STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **11**, September 2023, pp 1027–1040.
Published online in International Academic Press (www.IAPress.org)

Variational Bayesian inference for exponentiated Weibull right censored survival data

Jibril Abubakar [1], Mohd Asrul Affendi Abdullah [1], Oyebayo Ridwan Olaniran [2,*]

[1] *Department of Mathematics and Statistics, Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Educational Hub, 84600 Pagoh, Malaysia.*
[2] *Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, PMB 1515, Nigeria.*

Abstract The exponential, Weibull, log-logistic and lognormal distributions represent the class of light and heavy-tailed distributions that are often used in modelling time-to-event data. The exponential distribution is often applied if the hazard is constant, while the log-logistic and lognormal distributions are mainly used for modelling unimodal hazard functions. The Weibull distribution is on the other hand well-known for modelling monotonic hazard rates. Recently, in practice, survival data often exhibit both monotone and non-monotone hazards. This gap has necessitated the introduction of Exponentiated Weibull Distribution (EWD) that can accommodate both monotonic and non-monotonic hazard functions. It also has the strength of adapting unimodal functions with bathtub shape. Estimating the parameter of EWD distribution poses another problem as the flexibility calls for the introduction of an additional parameter. Parameter estimation using the maximum likelihood approach has no closed-form solution, and thus, approximation techniques such as Newton-Raphson is often used. Therefore, in this paper, we introduce another estimation technique called Variational Bayesian (VB) approach. We considered the case of the accelerated failure time (AFT) regression model with covariates. The AFT model was developed using two comparative studies based on real-life and simulated data sets. The results from the experiments reveal that the Variational Bayesian (VB) approach is better than the competing Metropolis-Hasting Algorithm and the reference maximum likelihood estimates.

Keywords Exponentiated Weibull Distribution, Survival Analysis, Accelerated Failure Time, Bayesian, Variational Approximation

AMS 2010 subject classifications 62N01, 62N02.

DOI: 10.19139/soic-2310-5070-1295

1. Introduction

The exponential, Weibull, log-logistic and lognormal distributions are the popularly used parametric time-to-event models [1–3]. These distributions are commonly applied in time-to-event analysis primarily due to modelling simplicity and common framework. The common framework used in this context implies that they share a similar log-location-scale family [4–6] for statistical inference. Also, the ability to model day-to-day commonly seen survival data is often considered. The main consideration is the ability to implement the procedures on standard off-shelf software readily. Commonly applied distributions for unimodal hazard shapes are log-logistic and lognormal, while the Weibull is often used when posed with monotone hazard functions [4].

[7] discussed the several extensions of the Weibull and log-logistic distributions that have been proposed for primarily fitting several forms of flexible hazards shapes. An example of such extension is the Exponentiated Weibull (EW) distribution which generalizes the Weibull by adding an additional shape parameter [8]. The

*Correspondence to: Oyebayo Ridwan Olaniran (Email: olaniran.or@unilorin.edu.ng). Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, PMB 1515, Nigeria.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

EW model simultaneously achieves flexibility and simplicity by accomodating both monotone (increasing and decreasing) and non-monotone (unimodal and bathtub shape) failure functions with compensation of introducing additional shape parameter. As a generalized approach, it can be used to confirm the adequacy of Weibull distribution, especially when the newly introduced shape parameter approaches 1.

The three-parameter generalized gamma distribution [9] can also be used for modelling these four common types of hazard shapes. As suggested by [10], the EW distribution was found to be a promising substitute to the generalized gamma distribution. Thus, an in-depth analysis of the distribution was sought to explore its capability in modelling lifetime data. The early application of EW distribution was reported by [8] in the analysis of survival data, and [11] described likelihood-based inference for the class of power distributions that include the EW as a special case. The data sets on hazard times don't typically include only observed information on the time-to-event (T) and censoring status, but also information on covariates. This in turn posed the needs to develop robust regression models for understanding the existing relationship between the response, T , and one or more covariates which may affect the distribution of T. A Bayesian study of EW distribution was first developed by [12], while a modification of the log-exponentiated-Weibull regression model within the Bayesian framework was proposed by [13] to address cure rate specifically.

In recent time, [14] provided an in-depth analysis of AFT EW regression models using the Maximum Likelihood Estimation (MLE) approach and Bayesian MCMC techniques. However, there is no study that has evaluated the performance of the Variational Bayes Approximation (VBA) for the exponentiated Weibull regression compared to the most commonly used techniques such as maximum likelihood estimation and MCMC approaches. The variational Bayes approach has been shown to be better than MCMC techniques under mild regularity conditions [15, 16]. In addition, variational Bayes techniques are not limited in application to the Bayesian paradigm alone, i.e. one need not be a Bayesian expert before one can use variational Bayes[2, 16].

Based on the aforementioned, we specifically focus on parametric regression models that require a distributional assumption for T in the presence of covariates vector x. In particular, we aim to propose a variational Bayesian (VB) regression methodology based using the exponentiated Weibull distribution. The main reason for using EW relies on its generalizability to accommodates both monotone and non-monotone hazard/failure functions while doing it an insignificant cost of only estimating one extra parameter. The performance of the VB method is evaluated by comparing it with MLE and Bayesian MCMC (Metropolis-Hastings techniques) using simulated and Lung cancer datasets.

2. Related Works

The recent updates in modelling time-to-event data have focused on mixing two distributions or adding extra parameters to the existing distribution. The commonly applied models in the time-to-event analysis are exponential (Poisson), Weibull, gamma, and lognormal. The approach of adding extra parameter improves the flexibility in modelling failure rates data [17]. The vast majority of these distributions have originated either from the domain of reliability engineering or biological sciences. The specific interests are to estimate the elapsed time (time elapsed since failure) and the residual time (time remaining to failure) of a product.

[18] reported that the Weibull distribution is the most popular and general probability model used in a time-to-event analysis. The Weibull distribution and its substitute distribution, such as Gamma and Lognormal, have been applied in many time-to-event modelling tasks. However, with the flexibility and extended applicability of the two-parameter or the three-parameter Weibull distribution, it still does not offer the non-monotonical failure rates shapes that are often observed in (medical sciences; survival analysis, cure rates etc.) or engineering (reliability, equipment failures). For example, in lung cancer survival analysis, there are basically three major stages: stage 1 (tumour development), stage 2 (organ damage or lung failure) and stage 3 (extension of tumour to another body part). These stages or phases are similarly experienced in engineering/human as early failure (infant mortality), intrinsic failure (random hazard) and wear-out or late failure (ageing hazard). These hazard shapes are also regarded as bathtub failure shapes. Thus, monotonical hazard shapes distribution is not adequate for such data type. These

main drawbacks was the main reason the exponentiated Weibull (EWD) was proposed among several competing generalized distributions for modelling bathtub shape time-to-event data.

The earlier development of EWD can be traced to [19] who introduced an extra shape parameter to the existing two-parameter Weibull distribution. The strength of the EWD family is its ability to accommodate monotonical and non-monotonical failure functions, such as the unimodal-shaped and the bathtub-shaped ones [8]. From the time it was proposed, the EWD and its several extended versions have been applied to a wide area of practical applications, such as environmental flood data analysis [20], bus motor failure [8], human mortality testing [21] as well as survival analysis of head and neck cancer patients [8].

2.1. Exponentiated Weibull Distribution

The exponentiated Weibull distribution (EWD) was developed by [19] as an extenson to the two parameter WD distribution. The EWD family distributions are designed to accomodates non-monotonically monotonically hazards. In most lifetime data analysis applications, the bathtub shape or downward bathtub shape hazard are often observed thus suggesting the applicability of the EWD for modelling hazards when compared to standard WD. This is the area where the EWD plays significant role in hazard modelling.The EWD has two shape parameters and one scale parameter, thus the probability density function (pdf) takes the form:

$$f(t) = \alpha\beta\gamma(\beta t)^{\alpha-1}\big(1 - \exp[-(\beta t)^{\alpha}]\big)^{\gamma-1} \exp[-(\beta t)^{\alpha}], \tag{1}$$

and the cumulative distribution function:

$$F(t) = \big(1 - \exp[-(\beta t)^{\alpha}]\big)^{\gamma}, \tag{2}$$

where $t > 0$ is the support of the distribution, and $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters. Note that $\gamma = 1$ reduces the exponentiated Weibull to the Weibull distribution for which the probability density function is

$$f(t) = \alpha\beta(\beta t)^{\alpha-1} \exp[-(\beta t)^{\alpha}], \tag{3}$$

The rth moment of the exponentiated Weibull distribution does not have a closed form expression. However, [14] derived the median survival time as:

$$M(t) = \frac{1}{\beta}\left[-\log\left(1 - 0.5^{\frac{1}{\gamma}}\right)\right]^{\frac{1}{\alpha}}. \tag{4}$$

The survivor function, hazard function and cumulative hazard function of the exponentiated Weibull distribution are, respectively,

$$S(t) = 1 - \big(1 - \exp[-(\beta t)^{\alpha}]\big)^{\gamma}, \tag{5}$$

$$h(t) = \frac{\alpha\beta\gamma(\beta t)^{\alpha-1}\big(1 - \exp[-(\beta t)^{\alpha}]\big)^{\gamma-1} \exp[-(\beta t)^{\alpha}]}{1 - \big(1 - \exp[-(\beta t)^{\alpha}]\big)^{\gamma}}, \tag{6}$$

$$H(t) = -\log\left\{1 - \big(1 - \exp[-(\beta t)^{\alpha}]\big)^{\gamma}\right\}, \tag{7}$$

The hazard is (a) monotone increasing for $\alpha \geq 1$ and $\alpha\gamma \geq 1$, (b) monotone decreasing for $\alpha \leq 1$ and $\alpha\gamma \leq 1$, (c) unimodal for $\alpha < 1$ and $\alpha\gamma > 1$, and (d) bathtub-shaped for $\alpha > 1$ and $\alpha\gamma < 1$ [14].

2.2. Accelerate Failure Time (AFT) EW regression models

Consider an ordinary regression model for log survival time T, of the form

$$Y = \log T = x'\theta + \sigma W; \tag{8}$$

where $x = (x_1, x_2 \ldots, x_p)$ be a column vector of p covariates, and $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ is the corresponding vector of regression coefficients, the error term W has a suitable distribution, e.g. extreme value, generalized extreme value, normal or logistic. This leads to Weibull, generalized gamma, log-normal or log-logistic models for T. For example if W is extreme value then T has a Weibull distribution with $\log\lambda = x'\theta$ and $p = \frac{1}{\sigma}$. Note that λ depends on the covariates but p is assumed the same for everyone.

This model has an accelerated life interpretation. In this formulation we view the error term σW as a standard or reference distribution that applies when $x = 0$. It will be convenient to translate the reference distribution to the time scale by defining $T_0 = \exp\{\sigma W\}$.

For EW AFT regression models with $Y = \log T$, the corresponding density and survival functions are:

$$f(y) = \frac{\gamma}{\tau}\left(1 - \exp\left\{-\exp\left[\frac{(y-\mu)}{\tau}\right]\right\}\right)^{\gamma-1} \exp\left\{\frac{(y-\mu)}{\tau} - \exp\left[\frac{(y-\mu)}{\tau}\right]\right\} \tag{9}$$

and

$$S(y) = 1 - \left(1 - \exp\left\{-\exp\left[\frac{(y-\mu)}{\tau}\right]\right\}\right)^{\gamma} \tag{10}$$

where $-\infty < y < \infty$, $\mu = -\log\beta$ and $\tau = \alpha^{-1}$. In the AFT regression framework, the assumption is that the probability of an individual (with covariates x) surviving to time t is the same as the probability of a reference individual (i.e., $x = 0$) surviving to time $t\exp(x'\theta)$ [14]. Formally, $S(t; x) = S_0(t\exp(x'\theta)$, where $S_0()$ is the baseline survivor function. This refers to the fact that the covariates act multiplicatively on time so that their effect is to accelerate (or decelerate) the time to failure. If we start with the exponentiated Weibull baseline survivor function, we get

$$S(t) = 1 - \left(1 - \exp\{-[\beta t\exp(x'\theta)]^{\alpha}\}\right)^{\gamma} = 1 - \left(1 - \exp[-[\beta^{*} t]^{\alpha}]\right)^{\gamma} \tag{11}$$

which is also an exponentiated Weibull survivor function with $\beta^{*} = \beta\exp(x'\theta)$. This shows that the exponentiated Weibull is closed under the AFT family. If we let T_0 be an exponentiated Weibull random variable corresponding to the lifetime when $x = 0$, so that the survivor function of T_0 is of the form $S_0(.)$. Then, $T_0 = T\exp(x'\theta)$ from (11). By taking the logarithm of both sides of equation (11), we get

$$Y = \theta_0 - x'\theta + \tau W; \tag{12}$$

where $\theta_0 = -\log\beta$, $\tau = \alpha^{-1}$ and $Y = \log T$ which follows (9) and $\alpha = \theta_0 - x'\theta$ and $W = (\log T_0 - \theta_0)/\tau$ is the random error component which is distributed as

$$f(w) = \gamma\left[1 - \exp\left(-e^{w}\right)\right]^{\gamma-1}\exp\left(w - e^{w}\right), -\infty < w < \infty. \tag{13}$$

Now, if we rewrite $\theta^{*} = (\theta_0, -\theta_1, \ldots, -\theta_p)$ and $x^{*} = (\mathbf{1}, x')'$, we get another simpler regression model

$$Y = x^{*'}\theta^{*} + \tau W. \tag{14}$$

2.3. Maximum Likelihood Estimation (MLE) of right censored AFT EW regression model

Suppose we have a right censored random sample consisting of data (y_i, δ_i, x_i'), $i = 1, 2, \ldots, n$, where $y_i = logt_i$ is a log-lifetime or log censoring time according to whether $\delta_i = 1$ (if the event occured at $t = t_0$) or $\delta_i = 0$ (if the event occured at $t > t_0$), respectively. The likelihood and log-likelihood function of the exponentiated Weibull regression model can be written as

$$L(\boldsymbol{\theta}) = \left(\frac{\gamma}{\tau}\right)^{r}\prod_{i=1}^{n}[a_i^{\gamma-1} + e^{w_i - e^{w_i}}]^{\delta_i}(1 - a_i^{\gamma})^{1-\delta_i} \tag{15}$$

$$l(\boldsymbol{\theta}) = r\log\gamma - r\log\tau + \sum_{i=1}^{n}\delta_i[(\gamma-1)\log a_i + (w_i - e^{w_i})] + \sum_{i=1}^{n}(1-\delta_i)\log(1 - a_i^{\gamma}) \tag{16}$$

where $\boldsymbol{\theta} = (\tau, \gamma, \theta^{*'})'$, $r = \sum_{i=1}^{n} \delta_i$, $w_i = (y_i - x^{*'}\theta^*)/\tau$ and $a_i = 1 - \exp(-e^{w_i})$. Suppose we let $b_i = (1 - a_i)\log(1 - a_i)$ and define

$$g_i = \frac{\partial l(\boldsymbol{\theta})}{\partial w_i} \tag{17}$$

$$= \frac{\partial}{\partial w_i}\{\sum_{i=1}^{n} \delta_i[(\gamma - 1)\log a_i + (w_i - e^{w_i})] + \sum_{i=1}^{n}(1 - \delta_i)\log(1 - a_i^{\gamma})\} \tag{18}$$

$$= -(\gamma - 1)\left(\frac{\delta_i b_i}{a_i}\right) + \delta_i[1 + \log(1 - a_i)] + \gamma\left[\frac{(1 - \delta_i)b_i}{a_i}\right]\left(\frac{a_i^{\gamma}}{1 - a_i^{\gamma}}\right). \tag{19}$$

Subsequently, the score functions of other parameters are

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \tau} = -\frac{r}{\tau} - \frac{1}{\tau}\sum_{i=1}^{n} g_i w_i \tag{20}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \gamma} = \frac{r}{\gamma} + \sum_{i=1}^{n} \delta_i \log a_i - \sum_{i=1}^{n}(1 - \delta_i)\left(\frac{a_i^{\gamma}\log a_i}{1 - \log a_i}\right) \tag{21}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_j} = -\frac{1}{\tau}\sum_{i=1}^{n} g_i x_{ij}, j = 0, 1, 2, \ldots, p. \tag{22}$$

The score functions are then solved numerically using Newton-Raphson procedure to accurately estimate the parameters.

2.4. Metropolis-Hastings Approach for AFT EW regression model

In this section, we present the Metropolis-Hastings algorithm procedure which is similar to the approach used in [2, 22–26] for the parameter estimation. We use the uninformative Uniform (c, d) prior since there is no prior information or elicitation may be difficult. Also, since there are three parameters, we suggest three independent Uniform (c, d) distributions. The joint density function for the prior of the three parameters $\boldsymbol{\theta} = (\tau, \gamma, \theta^{*'})'$ can be defined as:

$$f(\tau, \gamma, \theta^{*'})|c_1, d_1; c_2, d_2; c_3, d_3) = \prod_{k=1}^{3}(d_k - c_k)^{-1} \tag{23}$$

where $c_1, d_1; c_2, d_2; c_3, d_3$ are the prior hyperparameters for the parameters $\tau, \gamma, \theta^{*'}$. The posterior distribution of the three parameters $\tau, \gamma, \theta^{*'}$ for the EW model can be defined as the product of the likelihood $L(\boldsymbol{\theta})$ and the prior density which is:

$$f(\boldsymbol{\theta}|y) = L(\boldsymbol{\theta}) \times \prod_{k=1}^{3}(d_k - c_k)^{-1} \tag{24}$$

The posterior distribution in (24) does not have a closed form as its an approximate distribution since the marginal distribution that ensures it scale to one have been dropped. One of the ways of sampling from this distribution is by using the Metropolis-Hastings algorithm (MH; Lee, 2012). The metropolis-hastings procedure for the EWD AFT regression model is:

1. Initialize Θ^0 such that $p(\Theta^0|y) > 0$.
2. For $i = 1, 2, \ldots$
3. Take a random sample $\tilde{\Theta}$ from a preferred proposal distribution (Preferably lognormal distribution).
4. Compute the accept/reject or moving probability by;
5. Take a random sample $U \sim U(0, 1)$
6.

$$\Theta^{i+1} = \begin{cases} \tilde{\Theta} & \text{if } U \leq \pi(\Theta^i, \tilde{\Theta}); \\ \Theta^i & \text{if } U > \pi(\Theta^i, \tilde{\theta}). \end{cases}$$

3. Variational Bayes Approach

Suppose we let $x = x_{1:n}$ represent a collection of observed variables and $z = z_{1:m}$ represent collection of latent variables, with joint density function $p(z, x)$. As expalined earlier in chapter 2, the constant of proportionality can be omitted. The inferential problem thus involves the computation of the conditional density for the latent variables using the observations, $p(z|x)$. Using the conditional density, the point and interval estimates of the latent variables can be estimated. The conditional density is often presented as

$$p(z|x) = \frac{p(z, x)}{p(x)} \tag{25}$$

The denominator part of $p(z|x)$ is referred to as the marginal of the random sample observed. This is usually calculated by integrating out the parameter of interest from the joint density,

$$p(x) = \int p(z, x)dz. \tag{26}$$

In most models, this marginal density is usually not available or computationally expensive[27]. The marginal density is what is required to calculate the conditional from the joint density and thus the main reason variational inference is difficult.

It is worthy of note that it is assumed that the unknown parameter values are random. These parameters encompass all that covers the data as often done in other Bayesian analysis. The parameters are also local to each observed data points.

Now in th case of the EW regression model, the desired posterior distribution is

$$p(\boldsymbol{\theta}|y, x) = \frac{L(\boldsymbol{\theta}|y, x)p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|y, x)p(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{27}$$

.

By variational inference, we want to approximate $p(\boldsymbol{\theta}|y, x)$ in 27 with a $q(\boldsymbol{\theta})$ by constructing the equality

$$\ln \int L(\boldsymbol{\theta}|y, x)p(\boldsymbol{\theta}) = \int q(\boldsymbol{\theta}) \ln \frac{L(\boldsymbol{\theta}|y, x)p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} + \int q(\boldsymbol{\theta}) \ln \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|y, x)} d\boldsymbol{\theta} \tag{28}$$

.

Next we define $q(\boldsymbol{\theta})$ as the product of independent densities using the mean-field assumption given as

$$q(\boldsymbol{\theta}) = q(\gamma)q(\tau)q(\theta^{*'}) \tag{29}$$

.

The variational objective $\mathcal{L}$ is then computed as:

$$\mathcal{L} = \int q(\boldsymbol{\theta}) \ln L(\boldsymbol{\theta}|y, x)d\boldsymbol{\theta} - \int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{30}$$

.

$$\begin{aligned} \mathcal{L} = &\int q(\gamma)q(\tau)q(\theta^{*'}) \Big(r\log\gamma - r\log\tau + \sum_{i=1}^{n} \delta_i[(\gamma - 1)\log a_i + (w_i - e^{w_i})] \\ &+ \sum_{i=1}^{n}(1-\delta_i)\log(1 - a_i^{\gamma}) \Big) d\gamma d\tau d\theta^{*'} \\ &- \int q(\gamma)q(\tau)q(\theta^{*'}) \ln q(\gamma)q(\tau)q(\theta^{*'}) d\gamma d\tau d\theta^{*'} \end{aligned} \quad . \tag{31}$$

Equation 31 is iterated until convergence is achieved.

4. Simulation, Results and Discussion

In this section, an empirical evaluation of the proposed method Variational Bayes (VB), Maximum Likelihood Estimation (MLE) and Metropolis-Hastings (MH) procedures was achieved using simulation and real-life dataset on Lung cancer treatment. The estimation methods were compared based on Bias, Standard Error or Standard Deviaion, Mean Square Error and Coverage probability.

4.1. Simulation studies

For the purpose of the study, we simulated two covariates using an AFT regression framework: the first covariate is a continuous covariate (x_1) which follows the standard normal distribution, and the other covariate (x_2) is binary which is assumed to follow a $Bernoulli(\pi = 0.5)$ distribution. The regression coefficient values are set to mimic the real-life Lung cancer dataset which result is presented in the later section. The MLE estimates for the covariates of variables Karnofsky performance score (100=good) and treatment in the Lung cancer dataset are $\boldsymbol{\theta} = [\theta^{*'} = (\theta_0^* = 3.5742793, \theta_1^* = 0.7547705, \theta_2^* = -0.1052200), \tau = 1/0.5868033, \gamma = 3.0987992]'$. The simulation process is made realistic using different censoring proportions simulated from exponential distribution. Although, the censoring proportion in the original dataset is 7%, we examine the behaviours of the methods at varying censoring proportions. The following censoring proportions 10%, $20\%, 30\%, 40\%$ and 50% corresponding to light to heavy censoring conditions are used.

The formula for the performance metrics of the various methods are as provided below:

$$Bias = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \tag{32}$$

$$Standard\ Error\ (SE) = \sqrt{\sum_{i=1}^{I} \frac{(\hat{\boldsymbol{\theta}} - \bar{\hat{\boldsymbol{\theta}}})^2}{I-1}} \tag{33}$$

$$Mean\ Square\ Error\ (MSE) = \sum_{i=1}^{I} \frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2}{I-1} \tag{34}$$

$$95\%\ Coverage\ Probability = \sum_{i=1}^{I} \frac{\left(\hat{\boldsymbol{\theta}} - Z_{0.025} \times SE(\hat{\boldsymbol{\theta}}) < \boldsymbol{\theta}\right) \bigcap \left(\hat{\boldsymbol{\theta}} + Z_{0.975} \times SE(\hat{\boldsymbol{\theta}}) > \boldsymbol{\theta}\right)}{I} \tag{35}$$

where $Z_{1-\alpha/2}$ is the quantile of standard normal distribution at the desired significance level, I is the number of replication of each simulation runs which is set as $I = 200$. The sample size n was fixed at $n = 137$.

4.2. Simulation results

Table 1 presents the simulation results for varying censoring proportions at fixed sample size n and replication set to be 200.

Table 1 results show that the estimates returned using the VB method is more consistent to the true value when compared to the other two methods. The estimates using MH is better than MLE estimates in terms of biasness and consistency. Overall, the MLE estimate is not consistent with the true value at high censoring proprotion.

Table 2 presents the standard error/ standard deviation and Mean Square Error (MSE) of the three methods. These metrics were used to assess the efficiency of the methods. The various results over the different censoring proportions show that the VB estimates are the most efficient. The MLE estimates are mostly inefficient for the parameter γ across the various censoring proportion.

Figure 1 shows that increasing the censoring proportion increases the biasness of the estimates especially for MLE. While the VB and MH exhibited some form of robustness, the MLE estimates are not robust at all which suggests that they shouldn't be used with high censored data. Similarly, Figure 2 shows that the MLE estimates are also not efficient for low to high censoring proportions. The two Bayesian approaches MH and VB are highly

Table 1. Simulation results for the estimates and bias at various censoring proportion p.

Censoring		Estimates				Bias		
		TRUE	MLE	MH	VB	MLE	MH	VB
$p = 0.1$	θ_0^*	3.574	3.438	3.361	**3.634**	-0.137	-0.213	**0.060**
	θ_1^*	0.755	0.747	0.742	**0.752**	-0.008	-0.013	**-0.003**
	θ_2^*	-0.105	-0.094	-0.106	**-0.086**	0.011	0.000	**0.020**
	α	0.587	0.606	0.553	**0.595**	0.019	-0.034	**0.008**
	γ	3.099	4.729	4.127	**3.242**	1.630	1.028	**0.143**
$p = 0.2$	θ_0^*	3.574	3.441	3.451	**3.736**	-0.133	-0.123	**0.162**
	θ_1^*	0.755	0.747	0.745	**0.750**	-0.008	-0.010	**-0.005**
	θ_2^*	-0.105	-0.096	-0.084	**-0.091**	0.009	0.022	**0.014**
	α	0.587	0.593	0.551	**0.583**	0.006	-0.036	**-0.004**
	γ	3.099	5.032	4.103	**3.247**	1.933	1.004	**0.148**
$p = 0.3$	θ_0^*	3.574	3.467	3.624	**3.816**	-0.108	0.049	**0.241**
	θ_1^*	0.755	0.750	0.759	**0.754**	-0.005	0.004	**-0.001**
	θ_2^*	-0.105	-0.091	-0.065	**-0.076**	0.014	0.041	**0.029**
	α	0.587	0.581	0.569	**0.562**	-0.006	-0.018	**-0.025**
	γ	3.099	5.319	3.954	**3.287**	2.220	0.855	**0.188**
$p = 0.4$	θ_0^*	3.574	3.445	3.757	**3.942**	-0.129	0.183	**0.367**
	θ_1^*	0.755	0.745	0.755	**0.749**	-0.010	0.000	**-0.006**
	θ_2^*	-0.105	-0.089	-0.066	**-0.080**	0.016	0.040	**0.026**
	α	0.587	0.559	0.540	**0.547**	-0.028	-0.047	**-0.040**
	γ	3.099	5.796	3.939	**3.318**	2.697	0.841	**0.219**
$p = 0.5$	θ_0^*	3.574	3.192	3.943	**4.052**	-0.383	0.368	**0.477**
	θ_1^*	0.755	0.756	0.781	**0.762**	0.001	0.026	**0.007**
	θ_2^*	-0.105	-0.103	-0.089	**-0.094**	0.003	0.016	**0.011**
	α	0.587	0.505	0.562	**0.518**	-0.082	-0.025	**-0.069**
	γ	3.099	7.024	3.990	**3.421**	3.925	0.891	**0.323**

effcient and robust to low through high censoring proportion. The combined effects of consistency and efficiency was measured using MSE. Similar behaviours as in variance of the estimates were observed for MSE in Figure 3. Again, the most efficient, consistent and robust estimates are VB estimates.

Figure 4 presents the result for the 95% coverage probability. The expected behaviour is that the estimates returned values that falls within the 95% confidence or credible intervals 95% of time. The MLE estimates exhibited robustness to censoring proportion here. While the coverage probability of MH increases with increase in censoring proportion until 0.4 before a sharp decline is observed, the VB exhibited a downward trend from low to high censoring proportion. Although, the estimated coverage probability for MH and VB varies between 90% to 98%, the coverage probability of MLE converges between 93% to 97%. This implies that approximeately 95% of time MLE produces estimates that conforms with nominal or target values while the estimates of MH and VB are less than the target by an error of 5% and more than the target by an error of 3% on the average.

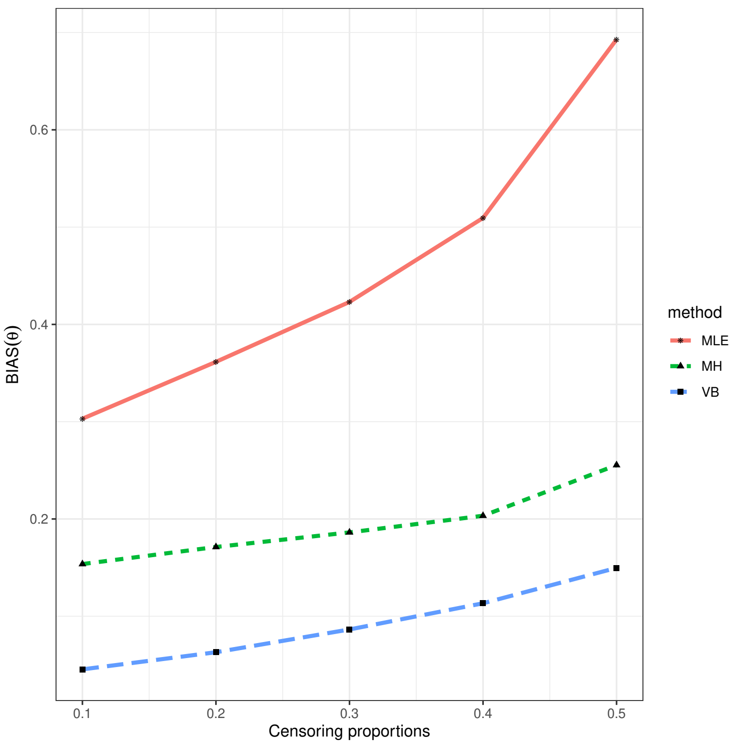

Figure 1. Average Bias of the estimates at varying censoring proportion.

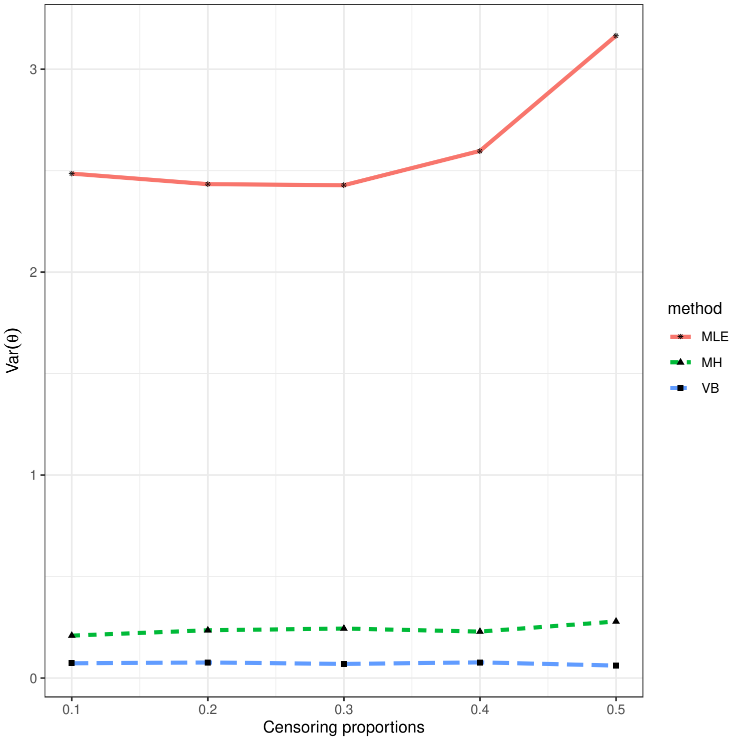

Figure 2. Average variance of the estimates at varying censoring proportion.

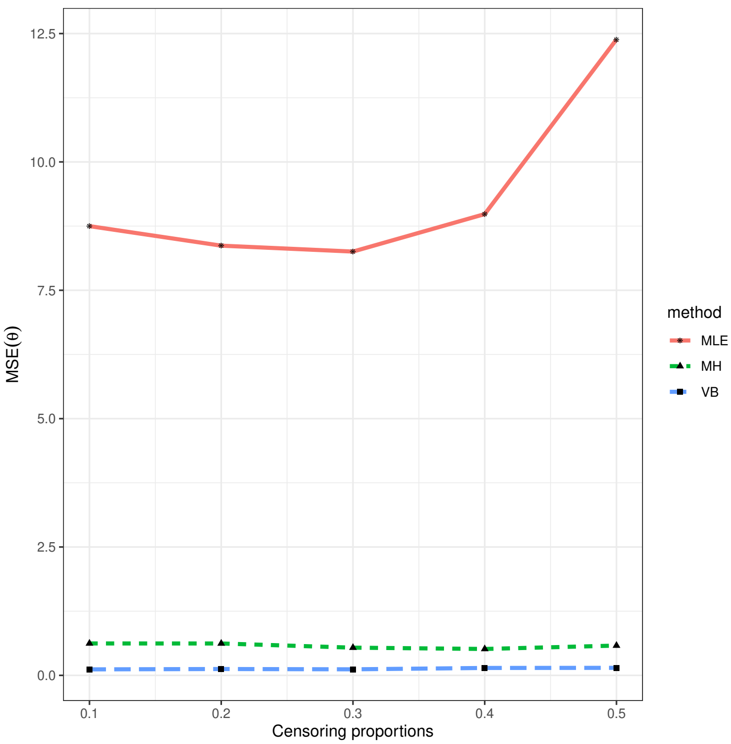

Figure 3. Average Mean Square Error of the estimates at varying censoring proportion.

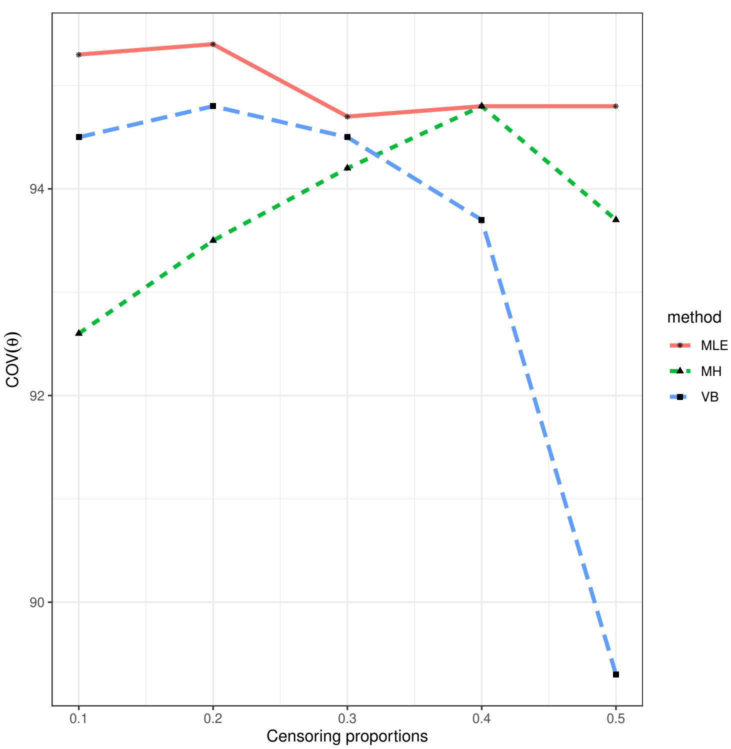

Figure 4. Average Coverage Probability at varying censoring proportion.

Table 2. Simulation results for the Standard Error (SE) and Mean Square Error (MSE) at various censoring proportion p.

Censoring		Standard Error			MSE		
		MLE	MH	VB	MLE	MH	VB
$p = 0.1$	θ_0^*	1.074	0.508	**0.345**	1.165	0.302	**0.122**
	θ_1^*	0.108	0.127	**0.098**	0.012	0.016	**0.010**
	θ_2^*	0.195	0.259	**0.194**	0.038	0.067	**0.038**
	α	0.178	0.104	**0.090**	0.032	0.012	**0.008**
	γ	6.328	1.289	**0.622**	42.501	2.710	**0.405**
$p = 0.2$	θ_0^*	1.216	0.531	**0.356**	1.488	0.296	**0.152**
	θ_1^*	0.119	0.138	**0.109**	0.014	0.019	**0.012**
	θ_2^*	0.208	0.273	**0.204**	0.043	0.074	**0.042**
	α	0.199	0.189	**0.099**	0.039	0.037	**0.010**
	γ	6.059	1.295	**0.618**	40.262	2.677	**0.402**
$p = 0.3$	θ_0^*	1.339	0.538	**0.329**	1.796	0.291	**0.166**
	θ_1^*	0.120	0.137	**0.108**	0.014	0.019	**0.012**
	θ_2^*	0.229	0.280	**0.237**	0.053	0.080	**0.057**
	α	0.221	0.289	**0.087**	0.048	0.084	**0.008**
	γ	5.883	1.225	**0.559**	39.360	2.224	**0.346**
$p = 0.4$	θ_0^*	1.491	0.525	**0.369**	2.228	0.307	**0.271**
	θ_1^*	0.132	0.156	**0.119**	0.017	0.024	**0.014**
	θ_2^*	0.243	0.302	**0.239**	0.059	0.092	**0.057**
	α	0.238	0.230	**0.097**	0.057	0.055	**0.011**
	γ	5.955	1.179	**0.567**	42.558	2.090	**0.367**
$p = 0.5$	θ_0^*	1.605	0.545	**0.310**	2.709	0.432	**0.323**
	θ_1^*	0.144	0.170	**0.138**	0.021	0.029	**0.019**
	θ_2^*	0.260	0.339	**0.258**	0.067	0.115	**0.066**
	α	0.264	0.421	**0.077**	0.076	0.177	**0.011**
	γ	6.623	1.166	**0.457**	59.050	2.147	**0.312**

5. Lung cancer survival data

Prentice [28] originally described a randomized clinical trial that involves 137 advanced lung cancer patients treated with a standard chemotherapy agent or a control drug [14]. The time to event were recorded from the study inception for each of the patients. Nine patients were censored as their event time were not known till the end of the study. The specific objective was to determine the cure rate of the chemotherapy on different tumour cell type. The four different tumour cell classification squamous, small, adeno and large. The other variables considered are performance status, months between diagnosis and entry into the study, age, and a history of previous therapy for lung cancer (prior).

Table 3 presents the estimates and standard deviation (SD) or error (SE) for the three methods. The two Bayesian methods (VB and MH) were found to be more stable (efficient: lower standard deviation) than the MLE. In addition, the interval estimates presented in Table 4 showed that VB and MH methods returned more significant estimates than MLE at 5% level of signficance.

Table 3. Real-life data results for the various methods.

	VB		MLE		MH	
	Estimate	SD	Estimate	SE	Estimate	SD
(Intercept)	0.810	0.641	2.787	0.818	1.767	0.006
trt	-0.170	0.202	-0.229	0.191	-0.237	0.091
celltypesmallcell	-0.838	0.235	-0.311	0.255	-0.529	0.002
celltypeadeno	-0.984	0.269	-0.657	0.283	-0.641	0.063
celltypelarge	-0.193	0.156	-0.123	0.262	0.160	0.066
karno	0.036	0.005	0.033	0.005	0.018	0.016
diagtime	0.002	0.010	-0.002	0.009	0.033	0.031
age	0.013	0.008	0.008	0.009	0.033	0.018
prior	-0.011	0.025	0.002	0.022	0.039	0.049
α	0.498	0.045	0.975	0.161	1.937	0.058
γ	4.525	0.822	1.136	0.329	2.060	0.055

Table 4. 95% credible and confidence intervals for the estimates

	VB		MLE		MH	
	2.5%LB	97.5%UB	2.5%LB	97.5%UB	2.5%LB	97.5%UB
(Intercept)	-0.471	2.091	1.184	4.390	1.761	1.774
trt	-0.574	0.233	-0.602	0.145	-0.321	-0.139
celltypesmallcell	-1.309	-0.367	-0.811	0.189	-0.531	-0.526
celltypeadeno	-1.522	-0.446	-1.211	-0.103	-0.700	-0.573
celltypelarge	-0.506	0.120	-0.636	0.390	0.099	0.231
karno	0.026	0.046	0.023	0.043	0.003	0.034
diagtime	-0.019	0.023	-0.019	0.015	0.000	0.062
age	-0.004	0.029	-0.009	0.026	0.014	0.050
prior	-0.061	0.038	-0.041	0.044	-0.014	0.084
α	0.407	0.589	0.659	1.290	1.883	2.000
γ	2.881	6.170	0.490	1.781	2.000	2.110

6. Conclusion

In this paper, the Variational Bayesian (VB) inference was developed for the Exponentiated Weibull (EW) right-censored survival data. The Accelerated Failure Time (AFT) model was used to determine the likelihood function. The MLE and MCMC estimate proposed in [14] was compared to the VB estimate using simulated and real-life data. Simulated results revealed that the VB estimates are more efficient than both MLE and MH procedures. However, the coverage probabilities of VB estimates are less precise than the MLE estimates. The efficient estimate results were replicated using the real-life Lung cancer dataset. The current study is limited to only right-censored time-to-event data; we plan to extend the methodology to other censoring schemes such as left, interval and party-interval in our future study.

REFERENCES

[1] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*, vol. 360. John Wiley & Sons, 2011.

[2] O. R. Olaniran and M. A. A. Abdullah, "Bayesian analysis of extended cox model with time-varying covariates using bootstrap prior," *Journal of Modern Applied Statistical Methods*, vol. 18, no. 2, p. 7, 2019.

[3] S. A. M. Jamil, M. A. A. Abdullah, S. L. Kek, O. R. Olaniran, and S. E. Amran, "Simulation of parametric model towards the fixed covariate of right censored lung cancer data," in *Journal of Physics: Conference Series*, vol. 890, p. 012172, IOP Publishing, 2017.

[4] J. Lawless, "Event history analysis and longitudinal surveys," *Analysis of Survey data*, pp. 221–243, 2003.

[5] J. Popoola, O. Popoola, and O. R. Olaniran, "An approximate performance of self-similar lognormal m 1 k internet traffic model," *Journal of Science and Technology*, vol. 11, no. 2, pp. 36–42, 2019.

[6] J. Popoola, W. B. Yahya, O. Popoola, and O. R. Olaniran, "Generalized self-similar first order autoregressive generator (gsfo-arg) for internet traffic," *Statistics, Optimization & Information Computing*, vol. 8, no. 4, pp. 810–821, 2020.

[7] D. Ghinolfi, J. Marti, P. De Simone, Q. Lai, D. Pezzati, L. Coletti, D. Tartaglia, G. Catalano, G. Tincani, P. Carrai, *et al.*, "Use of octogenarian donors for liver transplantation: a survival analysis," *American Journal of Transplantation*, vol. 14, no. 9, pp. 2062–2071, 2014.

[8] G. S. Mudholkar, D. K. Srivastava, and M. Freimer, "The exponentiated weibull family: A reanalysis of the bus-motor-failure data," *Technometrics*, vol. 37, no. 4, pp. 436–445, 1995.

[9] E. W. Stacy *et al.*, "A generalization of the gamma distribution," *The Annals of mathematical statistics*, vol. 33, no. 3, pp. 1187–1192, 1962.

[10] C. Cox and M. Matheson, "A comparison of the generalized gamma and exponentiated weibull distributions," *Statistics in medicine*, vol. 33, no. 21, pp. 3772–3780, 2014.

[11] A. Pewsey, H. W. Gómez, and H. Bolfarine, "Likelihood-based inference for power distributions," *Test*, vol. 21, no. 4, pp. 775–789, 2012.

[12] V. Cancho, H. Bolfarine, and J. Achcar, "A bayesian analysis for the exponentiated-weibull distribution," *Journal of Applied Statistical Science*, vol. 8, no. 4, pp. 227–242, 1999.

[13] V. G. Cancho, J. Rodrigues, and M. de Castro, "A flexible model for survival data with a cure rate: a bayesian approach," *Journal of Applied Statistics*, vol. 38, no. 1, pp. 57–70, 2011.

[14] S. A. Khan, "Exponentiated weibull regression for time-to-event data," *Lifetime data analysis*, vol. 24, no. 2, pp. 328–354, 2018.

[15] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[16] O. R. Olaniran and M. A. A. Abdullah, "Subset selection in high-dimensional genomic data using hybrid variational bayes and bootstrap priors," in *Journal of Physics: Conference Series*, vol. 1489, p. 012030, IOP Publishing, 2020.

[17] S. Pasari and O. Dikshit, "Earthquake interevent time distribution in kachchh, northwestern india," *Earth, Planets and Space*, vol. 67, no. 1, p. 129, 2015.

[18] S. Pasari and O. Dikshit, "Stochastic earthquake interevent time modeling from exponentiated weibull distributions," *Natural hazards*, vol. 90, no. 2, pp. 823–842, 2018.

[19] G. S. Mudholkar and D. K. Srivastava, "Exponentiated weibull family for analyzing bathtub failure-rate data," *IEEE transactions on reliability*, vol. 42, no. 2, pp. 299–302, 1993.

[20] S. Nadarajah, “Bathtub-shaped failure rate functions,” *Quality & Quantity*, vol. 43, no. 5, pp. 855–863, 2009.

[21] M. Bebbington, C.-D. Lai, and R. Zitikis, “A flexible weibull extension,” *Reliability Engineering & System Safety*, vol. 92, no. 6, pp. 719–726, 2007.

[22] W. Yahya, O. Olaniran, and S. Ige, “On bayesian conjugate normal linear regression and ordinary least square regression methods: A monte carlo study,” *Ilorin Journal of Science*, vol. 1, no. 1, pp. 216–227, 2014.

[23] O. Olaniran, S. Olaniran, W. Yahya, A. Banjoko, M. Garba, L. Amusa, and N. Gatta, “Improved bayesian feature selection and classification methods using bootstrap prior techniques,” *Annals. Computer Science Series*, vol. 14, no. 2, 2016.

[24] O. R. Olaniran and M. A. A. B. Abdullah, “Gene selection for colon cancer classification using bayesian model averaging of linear and quadratic discriminants,” *Journal of Science and Technology: Special Issue on the Application of Science and Mathematics*, vol. 9, no. 3, pp. 140–144, 2017.

[25] O. R. Olaniran and M. A. A. B. Abdullah, “Bayesrandomforest: An r implementation of bayesian random forest for regression analysis of high-dimensional data,” in *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*, pp. 269–275, Springer, 2019.

[26] O. R. Olaniran and W. B. Yahya, “Bayesian hypothesis testing of two normal samples using bootstrap prior technique,” *Journal of Modern Applied Statistical Methods*, vol. 16, no. 2, p. 34, 2017.

[27] N. M. Bala and S. bin Safei, “A hybrid harmony search and particle swarm optimization algorithm (hspso) for testing non-functional properties in software system,” *Statistics, Optimization & Information Computing*, 2021.

[28] R. L. Prentice, “Exponential survivals with censoring and explanatory variables,” *Biometrika*, vol. 60, no. 2, pp. 279–288, 1973.