

Classical and Bayes Analysis of A Competing Risk Model Based on Two Weibull Distributions with Increasing and Decreasing Hazards

Ankita Gupta^{a,c}, Rakesh Ranjan^b and Satyanshu K. Upadhyay^{a,b}

Department of Statistics^a
DST Centre for Interdisciplinary Mathematical Sciences^b
Banaras Hindu University, Varanasi-221 005, India
Department of Mathematics and Statistics^c
Banasthali Vidyapith, Banasthali -304022, India.
ankitagupta1611r@gmail.com

Abstract

The paper considers a competing risk model based on two Weibull distributions, one with increasing and the other with decreasing hazard rate. It then considers both classical and Bayesian analysis of the model, the later development utilizes the informative but weak priors for the parameters. The analysis is facilitated by the fact that a competing risk model can be considered as an incomplete data model even if the situation allows all the observations on the test to be made available although the results are extended for censored data cases as well. The paper uses the expectation-maximization algorithm for classical maximum likelihood estimation and Gibbs sampler algorithm for posterior based inferences. It is shown that the likelihood function offers unique and consistent maximum likelihood estimates. The results are illustrated based on a real data example. Finally, the compatibility of the model is examined for the considered real data set using some standard tools of Bayesian paradigm.

Keywords: Competing risk, Weibull distribution, Right censoring, Expectation-maximization algorithm, Gibbs sampler, Predictive p-value.

1 Introduction

In real life situations, it is quite frequent that an item is exposed to experience failure due to more than one cause at the same point of time. Say for instance, in reliability experiment, an item may fail due to one of several possible causes, such as breakdown, manufacturing defects, etc. Similarly, in medical experiment, a patient suffering from several diseases may die because of the one that relapses first. Such situations usually come under the purview of competing risk scenario where an item or organism is subject to several competing causes and the failure may occur because of any cause that arises first. A traditional approach for modelling failure time data in presence of competing risks is to assume that there is a latent failure time associated with each of the causes to which the item is exposed and the realized failure time of the item is lowest among these latent failure times. Moreover, these latent failure times are assumed to be independent of

each other following some distributions, either same or different (see [5]).

The simplest case of competing risk model arises when failure of an item is subject to two possible causes. In this case, a competing risk model is defined considering the distribution of minimum of two different failure times. The analysis of such competing risk models based on two failure time distributions is considered by several authors. [10] studied a competing risk model defined on the basis of exponential and Weibull failures to model failures due to shock and wear out, respectively, and discussed the properties of maximum likelihood (ML) estimates of resulting model parameters. [21] proposed both parametric and non-parametric estimation techniques for a two-component competing risk model under the assumption that the component failure times are exponentially distributed. [4] proposed a competing risk model for a situation where the population is exposed to wear out failures but a fraction of population is also exposed to early failures. The authors obtained ML estimates of model parameters under the assumption that failure modes follow either Weibull or lognormal distributions. [2] considered both classical and Bayesian analyses of a model based on minimum of Weibull and exponential failures assuming that former results in failures due to ageing and the latter results in accidental failures as the two competing causes. More recently, [22] suggested an alternative competing risk model based on gamma and exponential distributions to model ageing and accidental failures, respectively, and provided complete classical as well as Bayesian analyses of the resulting competing risk model.

This paper models a situation where infancy and ageing work together to induce failures. Such a situation may occur where an item having an initial birth defect is also subject to failure due to ageing. Many real life situations can be found in practice that may include items from automobile segment, high-power transmitting tubes and computer disk-drives, etc. Obviously, to deal with such situations, one may consider, among various other choices, a model based on two distributions, one with decreasing hazard rate and other with increasing hazard rate. The resulting failure time can then be considered as the minimum of two failure times, one corresponding to decreasing hazard rate distribution and the other corresponding to increasing hazard rate distribution. One can, of course, consider a number of models to define decreasing and increasing hazard rate behaviour. We, however, consider a competing risk model defined on the basis of two Weibull distributions, one corresponding to decreasing hazard rate situation and the other corresponding to increasing hazard rate situation.

The Weibull distribution is an important failure time distribution that encompasses both increasing and decreasing hazard rate behaviour and, perhaps because of its enormous scope and flexibility, it has been used to describe both initial as well as ageing failures (see, for example, Lawless (2002)). The probability density function (pdf) of its simplest two-parameter form ($\mathcal{W}(\theta, \beta)$) can be written as

$$f_w(t|\theta, \beta) = \frac{\beta}{\theta} \left(\frac{t}{\theta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\theta}\right)^\beta\right]; \quad t > 0, \quad \theta > 0, \quad \beta > 0, \quad (1)$$

where θ and β are scale and shape parameters, respectively. It is actually the shape parameter that results in different characteristics of the model. Say, for instance, $\beta < 1.0$ defines a decreasing hazard rate distribution that can be considered to model early birth defects. Similarly, $\beta > 1.0$ defines increasing hazard rate behavior, a situation that can be very well used for defining failures due to ageing. Although, not of importance in the present work, the distribution reduces to constant hazard-rate exponential distribution when $\beta = 1.0$. The important reliability characteristics such as the reliability at time t , the hazard rate and the mean time to failure (MTF) for $\mathcal{W}(\theta, \beta)$ can be written as

$$R_w(t) = \exp\left[-\left(\frac{t}{\theta}\right)^\beta\right], \quad (2)$$

$$h_w(t) = \frac{\beta}{\theta} \left(\frac{t}{\theta}\right)^{\beta-1}, \quad (3)$$

and

$$MTF_{\mathcal{W}} = \theta\Gamma(1 + 1/\beta), \quad (4)$$

respectively.

As mentioned, the proposed model is defined on the basis of two Weibull distributions, one with $\beta < 1.0$ and the other with $\beta > 1.0$. A similar model, named as Bi-Weibull (\mathcal{BW}) distribution, was also entertained by [1] but they did not impose any restriction on the corresponding shape parameters. The authors considered the model as a particular case of poly-Weibull model and analyzed in a Bayesian framework using Gibbs sampler algorithm. The illustration was, however, based on a simulated data set. The unrestricted model was later analyzed by a number of authors including [16], [6] and [17]). Whereas [16] provided complete parametric characterization of the model in a three dimensional parameter space, [6] analyzed the model for real as well as generated data sets using both classical and Bayesian tools. In fact, [6] mentioned that the likelihood of the model could be obtained in a simple manner and there was no need of using Gibbs sampler algorithm. The authors rather used standard likelihood method to obtain classical inferences and both Laplace's method and sampling-importance resampling technique for Bayesian inferences.

Recently [17] considered various characterizations of (\mathcal{BW}) model in its four-parameter setup and obtained ML estimates of model parameters with observed information matrix. The authors then used the results on asymptotic normality of ML estimators to derive approximate confidence intervals and confidence regions for the model parameters. The results obtained by [17] are certainly extensive but inferential aspects are comparatively meagre as compared to various characterizations of the model. Moreover, the model considered by the authors is a simplified version that separates the two parameters of the Weibull model and it is often considered for mathematical convenience.

The present paper can be considered as an extension of previous work where two Weibull distributions with form given in (1) are used to define the competing risk model. Since the shape parameters of the corresponding components are restricted, we shall call the resulting competing risk model as restricted Bi-Weibull ($\mathcal{BW}_{\mathcal{R}}$) model. The paper then considers not only the complete Bayes analysis using weak proper priors but also the ML estimation of model parameters using expectation-maximization (EM) algorithm. It is shown that there exists a unique consistent solution of the likelihood function, a result that may be considered significant for the likelihood form arising from $\mathcal{BW}_{\mathcal{R}}$ model. Throughout the inferential developments are done for both complete and censored data cases, the latter situation is, of course, important in failure time data analysis but not considered in any of the previous references.

The plan of the paper is as follows. The next section introduces the proposed $\mathcal{BW}_{\mathcal{R}}$ model and provides a few important characteristics for the same. Some of the results provided in this section are given in slightly different forms in [17] but we have reproduced them for a ready reference and also because of the fact that the model form for the Weibull distribution used by [17] is different from the one considered by us in this paper. Section 3 details the ML estimation of the model parameters for the considered $\mathcal{BW}_{\mathcal{R}}$ model using EM algorithm. Since the competing risk model as considered in the paper is an incomplete data model in the sense that we do not have the actual cause of failures, the EM algorithm happens to be an important choice for such situations. Section 4 provides the Bayesian model formulation for the considered likelihood and prior combinations. The resulting posterior is not easy to deal with and, therefore, we have considered the use of Gibbs sampler algorithm for drawing the posterior based inferences. The section also provides a brief discussion of the Gibbs sampler algorithm and its implementation details for the model in hand. Throughout the censored data cases are also considered and appropriate implementation details for the same are given. Section 5 is given for completeness that provides a few important tools for model compatibility study in Bayesian paradigm. These tools are used in the next section where we have considered a real data set for numerical illustration and provided the compatibility of the data with $\mathcal{BW}_{\mathcal{R}}$ to justify our analysis. Finally, a conclusion is given in the last section along with the proof of theorems in the Appendix.

2 The Proposed \mathcal{BW}_R Model

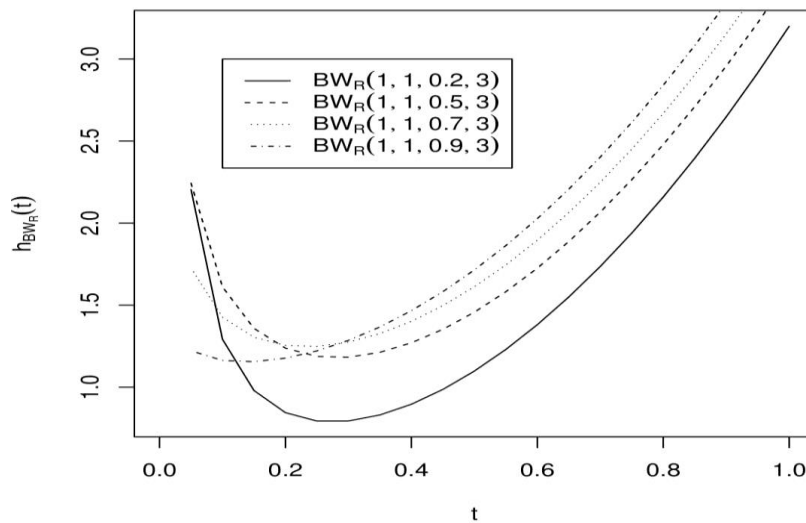
Let T_1 be a random variable following $\mathcal{W}(\theta_1, \beta_1)$, $\beta_1 < 1.0$, and T_2 be another random variable following $\mathcal{W}(\theta_2, \beta_2)$, $\beta_2 > 1.0$. A competing risk \mathcal{BW}_R model based on random variable $T = \min(T_1, T_2)$ can be characterized by four parameters $\theta_1, \theta_2, \beta_1$ and β_2 where $\theta_1 > 0, \theta_2 > 0, \beta_1 < 1.0$ and $\beta_2 > 1.0$. The hazard rate of $\mathcal{BW}_R(\theta_1, \theta_2, \beta_1, \beta_2)$ model can be expressed as sum of hazard rates of its components, which can be given as

$$h_{\mathcal{BW}_R}(t) = \frac{\beta_1}{\theta_1} \left(\frac{t}{\theta_1}\right)^{\beta_1-1} + \frac{\beta_2}{\theta_2} \left(\frac{t}{\theta_2}\right)^{\beta_2-1}. \quad (5)$$

The hazard rate of general \mathcal{BW} model is discussed by several authors. A few important references include [16], [6] and [17]. Truly speaking, the hazard rate of \mathcal{BW} model may be characterized by its shape parameters β_1 and β_2 . If $\min(\beta_1, \beta_2) > 1.0$, the hazard rate is increasing; if $\max(\beta_1, \beta_2) < 1.0$, the hazard rate is decreasing; and for $\beta_1 < 1.0$ and $\beta_2 > 1.0$, the hazard rate is bathtub shaped. On the other hand, the hazard rate curve for \mathcal{BW}_R model is always bathtub shaped since one of its shape parameters is less than unity. The hazard rate curve, however, changes its shape with different choices of β_1 and β_2 . The change-point for bathtub shaped hazard rate for the model \mathcal{BW}_R can be obtained as

$$t^* = \left[\frac{\beta_1(1-\beta_1) \times \theta_2^{\beta_2}}{\beta_2(\beta_2-1) \times \theta_1^{\beta_1}} \right]^{\frac{1}{\beta_2-\beta_1}}. \quad (6)$$

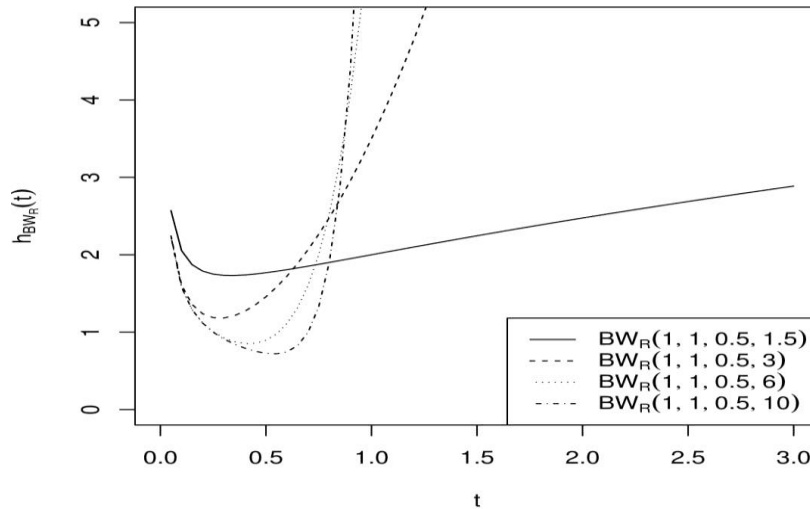
Figure 1: Hazard rate curves corresponding to \mathcal{BW}_R model for different values of β_1 .



It can be seen that t^* is always positive since β_1 is less than unity and the other shape parameter β_2 is greater than unity. Moreover, it can be seen that the change point increases as β_1 approaches towards zero. The first derivative of $h(t)$ can be shown to be negative(positive) for values of t less(greater) than t^* and, therefore, the hazard rate is always decreasing(increasing) for values of t less(greater) than t^* and finally approaching to infinity both when t approaches to zero or infinity.

The hazard rate curves corresponding to \mathcal{BW}_R model are shown in Figures 1-2 for some arbitrary choices of the parameters. Since the shape parameters are only responsible for changing the hazard rates shapes, both the scale parameters θ_1 and θ_2 are fixed at unity. It is obvious from the figures that the curves exhibit convex shapes in every case with a unique minimum given by (6). Rest of the conclusions are same that have been detailed in the preceding paragraph.

Figure 2: Hazard rate curves corresponding to \mathcal{BW}_R model for different values of β_2 .



The reliability function corresponding to \mathcal{BW}_R model can be obtained as product of reliability functions of its component models. The expression for the same can be written as

$$R_{\mathcal{BW}_R}(t) = \exp \left[- \left(\frac{t}{\theta_1} \right)^{\beta_1} - \left(\frac{t}{\theta_2} \right)^{\beta_2} \right]. \quad (7)$$

Using (5) and (7), one can easily obtain the pdf corresponding to \mathcal{BW}_R model. The same can be written as

$$f_{\mathcal{BW}_R}(t) = \left[\frac{\beta_1}{\theta_1} \left(\frac{t}{\theta_1} \right)^{\beta_1-1} + \frac{\beta_2}{\theta_2} \left(\frac{t}{\theta_2} \right)^{\beta_2-1} \right] \times \exp \left[- \left(\frac{t}{\theta_1} \right)^{\beta_1} - \left(\frac{t}{\theta_2} \right)^{\beta_2} \right]. \quad (8)$$

Fortunately, the moments of \mathcal{BW}_R distribution exist in closed forms and the corresponding expression for the s^{th} moment (see also [17]) can be written as

$$\begin{aligned} \mu'_s &= \frac{\beta_1 \theta_2^s}{\beta_2} \sum_{m=0}^{\infty} (-1)^m \left(\frac{\theta_2}{\theta_1} \right)^{(m+1)\beta_1} \Gamma \left(\frac{s+(m+1)\beta_1}{\beta_2} \right) \\ &+ \theta_2^s \sum_{m=0}^{\infty} (-1)^m \left(\frac{\theta_2}{\theta_1} \right)^{m\beta_1} \Gamma \left(\frac{s+m\beta_1}{\beta_2} + 1 \right). \end{aligned} \quad (9)$$

Another important characteristic of interest is the probability of failure due to one of the causes. The corresponding expression for the probability of failure due to early birth defect can be written as

$$P_r(t = t_1) = \int_0^{\infty} \frac{\beta_1}{\theta_1} \left(\frac{t}{\theta_1} \right)^{\beta_1-1} \exp \left[- \left(\frac{t}{\theta_1} \right)^{\beta_1} - \left(\frac{t}{\theta_2} \right)^{\beta_2} \right] dt,$$

which, on simplification, reduces to

$$P_r(t = t_1) = \frac{\beta_1}{\beta_2} \sum_{m=0}^{\infty} (-1)^m \left(\frac{\theta_2}{\theta_1} \right)^{(m+1)\beta_1} \Gamma \left(\frac{(m+1)\beta_1}{\beta_2} \right). \quad (10)$$

Obviously, the probability of failure due to ageing can be written as complimentary probability of (10).

3 The ML Estimation

Let n items having failure time distribution given in (8) be put on test and let $y: (y_i = (t_i, \delta_i); i = 1, 2, \dots, n)$ denote the corresponding observations. We use t_i to denote the failure time of i^{th} unit and δ_i as an associated censoring indicator such that $\delta_i = 0$ indicates that i^{th} observation is right censored at time t_i and $\delta_i = 1$ indicates that t_i is the observed failure time of i^{th} item. The corresponding likelihood function (LF) can be written as

$$L(y|\boldsymbol{\theta}) = \prod_{i=1}^n [f_{\mathcal{W}_R}(t_i)]^{\delta_i} [R_{\mathcal{W}_R}(t_i)]^{1-\delta_i},$$

which, using (7) and (8), becomes

$$L(y|\boldsymbol{\theta}) = \prod_{i=1}^n \left\{ \frac{\beta_1}{\theta_1} \left(\frac{t_i}{\theta_1} \right)^{\beta_1-1} + \frac{\beta_2}{\theta_2} \left(\frac{t_i}{\theta_2} \right)^{\beta_2-1} \right\}^{\delta_i} \exp \left[- \left(\frac{t_i}{\theta_1} \right)^{\beta_1} - \left(\frac{t_i}{\theta_2} \right)^{\beta_2} \right], \quad (11)$$

where $\boldsymbol{\theta}$ is used to denote the parameter vector $(\theta_1, \theta_2, \beta_1, \beta_2)$. It is important to mention here that the LF corresponding to complete data case can be written as a special case of (11) when all δ_i s are taken to be unity. The LF for complete data case can, therefore, be written as

$$L(y|\boldsymbol{\theta}) = \prod_{i=1}^n \left[\frac{\beta_1}{\theta_1} \left(\frac{t_i}{\theta_1} \right)^{\beta_1-1} + \frac{\beta_2}{\theta_2} \left(\frac{t_i}{\theta_2} \right)^{\beta_2-1} \right] \exp \left[- \left(\frac{t_i}{\theta_1} \right)^{\beta_1} - \left(\frac{t_i}{\theta_2} \right)^{\beta_2} \right]. \quad (12)$$

The ML estimates are usually obtained by maximizing the LF, but in high dimensional case direct maximization of LF may be often difficult and, sometimes, may also lead to unstable results. Fortunately, in our case, we are in a position to verify that the corresponding likelihood equations offer unique and consistent solutions both for complete and censored data cases (the details of our proof are given in the Appendix). Instead of direct maximization of the LF, we propose to obtain ML estimates of model parameters using EM algorithm. It may be noted that the EM algorithm is an iterative method proposed by [7] for computing ML estimates of model parameters, especially in the situations where data or model can be viewed as incomplete. This is perhaps the reason that the implementation of EM algorithm is facilitated in a competing risk scenario simply because the data corresponding to a competing risk model are always incompletely specified in the sense that we do not know exact cause of failure for a particular item (see also [23]). The algorithm is lucid and simple in the sense that it does not require calculation of the Jacobian matrix which is normally required in other optimization techniques like Newton-Raphson, etc. (see also [20] and [18]).

For implementing the EM algorithm, we rather work with an alternative formulation of the LF by introducing missing observations in the form of indicator variables (see also [2] since each observation t_i has a missing component in the sense that it is not known exactly which of the two causes, early birth defect or ageing, is responsible for producing t_i . Similarly, each observation t_i can either be a failure time or censoring time as governed by the censoring indicator δ_i . Thus to provide an alternative formulation of the LF, we associate with each t_i a missing cause indicator z_i with components of z_i as (z_i^1, z_i^2) , $i = 1, 2, \dots, n$. We may then define $z_i^1=1$ and $z_i^2=0$ as an indication that t_i is an observed failure time arising from $\mathcal{W}(\theta_1, \beta_1)$ and, similarly, $z_i^1 = 0$ and $z_i^2 = 1$ as an indication that t_i is an observed failure time arising from $\mathcal{W}(\theta_2, \beta_2)$. Obviously, for each censored observation t_i ($\delta_i = 0$), we can define associated z_i^1 and z_i^2 both equal to zero, $i = 1, 2, \dots, n$. Moreover, since failure can arise due to only one of the two competing causes, both z_i^1 and z_i^2 cannot take value unity simultaneously.

With the assumptions as given above, the pdf for each observation can be written as

$$\begin{aligned} f(x_i) &= [f_{\mathcal{W}}(t_i, \theta_1, \beta_1)]^{z_i^1} [R_{\mathcal{W}}(t_i, \theta_1, \beta_1)]^{1-z_i^1} [f_{\mathcal{W}}(t_i, \theta_2, \beta_2)]^{z_i^2} [R_{\mathcal{W}}(t_i, \theta_2, \beta_2)]^{1-z_i^2} \\ &= [h_{\mathcal{W}}(t_i, \theta_1, \beta_1)]^{z_i^1} [h_{\mathcal{W}}(t_i, \theta_2, \beta_2)]^{z_i^2} R_{\mathcal{W}}(t_i, \theta_1, \beta_1) R_{\mathcal{W}}(t_i, \theta_2, \beta_2), \end{aligned} \quad (13)$$

where $f_{\mathcal{W}}$, $h_{\mathcal{W}}$ and $R_{\mathcal{W}}$ denote the pdf, hazard function and the reliability function of Weibull model, respectively, and $x_i = (t_i, z_i)$, $i = 1, 2, \dots, n$. Obviously, the corresponding LF based on a set of n observations can be written as

$$L(x|\boldsymbol{\theta}) \propto \prod_{i=1}^n [h_{\mathcal{W}}(t_i, \theta_1, \beta_1)]^{z_i^1} [h_{\mathcal{W}}(t_i, \theta_2, \beta_2)]^{z_i^2} R_{\mathcal{W}}(t_i, \theta_1, \beta_1) R_{\mathcal{W}}(t_i, \theta_2, \beta_2), \quad (14)$$

and the associated log likelihood can be written as

$$l(x|\boldsymbol{\theta}) = \sum_{i=1}^n \{z_i^1 \ln[h_{\mathcal{W}}(t_i, \theta_1, \beta_1)] + z_i^2 \ln[h_{\mathcal{W}}(t_i, \theta_2, \beta_2)] + \ln[R_{\mathcal{W}}(t_i, \theta_1, \beta_1)] + \ln[R_{\mathcal{W}}(t_i, \theta_2, \beta_2)]\}. \quad (15)$$

where $x = (x_1, x_2, \dots, x_n)$. Once the log likelihood is specified as in (15), the EM algorithm can proceed as usual in two steps, the expectation (E) step and the maximization (M) step. The E step may involve the calculation of expected value of the log likelihood based on the complete data and the current realization of model parameters. The M step, on the other hand, intends to find the values of model parameters that maximizes the expected log likelihood function evaluated at E step.

To clarify, suppose $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\beta}_1, \tilde{\beta}_2)$ is the current realization of parameter vector $\boldsymbol{\theta}$ at a particular iteration then the expected log LF $\{(= E[l(x|\boldsymbol{\theta})|t, \tilde{\boldsymbol{\theta}}]), t = (t_1, t_2, \dots, t_n)\}$ denoted by $A_l(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$, say, can be obtained as

$$A_l(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}}) = \sum_{i=1}^n \{E(z_i^1|t_i, \tilde{\boldsymbol{\theta}}) \ln[h_{\mathcal{W}}(t_i, \theta_1, \beta_1)] + E(z_i^2|t_i, \tilde{\boldsymbol{\theta}}) \ln[h_{\mathcal{W}}(t_i, \theta_2, \beta_2)] + \ln[R_{\mathcal{W}}(t_i, \theta_1, \beta_1)] + \ln[R_{\mathcal{W}}(t_i, \theta_2, \beta_2)]\}, \quad (16)$$

where $E(z_i^1|t_i, \tilde{\boldsymbol{\theta}}) = \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1))$ and $E(z_i^2|t_i, \tilde{\boldsymbol{\theta}}) = \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2))$. $\tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1))$ [$\tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2))$] gives the probability that the observation t_i is arising from $\mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1)$ [$\mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2)$] given the current realization of parameters. These probabilities can be obtained as

$$\tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1)) = P_r(z_i^1 = 1|t_i, \tilde{\boldsymbol{\theta}}) = \begin{cases} 0 & \text{if } \delta_i = 0 \\ \frac{h_{\mathcal{W}}(t_i, \tilde{\theta}_1, \tilde{\beta}_1)}{h_{\mathcal{W}}(t_i, \tilde{\theta}_1, \tilde{\beta}_1) + h_{\mathcal{W}}(t_i, \tilde{\theta}_2, \tilde{\beta}_2)} & \text{if } \delta_i = 1, \end{cases} \quad (17)$$

and

$$\tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2)) = P_r(z_i^2 = 1|t_i, \tilde{\boldsymbol{\theta}}) = \begin{cases} 0 & \text{if } \delta_i = 0 \\ \frac{h_{\mathcal{W}}(t_i, \tilde{\theta}_2, \tilde{\beta}_2)}{h_{\mathcal{W}}(t_i, \tilde{\theta}_1, \tilde{\beta}_1) + h_{\mathcal{W}}(t_i, \tilde{\theta}_2, \tilde{\beta}_2)} & \text{if } \delta_i = 1 \end{cases} \quad (18)$$

respectively. From (16), one can see that $A_l(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$ can be split into two parts as given below

$$A_l(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}}) = A_l^1(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}}) + A_l^2(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}}), \quad (19)$$

where

$$A_l^1(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}}) = \sum_{i=1}^n \{\tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1)) \ln[h_{\mathcal{W}}(t_i, \theta_1, \beta_1)] + \ln[R_{\mathcal{W}}(t_i, \theta_1, \beta_1)]\}, \quad (20)$$

and

$$A_l^2(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}}) = \sum_{i=1}^n \{\tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2)) \ln[h_{\mathcal{W}}(t_i, \theta_2, \beta_2)] + \ln[R_{\mathcal{W}}(t_i, \theta_2, \beta_2)]\}. \quad (21)$$

From (20) and (21), it is obvious that both $A_l^1(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$ and $A_l^2(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$ depend separately on (θ_1, β_1) and (θ_2, β_2) , respectively, and, therefore, $A_l(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$ in (19) can be maximized by maximizing separately the two terms $A_l^1(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$ and $A_l^2(\boldsymbol{\theta}|t, \tilde{\boldsymbol{\theta}})$. This partitioning has an obvious advantage in the sense that it reduces a four-dimensional optimization problem into a two-dimensional optimization problem and thereby resulting into an increased efficiency of the algorithm. Thus differentiating (20) with respect to θ_1 and β_1 and (21) with respect to θ_2 and β_2 and equating all the derivatives to zero, we get the following set of equations

$$\frac{1}{\tilde{\beta}_1} + \frac{\sum_{i=1}^n \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1)) \ln(t_i)}{\sum_{i=1}^n \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1))} - \frac{\sum_{i=1}^n t_i^{\tilde{\beta}_1} \ln(t_i)}{\sum_{i=1}^n t_i^{\tilde{\beta}_1}} = 0, \quad (22)$$

$$\hat{\theta}_1 = \left(\frac{\sum_{i=1}^n t_i^{\tilde{\beta}_1}}{\sum_{i=1}^n \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_1, \tilde{\beta}_1))} \right)^{\frac{1}{\tilde{\beta}_1}}, \quad (23)$$

$$\frac{1}{\beta_2} + \frac{\sum_{i=1}^n \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2)) \ln(t_i)}{\sum_{i=1}^n \tilde{p}(t_i, \mathcal{W}(\tilde{\theta}_2, \tilde{\beta}_2))} - \frac{\sum_{i=1}^n t_i^{\beta_2} \ln(t_i)}{\sum_{i=1}^n t_i^{\beta_2}} = 0, \quad (24)$$

$$\hat{\theta}_2 = \left(\frac{\sum_{i=1}^n t_i^{\hat{\beta}_2}}{\sum_{i=1}^n \tilde{p}(t_i, \mathcal{W}(\hat{\theta}_2, \hat{\beta}_2))} \right)^{\frac{1}{\hat{\beta}_2}}. \quad (25)$$

The value of $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\beta}_1, \hat{\beta}_2)$ can be obtained by solving the above system of equations. Moreover, once $\hat{\theta}$ is obtained, we update $\tilde{\theta}$ by $\hat{\theta}$ for the next iterations. The process is repeated unless the systematic pattern of convergence or desired level of accuracy is achieved.

4 Bayesian Model Formulation

In order to provide the Bayesian model formulation for analyzing the model in (8), let us begin with specifying the prior distributions for the four parameters involved in the model. We consider independent uniform priors for the two shape parameters β_1 and β_2 and independent conditional priors for θ_1 and θ_2 such that for given β_1 and β_2 , the parameters $\theta_1^{\beta_1}$ and $\theta_2^{\beta_2}$ have inverted gamma $\mathcal{IG}(a, b)$ and $\mathcal{IG}(c, d)$ distributions, respectively (see also [1]). The first term in the parenthesis represents the shape parameter of \mathcal{IG} distribution while the second term is the scale parameter. The considered priors for the parameters can be written as

$$\begin{aligned} \pi_1(\beta_1) &\propto 1; & 0 < \beta_1 \leq 1 \\ \pi_2(\beta_2) &\propto \frac{1}{\beta_{2u}^{\beta_2-1}}; & 1 \leq \beta_2 \leq \beta_{2u} \\ \pi_3(\theta_1 | \beta_1) &\propto \frac{1}{\theta_1^{\beta_1 a + 1}} \exp\left[\frac{-b}{\theta_1^{\beta_1}}\right]; & \theta_1, a, b > 0 \\ \pi_4(\theta_2 | \beta_2) &\propto \frac{1}{\theta_2^{\beta_2 c + 1}} \exp\left[\frac{-d}{\theta_2^{\beta_2}}\right]; & \theta_2, c, d > 0. \end{aligned} \quad (26)$$

where a, b, c, d and β_{2u} are the hyperparameters. The choice of prior hyperparameters is always crucial and in case of non-availability of any specific information, either subjective or objective, one can consider vague or weak priors for the parameters and allow inferences to remain data dependent. This can be done, for instance, by taking a large choice of β_{2u} in case of β_2 . Similarly, in case of θ_1 and θ_2 , one can consider choices of a, b, c and d in such a way that resulting \mathcal{IG} distribution becomes more or less flat. It is important to mention here that the standard non-informative priors suggested for the parameters of Weibull distribution cannot be used here as such priors lead to improper posterior in a situation where all the observations arise from one of the two Weibull components in the model (8) (see also [1]).

The prior hyperparameters can also be specified on the basis of experts' opinion if the same are available. Say, for instance, while specifying β_{2u} associated with β_2 of $\mathcal{W}(\theta_2, \beta_2)$ model, the expert conveys that the hazard rate is not increasing abruptly with a steep rise in its behaviour rather rises in such a way that the slope of the hazard rate curve increases from the beginning itself, say at a constant rate. The expert also conveys that the choice of β_2 is such that the corresponding Weibull distribution is exhibiting skewed behaviour with slightly high variability. It is to be noted that for large values of shape parameter, the Weibull distribution approaches to symmetry (close to normality) with very low variability. As such, the expert feels that β_2 cannot be too large and in any case cannot go beyond (say) 10.0, that is, a value suggested for β_{2u} . Similarly, for the \mathcal{IG} prior, the expert can be asked to specify at least two characteristics of the prior model so that two of its hyperparameters can be made known using these two specified characteristics. The characteristics can be simply mean and variance of the \mathcal{IG} distribution or two of its quantiles, etc. Finally, it is essential to mention that the prior of β_1 does not involve any hyperparameter so we do not need any opinion from the expert in this case.

To proceed further, let n items with failure time distribution given in (8) be put on test and let

$t: t_1, t_2, \dots, t_n$ be the observed failure times. Combining the priors in (26) with the LF in (12) via Bayes theorem yields the corresponding posterior distribution and the same can be written up to proportionality as

$$p(\theta_1, \theta_2, \beta_1, \beta_2 | t) \propto \left\{ \prod_{i=1}^n \left[\frac{\beta_1}{\theta_1} \left(\frac{t_i}{\theta_1} \right)^{\beta_1-1} + \frac{\beta_2}{\theta_2} \left(\frac{t_i}{\theta_2} \right)^{\beta_2-1} \right] \right\} \exp \left[-\sum_{i=1}^n \left(\frac{t_i}{\theta_1} \right)^{\beta_1} - \sum_{i=1}^n \left(\frac{t_i}{\theta_2} \right)^{\beta_2} \right] \\ \times \frac{1}{\theta_1^{\beta_1 a+1}} \exp \left[\frac{-b}{\theta_1^{\beta_1}} \right] \times \frac{1}{\theta_1^{\beta_1 a+1}} \exp \left[\frac{-b}{\theta_1^{\beta_1}} \right] \times I_{(0,1)}(\beta_1) \times I_{(1,\beta_{2u})}(\beta_2); \quad (27) \\ \theta_1 > 0, \theta_2 > 0,$$

where $I_{(v_1, v_2)}(\xi)$ is an indicator function that takes value unity if ξ lies in the interval (v_1, v_2) and zero otherwise. It can be easily seen that the form of the posterior given in (27) is difficult to offer closed form solution and, therefore, one can proceed with sample based approaches. The Gibbs sampler is, however, difficult since the corresponding full conditionals from (27) are not available for easy sample generation. We may, therefore, recommend Metropolis-Hastings algorithm by defining a four-dimensional appropriately centred and scaled candidate generating density (see, for example, [25]) and proceed with ML estimates and the corresponding Hessian based approximation as the initial values for running the chain.

Alternatively, one can use the augmented data structure as given in Section 3 and resort to implementation of the Gibbs sampler algorithm in a routine manner. Following (14), the corresponding likelihood for augmented data structure can be written as

$$L(x|\theta) \propto \frac{\beta_1^{\sum_{i=1}^n z_i^1} \beta_2^{\sum_{i=1}^n z_i^2}}{\theta_1^{\beta_1 \sum_{i=1}^n z_i^1} \theta_2^{\beta_2 \sum_{i=1}^n z_i^2}} \prod_{i=1}^n [t_i^{\beta_1-1}]^{z_i^1} \prod_{i=1}^n [t_i^{\beta_2-1}]^{z_i^2} \exp \left[-\frac{\sum_{i=1}^n t_i^{\beta_1}}{\theta_1^{\beta_1}} - \frac{\sum_{i=1}^n t_i^{\beta_2}}{\theta_2^{\beta_2}} \right], \quad (28)$$

where $z_i (= (z_i^1, z_i^2))$ is a binary indicator vector associated with t_i , $i = 1, 2, \dots, n$, such that $z_i = (1, 0)$ conveys that associated t_i is coming from $\mathcal{W}(\theta_1, \beta_1)$ and $z_i = (0, 1)$ results in t_i coming from $\mathcal{W}(\theta_2, \beta_2)$. Also, in case of no censoring z_i^1 and z_i^2 are related as $z_i^1 = 1 - z_i^2$ since our assumption confirms that an observed t_i will certainly come from one of the two Weibull distributions. Next, let us consider the same priors as given in (26) and obtain the joint posterior up to proportionality by combining (26) with (28) via Bayes theorem. This posterior can be written as

$$p_1(\theta_1, \theta_2, \beta_1, \beta_2 | t, z) \propto \frac{\beta_1^{\sum_{i=1}^n z_i^1}}{\theta_1^{(\sum_{i=1}^n z_i^1 + a)\beta_1 + 1}} \exp \left[-\frac{\sum_{i=1}^n t_i^{\beta_1 + b}}{\theta_1^{\beta_1}} \right] \prod_{i=1}^n [t_i^{\beta_1-1}]^{z_i^1} \\ \times \frac{\beta_2^{\sum_{i=1}^n z_i^2}}{\theta_2^{(\sum_{i=1}^n z_i^2 + c)\beta_2 + 1}} \exp \left[-\frac{\sum_{i=1}^n t_i^{\beta_2 + d}}{\theta_2^{\beta_2}} \right] \prod_{i=1}^n [t_i^{\beta_2-1}]^{z_i^2}; \quad (29) \\ \theta_1 > 0, \theta_2 > 0, 0 \leq \beta_1 \leq 1, 1 \leq \beta_2 \leq \beta_{2u},$$

where $z = (z_1, z_2, \dots, z_n)$.

Before we proceed further, let us briefly comment on the Gibbs sampler algorithm. Gibbs sampler is a Markovian updating mechanism for extracting samples from (often) high-dimensional posteriors specified up to proportionality by extracting samples from all univariate (or lower dimensional) full conditionals. The generation begins with some appropriately chosen initial values and proceeds in a cyclic frame-work covering all the full conditionals and each time using the most recent values of conditioning variates. The values obtained after one complete cycle represent a state of a Markov chain. This process is repeated until a systematic pattern of convergence is achieved by the generating chain. Once the convergence is achieved, one can either pick up equidistant observations in a single long run of the chain or pick up observations from various parallel chains to form independent and identically distributed samples from the concerned posterior. For details about the Gibbs sampler algorithm, its implementation and convergence diagnostic issues, one can refer to [12], [24] and [25], etc.

Thus in order to apply the Gibbs sampler algorithm, we need to specify all possible full conditionals corresponding to posterior (29). The incomplete specification of data can be resolved by using additional full conditionals corresponding to different indicator variables. We shall begin

with a comment on this latter full conditionals and the corresponding generations. Since z_i^1 , $i = 1, 2, \dots, n$, is a binary indicator variable, it can be considered to follow a Bernoulli distribution with parameter \tilde{p}_i , where

$$\tilde{p}_i = P_r[z_i^1 = 1] = \frac{h_W(t_i, \theta_1, \beta_1)}{h_W(t_i, \theta_1, \beta_1) + h_W(t_i, \theta_2, \beta_2)}.$$

Moreover, using the fact that each z_i^1 is independent of all other z_i^1 s, we can generate it independently corresponding to each t_i at the current realization of parameters $(\theta_1, \theta_2, \beta_1, \beta_2)$ and for each generated z_i^1 , we can obtain corresponding z_i^2 using the relationship $z_i^2 = 1 - z_i^1$. Rest of the full conditionals corresponding to (29) can be obtained up to proportionality as

$$p_2^*(\theta_1 | \theta_2, \beta_1, \beta_2, t, z) \propto \frac{1}{\theta_1^{(\sum_{i=1}^n z_i^1 + a)\beta_1 + 1}} \exp\left[-\frac{\sum_{i=1}^n t_i^{\beta_1 + b}}{\theta_1^{\beta_1}}\right]; \quad \theta_1 > 0 \quad (30)$$

$$p_3^*(\theta_2 | \theta_1, \beta_1, \beta_2, t, z) \propto \frac{1}{\theta_2^{(\sum_{i=1}^n z_i^2 + c)\beta_2 + 1}} \exp\left[-\frac{\sum_{i=1}^n t_i^{\beta_2 + d}}{\theta_2^{\beta_2}}\right]; \quad \theta_2 > 0 \quad (31)$$

$$p_4^*(\beta_1 | \theta_1, \theta_2, \beta_2, t, z) \propto \frac{\beta_1^{\sum_{i=1}^n z_i^1}}{\theta_1^{(\sum_{i=1}^n z_i^1 + a)\beta_1}} \exp\left[-\frac{\sum_{i=1}^n t_i^{\beta_1 + b}}{\theta_1^{\beta_1}}\right] \prod_{i=1}^n [t_i^{\beta_1}]^{z_i^1}; \quad 0 < \beta_1 \leq 1 \quad (32)$$

$$p_5^*(\beta_2 | \theta_1, \theta_2, \beta_1, t, z) \propto \frac{\beta_2^{\sum_{i=1}^n z_i^2}}{\theta_2^{(\sum_{i=1}^n z_i^2 + c)\beta_2}} \exp\left[-\frac{\sum_{i=1}^n t_i^{\beta_2 + d}}{\theta_2^{\beta_2}}\right] \prod_{i=1}^n [t_i^{\beta_2}]^{z_i^2}; \quad 1 < \beta_2 \leq \beta_{2u}. \quad (33)$$

Full conditionals (30) and (31) appear to be the kernels of \mathcal{JG} distribution. Hence by considering transformation $\lambda_1 = \theta_1^{\beta_1}$, one can show that $\lambda_1 \sim \mathcal{JG}(\sum_{i=1}^n z_i^1 + a, \sum_{i=1}^n t_i^{\beta_1 + b})$. Similarly, $\lambda_2 = \theta_2^{\beta_2}$ can be shown to follow $\mathcal{JG}(\sum_{i=1}^n z_i^2 + c, \sum_{i=1}^n t_i^{\beta_2 + d})$. Hence both θ_1 and θ_2 can be generated using any standard routine for inverse gamma generator (see, for example, [8]). The full conditionals (32) and (33) can be shown to be logconcave hence both β_1 and β_2 can be simulated easily using (say) adaptive rejection sampling (ARS) algorithm proposed by [14].

4.1 Implementation of Gibbs sampler in case of censored data

An appreciable property of Gibbs sampler algorithm is that it can be easily extended to deal with censored data situations. The idea is very simple in the sense that the algorithm proceeds with exactly the same posterior as specified for the complete data case but assumes censored observations as further unknowns. Thus, in censored data case, the Gibbs sampler algorithm has additional full conditionals corresponding to censored observations. Obviously, the full conditionals for the other parameters will remain same to those obtained for the complete data case whereas the full conditionals for the independent censored observations will be the parent sampling distributions truncated in the appropriate regions (for details, see [26]).

As mentioned in Section 3, each t_i is associated with an indicator variable δ_i such that if $\delta_i = 0$, t_i is the right censoring time and if $\delta_i = 1$, t_i is the observed failure time corresponding to the item i , $i = 1, 2, \dots, n$. Obviously, the full conditionals corresponding to $z, \theta_1, \theta_2, \beta_1$ and β_2 are same that were obtained earlier for complete data case. The additional full conditionals correspond to censored observations that can be generated from the left truncated \mathcal{BW}_R distribution. The generation can be simple and the variate value t_i can be retained if it lies in the constrained region (t_i, ∞) . For the generation from truncated \mathcal{BW}_R model, a simple two-step algorithm can be designed based on the following theorem.

Theorem 1 Suppose T_1 and T_2 be two random variables following $\mathcal{W}(\theta_1, \beta_1)$, $\beta_1 < 1.0$ and $\mathcal{W}(\theta_2, \beta_2)$, $\beta_2 > 1.0$, respectively, and suppose both are truncated from left at the same point c . If we define a random variable $T = \min(T_1, T_2)$, then T will follow left truncated \mathcal{BW}_R distribution with parameters $(\theta_1, \theta_2, \beta_1, \beta_2)$, the point of truncation from left is at c .

Proof. The proof of the Theorem (1) is added to Appendix given at the end.

The generation of random variable T from truncated \mathcal{BW}_R distribution involves the following two steps. First, generate T_1 and T_2 from left truncated $\mathcal{W}(\theta_1, \beta_1)$, $\beta_1 < 1.0$ and $\mathcal{W}(\theta_2, \beta_2)$, $\beta_2 > 1.0$, respectively. Second, take $T = \min(T_1, T_2)$. The resulting T will follow left truncated \mathcal{BW}_R distribution.

5 Model Compatibility

Model compatibility study is an important concept in any statistical data analysis which provides an assurance that the entertained model is rightly used for the data in hand. If the model is compatible with the data, the analysis is of course justified. In case of poor resemblance with the data, it is desired to consider an alternative model that best represents the entertained data. The model compatibility study can be performed in a variety of ways. The classical statisticians, of course, use tail area probabilities based on a discrepancy measuring statistic to provide agreement or disagreement of data with the model.

Bayesian statistics offers a number of tools for studying model compatibility. The simplest and an informal approach may involve checking predictive capability of the model with regard to some of its important characteristics possibly using graphical tools (see, for example, [19]). A practical approach to implement this idea in reliability studies may entail investigating the empirical plots of observed data based and some of the posterior predictive data based reliability characteristics on the same graphical scale. Some of the important reliability characteristics in this context may be considered as hazard rate function, reliability function, mean time to failure, etc. Thus one can consider plotting the observed data based entertained characteristic and correspondingly the predictive data based same characteristic where predictive data are generated from the posited model. Such a graphical tool will not only provide an informal assessment of discrepancy between the model and the data but also sometimes help in improving the model (see also Upadhyay et al. (2001)).

Bayesian study on model compatibility can also be extended in an objective manner using tail area probability or the p-value based on a discrepancy measuring statistic under the assumption that the considered model is true for data. A number of versions of Bayesian p-values are defined in the literature based on several considerations, each having its own merit or demerit. We shall not discuss these details here due to space restriction rather refer to Bayarri and Berger (1998) for a systematic accountability. In this paper, we shall use posterior predictive p-value based on chi-square discrepancy measure (see also Upadhyay et al. (2001)). Although the posterior predictive p-value (PPV) has its own disadvantages, the most important being double use of data, we shall use it for its inherent simplicity and also because of the fact that it is easily computable for any choice of prior. A brief review about the PPV is given in the next subsection.

5.1 Posterior predictive p-value

[15] proposed the use of p-value based on the posterior predictive distribution of model

departure statistics (see also [26]). The idea is very simple. To begin with, let us define the chi-square statistic as a measure of discrepancy given by

$$\chi^2 = \frac{\sum_{i=1}^n (t_i - E(t_i|\theta))^2}{V(t_i|\theta)}, \quad (34)$$

where $E(\cdot)$ denotes the operation of taking expectation and $V(\cdot)$ denotes the variance. Accordingly, the PPV based on χ^2 discrepancy measure can be obtained as

$$PPV = \int P_r[\chi_2^2 > \chi_1^2 | f, \theta] p(\theta|t) d\theta, \quad (35)$$

where χ_1^2 and χ_2^2 are the calculated values of χ^2 for the observed and predictive data sets, respectively, $p(\theta|t)$ is posterior distribution of θ and f is the entertained model. For complete data case, PPV can be calculated using the procedure suggested by [26] (see also [13]) which consists of two steps. The first step is to draw θ from $p(\theta|t)$ and calculate χ_1^2 based on the given data set. The second step is to extract predictive data sets each of same size as that of given data from the model f using the simulated θ and calculate χ_2^2 based on these predictive data sets. We then calculate $P_r[\chi_2^2 > \chi_1^2 | f, \theta]$ as the number of times χ_2^2 exceeds χ_1^2 . These steps are repeated a number of times with different simulated θ and PPV is estimated as the posterior expectation of $P_r[\chi_2^2 > \chi_1^2 | f, \theta]$.

In order to evaluate PPV in situation where data set has some right censored observations, one can first complete the data set by replacing all censored observations with the maximum of their respective censoring times and predictive means (see [11]). The PPV can then be calculated using this completed data set in the same way as described above for complete data case.

6 Numerical Illustration

For analyzing the proposed \mathcal{BW}_R model, we considered a real data set reported initially by [9]. The dataset consists of failure times of 58 electrodes (segments cut from bars) which were put on a high-stress voltage endurance life test. Observations on failure time from voltage endurance test are given in Table 1. First, fourth and seventh columns of the table list the observations on failure times of electrodes in hours whereas second, fifth and eighth column represent failure modes, that is, the causes of failures. The failures were attributed to one of two modes (causes) which are as under. The first cause is the insulation defect due to a processing problem (mode E) which tends to occur early in life. The second cause, on the other hand, is degradation of the organic material (mode D) which typically occurs at a later stage. Third, sixth and ninth columns of the table indicate completely observed or censored failure times. $\delta = 0$ indicates observed failure times while $\delta = 1$ indicates censored failure times. Since, for censored observations, failure cause is unknown, we have denoted the missing cause by '*' (see Table 1).

In order to analyze the model for the assumed dataset, we first considered only those observations which were completely observed and left those observations which were censored. As such, we formed a new sample where all the failure times are completely observed. Besides, we also omitted the two causes of failures (E and D) so that the observations can be treated appropriate for the considered competing risk model with latent (unknown) causes of failures. We next considered the entire sample treating it as a case of censored data problem but omitted the two causes of failures (E and D) for the appropriateness of proposed competing risk model with latent causes of failures.

In the first part of our analysis, we considered obtaining ML estimates of \mathcal{BW}_R parameters using EM algorithm after visualizing the model as an incomplete data model and introducing missing indicator variables z^1 for failure mode E and z^2 for failure mode D as described in Section 3. The ML estimates for the parameters θ_1 , θ_2 , β_1 and β_2 were found to be 885.030(1209.506), 341.553(343.841), 0.613(0.629) and 5.545(5.592), respectively. The values in parenthesis correspond to the estimates of parameters for censored data case.

Table 1: Voltage endurance life test results of 58 electrodes.

Hours	Failure Mode	δ	Hours	Failure Mode	δ	Hours	Failure Mode	δ
2	E	0	144	E	0	303	E	0
3	E	0	157	*	1	314	D	0
5	E	0	160	E	0	317	D	0
8	E	0	168	D	0	318	D	0
13	*	1	179	*	1	320	D	0
21	E	0	191	D	0	327	D	0
28	E	0	203	D	0	328	D	0
31	E	0	211	D	0	328	D	0
31	*	1	221	E	0	348	*	1
52	*	1	226	D	0	348	D	0
53	*	1	236	E	0	350	D	0
64	E	0	241	*	1	360	D	0
67	*	1	257	*	1	369	D	0
69	E	0	261	D	0	377	D	0
76	E	0	264	D	0	387	D	0
78	*	1	278	D	0	392	D	0
104	E	0	282	E	0	412	D	0
113	*	1	284	D	0	446	D	0
119	E	0	286	D	0			
135	*	1	298	D	0			

For Bayesian analysis of the model, we first consider specification of prior hyperparameters of the considered prior distributions. In this context, we begin with the choice of $\beta_{2u} = 7.0$ following the argument given in Section 4 so that the shape of the hazard rate curve is increasing at a constant rate from the beginning itself and the Weibull distribution does not approach to normality in a real sense. We next assume that given β_1 , the expert has provided two characteristics of $\mathcal{IG}(a, b)$ distribution, that is, the expert specifies the mean and variance of $\theta_1^{\beta_1}$ to be 200 and 1000, respectively. Similarly, we assume that given β_2 , the expert specifies both mean and variance of $\theta_2^{\beta_2}$ to be very large, say of order 1.0×10^{10} . It may be noted that since the expert is not available in a real sense when specifying mean and variance of $\theta_1^{\beta_1}$ and $\theta_2^{\beta_2}$, we have taken help of ML estimates of various parameters as well. Utilizing ML estimates is certainly an objective consideration but there is no harm if data based information is used in forming the appropriate priors. Also, the large variability for both $\theta_1^{\beta_1}$ and $\theta_2^{\beta_2}$ convey a kind of vagueness in the choice of priors. Based on these choices, the prior hyperparameters a and b were evaluated to be 5.0 and 600.0, respectively. Similarly, the prior hyperparameters c and d were found to be 6.0 and 5.0×10^8 , respectively.

In order to provide the posterior based inferences of model (8), the posterior samples were extracted using the Gibbs sampler algorithm as described in Section (4) for complete data case and Subsection 4.1 for censored data case. In each case, we considered a single long run of Gibbs chain using the ML estimates of the parameters as the starting values for running the chain. For censored case, we also used known censoring time for each censored observation as the initial value of the corresponding censored observation. Convergence of the chain was monitored using the ergodic averages at about 20K iterations in each case. Once the convergence was achieved a sample of size 2K was taken from the corresponding posterior distribution by picking up equidistant observations (at a gap of 10). The gap was chosen to make the serial correlation among the generating variates negligibly small (see also [26]). It may be noted that the assessment of a specific risk in the presence of other risk factors is of particular interest in a competing risk scenario and, therefore, we also obtained a sample of size 2K for the probability of failure due to insulation

defects (mode E) at each iterated value of posterior variates (see also (10)).

Some of the important sample based estimated posterior characteristics were obtained based on the final sample of size 2K. We do not report here all the characteristics except the estimated posterior modes and the highest posterior density (HPD) intervals with coverage probability 0.95 for all the parameters considering both complete and censored data cases (see Table 2). However, while writing the conclusion, some other characteristics, not reported in the paper, may be taken into account as well. Posterior modes are given because these are the most probable values and can be reasonably considered to be the Bayes estimates. Similarly, HPD limits will provide to a large extent the overall idea of the estimated posterior densities. Table 2 also provides the corresponding estimates for censored data case and these estimates are shown in parentheses. Besides the model parameters, the table also exhibits the estimated posterior modes and HPD limits for the probability of failure due to mode E (see (10)).

Table 2: Estimated posterior modes and HPD limits with coverage probability 0.95 for \mathcal{BW}_R parameters

Parameters	Estimated posterior mode	HPD limits	
		lower	upper
θ_1	702.131 (950.466)	289.561 (345.505)	3819.390 (5734.279)
β_1	0.663 (0.610)	0.497 (0.500)	0.823 (0.845)
θ_2	340.758 (342.718)	317.392 (319.954)	375.069 (372.495)
β_2	5.311 (5.594)	5.138 (5.415)	5.562 (5.824)
$P_r(t = t_1)$	0.349 (0.310)	0.189 (0.199)	0.546 (0.557)

Values in parentheses correspond to censored data case.

It can be seen that the Bayes estimates in both complete and censored data cases are more or less similar to the classical ML estimates except in case of θ_1 for censored data case where ML estimate appears to be slightly overestimated value. The finding, therefore, confirms the vague consideration of priors with choice of hyperparameters guided to some extent by ML estimates. Based on the HPD limits and the estimated posterior modes, it can be concluded that the parameters θ_2 , β_1 and β_2 are more or less symmetric with small variability whereas the parameter θ_1 is highly positively skewed with very large variability (see Table 2). This last conclusion was also confirmed by the characteristics such as estimated posterior means and medians, the values of which are not shown in the paper. A word of remark: since the posterior variability of θ_1 is quite large and the variability is appreciably increased for censored data case, the difference between Bayes and ML estimates, especially for censored data case, can be considered marginal only. Another important and striking conclusion is that the censoring does not cause appreciable loss of information as the estimates corresponding to complete and censored data case are quite close to each other although we have considered only 22 % observations to be censored. The Table 2 also shows the estimated posterior mode and HPD limits with coverage probability 0.95 for $P_r(t = t_1)$. It can be seen that nearly 35% of the observations (31% for censored data case) are failed due to initial birth defect, a conclusion that appears to be quite close to the true entertained values reported in Table 1.

Figure 3: Estimated posterior densities and bi-variate characteristics for complete data case.

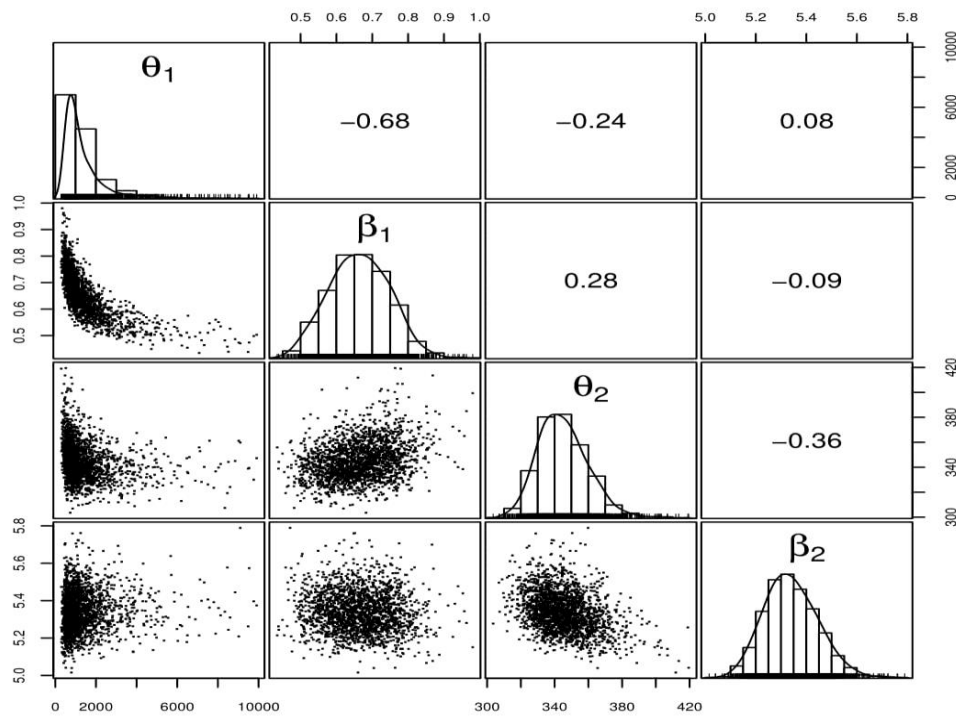
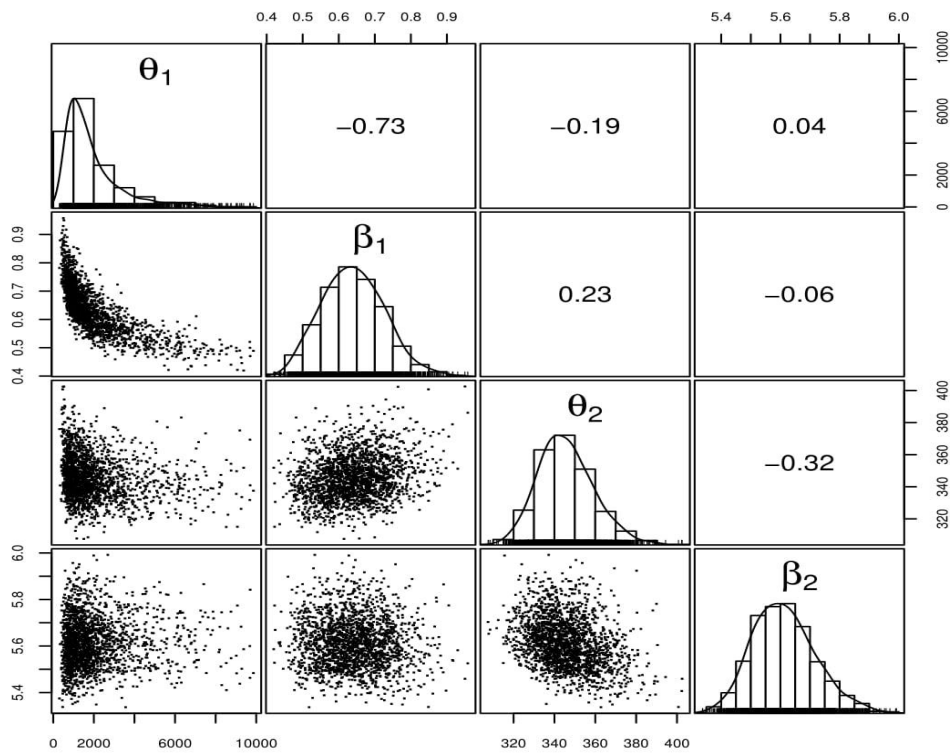


Figure 4: Estimated posterior densities and bi-variate characteristics for censored data case.



We also worked out for the bivariate posterior characteristics for both complete and censored data cases, which are shown in Figures 3 and 4 in the form of scatter plots and estimated posterior correlations although the figures also display the estimated marginal posterior densities. As far as

the marginal densities are concerned, the conclusions are same that are discussed in the previous paragraph and, therefore, we do not need to interpret these estimated densities anymore. Based on the scatter plots and the estimated correlations, it can be concluded that the parameters θ_1 and β_1 are highly correlated *a posteriori*, the estimated correlation being -0.68. Similarly, the parameters θ_2 and β_2 are also exhibiting good correlation, the estimated value for the same being -0.33. In this very sense the conditional priors for θ_1 and θ_2 given in (26) are justified to some extent. The remaining parameter combinations (θ_1, θ_2) , (θ_1, β_2) , (β_1, θ_2) and (β_1, β_2) are, however, having relatively low estimated correlations, the estimated values being -0.21(-0.19), 0.03(0.04), 0.26(0.23) and -0.03(-0.06), respectively. Again the values within parentheses correspond to censored data case.

To proceed further, we estimated the change point as given in (6) based on the final posterior samples of size 2K. The change point is certainly an important characteristic which differentiates between the two modes of failure. That is, the units having failure times less than the change point can be said to have failed due to early birth defect and the failure of units with failure times exceeding it can be attributed to degradation failure. The estimated posterior mode and HPD limits with coverage probability 0.95 for t^* were found to be 107.240{109.467} and [92.269, 125.235]{[95.320, 125.764]}, respectively, where the values in curly parentheses correspond to censored data case. In addition to these estimates, we also obtained the posterior estimates of hazard rate at different time points including at the change point. The estimates in the form of posterior modes based on a sample of size 2K are given in Table 3. It can be seen that the estimated hazard rate is least at the change point time and increasing as we move away from the change point time in either direction, a conclusion that was expected too.

Table 3: Posterior estimates of hazard rate at different times

t=80	t=100	t=t*	t=150	t=200
0.121e-02	0.115e-02	0.113e-02	0.125e-02	0.224e-02
(0.147e-02)	(0.141e-02)	(0.139e-02)	(0.147e-02)	(0.254e-02)

Values in parentheses correspond to censored data case.

Before we end the section, let us examine the compatibility of the model with the entertained data based on the ideas discussed in Section 5. For this purpose, we first generated a posterior sample of size 20 using the Gibbs sampler algorithm (subsection 4.1) and correspondingly obtained 20 predictive samples each of same size as that of informative data. We then considered the non-parametric empirical hazard rate estimates for both observed and predictive data sets at the time corresponding to the two data sets. Since some of the observations in the original data are censored, we implemented our procedure after completing the data by replacing the censored observations by the maximum of their predictive means and censoring times. The corresponding plots are shown in Figure 5 where solid line represents the informative data based estimated hazard rate and the dotted lines represent the predictive data based estimated hazard rate.

A similar strategy was used to draw observed data and correspondingly predictive data based estimates of reliability function. This latter plot is shown in Figure 6. It can be seen that in both the cases the solid line is well superimposed by the dotted lines (see Figures 5-6) giving us a clear cut conclusion that the model is justified for the data in hand.

We next obtained the numerical summary of model compatibility study in the form of PPV discussed in Section 5. To calculate the same, we first considered 100 posterior samples corresponding to the model \mathcal{BW}_R using the Gibbs sampler algorithm (subsection 4.1) and then obtained χ_1^2 for the given data set for each generated posterior sample of θ . As a second step, we simulated 1K predictive samples from (8), each of size exactly similar to that of the observed data, for each value of θ and calculated χ_2^2 based on these predictive data sets. Our next step calculated $P_r[\chi_2^2 > \chi_1^2]$ for each given θ . Finally, the above steps were repeated to calculate PPV as described in Section 5. The estimated PPV was found to be 0.331, a value that again confirms the

compatibility of the model (8) with the observed data set. A word of remark: for the compatibility study considered here, we do not report the results corresponding to complete data case by omitting the censored observations as it was done earlier in obtaining other estimates. The results for the complete data case were more or less similar and, therefore, avoided due to space restriction.

Figure 5: Estimated hazard rate plots corresponding to observed (solid line) and predictive data sets.

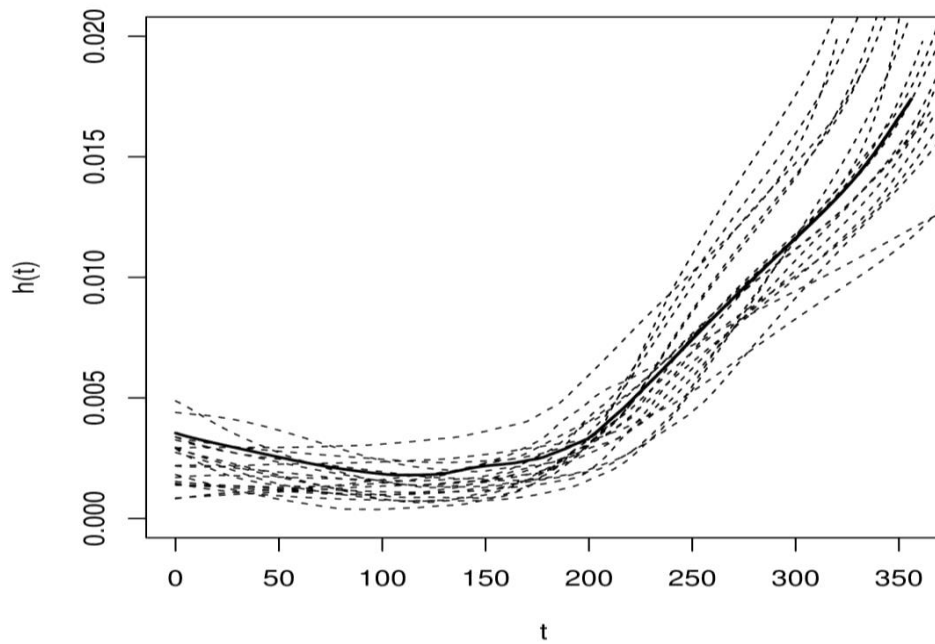
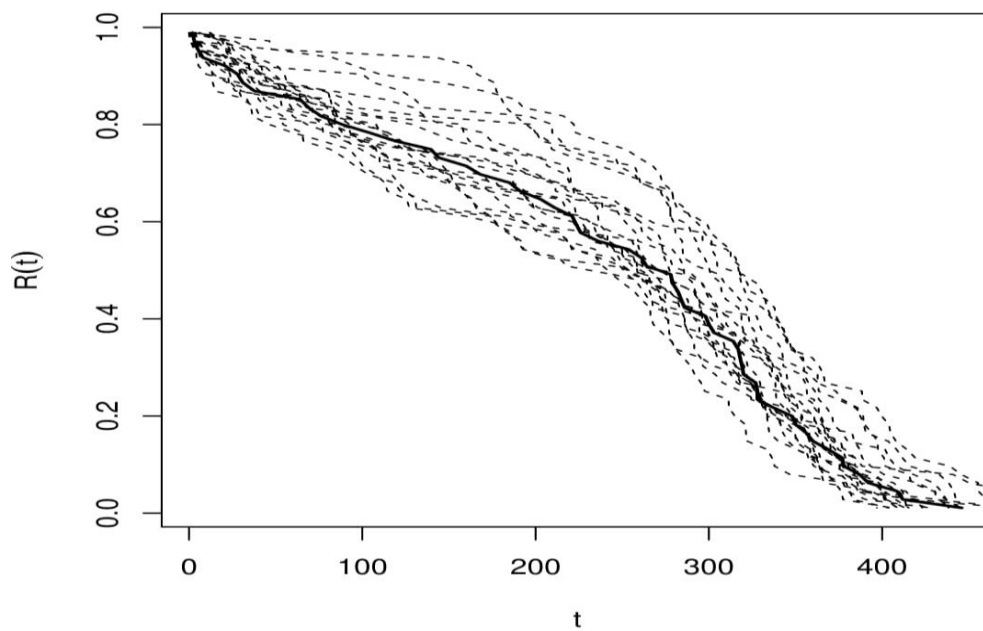


Figure 6: Estimated reliability plots corresponding to observed (solid line) and predictive data sets.



7 Conclusion

The paper considers a BW_R model, a model that is quite popular in a competing risk scenario. Our assumption, however, considers two competing causes, the initial birth defect and ageing effect, that work together to induce failure of a unit. The considered model is analyzed in both classical and Bayesian frameworks although the classical analysis focuses on ML estimates only. The important feature of the corresponding likelihood function is that it offers unique consistent solution in the form of ML estimates. The paper finally shows how the idea of visualizing the competing risk model as an incomplete data model facilitates both classical and Bayesian analyses of the considered model not only for complete case but also for censored data situation. The applicability of model is also justified by conducting a Bayesian compatibility study of the model with a real data set involving a voltage endurance life test with two different failure modes.

Acknowledgements

Research work of Ankita Gupta is financially supported by Council of Scientific and Industrial Research, New Delhi, India, in the form of Junior Research Fellowship.

References

- [1] James O Berger and Dongchu Sun. Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, 88(424):1412–1418, 1993.
- [2] Nicolas Bousquet, Henri Bertholon, and Gilles Celeux. An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis. *Lifetime Data Analysis*, 12(4):481–504, 2006.
- [3] K. C. Chanda. A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, 41(1/2):56–61, 1954.
- [4] Victor Chan and William Q Meeker. A failure-time model for infant-mortality and wearout failure modes. *IEEE Transactions on Reliability*, 48(4):377–387, 1999.
- [5] Herbert Aron David and Melvin L Moeschberger. *The Theory of Competing Risks: HA David, ML Moeschberger*. C. Griffin, 1978.
- [6] A. C. Davison and F Louzada-Neto. Inference for the Poly-Weibull model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(2):189–196, 2000.
- [7] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.
- [8] Luc Devroye. Nonuniform random variate generation. *Handbooks in operations research and management science*, 13:83–121, 2006.
- [9] Necip Doganaksoy, Gerald J Hahn, and William Q Meeker. Reliability analysis by failure mode. *Quality Progress*, 35(6):47, 2002.
- [10] Lea Friedman and Ilya B Gertsbakh. Maximum likelihood estimation in a minimum-type model with exponential and Weibull failure modes. *Journal of the American Statistical Association*, 75(370):460–465, 1980.
- [11] Alan E Gelfand and Sujit K Ghosh. Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11, 1998.
- [12] Alan E Gelfand and Adrian FM Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.
- [13] Andrew Gelman, Xiao-Li Meng, and Hal Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760, 1996.
- [14] Walter R Gilks and Pascal Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, pages 337–348, 1992.
- [15] Irwin Guttman. The use of the concept of a future observation in goodness-of-fit

- problems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 83–100, 1967.
- [16] R Jiang and D. N. P. Murthy. Parametric study of competing risk model involving two Weibull distributions. *International Journal of Reliability, Quality and Safety Engineering*, 4(01):17–34, 1997.
- [17] Artur J Lemonte, Gauss M Cordeiro, and Edwin M. M. Ortega. On the additive Weibull distribution. *Communications in Statistics-Theory and Methods*, 43(10-12):2066–2080, 2014.
- [18] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.
- [19] Puja Makkar, SK Upadhyay, VK Shukla, and RS Singh. Examining biliary acid constituents among gall bladder patients: A bayes study using the generalized linear model. *International Journal of Statistics in Medical Research*, 4(2):224–239, 2015.
- [20] Geoffrey McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.
- [21] Masami Miyakawa. Analysis of incomplete data in competing risks model. *IEEE Transactions on Reliability*, 33(4):293–296, 1984.
- [22] Rakesh Ranjan, Sonam Singh, and Satyanshu K Upadhyay. A Bayes analysis of a competing risk model based on gamma and exponential failures. *Reliability Engineering & System Safety*, 144:35–44, 2015.
- [23] Rakesh Ranjan and Satyanshu K Upadhyay. Classical and Bayesian estimation for the parameters of a competing risk model based on minimum of exponential and gamma failures. *IEEE Transactions on Reliability*, 65(3):1522–1535, 2016.
- [24] Adrian F. M. Smith and Gareth O Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23, 1993.
- [25] Satyanshu K Upadhyay and Adrian F. M. Smith. Modelling complexities in reliability, and the role of simulation in Bayesian computation. *International Journal of Continuing Engineering Education and Life Long Learning*, 4(1-2):93–104, 1994.
- [26] S. K. Upadhyay, N Vasishta, and A. F. M. Smith. Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 63(1):15–40, 2001.

Appendix

Proof of Theorem 1

Proof. Let T_1 and T_2 follow $\mathcal{W}(\theta_1, \beta_1)$, $\beta_1 < 1.0$ and $\mathcal{W}(\theta_2, \beta_2)$, $\beta_2 > 1.0$, respectively, where both are truncated from left at a point c . The densities of T_1 and T_2 can be written as

$$f_{W_1}^c(t_1|t_1 > c) = \frac{\beta_1}{\theta_1} \left(\frac{t_1}{\theta_1}\right)^{\beta_1-1} \exp\left[-\left(\frac{t_1}{\theta_1}\right)^{\beta_1} + \left(\frac{c}{\theta_1}\right)^{\beta_1}\right],$$

$$f_{W_2}^c(t_2|t_2 > c) = \frac{\beta_2}{\theta_2} \left(\frac{t_2}{\theta_2}\right)^{\beta_2-1} \exp\left[-\left(\frac{t_2}{\theta_2}\right)^{\beta_2} + \left(\frac{c}{\theta_2}\right)^{\beta_2}\right],$$

respectively. Similarly, the cumulative distribution functions of T_1 and T_2 can be written as

$$F_{W_1}^c(t_1|t_1 > c) = 1 - \exp\left[-\left(\frac{t_1}{\theta_1}\right)^{\beta_1} + \left(\frac{c}{\theta_1}\right)^{\beta_1}\right],$$

$$F_{W_2}^c(t_2|t_2 > c) = 1 - \exp\left[-\left(\frac{t_2}{\theta_2}\right)^{\beta_2} + \left(\frac{c}{\theta_2}\right)^{\beta_2}\right].$$

If we define a random variable $T = \min(T_1, T_2|T_1, T_2 > c)$ then the cumulative distribution function of T can be obtained as

$$F_{\mathcal{B}\mathcal{W}_{\mathcal{R}}}^c(t) = 1 - [1 - F_{W_1}^c(t|t > c)][1 - F_{W_2}^c(t|t > c)],$$

which simplifies to

$$F_{\mathcal{B}\mathcal{W}_{\mathcal{R}}}^c(t) = 1 - \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} + \left(\frac{c}{\theta_1}\right)^{\beta_1}\right] \exp\left[-\left(\frac{t}{\theta_2}\right)^{\beta_2} + \left(\frac{c}{\theta_2}\right)^{\beta_2}\right].$$

Now differentiating $F_{\mathcal{B}\mathcal{W}_{\mathcal{R}}}^c(t)$ with respect to t , we can obtain the density function of T as

$$f_{\mathcal{B}\mathcal{W}_{\mathcal{R}}}^c(t) = \frac{\left[\frac{\beta_1}{\theta_1} \left(\frac{t}{\theta_1}\right)^{\beta_1-1} + \frac{\beta_2}{\theta_2} \left(\frac{t}{\theta_2}\right)^{\beta_2-1}\right] \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right]}{\exp\left[-\left(\frac{c}{\theta_1}\right)^{\beta_1} - \left(\frac{c}{\theta_2}\right)^{\beta_2}\right]}.$$

which is the density of $\mathcal{B}\mathcal{W}_{\mathcal{R}}$ distribution with parameters $(\theta_1, \theta_2, \beta_1, \beta_2)$, truncated from left at point c . Hence we can say that $T = \min(T_1, T_2)$ follows $\mathcal{B}\mathcal{W}_{\mathcal{R}}(\theta_1, \theta_2, \beta_1, \beta_2)$, left truncated at point c .

Existence of Unique and Consistent Roots of Likelihood Equations

Theorem 2 (see [3]). Let $f(t|\nu)$ be the pdf with parameter vector $\nu = (\nu_1, \nu_2, \dots, \nu_k)$, then the solution of the likelihood equation for the observation vector t

$$\frac{\partial \sum_{i=1}^n \ln(f(t_i|\nu))}{\partial \nu_r} = 0,$$

will be unique and consistent if the following three conditions hold. For the sake of brevity, we shall write f for $f(t|\nu)$ and f_i for $f(t_i|\nu)$.

(i). For almost all t and $\nu \in \Theta$, $\frac{\partial \ln(f)}{\partial \nu_r}$, $\frac{\partial^2 \ln(f)}{\partial \nu_r \partial \nu_s}$, $\frac{\partial^3 \ln(f)}{\partial \nu_r \partial \nu_s \partial \nu_w}$ exist for all $r, s, w = 1, 2, \dots, k$.

(ii). For almost all t and $\nu \in \Theta$, $\left|\frac{\partial f}{\partial \nu_r}\right| < F_r(t)$, $\left|\frac{\partial^2 f}{\partial \nu_r \partial \nu_s}\right| < F_{rs}(t)$, $\left|\frac{\partial^3 f}{\partial \nu_r \partial \nu_s \partial \nu_w}\right| < H_{rsw}(t)$, where $H_{rsw}(t)$ is such that $\int_{-\infty}^{\infty} H_{rsw}(t) f dt < C_M < \infty$ and $F_r(t)$ and $F_{rs}(t)$ are bounded for all t .

(iii). For all $\nu \in \Theta$, the matrix $J = (J_{rs}(\nu))$, where

$$J_{rs}(\nu) = \int_{-\infty}^{\infty} \frac{\partial \ln(f)}{\partial \nu_r} \frac{\partial \ln(f)}{\partial \nu_s} f dt$$

is positive-definite and that $|J|$ is finite.

Now the conditions of above theorem may be verified for the likelihood equations corresponding to $\mathcal{B}\mathcal{W}_{\mathcal{R}}$ model as follows. (It may be noted that the following proof verifies

Chanda's theorem for the likelihood equations corresponding to right censored data case as well since for the censored observation, the expression of density for \mathcal{BW}_R model becomes $f(c_1) = \exp\left[-\left(\frac{c_1}{\theta_1}\right)^{\beta_1} - \left(\frac{c_1}{\theta_2}\right)^{\beta_2}\right]$ instead of (8) where c_1 denotes the censoring time (see also [2]).

Let us denote f for $f_{\mathcal{BW}_R}(t)$. The function f is differentiable with respect to all the parameters $\theta_1, \theta_2, \beta_1$ and β_2 any number of times. Under the assumption that all these parameters are positive and finite, any arbitrary ℓ^{th} order partial derivative of f (let it be denoted by g) can only produce term of the following form

$$g = \left[\sum_{j,k=0, j+k \leq \ell}^{\ell} \left(\ln\left(\frac{t}{\theta_1}\right)\right)^j \left(\ln\left(\frac{t}{\theta_2}\right)\right)^k \Phi_{jk}(t)\right] \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right]; \tag{36}$$

$$j = 0, 1, \dots, \ell; k = 0, 1, \dots, \ell,$$

where $\Phi_{jk}(t)$ s are polynomials in t , hence continuous and bounded in any closed interval. $\ln(t)$ is also continuous for $t > 0$, which indicates that the term $\left[\sum_{j,k=0, j+k \leq \ell}^{\ell} \left(\ln\left(\frac{t}{\theta_1}\right)\right)^j \left(\ln\left(\frac{t}{\theta_2}\right)\right)^k \Phi_{jk}(t)\right]$ is continuous and bounded for t in any closed interval. For large values of t , behavior of $\left[\sum_{j,k=0, j+k \leq \ell}^{\ell} \left(\ln\left(\frac{t}{\theta_1}\right)\right)^j \left(\ln\left(\frac{t}{\theta_2}\right)\right)^k \Phi_{jk}(t)\right]$ is dominated by the term $\exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right]$ and also $\exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right] \rightarrow 0$ as $t \rightarrow \infty$. Hence,

$$\left|\left[\sum_{j,k=0, j+k \leq \ell}^{\ell} \left(\ln\left(\frac{t}{\theta_1}\right)\right)^j \left(\ln\left(\frac{t}{\theta_2}\right)\right)^k \Phi_{jk}(t)\right] \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right]\right| < \infty.$$

With this we can say that all partial derivatives of density f exist. Now for verifying condition (i) of Theorem 2, we have to show that all the first, second and third order partial derivatives of $\ln(f)$ exist. These partial derivatives can be written as

$$\frac{\partial \ln(f)}{\partial v_r} = \frac{1}{f} \frac{\partial f}{\partial v_r} \tag{37}$$

$$\frac{\partial^2 \ln(f)}{\partial v_r \partial v_s} = \frac{1}{f} \frac{\partial^2 f}{\partial v_r \partial v_s} - \frac{1}{f^2} \frac{\partial f}{\partial v_r} \frac{\partial f}{\partial v_s} \tag{38}$$

$$\begin{aligned} \frac{\partial^3 \ln(f)}{\partial v_r \partial v_s \partial v_w} &= 2 \frac{1}{f^3} \frac{\partial f}{\partial v_r} \frac{\partial f}{\partial v_s} \frac{\partial f}{\partial v_w} - \frac{1}{f^2} \frac{\partial^2 f}{\partial v_r \partial v_s} \frac{\partial f}{\partial v_w} - \frac{1}{f^2} \frac{\partial^2 f}{\partial v_s \partial v_w} \frac{\partial f}{\partial v_r} \\ &- \frac{1}{f^2} \frac{\partial^2 f}{\partial v_r \partial v_w} \frac{\partial f}{\partial v_s} + \frac{1}{f} \frac{\partial^3 f}{\partial v_r \partial v_s \partial v_w} \end{aligned} \tag{39}$$

All the partial derivatives of f that are involved in the above expressions are earlier shown to exist and since $0 < f < \infty$ for $t > 0$. Hence the derivatives in (37)-(39) exist and the condition (i) is verified.

Since for $t > 0$, $\ln(t) < t$, hence by replacing $\ln(t)$ by t and negative signs in $\Phi_{jk}(t)$ s by positive signs in expression (36), we can always find a function $A(t) = [a_1 t^{\alpha_1} + \dots + a_k t^{\alpha_k}]$ and a positive number N such that

$$\left|\left[\sum_{j,k=0, j+k \leq \ell}^{\ell} \left(\ln\left(\frac{t}{\theta_1}\right)\right)^j \left(\ln\left(\frac{t}{\theta_2}\right)\right)^k \Phi_{jk}(t)\right]\right| < A(t)$$

and

$$\exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right] \leq e^{-Nt}.$$

From these, we have

$$|g| < A(t)e^{-Nt}$$

and $A(t)e^{-Nt}$ is bounded for all $t > 0$. Therefore, the two parts of condition (ii) are satisfied. For proving the third part of condition (ii), we have the third order partial derivatives of f with respect to the parameters which are of the following form

$$\frac{\partial^3 f}{\partial v_1 \partial v_2 \partial v_3}; 2.0cmv_j \varepsilon \{\theta_1, \theta_2, \beta_1, \beta_2\}, 0.5cm \text{ for } 0.1cm j = 1, 2, 3.$$

On simplification, it can produce a form

$$[\sum_{j,k=0, j+k \leq 3} \left(\ln\left(\frac{t}{\theta_1}\right)\right)^j \left(\ln\left(\frac{t}{\theta_2}\right)\right)^k \Phi_{jk}(t)] \times \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right].$$

Again, corresponding to each $\frac{\partial^3 f}{\partial v_1 \partial v_2 \partial v_3}$, we can easily find a polynomial $A_1(t)$ such that

$$\left| \frac{\partial^3 f}{\partial v_1 \partial v_2 \partial v_3} \right| < A_1(t) \times \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right].$$

Let us denote $A_1(t) \times \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1} - \left(\frac{t}{\theta_2}\right)^{\beta_2}\right]$ by $H(t)$. For $H(t)$, we can write

$$\int_0^\infty H(t) f(t) dt = \int_0^\infty A_1'(t) \exp\left[-2\left(\frac{t}{\theta_1}\right)^{\beta_1} - 2\left(\frac{t}{\theta_2}\right)^{\beta_2}\right] dt,$$

where $A_1'(t) = [b_1 t^{\gamma_1} + \dots + b_m t^{\gamma_m}]$. Since $t \geq 0$ and $\theta_1 > 0, \theta_2 > 0$, hence replacing $\exp\left[-2\left(\frac{t}{\theta_2}\right)^{\beta_2}\right]$ by its maximum value, which is 1, we can write

$$\int_0^\infty H(t) f(t) dt < \int_0^\infty [b_1 t^{\gamma_1} + \dots + b_m t^{\gamma_m}] \exp\left[-2\left(\frac{t}{\theta_1}\right)^{\beta_1}\right] dt. \tag{40}$$

Let us consider a integral of the following type

$$\int_0^\infty t^{\gamma_j} \exp\left[-2\left(\frac{t}{\theta_1}\right)^{\beta_1}\right] dt = \frac{\theta_1^{\gamma_j+1}}{\beta_1 2^{\beta_1}} \Gamma\left(\frac{\gamma_j+1}{\beta_1}\right) = \delta_j \text{ (say)},$$

where δ_j is a moment of generalized *gamma* density and exists for all $\gamma_j > -1, \beta_1 > 0$ and $\theta_1 > 0$. Hence, each term in integral of right hand side of inequality (40) results in constant multiple of moment of generalized *gamma* density, which exist. If we denote $\sum_{j=1}^m b_j \delta_j$ by C_M , we can write

$$\int_0^\infty H(t) f(t) dt < C_M < \infty.$$

With this, third part of condition(ii) is verified.

For verifying condition (iii), we need to prove the matrix $J = ((J_{rs}(v)))$ positive definite.

The matrix J can be written as

$$J = \int_{t=0}^\infty \begin{bmatrix} \left(\frac{\partial \ln(f)}{\partial \theta_1}\right)^2 & \left(\frac{\partial \ln(f)}{\partial \theta_1}\right)\left(\frac{\partial \ln(f)}{\partial \theta_2}\right) & \left(\frac{\partial \ln(f)}{\partial \theta_1}\right)\left(\frac{\partial \ln(f)}{\partial \beta_1}\right) & \left(\frac{\partial \ln(f)}{\partial \theta_1}\right)\left(\frac{\partial \ln(f)}{\partial \beta_2}\right) \\ \left(\frac{\partial \ln(f)}{\partial \theta_2}\right)\left(\frac{\partial \ln(f)}{\partial \theta_1}\right) & \left(\frac{\partial \ln(f)}{\partial \theta_2}\right)^2 & \left(\frac{\partial \ln(f)}{\partial \theta_2}\right)\left(\frac{\partial \ln(f)}{\partial \beta_1}\right) & \left(\frac{\partial \ln(f)}{\partial \theta_2}\right)\left(\frac{\partial \ln(f)}{\partial \beta_2}\right) \\ \left(\frac{\partial \ln(f)}{\partial \theta_1}\right)\left(\frac{\partial \ln(f)}{\partial \beta_1}\right) & \left(\frac{\partial \ln(f)}{\partial \beta_1}\right)\left(\frac{\partial \ln(f)}{\partial \theta_2}\right) & \left(\frac{\partial \ln(f)}{\partial \beta_1}\right)^2 & \left(\frac{\partial \ln(f)}{\partial \beta_1}\right)\left(\frac{\partial \ln(f)}{\partial \beta_2}\right) \\ \left(\frac{\partial \ln(f)}{\partial \theta_1}\right)\left(\frac{\partial \ln(f)}{\partial \beta_2}\right) & \left(\frac{\partial \ln(f)}{\partial \beta_2}\right)\left(\frac{\partial \ln(f)}{\partial \theta_2}\right) & \left(\frac{\partial \ln(f)}{\partial \beta_2}\right)\left(\frac{\partial \ln(f)}{\partial \beta_1}\right) & \left(\frac{\partial \ln(f)}{\partial \beta_2}\right)^2 \end{bmatrix} f dt \tag{41}$$

$$J = \int_{t=0}^\infty \left\{ \begin{bmatrix} \frac{\partial \ln(f)}{\partial \theta_1} \\ \frac{\partial \ln(f)}{\partial \theta_2} \\ \frac{\partial \ln(f)}{\partial \beta_1} \\ \frac{\partial \ln(f)}{\partial \beta_2} \end{bmatrix} \times \begin{bmatrix} \frac{\partial \ln(f)}{\partial \theta_1} & \frac{\partial \ln(f)}{\partial \theta_2} & \frac{\partial \ln(f)}{\partial \beta_1} & \frac{\partial \ln(f)}{\partial \beta_2} \end{bmatrix} \right\} f dt. \tag{42}$$

Let $\frac{\partial \ln(f)}{\partial v}$ denotes the vector of all first order partial derivatives of $\ln(f)$ with respect to parameters $v = (\theta_1, \theta_2, \beta_1, \beta_2)$ and can be written as

$$\frac{\partial \ln(f)}{\partial v} = \begin{bmatrix} \frac{\partial \ln(f)}{\partial \theta_1} \\ \frac{\partial \ln(f)}{\partial \theta_2} \\ \frac{\partial \ln(f)}{\partial \beta_1} \\ \frac{\partial \ln(f)}{\partial \beta_2} \end{bmatrix}.$$

Mathematically, any matrix A that can be written as $A = BB'$, where B may be square or rectangular, is at least positive semidefinite. Hence matrix in the integral of equation (41) is at least positive semidefinite. The matrix J is expected value of the matrix $(\frac{\partial \ln(f)}{\partial v})(\frac{\partial \ln(f)}{\partial v})'$. Thus J has a covariance structure and such a structure can be singular only when two or more elements in vector $\frac{\partial \ln(f)}{\partial v}$ are linearly related. But in our case, the elements are partial derivatives of same function with respect to different parameters, that is, $\frac{\partial \ln(f)}{\partial v_j}$; $j = 1, 2, 3, 4$ and all v_j s are independent of each other. Hence there is no linear relationship among the elements $\frac{\partial \ln(f)}{\partial v_j}$. Obviously, J is nonsingular or in other words J is positive definite. Again since all the first order partial derivatives of $\ln(f)$ are shown to exist and finite, hence determinant of matrix J is also finite.

Therefore, we can say that the density f offers unique and consistent ML estimates of the parameters.