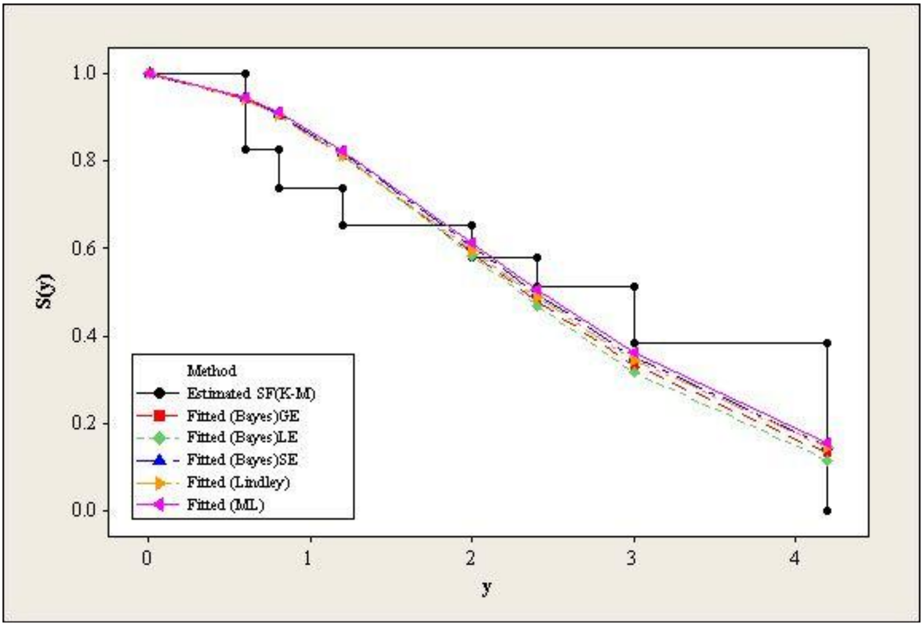

Figure 1: The Burr Type XII Survival Function Fitted to K-M Survival Curve using Different Methods of Estimation

6. POSTERIOR PREDICTIVE ASSESSMENT OF MODEL FITNESS

The nice development of Bayesian computation enables us to fit more and more realistic and sophisticated models which were not previously possible. Therefore, there is a need of general methods to assess the fitness of these models when classical methods are not applicable. One such method is the posterior predictive assessment of model fitness that connects well to classical goodness-of-fit methods. The concept of posterior predictive assessment was introduced by Guttman (1967) and applied by several authors including Rubin (1981), Gelman et al. (1996) and Upadhyay et al. (2004). The method is based on the comparison of predicted data, say $z = z_1, ..., z_n$, with the observed data

$y = y_1, \ldots, y_n$, and can be easily implemented within any MCMC sampler. If the comparisons are reasonable, we can say that both the data sets appear to have come from the same distribution. The future and observed data sets can be compared in a variety of ways. In this article we use the future and observed data sets to compute the tail area probability also known as posterior predictive *p*-value (PPP).

Suppose that $D(y, \theta)$ is any discrepancy between the observed data y and the assumed model quantities θ. The PPP is defined as

$$\begin{aligned} p &= P\left(D(z,\theta) \geq D(y,\theta) \middle| (f, y)\right) \\ &= \int_\theta P\left(D(z,\theta) \geq D(y,\theta) \middle| (f,\theta)\right) \pi\left(\theta \middle| f, y\right) d\theta. \end{aligned} \tag{20}$$

For the chi-square discrepancy measure, one can use

$$\mathrm{X}^2(y,\theta) = \sum_{i=1}^{n} \frac{\left(y_i - E\left(y_i | \theta\right)\right)^2}{\mathrm{Var}\left(y_i | \theta\right)}. \tag{21}$$

To compute the PPP based on (21), first obtain the posterior sample of θ from posterior distribution $\pi\left(\theta \middle| f, y\right)$ using MCMC Gibbs sampler. Corresponding to each generated value of θ from its posterior distribution, obtain the predictive data z from $f(z;\theta)$. Then calculate $X^2(\cdot,\theta)$ using both the predicted data $z = z_1, \ldots, z_n$ and the observed data $y = y_1, \ldots, y_n$, and compare $X^2(y,\theta)$ and $X^2(z,\theta)$.

In order to provide the numerical illustration based on the data analyzed in Section 5, the posterior distribution given in (14) is simulated using Gibbs sampler to get a sample of size 1000. Corresponding to each generated value of θ, we obtain a predictive data z from $f(z;\theta)$. Thus we have 1000 predictive data sets each of size equal to the observed data set. Figure 2 shows a scatter plot of the realized chi-square $X^2(y,\theta)$ and the predicted chi-square $X^2(z,\theta)$ based on these 1000 MCMC samples. The PPP is estimated by the proportion of points above the 45^0 line and it is 0.763. This high PPP further supports the appropriateness of the proposed model to analyze this data.

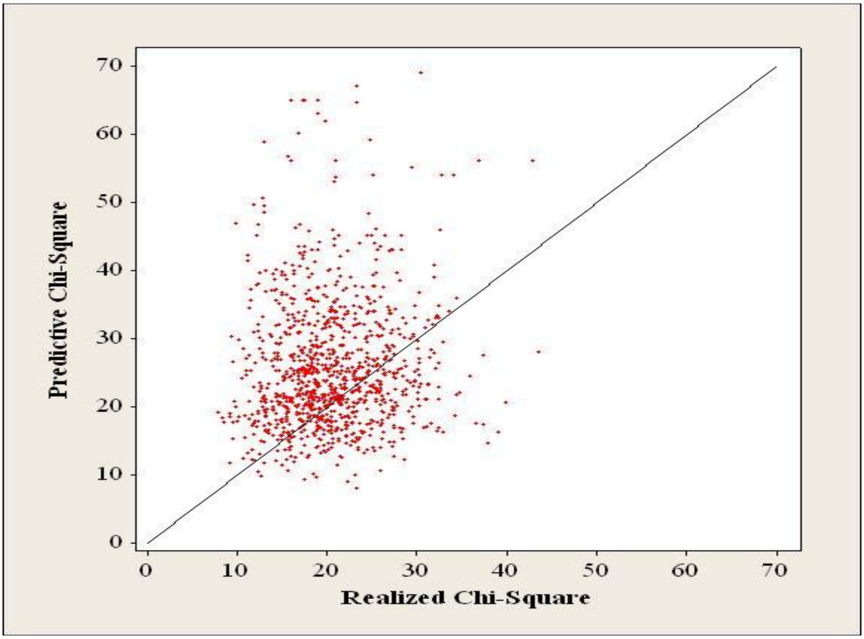

Figure 2: The Scatter Plot of Predicted Versus Realized Chi-Square Discrepancies

7. CONCLUSION

The present article provides the Bayesian analysis of Burr type XII distribution under general model of random censorship. The explicit expressions for the Bayes estimators are not possible, we use two different methods of Bayesian computation, namely, Gibbs sampling and Lindley's approximation to obtain the approximate Bayes estimates. The ML estimates are also obtained for comparison purposes. The properties of the proposed estimators are observed via numerical experiments for different combinations of sample sizes, priors and censoring rates. It is observed that with increasing sample size, the MSEs and the average confidence/credible interval lengths of the estimates decrease substantially. It is further observed that the Bayes estimates under non-informative gamma priors based MCMC Gibbs sampling perform slightly better than the ML estimates for small sample sizes and their performance is very similar for large sample sizes. The Bayes estimates under informative gamma priors based on the Lindley's approximation perform relatively better than both the ML estimates and the Bayes estimates under non-informative gamma priors based on MCMC sampling. A similar behavior is observed under the informative priors. However, the Bayes estimates under informative gamma priors perform quite better than both the ML estimates and the Bayes estimates under non-informative gamma priors. We apply proposed methods to analyze a real data set and it is seen that all the methods fit the data quite well in terms of Kolomogorov-Smirnov test of fit. The proposed model is validated using the predictive simulation ideas based on chi-square discrepancy measure.

REFERENCES

1. Berger, J.O. and Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *J. Am. Stat. Assoc.*, 88, 1412-1418.
2. Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist*., 2, 437-453.
3. Burr, I.W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, 1(2), 215-232.
4. Danish, M.Y. and Aslam, M. (2013). Bayesian estimation for randomly censored generalized exponential distribution under asymmetric loss functions. *J. Appl. Statist.*, 40 (5), 1106-1119.
5. Danish, M.Y. and Aslam, M. (2014). Bayesian inference for the randomly censored Weibull distribution. *J. Statist. Comp. Simul.*, 84(1), 215-230.
6. Devroye, L. (1984). A simple Algorithm for generating random variates with a log-concave density. *Computing,* 33, 247-257.
7. Effron, B. (1967). The two-sample problem with censored data. *Proceeding of the 5th Berkeley Symposium*, 831-853.
8. Gelman, A., Meng, X.L. and Stern, H. (1996). Posterior predicted assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733-807.
9. Geman, S. and Geman, A. (1984).Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pat. Ana. Mach. Intelli.,* 6, 721-740.
10. Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *Journal of the Royal Statistical Society, Series B,* 29, 83-100.
11. Joarder, A., Krishna, H. and Kundu, D. (2011). Inference on Weibull parameters with conventional type I censoring. *Comput. Stat. Data Anal.*, 55, 1-11.
12. Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, 53, 457-481.
13. Kundu, D. and Pradhan, B. (2009). Bayesian inference and life testing plans for GE distribution. *Science in China, Series A: Mathematics*, 52, 1373-1388.
14. Lindley, D.V. (1980). Approximate Bayes methods. *Trabajos de Estadistica*, 31, 223-237.
15. McIllmurray, M. B. and Turkie, W. (1987). Controlled trial of γ linolenic acid in Dukes' C colorectal cancer. *British Medical Journal*, 294, 1260.
16. Rodriguez, R. N. (1977). A guide to the Burr type XII distributions. *Biometrika*, 64(1), 129-134.
17. Rubin, D.B. (1981). Estimation in parallel randomized experiments. *J. Edu. Stat.*, 6, 377-400.
18. Upadhyay, S.K., Javed, I.A. and Peshwani, M. (2004). Bayesian analysis of generalized four-parameter Burr distribution via Gibbs sampler. *Metron,* LXII, 115-135.

APPENDIX

I. Expected Fisher Information Matrix

The second order partial derivatives of the log-likelihood function in (8) are

$$\frac{\partial^2 L}{\partial\theta^2}=-\frac{1}{\theta^2}\sum_{i=1}^{n}d_i,\quad \frac{\partial^2 L}{\partial\theta\partial\lambda}=-\sum_{i=1}^{n}\frac{y_i^{\lambda}\ln y_i}{1+y_i^{\lambda}},\quad \frac{\partial^2 L}{\partial\theta\partial\beta}=0,$$

$$\frac{\partial^2 L}{\partial\lambda^2}=-\frac{n}{\lambda^2}-\left(\theta+\beta+1\right)\sum_{i=1}^{n}\frac{y_i^{\lambda}\left(\ln y_i\right)^2}{\left(1+y_i^{\lambda}\right)^2},$$

$$\frac{\partial^2 L}{\partial\lambda\partial\beta}=-\sum_{i=1}^{n}\frac{y_i^{\lambda}\ln y_i}{1+y_i^{\lambda}},\quad \frac{\partial^2 L}{\partial\beta^2}=-\frac{1}{\beta^2}\left(n-\sum_{i=1}^{n}d_i\right).$$

The elements of the Fisher information matrix $\mathrm{I}\left(\theta\right)$ are

$$\mathrm{I}_{11}\left(\theta\right)=-E\left(\frac{\partial^2 L}{\partial\theta^2}\right)=\frac{n}{\theta\left(\theta+\beta\right)},\ \mathrm{I}_{12}\left(\theta\right)=-E\left(\frac{\partial^2 L}{\partial\theta\partial\lambda}\right)=\frac{n\left\{\Psi(2)-\ln(\theta+\beta)\right\}}{\lambda\left(\theta+\beta+1\right)},$$

$$\mathrm{I}_{13}\left(\theta\right)=-E\left(\frac{\partial^2 L}{\partial\theta\partial\beta}\right)=0,\ \mathrm{I}_{33}\left(\theta\right)=-E\left(\frac{\partial^2 L}{\partial\beta^2}\right)=\frac{n}{\beta\left(\theta+\beta\right)},$$

$$\mathrm{I}_{22}\left(\theta\right)=\frac{n}{\lambda^2}\left[1+\frac{\left(\theta+\beta\right)}{\left(\theta+\beta+2\right)}\left\{\left(\Psi(2)-\Psi(\theta+\beta+1)\right)^2+\Psi'(2)+\Psi'(\theta+\beta+1)\right\}\right],$$

$$\mathrm{I}_{23}\left(\theta\right)=-E\left(\frac{\partial^2 L}{\partial\lambda\partial\beta}\right)=\frac{n\left\{\Psi(2)-\ln(\theta+\beta)\right\}}{\lambda\left(\theta+\beta+1\right)},$$

where $\Psi(\cdot)$ is digamma function and $\Psi'(\cdot)$ is its derivative.

II. Lindley's Approximation

The posterior expectation (15), using the notations $\left(\theta,\lambda,\beta\right)=\left(\theta_1,\theta_2,\theta_3\right)$ and $\rho\left(\theta_1,\theta_2,\theta_3\right)=\ln\pi\left(\theta_1,\theta_2,\theta_3\right)$, can be written as

$$E\left(U\left(\theta_1,\theta_2,\theta_3\right)\middle|y,d\right)=\frac{\int_{(\theta_1,\theta_2,\theta_3)}U\left(\theta_1,\theta_2,\theta_3\right)e^{L\left(\theta_1,\theta_2,\theta_3\right)+\rho\left(\theta_1,\theta_2,\theta_3\right)}d\left(\theta_1,\theta_2,\theta_3\right)}{\int_{(\theta_1,\theta_2,\theta_3)}e^{L\left(\theta_1,\theta_2,\theta_3\right)+\rho\left(\theta_1,\theta_2,\theta_3\right)}d\left(\theta_1,\theta_2,\theta_3\right)}.\quad \text{(A.1)}$$

For large n, the expression (19) is evaluated by the Lindley's approximation as

$$\hat{U}_B\left(\theta_1,\theta_2,\theta_3\right)=\left(U_1d_1+U_2d_2+U_3d_3+d_4+d_5\right)+\frac{1}{2}\Big[A_1\left(U_1\sigma_{11}+U_2\sigma_{12}+U_3\sigma_{13}\right)$$
$$+A_2\left(U_1\sigma_{21}+U_2\sigma_{22}+U_3\sigma_{23}\right)+A_3\left(U_1\sigma_{31}+U_2\sigma_{32}+U_3\sigma_{33}\right)\Big]+U\left(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3\right).\quad \text{(A.2)}$$

where

$$d_i = \rho_1\sigma_{i1} + \rho_2\sigma_{i2} + \rho_3\sigma_{i3},\; U_i = \frac{\partial U(\theta_1,\theta_2,\theta_3)}{\partial\theta_i},\; \rho_i = \frac{\partial\rho(\theta_1,\theta_2,\theta_3)}{\partial\theta_i},$$

$$d_4 = U_{12}\sigma_{12} + U_{13}\sigma_{13} + U_{23}\sigma_{23},\; d_5 = \frac{1}{2}\left(U_{11}\sigma_{11} + U_{22}\sigma_{22} + U_{33}\sigma_{33}\right),$$

$$A_1 = \sigma_{11}L_{111} + 2\sigma_{12}L_{121} + 2\sigma_{13}L_{131} + 2\sigma_{23}L_{231} + \sigma_{22}L_{221} + \sigma_{33}L_{331},$$

$$A_2 = \sigma_{11}L_{112} + 2\sigma_{12}L_{122} + 2\sigma_{13}L_{132} + 2\sigma_{23}L_{232} + \sigma_{22}L_{222} + \sigma_{33}L_{332},$$

$$A_3 = \sigma_{11}L_{113} + 2\sigma_{12}L_{123} + 2\sigma_{13}L_{133} + 2\sigma_{23}L_{233} + \sigma_{22}L_{223} + \sigma_{33}L_{333},$$

and $L_{ijk} = \dfrac{\partial^3 L(\theta_1,\theta_2,\theta_3)}{\partial\theta_i\partial\theta_j\partial\theta_k}, \quad i, j, k = 1, 2, 3,$

Moreover σ_{ij} is ijth element of minus the inverse of Fisher information matrix and all the quantities are evaluated at ML estimates $\left(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3\right)$. In our case

$$\rho_1 = \frac{a_1 - 1}{\hat{\theta}} - b_1,\; \rho_2 = \frac{a_2 - 1}{\hat{\lambda}} - b_2,\; \rho_3 = \frac{a_3 - 1}{\hat{\beta}} - b_3,$$

$$A_1 = \frac{2\sigma_{11}}{\hat{\theta}^3}\sum_{i=1}^{n} d_i - \sigma_{22}\sum_{i=1}^{n}\frac{y_i^{\hat{\lambda}}(\ln y_i)^2}{\left(1 + y_i^{\hat{\lambda}}\right)^2},$$

$$A_2 = \frac{2n\sigma_{22}}{\hat{\lambda}^3} - 2(\sigma_{12} + \sigma_{23})\sum_{i=1}^{n}\frac{y_i^{\hat{\lambda}}(\ln y_i)^2}{\left(1 + y_i^{\hat{\lambda}}\right)^2} - \left(\hat{\theta} + \hat{\beta} + 1\right)\sigma_{22}\sum_{i=1}^{n}\frac{y_i^{\hat{\lambda}}\left(1 - y_i^{\hat{\lambda}}\right)(\ln y_i)^3}{\left(1 + y_i^{\hat{\lambda}}\right)^3}$$

And $A_3 = \dfrac{2\sigma_{33}}{\hat{\beta}^3}\left(n - \sum_{i=1}^{n} d_i\right) - \sigma_{22}\sum_{i=1}^{n}\dfrac{y_i^{\hat{\lambda}}(\ln y_i)^2}{\left(1 + y_i^{\hat{\lambda}}\right)^2}.$

To find the Bayes estimate of θ under SELF, take $U(\theta,\lambda,\beta) = \theta$ in (A.2). Since $U_1 = 1$ and all other U-terms are zero, we have

$$\hat{\theta}_{BL} = \hat{\theta} + d_1 + \frac{1}{2}\left(A_1\sigma_{11} + A_2\sigma_{21} + A_3\sigma_{31}\right). \tag{A.3}$$

Now the expression (19) for the Bayes estimate of θ using the Lindley's method follows from (A.3) by substituting

$$d_1 = \left(\frac{a_1 - 1}{\hat{\theta}}\right)\sigma_{11} + \left(\frac{a_2 - 1}{\hat{\lambda}}\right)\sigma_{21} + \left(\frac{a_3 - 1}{\hat{\beta}}\right)\sigma_{31}.$$

The expressions 20-23 can be obtained similarly.