VARIATIONAL BAYESIAN INFERENCE FOR EXPONENTIATED WEIBULL RIGHT-CENSORED SURVIVAL DATA

JIBRIL ABUBAKAR

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Master of Science

Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia

SEPTEMBER, 2023

I dedicated this research work to Allah subuhanahu Wata' ala, my late parent Mallam Abubakar Achor Idah, Amina Inusa Ebaju and All My Family Members.o my father Alhaji,

ACKNOWLEDGMENT

All praise and adoration be to Allah SWT, the lord of the world, who grant me this ample opportunity to stand all the strains and problem encountered during the self-sponsored research work. I am particularly indebted to my supervisor Dr Mohd Asrul Affendi Abdullah from Faculty of Applied Sciences and Technology (FAST), Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia and the Co-supervisor, Dr Ridwan Oyebayo Olaniran of the Department of Statistics, University of Ilorin, Ilorin Kwara State, Nigeria for their invaluable contributions and greater support throughout this research work. I thank other Lecturers from FAST, especially my Research Methodology coordinator, Dr Suliadi Firdaus Sufahani, Professor Sufiyan, Dr Suzi Salwah Binti Jikan and my exammers. I also thank the FAST administrator En Suhaimi Bin Kamit and Miss Falatihat at CGS. Similarly, my gratitude goes to the Chief Executive, Dr Hamid Bobboye, the Executive Secretary Universal Basic Education Commission (UBEC), Abuja, for his personal support and endorsement of the study and his PA Barrister Suleiman Modibbo, Dr. (Engr.) Al-Hassan Garba, Faculty of Electrical and Electronic UTHM and Dr. Ibrahim Bala Bakari, Faculty of ICT, UTHM and Mr Yakubu Achimugu HOU Education Management Information System (EMIS)/Statistics Unit. UBEC Abuja. Finally, my deepest gratitude goes to Almighty Allah, who gave me the ability and courage to undergo this research successfully.

ABSTRACT

The Weibull, log-logistic and log-normal distributions represent the heavy-tailed distributions that are often used in modelling time-to-event data. While the loglogistic and log-normal distributions are mainly used for modelling unimodal hazard functions, the Weibull distribution is well-known for modelling monotonic hazard rates. The commonly applied estimation technique for this class of model is the Maximum Likelihood Estimator (MLE). However, previous studies have established the inadequacy of this technique for the exponentiated class of models, such as the exponentiated-Weibull model. Thus, in this thesis, we revisited the parameter estimation for the exponentiated-Weibull model class by introducing a new Bayesian technique called Variational Bayes. We considered the case of accelerated failure time (AFT) exponentiated-Weibull regression model with covariates. The AFT model was developed using two comparative studies based on real-life Lung cancer and simulated datasets. The AFT model parameters were estimated using the MLE, Bayesian Metropolis-Hasting and Variational Bayes procedure. The data calibration results showed that the exponentiated Weibull regression adequately describes the time-toevent data. In addition, the Variational Bayesian procedure was found to be the most efficient among the three estimation techniques considered.

ABSTRAK

Pengagihan Weibull, log-logistik dan log-normal mewakil I taburan penghujung yang sering digunakan dalam pemodelan data masa ke peristiwa. Walaupun taburan log-logistik dan log-normal digunakan terutamanya untuk pemodelan fungsi bahaya unimodal, taburan Weibull terkenal dengan pemodelan kadar bahaya monotonik. Teknik anggaran yang biasa digunakan untuk kelas model ini ialah Estimator Kemungkinan Maksimum (MLE). Walau bagaimanapun, kajian terdahulu telah menubuhkan kekurangan teknik Ini untuk kelas model eksponen, seperti model eksponen-Weibull. Oleh Itil, dalam tesis ini, kami menyemak semula anggaran parameter untuk kelas model eksponen-We Ibull dengan memperkenalkan teknik Bayesian baru yang dipanggil Variational Bayes. Kami menganggap kes masa kegagalan dipercepatkan (AFT) model ırgres I eksponen-We Ibull dengan covariat. Model AFT dibangunkan menggunakan dua kajian perbandingan berdasarkan kanser paru-paru kehidupan sebenar dan set data simulasi. Parameter model AFT dianggarkan menggunakan prosedur MLE, Bayesian Metropolis-Hasting dan Variational Bayes. Hasil penentukuran data menunjukkan bahawa regres I Weibull yang diekspasi menerangkan data masa ke acara dengan secukupnya. D I s¸amping I tu, prosedur Variational Bayesian didapati paling cekap di antara liga teknik anggaran yang dipertimbangkan.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF SYMBOLS AND ABBREVIATIONS

AFT	-	Accelerated Failure Time
Bes	-	Byes Estimations
BSEL	-	Balanced Square Error Loss
EW	-	Exponentiated Weibull
FGN	-	Federal Government of Nigeria
GTM	-	Grammar Translation Method
KL	-	Kullback -Leibler
LGEA	-	Local Government Education Authority
MCMC	-	Morkov Chain Monte-Carlos
MLE	-	Maximum Likehood Estimation
MOE	-	Ministry of Education
NCCTG	-	North Central Centre treatment Group
NCE	-	Nigerian Certificate of Education
NERDC	-	National Educational Research and Development Council
NPE	-	National Policy on Education
NTI	-	National Teachers Institute
SUBEB	-	State Universal Basic Education Board
T	-	Time
TK	-	Tremey and Kadane
UBEC	-	Universal Basic Education Commission

CHAPTER 1

INTRODUCTION

1.1 Background of the study

The Weibull, log-logistic and lognormal distributions are the most popular parametric time-to-event models (Kalbfleisch and Prentice, 2011). Due to modelling simplicity and common framework, these distributions are commonly applied in the time-to-event analysis. Common framework implies that the distributions share a similar log-location-scale family (Lawless, 2003) for statistical inference. Also, the ability to model day-to-day commonly seen survival data is often considered. The primary consideration is the ability to implement the procedures on standard off-shelf softwares readily. The commonly applied distributions for unimodal hazard shapes are log-logistic and log-normal, while the Weibull is often used when posed with monotone hazard functions (Lawless, 2003). Prentice (1973) as shown in Table 1.1 presents a typical structure of time-to-event data for the Lung cancer dataset; a full description of variable names can be found in Appendix B.

Ghinolfi *et al.* (2014) 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 extra shape parameter (Mudholkar *et al.*, 1995). The EW model simultaneously achieves flexibility and simplicity by accommodating both monotone (increasing and decreasing) and non-monotone (unimodal and bathtub shape) failure functions by introducing additional shape parameters. As this is a generalized approach, it can be used to confirm the adequacy of Weibull distribution, especially when the newly introduced shape parameter approaches unity.

Table 1.1: Survival of Patients with Advanced Lung Cancer from North Central Cancer Treatment Group (NCCTG) Minnesota, USA

Patient id	inst	time	status	age	sex	ph.ecog	ph.karno	pat.karno	meal.cal	wt.loss
1	3	306	2	74	1	1	90	100	1175	NA
2	3	455	2	68	1	0	90	90	1225	15
3	3	1010	1	56	1	0	90	90	NA	15
4	5	210	2	57	1	1	90	60	1150	11
5	1	883	2	60	1	0	100	90	NA	0
6	12	1022	1	74	1	1	50	80	513	0
7	7	310	2	68	2	2	70	60	384	10
8	11	361	2	71	2	2	60	80	538	1
9	1	218	2	53	1	1	70	80	825	16
10	7	166	2	61	1	2	70	70	271	34

Table 1.1 presents a typical structure of time-to-event data for the Lung cancer dataset according to Prentice (1973). The data were on survival of patients with advanced lung cancer from the North Central Cancer Treatment Group (NCCTG), Rochester, Minnesota, United States.

1.2 Problem statement

The three-parameter generalized gamma distribution (Stacy *et al.*, 1962) can also be used for modelling the four common types of hazard shapes. As suggested by Cox and Matheson (2014), 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 made by Mudholkar *et al.* (1995) in the analysis of survival data, and Pewsey *et al.* (2012) described likelihood-based inference for the class of power distributions that include the EW as a special case. The data sets on hazard times do not typically include only observed information on the time-to-event (T) and censoring status, but also on covariates. These in-turn posed the need to develop robust regression models to understand 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 Cancho *et al.* (1999), while a modification of the Log-Exponentiated-Weibull regression model within the Bayesian framework was proposed by Cancho *et al.* (2011) to specifically address cure rate.

In recent times, Khan (2018) provided an in-depth analysis of Accelerated Failure Time (AFT) EW regression models by using teh Maximum Likelihood Estimation (MLE) approach and Bayesian Markov Chain Monte-Carlo (MCMC) techniques. However, no study has evaluated the performance of the variational Bayes approximation for the EW regression in comparison with the most used techniques such as MLE and MCMC approaches. The variational Bayes approach is better than MCMC techniques under mild regularity conditions Blei *et al.* (2017). 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.

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. We developed a variational Bayesian (VB) regression using the EW 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 the MLE and Bayesian MCMC (Metropolis-Hasting techniques) using simulation and Lung cancer datasets used in Khan (2018).

1.3 Research objectives

The objectives of this research are to:

i develop the maximum likelihood estimation procedure for Accelerated Failure Time (AFT) EW survival regression model.

ii develop Metropolis-Hastings posterior sampling procedure for Accelerated Failure Time (AFT) EW survival regression model.

iii develop Variational Bayesian (VB) procedure for Accelerated Failure Time (AFT) EW survival regression model.

iv compare the performances of the procedure in objective 1 - 3 using simulated and lung cancer survival data.

1.4 Significance of study

This study is significant in the application of Bayesian and frequentist techniques for the estimation of monotone and non-monotone hazard functions of time to event data using EW distribution.

1.5 Scope and limitation

This study covers the survival analysis of monotone and non-monotone hazard functions of time-to-event data by using EW distribution. In addition, only right-censored survival datasets were used to test the distribution fitness and efficiency respectively.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a brief intoduction to time-to-event data, reviewed the exponential Weibull distribution regression model, Bayesian modelling and Bayesian computational techniques.

2.2 Time-to-event data

Multi-state models describe how individuals move between a finite number of states. The simplest example is the survival model with one transient state '0: alive' and one absorbing state '1: dead'. This model is characterized by the distribution of the survival time T, representing the time from a given origin (time 0) to the occurrence of the event 'death'. The distribution of T may be characterized by the distribution $F(t) = Prob(T \leq t)$ or equivalently, by the Survival function $S(t) = 1 - F(t) = Prob(T \leq t)$. It is seen that $S(t)$ and $F(t)$, respectively, correspond to the probabilities of being in state 0 or 1 at time t. If every individual is assumed to be in state 0 at the time 0 then $F(t)$ is also the transition probability from state 0 to state 1 for the time interval from 0 to t. In continuous time, the distribution of T may also be characterized by the hazard rate function $\alpha(t) = -(d \log S(t))/dt$, that is, $S(t) = \exp(-A(t))$ with $A(t) = \int_0^0 \alpha(u)du$ (Khan, 2018).

2.3 Exponential Weibull distribution regression model

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 parameters adds to the flexibility and has a better fit in modelling failure rates data (Pasari and Dikshit, 2015). The majority of these distributions have originated either from the domain of reliability engineering or biological sciences, where the specific interests are to estimate the elapsed time (time elapsed since failure) and the residual time (time remaining to failure) of a product.

(Pasari and Dikshit, 2018) reported that the Weibull distribution is the most popular and general probability model used in the 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, even with the flexibility and extensive application of the two-parameter or the three-parameter Weibull distribution, it still does not offer the non-monotonical failure rate shapes often observed in (medical sciences; survival analysis, cure rates etc.) or engineering (reliability, equipment failures).

In lung cancer survival analysis, there are three major stages: stage 1 (tumour development), stage 2 (organ damage or lung failure) and stage 3 (extension of tumour to other body parts) (Khan, 2018). 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 bath-tub failure shapes. Thus, monotonical hazard shape distribution is not adequate for such data type. These main drawbacks were the main reason the EW (EW) was proposed among several competing generalized distributions for modelling bath-tub shape time-to-event data.

The earlier development of EW can be traced to Mudholkar and Srivastava (1993) who introduced an extra shape parameter to the existing two-parameter Weibull distribution. The strength of the EW family is its ability to accommodate monotonical as well as non-monotonical failure functions, such as the unimodal-shaped and the bathtub-shaped ones (Mudholkar *et al.*, 1995). From the time it was proposed, the EW and its several extended versions have been applied to a wide area of practical applications, such as environmental flood data analysis (Nadarajah, 2009), bus motor failure (Mudholkar *et al.*, 1995), human mortality testing (Bebbington *et al.*, 2007) as well as survival analysis of head and neck cancer patients (Mudholkar *et al.*, 1995).

More recently, Mansour *et al.* (2020) worked on the generalization of EW distributions in Copula modelling. Their paper introduced and studied, a new flexible version of the EW model. The new model generalizes many new and well-known models. The new models were motivated through the introduction and studying of its density, which exhibits various important shapes such as decreasing, unimodal, bimodal, inverse-N shaped, right-skewed and left-skewed. The failure rate of the proposed model is very attractive to define many special models with different types of failure rates such as decreasing, increasing, unimodal, N-shaped and bathtub shaped hazard rates. The maximum likelihood method was employed to estimate the unknown parameters.

El-Din *et al.* (2020) worked on Bayesian inference on progressive-stress accelerated life testing for the EW distribution under progressive type-II censoring. In their paper, a progressive-stress accelerated life test (ALT) under progressive type-II censoring is considered. The cumulative exposure model is used when the lifetime of test units follows the EW distribution (EW). Moreover, the maximum likelihood estimates (MLEs) and Bayes Estimates (BEs) of the model parameters are obtained. Furthermore, the estimators' approximate and credible confidence intervals (CIs) are derived. The accuracy of the MLEs and BEs for the model parameters is investigated through simulation studies. Finally, the simulation studies are used to compare two different designs of the progressive-stress test (simple ramp-stress test and multiple ramp-stress test). In most cases, the BEs estimates were found to be more consistent and efficient than MLEs.

Cheema and Aslam (2020) studied Bayesian analysis for the 3-component mixture of EW distribution assuming non-informative priors. In their paper, Bayesian analysis of 3-component mixture model, of EW distribution under type-I right censoring scheme, is explored. With the help of non-informative (uniform and Jeffreys) priors and loss functions (e.g., squared error loss function, quadratic loss function, precautionary loss function and DeGroot loss function), Bayes estimators and posterior risks are derived. The Bayes estimators and posterior risks are observed as a function of the test termination time. A simulation study as well as a practical example, is given in this study.

2.4 Review of Bayesian modelling

The general procedure for estimating unknown parameter θ defined over a functional model f about data D using Bayesian techniques involves the specification of the data likelihood and prior (Gelman *et al.*, 2013). The maximum likelihood method is a frequentist-based procedure. Frequentist-based method of the opinion that unknown parameters are fixed while the dataset to be used in estimating them is random. Thus, it implies that the frequentists procedure focused on estimating the data likelihood $f(D)$, pretending that the parameter only depends on the data at hand.

Gelman *et al.* (2013) defined this type of probability $p(D|\theta)$ as the likelihood of the parameter after observing a set of random samples. On the other hand, the Bayesian expert believes the parameter θ is random while the data D is fixed. The important probability in Bayesian inference is the posterior distribution defined as $p(\theta|D)$. This probability distribution is

$$p(\theta|D) = \frac{p(D|\theta) \times p(\theta)}{p(D)} \tag{2.1}$$

where $p(\theta)$ is simply referred to as the prior distribution while $p(D)$ is the observed data samples (Lesaffre and Lawson, 2012). The denominator $p(D) = \int p(D|\theta) \times p(\theta)d\theta$ is usually referred to as marginal probability or normalizing constant which ensures $\int p(\theta|D) = 1$. This denominator is usually dropped in most Bayesian analysis such that the proportionality constant subsumes the equality sign. Thus, $p(\theta|D)$ is obtained using:

$$p(\theta|D) \propto p(D|\theta) \times p(\theta) \tag{2.2}$$

Equation (2.2) shows that the $p(\theta|D)$ can be expressed as the product data likelihood and prior distribution (Gelman *et al.*, 2013). In real-life analysis, Bayesian statistics involves the combination of prior distribution and data likelihood when drawing statistical inferences.

2.4.1 Prior distribution

According to Lesaffre and Lawson (2012), the prior distribution is regarded as an essential aspect of Bayesian analysis and in fact the main distinction between frequentist and Bayesian statistical inference. There are primarily two classes of prior distributions namely informative and non-informative priors (Gelman *et al.*, 2013; Lesaffre and Lawson, 2012; Lee, 2012).

The on-informative prior distributions also known as flat or vague or diffuse, contribute minimal information to posterior inference. According to Gelman *et al.* (2013), posterior distributions are known to be responsive to the type of prior distribution used. It was further recommended that non-informative should only be used when prior information is difficult to be elicited. Non-informative priors are sometimes known as objective Bayes (Rouder *et al.*, 2009).

On the other hand, informative prior also known as the subjective Bayesian approach is based on the process of elicitation of priors from related past data or expert opinions. Informative priors play significant roles in posterior distribution estimation and generally in Bayesian inference (Gelman *et al.*, 2013).

The conjugate prior, Jeffrey's prior and data-induced prior are other commonly prior in Bayesian inference (Gelman *et al.*, 2013; Ibrahim *et al.*, 2001). Robbins (1956) introduced data-induced prior and it was defined as Empirical Bayes (EB) method. The process of estimating the prior hyperparameters using the actual data is termed EB (Lee, 2012). Generally, EB methods are usually classified into two; parametric and non-parametric (Lee, 2012). The parametric EB techniques involves the estimation of prior hyperparameters using the denominator of the posterior formula i.e., $p(D)$.

Furthermore, within EW distribution, Ali and Kanani (2021) worked on Bayesian Methods to Estimate the Parameters of EW Distribution. Their paper introduced some properties of the EW distribution. Tierney and Lindely estimator methods are proposed to estimate all the unknown parameters (α, β, γ) of the EW distribution. The simulation procedure is used to generate some sample sizes and mean squares error (MSE) measure, and when we compared between the above two methods, we found that Tierney method has the less (MSE).

Yoon-sik and Sang-hoon (2020) worked on Bayesian Estimation of Inverted EW Distribution under Progressive Type II Censoring with Binomial Removal. Their paper conducted the experiment to estimate the three parameters of the inverted EW (IEW) distribution. The prior distribution of the model parameters is the gamma distribution. The tests are carried out under progressive Type II censoring with binomial removal.

Maximum likelihood estimates (MLEs), Bayes estimates are obtained by the Newton-Raphson algorithm and the Bayes methods. Also, we take the survival function and the hazard function of the IEW model. The Bayesian estimates are derived by the hybrid Markov chain Monte Carlo (MCMC) method using Gibbs sampling with Metropolis-Hastings algorithm and Tierney and Kadane (T-K) approximation. Bayes procedures have loss functions such as the squared error loss (SEL) and the balanced squared error loss (BSEL) function. To compare the proposed method results, some simulation experiments are performed with the different censoring schemes.

2.4.2 Bayesian computation

The Bayesian analysis procedures involves sampling from the posterior distribution. It is easy to achieve if the posterior distribution is easy to compute such that the posterior distribution has a closed-form expression. This case often occurs if the prior distribution is a conjugate of the posterior such that the prior and posterior distribution belong to the same family of distribution. However, as observed in most medical and engineering problems, posterior estimation is usually difficult to estimate. In most situations, the resulting way out is to use the Markov Chain Monte Carlo (MCMC) sampling algorithms (Lesaffre and Lawson, 2012; Lee, 2012). According to Lesaffre and Lawson (2012), popular MCMC procedures are the Gibbs sampler and Metropolis-Hastings algorithm.

2.5 Metropolis-Hastings algorithm (MH)

The Metropolis-Hastings algorithm is a widely used Markov chain Monte Carlo (MCMC) algorithm that allows generating samples from a target distribution that is difficult or impossible to directly sample from. The algorithm was first proposed by Nicholas Metropolis et al. in 1953 and later extended by Nicholas Hastings in 1970 (Hassan and Alharbi, 2023; Alexopoulos *et al.*, 2023; Du *et al.*, 2022; Wang and Nishi, 2022). The Metropolis-Hastings algorithm generates a sequence of samples from a target distribution using a proposal distribution, which is a distribution that is easy to sample from. The algorithm proceeds as follows:

i Initialize the algorithm with an initial state x_0.

ii For each iteration t, generate a candidate sample y from the proposal distribution $q(y|x_{t-1})$.

iii Compute the acceptance probability $\alpha(x_{t-1}, y)$, which is defined as:

$$\alpha(x_{t-1}, y) = \min\left(1, \frac{p(y)q(x_{t-1}|y)}{p(x_{t-1})q(y|x_{t-1})}\right) \tag{2.3}$$

where $p(x)$ is the target distribution and $q(x|y)$ is the proposal distribution.

iv Generate a uniform random variable u from the interval $[0, 1]$ and accept the candidate sample with probability $\alpha(x_{t-1}, y)$ if $u \leq \alpha(x_{t-1}, y)$. If the candidate sample is rejected, set $x_t = x_{t-1}$; otherwise, set $x_t = y$.

The algorithm generates a sequence of samples $x_1, x_2, \ldots, x_n$, which can be used to estimate the moments of the target distribution. The algorithm is guaranteed to converge to the target distribution as the number of iterations goes to infinity, provided that the proposal distribution satisfies certain conditions (Alexopoulos *et al.*, 2023; Du *et al.*, 2022; Wang and Nishi, 2022).

One important aspect of the algorithm is the choice of the proposal distribution. A good proposal distribution should have a high acceptance rate, which can be achieved by choosing a distribution that is similar to the target distribution. However, it should also be easy to sample from, which can be achieved by choosing a simple distribution that covers the support of the target distribution (Alexopoulos *et al.*, 2023; Du *et al.*, 2022; Wang and Nishi, 2022).

2.6 Variational Bayesian approximation

The aim of proposing the variational Bayesian inference is to provide an approximate approach for computing the conditional density of posterior parameters given the observed random samples. The main procedure involves solving the tasks using an optimization technique. Firstly, a family of distribution is defined for the desired unknown parameters which are to be estimated. The next step involves using the optimization procedure to determine the most plausibility values for the parameter set. The Kullback-Leibler (KL) divergence is the optimization criteria which minimum is sought when determining plausible values for the parameter of interest. The resulting fitted variational Bayesian density is then used as a reference point for the desired conditional density.

The earliest introduction of variational techniques for Bayesian inference can be traced to two different originating tracks. Anderson and Peterson (1987) is undoubtedly the first variational procedure developed for the neural network technique. The paper alongside contributions from statistical mechanics gave birth to different shades of variational inference procedures for other classes of models. In another work by Hinton and Van Camp (1993), another form of variational Bayesian inference was proposed for the neural network model. A significant linkage between variational Bayes and the expectation-maximization approach (Dempster *et al.*, 1977) was established by Neal and Hinton (1998). This exploration led to the recent development of several forms of variational Bayes algorithms for several models seen today (Blei *et al.*, 2017).

Recent studies on variational Bayesian inference focus on several aspects which are: involving Bayesian inference tasks that includes Big data (Blei *et al.*, 2017), using hybrid optimization methods for solving Kullback-Leibler (KL) divergence (which is usually subject to local minima); developing generic variational inference, algorithms that are easy to apply to a wide class of models; and increasing the accuracy of variational inference, e.g., by stretching the boundaries of Q while managing complexity in optimization .

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter introduces the methodology presented in this thesis. These components consist of the research process, research framework, the dataset to be used and presentation of the variational Bayes procedure for estimating the parameters of Exponentiated Weibull (EW) distribution.

3.2 Research process

The research process covers the following stages:

i Literature review analysis.

ii Model design: presentation of maximum likelihood procedure, Metropolis-Hastings approach and development of variational Bayes procedure for estimating the parameters of Exponentiated Weibull (EW) distribution.

iii Implementation of the variational Bayesian method for estimating the parameters of Exponentiated Weibull (EW) distribution.

iv Performance comparison between maximum likelihood, Metropolis-Hastings and variational Bayes procedure for estimating the parameters of Exponentiated Weibull (EW) distribution.

3.3 Research framework

The research framework is a structure that constitutes the overall flow process of the research as represented in Figure 3.1. The framework presents three essential stages, namely; (1) modelling time to event data using EW distribution, (2) AFT EW regression model parameter estimation using maximum likelihood estimator, Metropolis-Hastings and Variational Bayes approach, and (3) simulation strategy, real-life data application to lung cancer data and performance comparisons results.

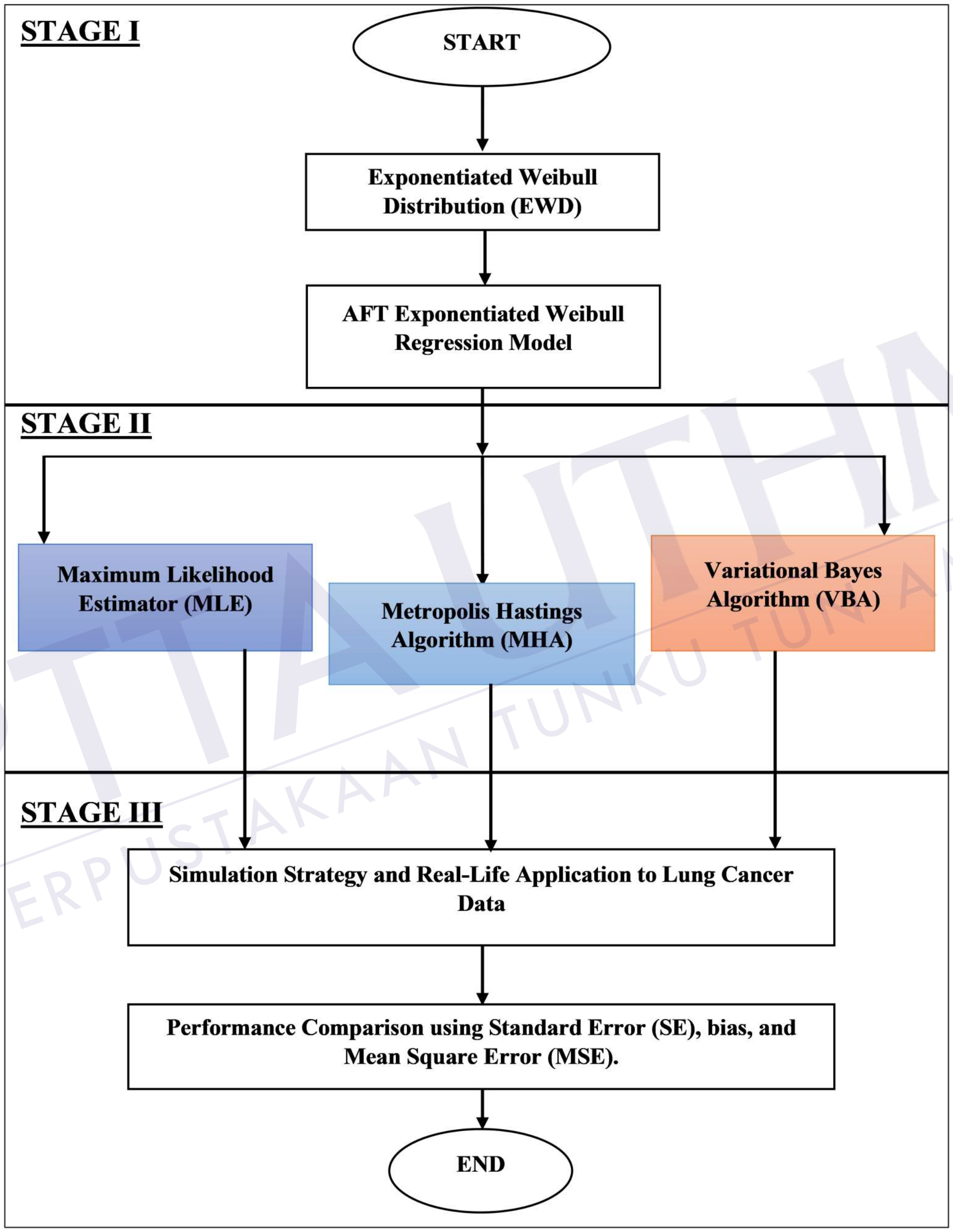

Figure 3.1: Research framework

3.4 Data descriptions

3.4.1 Simulated data

In this study, we simulated two covariates using an AFT regression framework: the first covariate will be continuous covariate (x_1) which will follow the standard normal distribution, and the other covariate (x_2) will be a binary that will follow a $Bernoulli(\pi = 0.5)$ distribution. The regression coefficient values are set to mimic the real-life dataset MLE estimated values for variables Karnofsky performance score (100=good) and treatment is $\boldsymbol{\theta} = [\theta^{*'} = (\theta_0^* = 3.5742793, \theta_1^* = 0.7547705, \theta_2^* = -0.1052200), \tau = 1/0.5868033, \gamma = 3.0987992]'$. The parameter vector corresponds to the covariate vector $x = (\mathbf{1}, x_1, x_2)$. The simulation process is made realistic using different censoring proportions simulated from an 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% which connote light to heavy censoring conditions are used.

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

$$Bias = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \tag{3.1}$$

$$Standard\ Error\ (SE) = \sqrt{\sum_{i=1}^{I} \frac{(\hat{\boldsymbol{\theta}} - \bar{\hat{\boldsymbol{\theta}}})^2}{I-1}} \tag{3.2}$$

$$Mean\ Square\ Error\ (MSE) = \sum_{i=1}^{I} \frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2}{I-1} \tag{3.3}$$

$$95\%CP = \sum_{i=1}^{I} \frac{\left(\hat{\boldsymbol{\theta}} - Z_{0.025} \times SE(\hat{\boldsymbol{\theta}}) < \boldsymbol{\theta}\right) \bigcap \left(\hat{\boldsymbol{\theta}} + Z_{0.975} \times SE(\hat{\boldsymbol{\theta}}) > \boldsymbol{\theta}\right)}{I} \tag{3.4}$$

where CP is coverage probability, $Z_{1-\alpha/2}$ 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 $n = 137$.

3.4.2 Real life lung cancer dataset

Prentice (1973) originally described a randomized clinical trial that involved 137 advanced lung cancer patients treated with a standard chemotherapy agent or a controlled drug (Khan, 2018). The time to event was recorded from the study inception for each patient. Nine patients were censored as their event times 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 types. The four different tumour cells are classified 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.

3.5 Exponentiated Weibull distribution

The Exponentiated Weibull distribution (EW) was developed by Mudholkar and Srivastava (1993) as an extension to the two-parameter Weibull distribution. The EW family distributions are designed to accommodate both non-monotonically and monotonically hazards. In most lifetime data analysis applications, the bathtub shape or downward bathtub shape hazard are often observed thus suggesting the EW's applicability for modelling hazards compared to standard Weibull distribution. This is where the EW plays a significant role in hazard modelling. The EW has two shape parameters and one scale parameter; thus, the probability density function (pdf) by

Khan (2018) takes the form:

$$f(t) = \alpha\beta\gamma(\beta t)^{\alpha-1}\big(1-\exp[-(\beta t)^{\alpha}]\big)^{\gamma-1}\exp[-(\beta t)^{\alpha}], \tag{3.5}$$

and the cumulative distribution function by Khan (2018) is:

$$F(t) = \big(1-\exp[-(\beta t)^{\alpha}]\big)^{\gamma}, \tag{3.6}$$

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

$$f(t) = \alpha\beta(\beta t)^{\alpha-1}\exp[-(\beta t)^{\alpha}], \tag{3.7}$$

The rth moment of the exponentiated Weibull distribution does not have a closed-form expression. However, Khan (2018) derived the median survival time as:

$$M(t) = \frac{1}{\beta}\left[-\log\left(1-0.5^{\frac{1}{\gamma}}\right)\right]^{\frac{1}{\alpha}}. \tag{3.8}$$

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

$$S(t) = 1-\big(1-\exp[-(\beta t)^{\alpha}]\big)^{\gamma}, \tag{3.9}$$

$$h(t) = \frac{\alpha\beta\gamma(\beta t)^{\alpha-1}\big(1-\exp[-(\beta t)^{\alpha}]\big)^{\gamma-1}\exp[-(\beta t)^{\alpha}]}{1-\big(1-\exp[-(\beta t)^{\alpha}]\big)^{\gamma}}, \tag{3.10}$$

$$H(t) = -\log\left\{1-\big(1-\exp[-(\beta t)^{\alpha}]\big)^{\gamma}\right\}, \tag{3.11}$$

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 $\alpha\gamma < 1$ (Khan, 2018).

3.6 Accelerate Failure Time exponentiated Weibull regression model

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

$$Y = \log T = x'\theta + \sigma W; \tag{3.12}$$

where $x = (x_1, x_2 \ldots, x_p)$ is 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 an 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. However, p is assumed the same for everyone.

This model has an accelerated life interpretation. This formulation views 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:

$$f(y) = \frac{\gamma}{\tau}\left(1 - \exp\left\{-\exp\left[\frac{(y-\mu)}{\tau}\right]\right\}\right)^{\gamma-1} \exp\left\{\frac{(y-\mu)}{\tau} - \exp\left[\frac{(y-\mu)}{\tau}\right]\right\} \tag{3.13}$$

and

$$S(y) = 1 - \left(1 - \exp\left\{-\exp\left[\frac{(y-\mu)}{\tau}\right]\right\}\right)^{\gamma} \tag{3.14}$$

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 $t\exp(x'\theta)$ (Jibril *et al.*, 2023). Formally, $S(t; x) = S_0(t\exp(x'\theta)$, where $S_0(\Delta)$ is the baseline survivor function. This formula refers to the fact that the covariates act multiplicatively on time so that their effect is to accelerate (or decelerate) the time to failure (Jibril *et al.*, 2023). If we start with the exponentiated Weibull baseline survivor function, we get

$$S(t) = 1 - \big(1 - \exp\{-[\beta t \exp(x'\theta)]^{\alpha}\}\big)^{\gamma} = 1 - \big(1 - \exp[-[\beta^{*} t]^{\alpha}]\big)^{\gamma} \tag{3.15}$$

which is also an exponentiated Weibull survivor function with $\beta^{*} = \beta \exp(x'\theta)$. It 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 $x = 0$, the survivor function of T_0 is of the form $S_0(.)$. Then, $T_0 = T \exp(x'\theta)$ from (3.15). By taking the logarithm of both sides of Equation (3.15), we get

$$Y = \theta_0 - x'\theta + \tau W; \tag{3.16}$$

where $\theta_0 = -\log \beta$, $\tau = \alpha^{-1}$ and $Y = \log T$ which follows (3.13) and $\alpha = \theta_0 - x'\theta$ and $W = (\log T_0 - \theta_0)/\tau$ is the random error component which is distributed as

$$f(w) = \gamma\big[1 - \exp\big(-e^{w}\big)\big]^{\gamma-1} \exp\big(w - e^{w}\big), -\infty < w < \infty. \tag{3.17}$$

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

$$Y = x^{*'}\theta^{*} + \tau W. \tag{3.18}$$

3.7 Maximum likelihood estimation of right-censored accelerated failure time exponentiated Weibull 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 occurred at $t = t_0$) or $\delta_i = 0$ (if the event occurred at $t > t_0$), respectively. The log-likelihood function of the exponentiated Weibull regression model by Jibril *et al.* (2023) can be written as

$$l(\boldsymbol{\theta}) = r\log\gamma - r\log\tau + \sum_{i=1}^{n}\delta_i[(\gamma-1)\log a_i + (w_i - e^{w_i})] + \sum_{i=1}^{n}(1-\delta_i)\log(1-a_i^{\gamma}) \tag{3.19}$$

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

$$g_i = \frac{\partial l(\boldsymbol{\theta})}{\partial w_i} \tag{3.20}$$

$$= \frac{\partial}{\partial w_i}\{\sum_{i=1}^{n}\delta_i[(\gamma-1)\log a_i + (w_i - e^{w_i})] + \sum_{i=1}^{n}(1-\delta_i)\log(1-a_i^{\gamma})\} \tag{3.21}$$

$$= -(\gamma-1)\left(\frac{\delta_i b_i}{a_i}\right) + \delta_i[1+\log(1-a_i)] + \gamma\left[\frac{(1-\delta_i)b_i}{a_i}\right]\left(\frac{a_i^{\gamma}}{1-a_i^{\gamma}}\right). \tag{3.22}$$

Subsequently, the score functions of other parameters are

$$\frac{\partial l(\boldsymbol{\theta})}{\partial\tau} = -\frac{r}{\tau} - \frac{1}{\tau}\sum_{i=1}^{n} g_i w_i \tag{3.23}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial\gamma} = \frac{r}{\gamma} + \sum_{i=1}^{n}\delta_i\log a_i - \sum_{i=1}^{n}(1-\delta_i)\left(\frac{a_i^{\gamma}\log a_i}{1-\log a_i}\right) \tag{3.24}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial\theta_j} = -\frac{1}{\tau}\sum_{i=1}^{n} g_i x_{ij}, j = 0, 1, 2, \ldots, p. \tag{3.25}$$

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

3.8 Metropolis-Hastings Approach for accelerated failure time exponentiated Weibull regression model

In this section, we discussed the Bayesian estimation of the parameters of EW. We use the uninformative Uniform (c, d) prior since 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:

$$f(\tau, \gamma, \theta^{*'})|c_1, d_1; c_2, d_2; c_3, d_3) = \prod_{k=1}^{3}(d_k - c_k)^{-1} \tag{3.26}$$

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:

$$f(\boldsymbol{\theta}|y) = L(\boldsymbol{\theta}) \times \prod_{k=1}^{3}(d_k - c_k)^{-1} \tag{3.27}$$

The posterior distribution in (3.27) does not have a closed-form as it is an approximate distribution since the marginal distribution that ensures its scale to one has been dropped. One of the ways of sampling from this distribution is by using the Metropolis-Hastings algorithm (Jibril *et al.*, 2023). The metropolis-hastings procedure for the EWD AFT regression model is:

i Initialize Θ^0 such that $p(\Theta^0|y) > 0$.

ii For $i = 1, 2, \ldots$

iii Take a random sample $\tilde{\Theta}$ from a preferred proposal distribution (Preferably lognormal distribution).

iv Compute the accept/reject or moving probability by; taking a random sample $U \sim U(0, 1)$ and computing

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

3.9 Variational Bayes approach for accelerated failure time exponentiated Weibull regression model

Suppose we let $x = x_{1:n}$ represent a collection of observed variables and $z = z_{1:m}$ represents collection of latent variables, with joint density function $p(z, x)$. As explained 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 conditional density, the point and interval estimates of the latent variables can be estimated (Jibril *et al.*, 2023). The conditional density is often presented as

$$p(z|x) = \frac{p(z,x)}{p(x)} \tag{3.28}$$

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

$$p(x) = \int p(z,x)dz. \tag{3.29}$$

In most models, this marginal density is usually not available or computationally expensive. The marginal density is required to calculate the conditional from the joint density and thus the main reason variational inference is difficult (Jibril *et al.*, 2023).

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 analyses. The parameters are also local to each observed data point (Jibril *et al.*, 2023).

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

$$p(\boldsymbol{\theta}|y,x) = \frac{L(\boldsymbol{\theta}|y,x)p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|y,x)p(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{3.30}$$

.

By variational inference, we want to approximate $p(\boldsymbol{\theta}|y,x)$ in 3.30 with a $q(\boldsymbol{\theta})$ by constructing the equality

$$\ln \int L(\boldsymbol{\theta}|y,x)p(\boldsymbol{\theta}) = \int q(\boldsymbol{\theta}) \ln \frac{L(\boldsymbol{\theta}|y,x)p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} + \int q(\boldsymbol{\theta}) \ln \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|y,x)} d\boldsymbol{\theta} \tag{3.31}$$

.

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

$$q(\boldsymbol{\theta}) = q(\gamma)q(\tau)q(\theta^{*'}) \tag{3.32}$$

.

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

$$\mathcal{L} = \int q(\boldsymbol{\theta}) \ln L(\boldsymbol{\theta}|y, x) d\boldsymbol{\theta} - \int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{3.33}$$

.

$$\begin{aligned} \mathcal{L} = \int q(\gamma)q(\tau)q(\theta^{*'}) \Bigg(r \log \gamma - r \log \tau + \sum_{i=1}^{n} \delta_i [(\gamma - 1) \log a_i + (w_i - e^{w_i})] \\ + \sum_{i=1}^{n} (1 - \delta_i) \log(1 - a_i^{\gamma}) \Bigg) d\gamma d\tau d\theta^{*'} \\ - \int q(\gamma)q(\tau)q(\theta^{*'}) \ln q(\gamma)q(\tau)q(\theta^{*'}) d\gamma d\tau d\theta^{*'} \end{aligned} \tag{3.34}$$

Equation (3.34) is iterated until convergence is achieved (Jibril *et al.*, 2023).

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, 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 Deviation, Mean Square Error and Coverage probability.

4.2 Simulation study results

This section presents the results of the simulation study described in chapter three. Table 4.1 results show that the estimates returned using the VB method is more consistent with the true value when compared to the other two methods. The estimates using MH are better than MLE estimates in terms of bias and consistency. Overall, the MLE estimate is inconsistent with the true value at high censoring proportions. Table 4.1 presents the simulation results for varying censoring proportions at fixed sample size n and replication set to be 200.

The results in Table 4.1 underscore the importance of choosing the appropriate estimation method when dealing with censored data. The VB method stands out as a robust and consistent choice, even under high censoring proportions. The MH method also offers advantages over MLE, emphasizing its suitability in scenarios involving incomplete data. These findings equip researchers with valuable insights into selecting the most suitable estimation method based on the characteristics of their data, ensuring the accuracy and reliability of their analyses.

Table 4.1: Simulation results for the estimates and bias over various censoring proportions 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**

Table 4.2 presents the three methods' standard error/ standard deviation and Mean Square Error (MSE). 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. Table 4.2 results emphasize that the VB method stands out as the most efficient choice for estimating the parameter γ across varying levels of data censoring.

Table 4.2: Simulation results for the standard error (SE) and mean square error (MSE) over various censoring proportions 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**

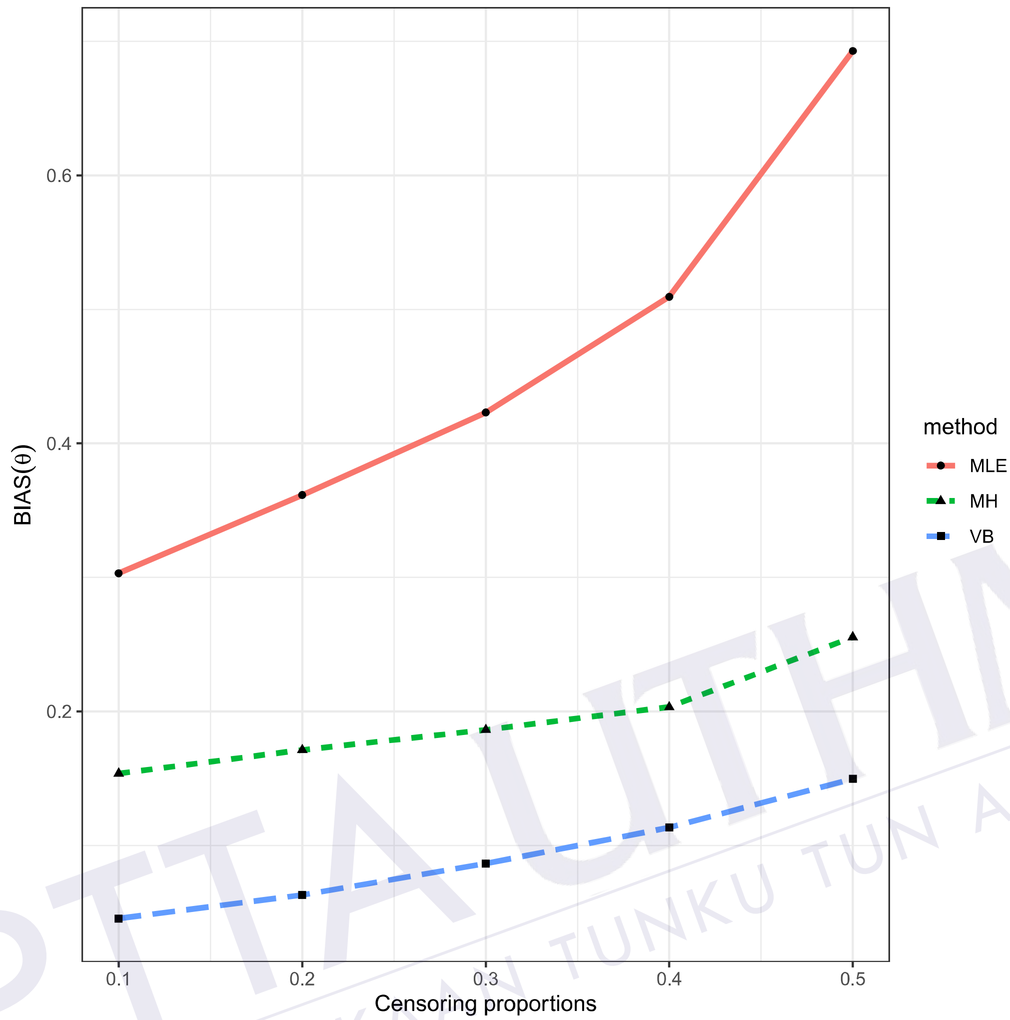

Figure 4.1: Average Bias of the estimates at varying censoring proportion.

Figure 4.1 shows that increasing the censoring proportion increases the biasedness of the estimates especially for MLE. While the VB and MH exhibited some form of robustness to increase in censoring proportion, the MLE estimates are not robust, which suggests that they should not be used with high censored data. Similarly, Figure 4.2 shows that the MLE estimates are inefficient for low to high censoring proportions. The two Bayesian approaches MH and VB are highly efficient and robust to low through high censoring proportion. The combined effects of consistency and efficiency were measured using MSE. Similar behaviours as in variance of the estimates were observed for MSE in Figure 4.3. Again, the most efficient, consistent and robust estimates are VB estimates. Robustness is determined based on the ability of the estimator to withstand some level censoring by maintaining the same efficiency. The maximum reasonable censoring is 0.5, becuase censoring implies missing data, if censoring is greater than 0.5, it implies more than 50% of the data is missing. Thus, the standadrd maximum threshold for censoring is 0.5.

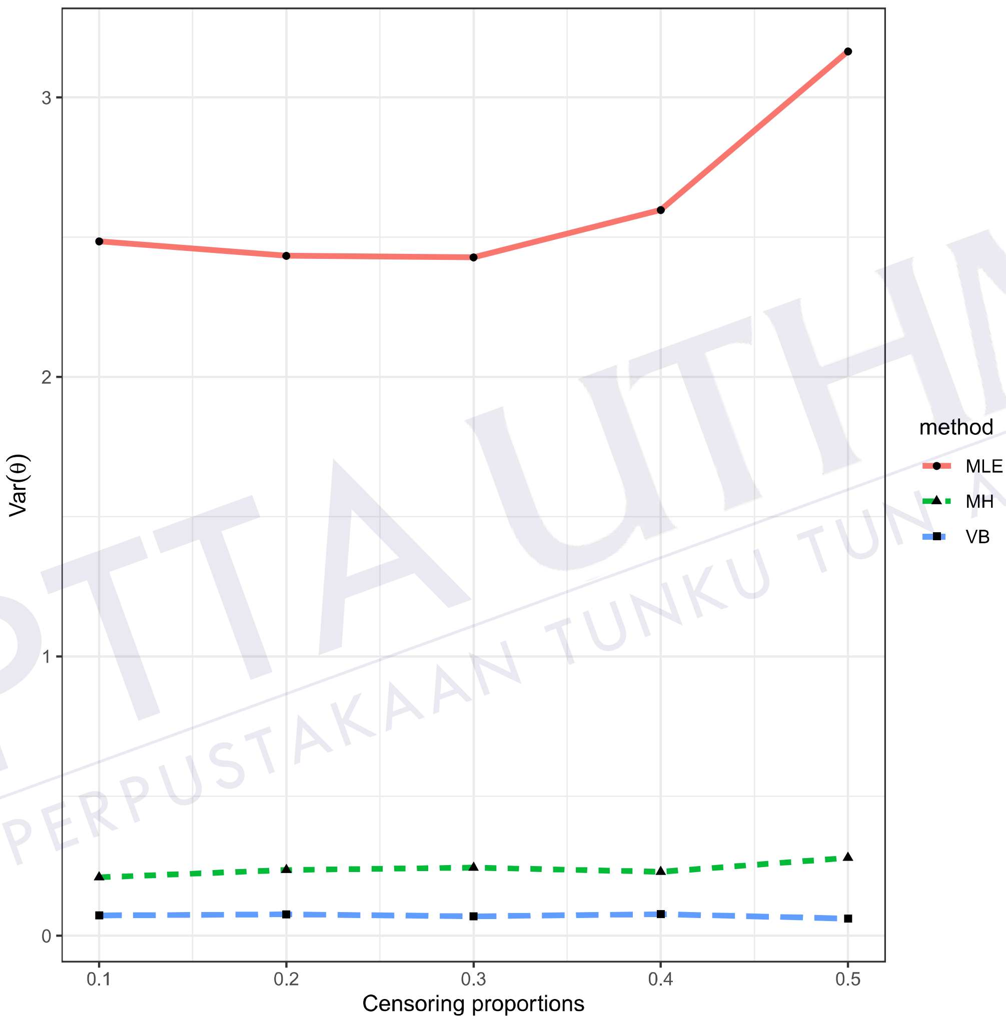

Figure 4.2: Average variance of the estimates at varying censoring proportions.

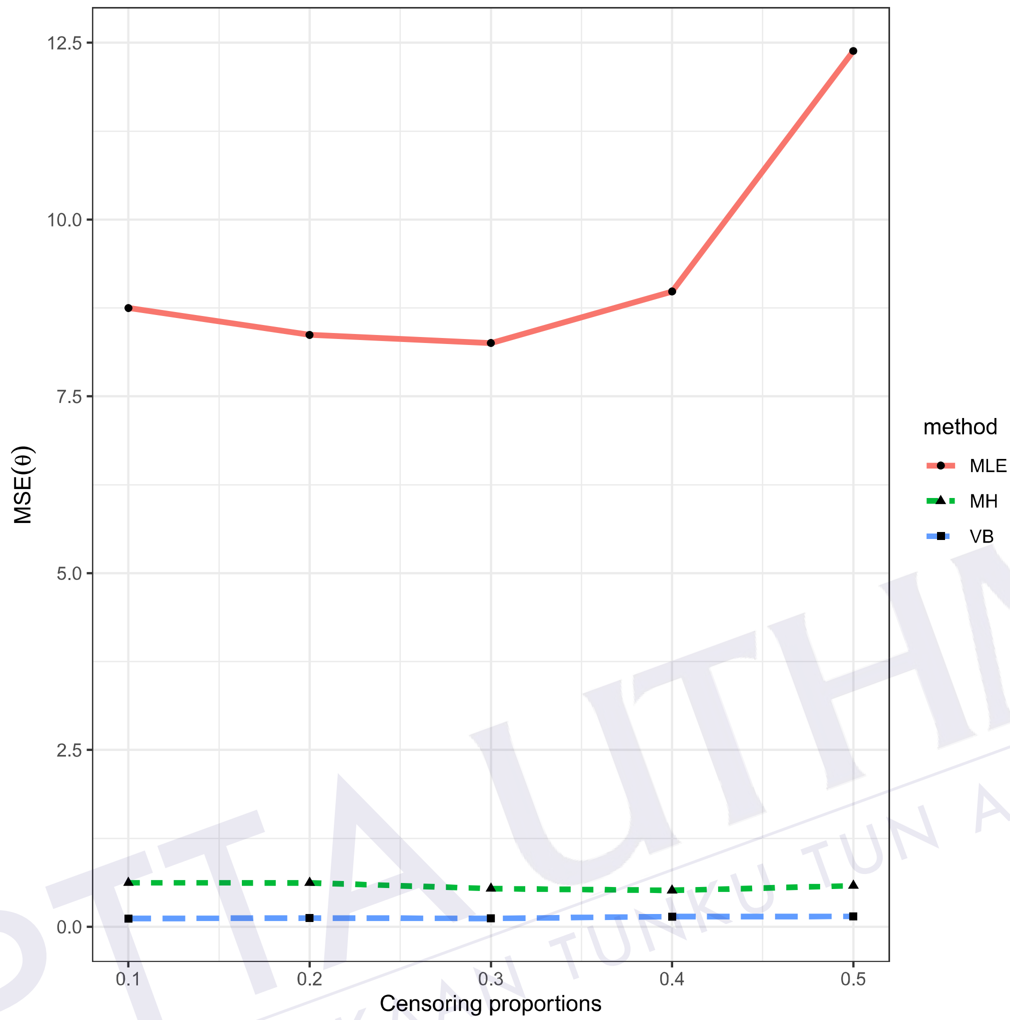

Figure 4.3: Average Mean Square Error of the estimates at varying censoring proportions.

Figure 4.4 presents the result for the 95% coverage probability. The expected behaviour is that the estimates returned values that fall 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 an 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%. It implies that approximately 95% of time MLE produces estimates that conform 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.

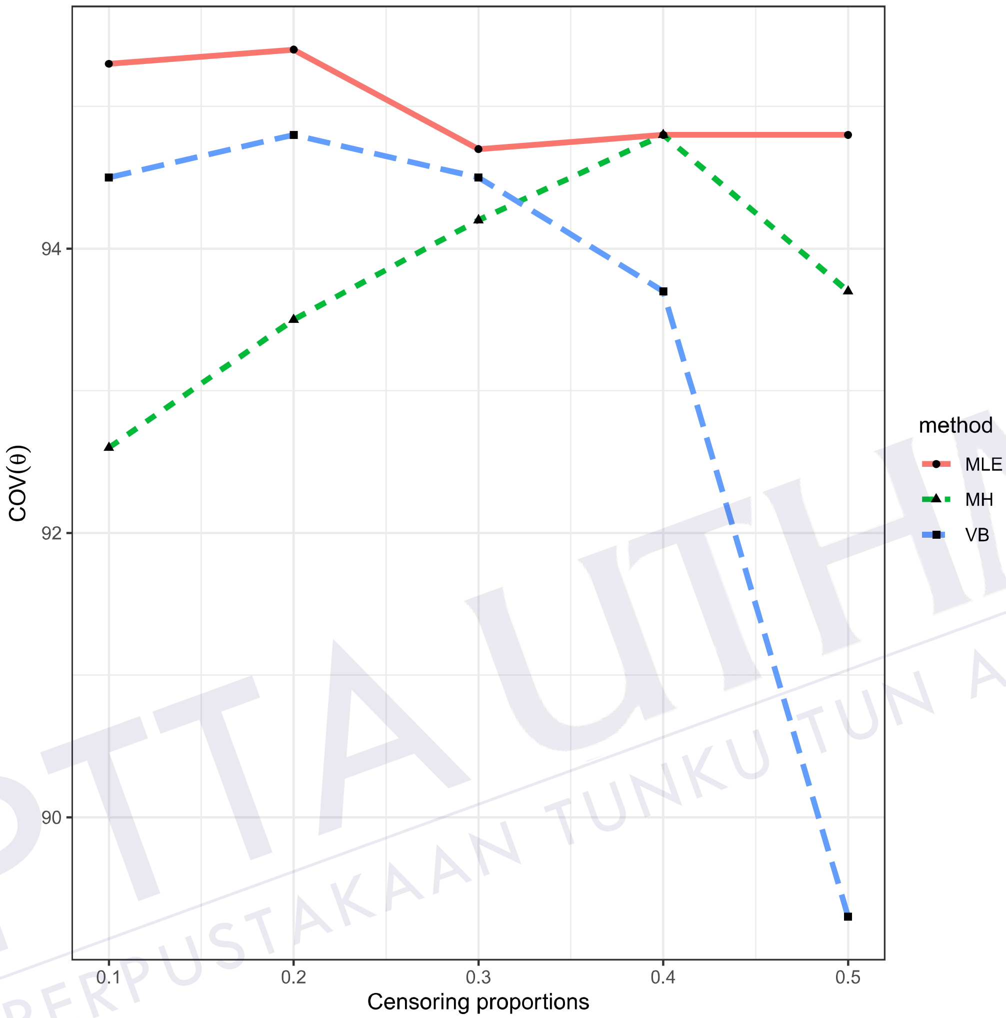

Figure 4.4: Average Coverage Probability at varying censoring proportions.

4.3 Lung cancer survival data analysis

This section presents the analysis of the lung cancer dataset described in chapter three. There were a total of 137 patiemts in the dataset. The time to event was recorded from the study inception for each patient. Nine patients were censored as their event times 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 types. The four different tumour cells are classified squamous, small, adeno and large. The other variables considered are performance status (karno), months between diagnosis and entry into the study (diagtime), age, and a history of previous therapy for lung cancer (prior).

Table 4.3: 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 4.4: 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

Table 4.3 presents the three methods' estimates and standard deviation (SD) or error (SE). 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 4.4 showed that VB and MH methods returned more significant estimates than MLE at 5% level of significance.

4.4 Chapter summary

In this chapter, 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 Khan (2018) was compared to the VB estimate using simulated and real-life data. Simulated results revealed that the VB estimates are more efficient than 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.

CHAPTER 5

CONCLUSIONS AND FUTURE WORKS

5.1 Research summary

This thesis specifically focused on parametric regression models requiring a distributional assumption for T in the presence of covariates vector x. It was aimed to develop a maximum likelihood estimator (MLE), metropolis-hastings (MH) algorithm and variational Bayesian (VB) for the Exponentiated Weibull (EW) accelerated failure time regression methodology. The EW distribution was used based on its apriori generalizability in accommodating both monotone and non-monotone hazard/failure functions while doing it an insignificant cost of only estimating one extra parameter. Thus, the objectives were to evaluate the efficiency and consistency performances of Maximum Likelihood Estimation (MLE) and Bayesian Markov Chain Monte-Carlo (MCMC) via the Metropolis-Hastings (MH) algorithm. In addition, the performance of the VB method is evaluated by comparing it with MLE and Bayesian MH using simulated and lung cancer datasets.

The main objectives were to develop MLE, MH and VB methods to estimate the parameters of the AFT EW regression model. The consistency and efficiency evaluation performance results using both the simulated and real-life datasets revealed that the MLE procedure is less efficient and consistent when compared to the VB and MH, especially at high censoring proportions. These results are replicated using the real-life Lung cancer dataset.

Based on the objective to evaluate the MH procedure in terms of consistency and efficiency using both simulated and real-life datasets, different results revealed that the MH is less efficient and consistent when compared to the VB, though better than MLE for the simulated dataset. However, the MH procedure competes favourably with the VB for the real-life Lung cancer dataset.

5.2 Research contribution

The following are the significant contribution of this research to the field of statistics and survival analysis;

i With right-censored data, a new consistent and efficient Bayesian estimation technique was developed for the accelerated failure time exponentiated Weibull regression survival model.

ii The new method named Variational Bayesian (VB) was tested on both simulated and real-life right-censored datasets.

iii The approach is flexible as one does not need to be a Bayesian expert to use it.

5.3 Future work

The VB procedure is flexible, but we observed some flaws while estimating categorical predictors such as treatment type and type of cancer. Thus, a re-examination of the strength of VB should be thoroughly conducted to identify the low efficiency in the estimation of categorical predictors.

5.4 Conclusion

The various analyses conducted in the thesis have established the applicability of the VB approach for estimating the parameters of the EW distribution with the right-censored dataset. In conclusion, the VB procedure presented in this thesis has been found to competes favourably with the existing methods in both simulated and real-life right-censored datasets. The technique was found to be better than the competing Bayesian MCMC (MH) procedure as well as the frequentist MLE method. The main strength of VB is in the estimation of shape parameters of the exponentiated Weibull distribution.

REFERENCES

Alexopoulos, A., Dellaportas, P. and Titsias, M. K. (2023). Variance reduction for Metropolis–Hastings samplers. *Statistics and Computing*. 33(1), 6.

Ali, A. H. and Kanani, I. H. A. (2021). Bayesian Methods to Estimate the Parameters of Exponentiated Weibull Distribution. In *Journal of Physics: Conference Series*, vol. 1818. IOP Publishing, 012143.

Anderson, J. R. and Peterson, C. (1987). A mean field theory learning algorithm for neural networks. *Complex Systems*. 1, 995–1019.

Bebbington, M., Lai, C.-D. and Zitikis, R. (2007). A flexible Weibull extension. *Reliability Engineering & System Safety*. 92(6), 719–726.

Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*. 112(518), 859–877.

Cancho, V., Bolfarine, H. and Achcar, J. (1999). A Bayesian analysis for the exponentiated-Weibull distribution. *Journal of Applied Statistical Science*. 8(4), 227–242.

Cancho, V. G., Rodrigues, J. and de Castro, M. (2011). A flexible model for survival data with a cure rate: a Bayesian approach. *Journal of Applied Statistics*. 38(1), 57–70.

Cheema, A. N. and Aslam, M. (2020). Bayesian analysis for 3-component mixture of exponentiated weibull distribution assuming non-informative priors. *Journal of Statistical Computation and Simulation*. 90(4), 586–605.

Cox, C. and Matheson, M. (2014). A comparison of the generalized gamma and exponentiated Weibull distributions. *Statistics in medicine*. 33(21), 3772–3780.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*. 39(1), 1–22.

Du, X., Du, C., Radolinski, J., Wang, Q. and Jian, J. (2022). Metropolis-hastings Markov Chain Monte Carlo approach to simulate van genuchten model parameters for soil water retention curve. *Water*. 14(12), 1968.

El-Din, M. M., Amein, M., Abd El-Raheem, A., Hafez, E. and Riad, F. H. (2020). Bayesian inference on progressive-stress accelerated life testing for the exponentiated Weibull distribution under progressive type-II censoring. *J. Stat. Appl. Pro. Lett.* 7, 109–126.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian data analysis*. Florida: CRC press.

Ghinolfi, D., Marti, J., De Simone, P., Lai, Q., Pezzati, D., Coletti, L., Tartaglia, D., Catalano, G., Tincani, G., Carrai, P. *et al.* (2014). Use of octogenarian donors for liver transplantation: a survival analysis. *American Journal of Transplantation*. 14(9), 2062–2071.

Hassan, A. S. and Alharbi, R. S. (2023). Different estimation methods for the unit inverse exponentiated weibull distribution. *Communications for Statistical Applications and Methods*. 30(2), 191–213.

Hinton, G. E. and Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the sixth annual conference on Computational learning theory*. 5–13.

Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*. 57(2), 383–388.

Jibril, A., Abdullah, M. A. A. and Olaniran, O. R. (2023). Variational Bayesian Inference for Exponentiated Weibull Right Censored Survival Data. *Statistics,* *Optimization & Information Computing*. -(-), 0–13.

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

Khan, S. A. (2018). Exponentiated Weibull regression for time-to-event data. *Lifetime data analysis*. 24(2), 328–354.

Lawless, J. (2003). Event history analysis and longitudinal surveys. *Analysis of Survey data*, 221–243.

Lee, P. M. (2012). *Bayesian statistics: an introduction*. New York City: John Wiley & Sons.

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian biostatistics*. John Wiley & Sons.

Mansour, M., Korkmaz, M. C., Ali, M. M., Yousof, H. M., Ansari, S. and Ibrahim, M. (2020). A generalization of the exponentiated Weibull model with properties, Copula and application. *Eurasian Bulletin of Mathematics*. 3(2), 84–102.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*. 42(2), 299–302.

Mudholkar, G. S., Srivastava, D. K. and Freimer, M. (1995). The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*. 37(4), 436–445.

Nadarajah, S. (2009). Bathtub-shaped failure rate functions. *Quality & Quantity*. 43(5), 855–863.

Neal, R. M. and Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*. (pp. 355–368). Springer.

Pasari, S. and Dikshit, O. (2015). Earthquake interevent time distribution in Kachchh, Northwestern India. *Earth, Planets and Space*. 67(1), 129.

Pasari, S. and Dikshit, O. (2018). Stochastic earthquake interevent time modeling from exponentiated Weibull distributions. *Natural hazards*. 90(2), 823–842.

Pewsey, A., Gómez, H. W. and Bolfarine, H. (2012). Likelihood-based inference for power distributions. *Test*. 21(4), 775–789.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*. 60(2), 279–288.

Robbins, H. (1956). *An empirical Bayes approach to statistics*. Technical report. COLUMBIA UNIVERSITY New York City United States.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*. 16(2), 225–237.

Stacy, E. W. *et al.* (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*. 33(3), 1187–1192.

Wang, Z. and Nishi, Y. (2022). Stochastic model for simulating levels of polychlorinated biphenyls in small tuna and planktons using Metropolis–Hastings algorithm. *Ecotoxicology and Environmental Safety*. 242, 113941.

Yoon-sik, J. and Sang-hoon, S. (2020). Bayesian Estimation of Inverted Exponentiated Weibull Distribution under Progressive Type II Censoring with Binomial Removal. *Journal of The Korean Data Analysis Society*. 22(1), 1–12.

VITA

The author was born in 24 th January 1974 in Kaduna state of Nigeria. He went to LEA Badarawa primary school in Kaduna, Kaduna state of Nigeria for education, went to Government Secondary School, Kurmin Mashi (Now Command Day Secondary School), opposite Nigeria Defence Academy (NDA), Kaduna for Secondary School education and obtained Secondary School Certificate of Certificate (SSCE). He pursued his National diploma (ND) at Federal Polytechnic Kaduna and graduated with the ND certificate in Statistics, 2002. He also pursuit Higher National Diploma (HND) Statistics in 2003 at Federal Polytechnic Kaduna, where he earns HND certificate in statistics, 2006. He worked as assistant statistical officer at National Primary Education Commission (NPEC) Headquarters, Kaduna from 1994 - 2000, He moved to Universal Basic Education Commission (UBEC) Headquarters, Abuja when NPEC metamorphosed to UBEC in 2000. He was promoted to the rank of Higher Statistical Officer in 2007. Subsequently, he have being promoted to a different rank until January, 2023 when the Commission after successful passed senior staff examination, he was promoted to the rank of Chief Statistical officer, the position he is up-till-date. Mr. Jibril Abubakar attended Universiti tun Hussein Onn Malaysia (UTHM), Malaysia for Master Degree of Science (MSC) in survival Analysis