VŠB - TECHNICAL UNIVERSITY OF OSTRAVA

DOCTORAL THESIS

Bayes approach to explore the mixture failure rate models

Study Program: Computer Science, Communication Technology and Applied Mathematics
Field of study: Computational and Applied Mathematics
Ph.D. Student: Tien Thanh THACH
Supervisor: Prof. Ing. Radim BRIS, CSc.

A thesis submitted
for the degree of Doctor of Philosophy

in the

Faculty of Electrical Engineering and Computer Science
Department of Applied Mathematics

Ostrava 2019

Declaration of Authorship

I, Tien Thanh THACH, declare that this thesis titled, "Bayes approach to explore the mixture failure rate models" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

Abstract

This thesis has two folds: Firstly, designing mixture failure rate functions by combing few other existing failure rate functions to obtain desirable mixture failure rate functions. The first proposed mixture failure rate is the non-linear failure rate. This failure rate is a mixture of the exponential and Weibull failure rate functions. It was designed for modeling data sets in which failures result from both random shock and wear out or modeling a series system with two components, where one component follows an exponential distribution and the other follows a Weibull distribution. The second proposed mixture failure rate is the additive Chen-Weibull failure rate. This failure rate is considered a mixture of the Chen and Weibull failure rates. It is decided for modeling lifetime data with flexible failure rate including bathtub-shaped failure rate. The final proposed mixture failure rate is the improvement of new modified Weibull failure rate. This failure rate is a mixture of the Weibull and modified Weibull failure rates. It is also decided for modeling lifetime data with flexible failure rate including bathtub-shaped failure rate. The superiority of the proposed models have been demonstrated by fitting to many well-known lifetime data sets. And secondly, applying modern methods and techniques from Bayesian statistics for analyzing failure time distributions which result from those mixture failure rate functions.

Acknowledgements

Firstly I would like to express my special thanks of gratitude to:

- Prof. Ing. Radim Bris, CSc., for his guidance, patience and support.

- Prof. Frank P. A. Coolen and Prof. Petr Volf, for their guidance and suggestion which helped me a lot in my research and I came to know about so many new things.

Secondly I would also like to thank my parents and friends for their love and support.

Contents

List of Figures

List of Tables

List of Abbreviations

T: The random variable denoting lifetime of the system
$h(t)$:   The failure rate function
$f(t)$:   The probability density function
$F(t)$:   The cumulative distribution function
$R(t)$:   The reliability/survival function
$H(t)$:   The cumulative failure rate function
MTTF: Mean Time To Failure

Chapter 1

Introduction

1.1 The Problem

Before discussing the main problem of the thesis, I explain what is the failure rate function and why it is important. Let T be the lifetime random variable with $F(t), f(t)$ and $R(t)$ being its cumulative distribution function (CDF), probability density function (PDF) and reliability/survival function, respectively. Failure rate function is one way that uses to specify the properties of the random variable T. Suppose that we are interested in the probability that a system will fail in the time interval $(t, t + \Delta t]$ when we know that the system is working at time t. That probability is defined as

$$\mathbb{P}(t < T \leq t + \Delta t | T > t) = \frac{\mathbb{P}(t < T \leq t + \Delta t)}{\mathbb{P}(T > t)} = \frac{F(t + \Delta t) - F(t)}{R(t)} \tag{1.1}$$

To define the failure rate function, we divide this probability by the length of the interval, Δt, and let $\Delta t \to 0$. This gives

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{\mathbb{P}(t < T \leq t + \Delta t | T > t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{R(t)} \\ &= \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)} \end{aligned} \tag{1.2}$$

Therefore, when Δt is sufficiently small,

$$\mathbb{P}(t < T \leq t + \Delta t | T > t) \approx h(t)\Delta t \tag{1.3}$$

$h(t)\Delta t$ gives (approximately) the probability of failure in $(t, t + \Delta t]$ given that the system has survived until time t. The failure rate function can be considered as the system's propensity to fail in the next short interval of time, given that it is working at time t. Fig. 1.1 shows four of the most common types of failure rate function which have been described by Hamada et al. [27].

1. *Increasing failure rate (IFR)*: the instantaneous failure rate increases as a function of time. We expect to see an increasing number of failures for a given period of time.

2. *Decreasing failure rate (DFR)*: the instantaneous failure rate decreases as a function of time. We expect to see the decreasing number of failures for a given period of time.

3. *Constant failure rate (CFR)*: the instantaneous failure rate is constant for the observed lifetime. We expect to see a relatively constant number of failures for a given period of time.
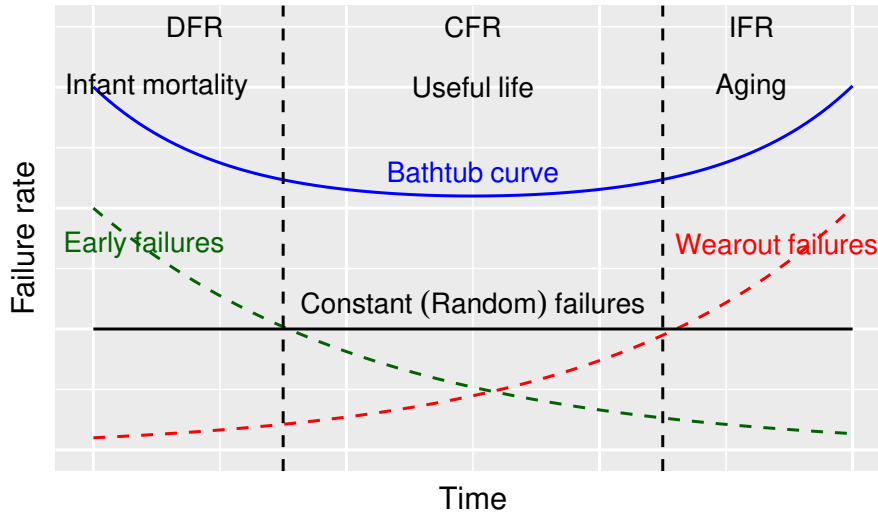
Figure 1.1: Four different classifications of failure rate functions

4. *Bathtub failure rate (BFR)*: the instantaneous failure rate begins high because of early failures ("infant mortality" or "burn-in" failures), levels off for a period of time ("useful life"), and then increases ("wearout" or "aging" failures).

In some cases, the failure rate function can be unimodal (unique mode), i.e. it has an upside-down bathtub shape.

The reliability function examines the chance that breakdowns (of people, of experimental units, of computer systems,...) occur beyond a given point in time, while the failure rate function monitors the lifetime of a component across the support of the lifetime distribution. We know that survival distributions may be equivalently specified by: the PDF, the CDF, the reliability/survival function and the failure rate function. This means that if one of these representations is specified, then the others may be derived from it. Therefore, a discussion of lifetime distributions needs only to concentrate on one of these expressions. The failure rate function is often modeled and specified because of its direct interpretation as imminent risk. It may even help identify the mechanism which underlies failures more effectively than the survival function. Researchers are likely to have first-hand knowledge of how the imminent risk changes with time for the lifetimes being studied. For example, light bulbs tend to break quite unexpectedly rather than because they are suffering from old age; people, on the other hand tend to wear out as they get older. We would expect the shapes of the failure rate functions for light bulbs and people to be different. People experience increasing failure rate, whereas light bulbs tend to exhibit constant failure rate [63].

The most wide used distribution for modeling lifetime data is the Weibull distribution due to the diversity shapes of its failure rate function. The failure rate can be increasing, decreasing or constant depending on its shape parameter. However, it fails to capture some complex situations where the data might result from multiple failure modes or represent non-monotonic failure rates. For example, let's consider a series system with two independent components (Fig. 1.2). Suppose T_1 and T_2 are the lifetimes of the two components. Then $T = \min(T_1, T_2)$ is the lifetime of the system. The reliability of the system is defined by

$$R(t) = \mathbb{P}(T > t) = \mathbb{P}(T_1 > t, T_2 > t)$$
$$= \mathbb{P}(T_1 > t)\mathbb{P}(T_2 > t)$$

$$= R_1(t)R_2(t)$$

and the failure rate function is derived by

$$h(t) = -\frac{d}{dt}\log(R(t))$$
$$= -\frac{d}{dt}\log(R_1(t)) - \frac{d}{dt}\log(R_2(t))$$
$$= h_1(t) + h_2(t)$$

This failure rate function can be considered a mixture of failure rates, $h(t) = p_1 h_1(t) + p_2 h_2(t)$, where $p_1 = p_2 = 1$. Such mixture failure rate will be considered throughout the thesis. Notice that the mixtures of failure rates do not need to be a convex combination of failure rates.
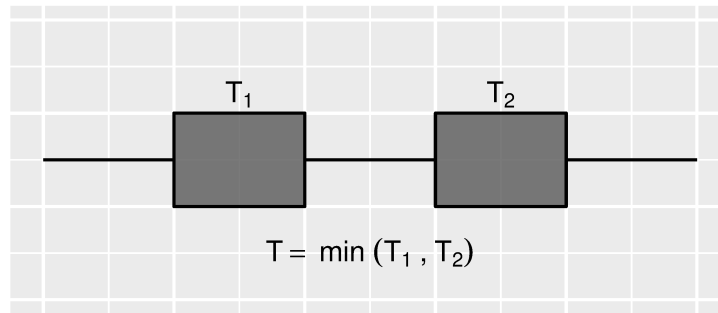
Figure 1.2: A series system with two independent components

The idea of combining failure rates of two or more well-known distributions is one of the solutions of such problems. Nevertheless, the increase of flexibility also comes along with the curse of dimension. As a result, the classical approach becomes too difficult for implementing. Therefore, the thesis focuses on advanced statistical tools for Bayesian inference of failure time distributions which result from the combination of failure rates or mixture failure rate. My early research works on these problems are given in [79, 80, 81, 82, 83, 84, 88].

1.2   Goals of the thesis

The main goal of the thesis is to develop mathematical and computational apparatus predominantly originating from Bayesian theory to describe special category of failure time distributions which can be considered as a mixture of failure rate functions. These distributions have in practice numerous applications, because engineering systems in operation are subject to different environmental stresses and shocks that can be mathematically described as a combination of different failure rates. To achieve the main goal, the following research activities (sub-goals) will be realized:

- Bayes theory of estimation must be studied in-depth, putting it into the context with classical approach for parameter estimation.

- Modern advanced statistical methods to estimate parameters of the mixture of failure rate models must be mastered, as for example the cross-entropy method for optimization, Markov chain Monte Carlo methods including adaptive MCMC and Hamiltonian Monte Carlo, etc.

- All advanced methods, as well as innovative theoretical findings must be implemented to an appropriate programming setting.

- New creative mixture models will be developed, analyzed and applied to real data sets.

- Different estimation approaches will be compared by mean of massive simulation study.

1.3 The method

The idea of combining failure rates will result in too many parameters of the proposed model so that the classical approach fails or becomes too difficult for practical implementation. Nonetheless, Bayesian approach makes these complex settings computationally feasible and straightforward due to the recent advances in Markov chain Monte Carlo (MCMC) methods. Since the number of parameters has increased, the conventional MCMC methods are hard to be implemented to find a good posterior sample. Therefore, the adaptive MCMC and Hamiltonian Monte Carlo (HMC) are exploited in order to empower the traditional methods of statistical estimation. In addition to the Bayes approach, the classical approach is also presented. The maximum likelihood estimation (MLE) is provided along with the Bayesian estimation using the cross-entropy method for optimizing the likelihood function. Traditional methods of maximization of likelihood functions of such mixture models sometimes do not provide the expected result due to multiple optimal points. For mixture modes, we usually rely on the expectation-maximization (EM) algorithm. However, the EM is a local search procedure and therefore there is no guarantee that it converges to the global maximum. As an alternative to EM algorithm, the cross-entropy (CE) algorithm is used which most of the time will provide the global maximum. In addition, the CE algorithm is not so sensitive to the choices of its initial values.

1.4 The outcomes

This thesis has developed new failure time distributions which result from mixture failure rate and provided originally Bayesian analysis of these proposed models. The first proposed model is the non-linear failure rate model with three parameters, the second is the additive Chen-Weibull model with four parameters and the third is the improvement of new modified Weibull model with five parameters. The thesis has also demonstrated the effect of parameterization on the Bayes estimators and has applied successfully some MCMC methods, known as adaptive MCMC and Hamiltonian Monte Carlo, for exploring the corresponding posterior distributions, and as a result, providing more accurate approximate Bayes estimates.

1.5 Outline of the thesis

The outline of the thesis is as follows. Chapter 2 provides a short literature survey of the area of study, the so-called "state of the art". Chapter 3 reviews the methods that are used for the following chapters. Mostly focus on MLE, MCMC and Bayesian inference methods. Chapter 4 devotes for a Bayes study of the proposed non-linear failure rate model which considers as the generalization of the linear failure rate model. Chapter 5 deals with the parameterizations of the Weibull model which help to understand the posterior correlations inside model parameters and to apply successfully Bayesian inference via MCMC

methods for the Weibull model as well as other models which induce from it. Chapter 6 provides a Bayes study of the proposed additive Chen-Weibull model. Chapter 7 provides a Bayes study of the improvement of the new modified Weibull model. And finally, Chapter 8 makes conclusions and proposes for future work.

Chapter 2

State of the Art

The exponential distribution is often used in reliability studies as the model for the time until failure of a device. For example, the lifetime of a semiconductor chip or a light bulb with failures caused by random shocks might be appropriately modeled as an exponential distribution. The lack of memory property of the exponential distribution implies that the device does not wear out. That is, regardless of how long the device has been operating, the probability of a failure in the next 1000 hours is the same as the probability of a failure in the first 1000 hours of operation [45]. However, the lifetime of human who suffers slow aging or a device that suffers slow mechanical wear, such as bearing wear, is better modeled by a distribution with increasing failure rate. One such distribution is the Weibull distribution. The Weibull distribution is one of the best known distribution and has wide applications in diverse disciplines, see for example Rinne [51]. However, its failure rate function has limitations in reliability applications which due to the following reasons.

- First, its failure rate function, $h(t) = bt^{k-1}$, equals 0 at time $t = 0$. This failure rate model might be only appropriate for modeling some physical systems that do not suffer from random shocks. Some physical systems where from the past experiences the random shocks have been studied, required corrections. The model, where initial failure rate equals 0 might be inappropriate for modeling some physical systems that require initial positive failure rate. That is the physical systems that suffer from random shocks and also from wear out failures.

- Second, its failure rate function can only be increasing, decreasing or constant. For many real life data sets, failure rate function could possibly exhibit some form of non-monotonic behavior. For example, the most popular non-monotonic failure rate function is the bathtub-shaped failure rate.

For these reasons, there have been several proposals to model such behaviors. One such proposal includes generalizing or modifying the Weibull distribution by adding an extra parameter into it as, for example, the exponentiated Weibull distribution [47], the modified Weibull distribution [39], the generalized modified Weibull distribution [14], etc.

Another proposal, also the main interest of the thesis, is to combine failure rate functions into a mixture of failure rates from which the failure time distribution is defined. On one hand, such models are useful for modeling a series system with independent components. On the other hand, such models are useful for modeling a physical system that suffers from independent competing risks. It is often seen that the first proposal fails to capture some important aspects of the data whereas the second is quite cumbersome. Therefore, Bayesian approach will make the second approach be easier and more practical.

Due to the first disadvantage of the Weibull failure rate, i.e $h(t) = bt^{k-1}$ equals 0 at time $t = 0$, the linear failure rate (LFR), $h(t) = a + bt$, was proposed as a remedy for this problem. The LFR was first introduced by Kodlin [34], and had been studied by

Bain [4] as a special case of polynomial failure rate model for type II censoring. It can be considered a generalization of the exponential model ($b = 0$) or the Rayleigh model ($a = 0$). It can also be regarded as a mixture of failure rates of an exponential and a Rayleigh distributions. However, because of the limitation of the Rayleigh failure rate, as well as the LFR, which is not flexible to capture the most common types of failure rate, new generalizations of LFR $h(t) = a + bt^{k-1}$, known as non-linear failure rate (NLFR), was developed. It is considered as a mixture of the exponential and Weibull failure rates. This mixture failure rate not only allows for an initial positive failure rate but also takes into account all shapes of Weibull failure rate. The first research work which attempts to solve the NLFR is given by Salem [54]. This model was also introduced later by Sarhan [56], Bousquet, Bertholon, and Celeux [10], and Sarhan and Zaindin [58], but with different name, motivation, parameterization, model explanation and purpose.

In spite of its flexibility, the NLFR still fails to represent the bathtub-shaped failure rate. Therefore, the additive Weibull failure rate was proposed by Xie and Lai [75] to meet this demand. It is a combination of two Weibull failure rates in which its failure rate function has bathtub-shaped. Many years later, a new combination of the Weibull failure rate and modified Weibull failure rate, namely the new modified Weibull distribution, was proposed by Almalki and Yuan [3] which possesses the bathtub-shaped failure rate. It was demonstrated as the most flexible model compare to all other existed models. However, short time later, the new model called additive modified Weibull distribution was proposed by He, Cui, and Du [28] by combining the modified Weibull failure rate and the Gompertz failure rate [25]. This new model was demonstrated to be better than the preceding model and all other existing models. In this thesis, we will see that the new modified Weibull model can be improved to be better than its original model and to some extent even better than the additive modified Weibull model. Recently, there is a new model have been proposed by combining the log-normal failure rate and the modified Weibull failure rate [60].

Bayesian methods have significantly increased over the past few decades that is due to advances in simulation-based computational tools and fast computing machines. The recent advances in Markov chain Monte Carlo (MCMC) methods, e.g. adaptive MCMC and Hamiltonian Monte Carlo (also know as Hybrid Monte Carlo), etc., have revolutionized the Bayesian approach. In case of failure distributions resulting from mixture failure rate, the models become so complicated so that the classical approaches fail or become too difficult for practical implementation. Bayesian approaches make these complex settings computationally feasible and straightforward. Since mixture failure rate will result in too many parameters of the proposed model, there is only two research papers so far which provide Bayesian approach for such models. The first paper was provided by Salem [54] which intended to provide Bayesian estimation of the non-linear failure rate model. However, the solutions were hard to obtain due to computational difficulties. The second paper was provided by Pandey, Singh, and Zimmer [50] which intended to provide the Bayesian estimation of the linear failure rate model with only two parameters. Of course, there are several papers which provided Bayesian analysis for lifetime models, but not such kind of mixture failure rate, and perhaps we have never seen any paper that provide Bayesian estimation of lifetime models with more than three parameters. This thesis attempts to explore such complicated models using Bayesian approach.

Chapter 3

Methodology

The chapter provides brief discussions of statistical methods that will be use in later chapters.

3.1 Total time on test

The scaled Total Time on Test (TTT) plot is used to identify the shapes of failure rate function of an observed data [1]. The scale TTT transform is defined as

$$G(u) = \frac{K^{-1}(u)}{K^{-1}(1)}, \quad 0 < u < 1 \tag{3.1}$$

where $K^{-1}(u) = \int_0^{F^{-1}(u)} R(t)dt$. The corresponding empirical version of the scaled TTT transform is defined as

$$G_n(r/n) = \frac{\sum_{i=1}^{r} y_{i:n} + (n-r)y_{r:n}}{\sum_{i=1}^{n} y_{i:n}}, \quad r = 1, \ldots, n \tag{3.2}$$

where $y_{i:n}$ denote the i-th order statistic of the sample. In practice, the scaled TTT transform is constructed by plotting $(r/n, G_n(r/n))$. It has been show by Aarset [1] that

- If the scaled TTT transform approximates a straight diagonal line, the failure rate is constant.

- If the scaled TTT transform is convex, the failure rate function is decreasing.

- If the scaled TTT transform is concave, the failure rate function is increasing.

- If the scaled TTT transform is first convex and then concave, the failure rate function is bathtub shaped.

- If the scaled TTT transform is first concave and then convex, the failure rate function is unimodal.

Fig. 3.1 displays the 5 different shapes of the scale TTT transform which correspond to the 5 different shapes of the failure rate function.

3.2 Maximum likelihood estimation

Let $\mathcal{D} : t_1, ..., t_n$ be the random failure times of n devices under test whose failure time distribution is determined by a pdf $f(t|\boldsymbol{\theta})$ with a d-dimensional vector parameter $\boldsymbol{\theta}$. If
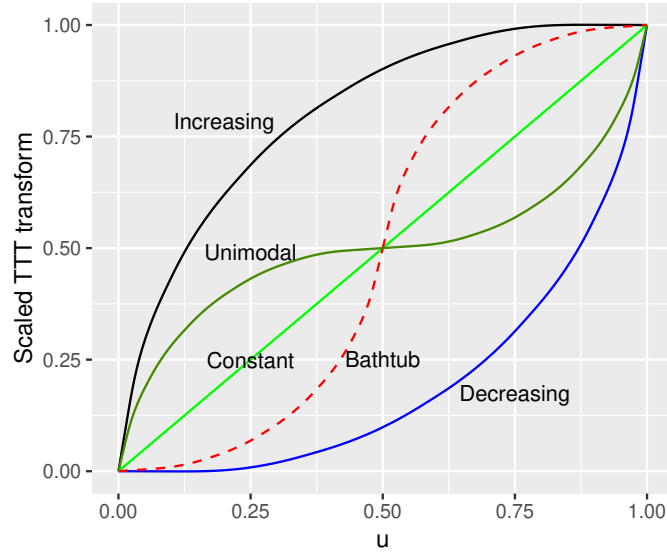
Figure 3.1: Scaled TTT transform plot

there is no censoring, the likelihood function takes the general form

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta}) \tag{3.3}$$

The log-likelihood function can be written as

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(t_i|\boldsymbol{\theta}) \tag{3.4}$$

If some observations are censored, we have to make an adjustment to the expression (3.3). For an observation of an observed failure, we put in the pdf as above. But for a right-censored observation, we put in the reliability function, indicating that observation is known to exceed a particular value. The likelihood function in general then takes the form

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta})^{\delta_i} R(t_i|\boldsymbol{\theta})^{1-\delta_i} \tag{3.5}$$

$$= \prod_{i=1}^{n} h(t_i|\boldsymbol{\theta})^{\delta_i} R(t_i|\boldsymbol{\theta}) \tag{3.6}$$

This expression means that when t_i is an observed failure, the censoring indicator is $\delta_i = 1$, and we enter a pdf factor. When t_i is a censored observation, we have $\delta_i = 0$, and we enter a reliability factor [46].

The log-likelihood function for censoring case is

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \log f(t_i|\boldsymbol{\theta}) + \sum_{i=1}^{n} (1-\delta_i) \log R(t_i|\boldsymbol{\theta}) \tag{3.7}$$

When some or all of uncensored, left censored, interval censored and right censored observations occur in the same dataset, the likelihood function where censoring mechanisms are independent of lifetimes will include products of terms:

$$\begin{array}{ll}
f(t) & \text{for observed failure at } t \\
R(t_R) & \text{for right censored at } t_R \\
1 - R(t_L) & \text{for left censored at } t_L \\
R(t_L) - R(t_R) & \text{for interval censored at } [t_L, t_R)
\end{array}$$

Maximizing the $L(\mathcal{D}|\boldsymbol{\theta})$ or $\log L(\mathcal{D}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ will give us the MLE of $\boldsymbol{\theta}$. In this thesis, this task is performed by using the cross-entropy method which is introduced below.

3.2.1 Observed and expected Fisher information matrices

A vector of first partial derivatives of the log-likelihood function, some time called the gradient vector is given as follows:

$$\nabla \log L(\mathcal{D}|\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_d} \end{pmatrix} \tag{3.8}$$

And a matrix of second partial derivatives is given by

$$\nabla^2 \log L(\mathcal{D}|\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_d} \\ \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_d \partial \theta_1} & \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_d \partial \theta_2} & \cdots & \frac{\partial^2 \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_d^2} \end{pmatrix} \tag{3.9}$$

The expected (Fisher) information matrix is defined as

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathbb{E}\left[\nabla^2 \log L(\mathcal{D}|\boldsymbol{\theta})\right] \tag{3.10}$$

However, it is not always possible to calculate expected information. But if we can evaluate the log likelihood, then we can calculate the other kind called observed (Fisher) information. The observed information matrix is defined as

$$\mathbf{J}(\boldsymbol{\theta}) = -\nabla^2 \log L(\mathcal{D}|\boldsymbol{\theta}) \tag{3.11}$$

Here we know the functions $\mathbf{I}$ and $\mathbf{J}$, but we do not know the true value of the parameter $\boldsymbol{\theta}$ where we should evaluate them. Hence we do not know either $\mathbf{I}(\boldsymbol{\theta})$ or $\mathbf{J}(\boldsymbol{\theta})$. However, using the plug-in theorem, both $\mathbf{I}$ and $\mathbf{J}$ can be evaluated at $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}) = -\mathbb{E}\left[\nabla^2 \log L(\mathcal{D}|\boldsymbol{\theta})\right]\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{3.12}$$

$$\mathbf{J}(\hat{\boldsymbol{\theta}}) = -\nabla^2 \log L(\mathcal{D}|\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{3.13}$$

Then the following approximations can be used to construct confident intervals based on maximum likelihood estimators [24]

$$\hat{\boldsymbol{\theta}} \approx N\left(\boldsymbol{\theta}, \mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}\right) \tag{3.14}$$

$$\hat{\boldsymbol{\theta}} \approx N\left(\boldsymbol{\theta}, \mathbf{J}(\hat{\boldsymbol{\theta}})^{-1}\right) \tag{3.15}$$

3.2.2 Asymptotic Confidence intervals

The $100(1 - \alpha)\%$ symmetric approximate normal confidence intervals of θ_i, $i = 1, \ldots, d$, are given by

$$\hat{\theta}_i \pm z_{1-\alpha/2} \sqrt{\left(\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}\right)_{ii}} \tag{3.16}$$

or

$$\hat{\theta}_i \pm z_{1-\alpha/2} \sqrt{\left(\mathbf{J}(\hat{\boldsymbol{\theta}})^{-1}\right)_{ii}} \tag{3.17}$$

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ percentage point of standard normal distribution.

3.2.3 Bootstrap standard error

For point estimator with complicated form, finding its standard error is difficult or impossible by using standard statistical methods. The bootstrap method can be used in such situation. To apply the bootstrap method, we draw bootstrap samples from $f(x|\theta = \hat{\theta})$ and calculate a bootstrap estimate $\hat{\theta}^*$ (MLE of θ based on bootstrap sample). This is called **parametric bootstrap**. In case the PDF $f(x|\theta)$ is unknown or difficult to generate data, the bootstrap method can be proceeded by considering the original sample data as a population and draw bootstrap samples from it with replacement. This is called **nonparametric bootstrap**. We repeat this process B times.

Bootstrap sample 1: $x_1^1, x_2^1, \ldots, x_n^1 \longrightarrow \hat{\theta}_1^*$
Bootstrap sample 2: $x_1^2, x_2^2, \ldots, x_n^2 \longrightarrow \hat{\theta}_2^*$
$\qquad \vdots$
Bootstrap sample B: $x_1^B, x_2^B, \ldots, x_n^B \longrightarrow \hat{\theta}_B^*$

Then we calculate the sample mean of the bootstrap estimates

$$\overline{\hat{\theta}}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}_i^* \tag{3.18}$$

The bootstrap standard error of $\hat{\theta}$ is just the sample standard deviation of the bootstrap estimates

$$SE_B(\hat{\theta}) = \sqrt{\frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{\theta}_i^* - \overline{\hat{\theta}}^*\right)^2} \tag{3.19}$$

B is usually chosen equal to 100 or 200 for obtaining bootstrap standard error and at least $10^4 - 1$ for calculating bootstrap confidence interval.

3.2.4 Bootstrap confidence intervals

Bootstrap method provides five types of confidence intervals. The **normal** confidence interval incorporates the estimated bootstrap bias and bootstrap standard error which is given as

$$\left[\hat{\theta} - \widehat{Bias}_B(\hat{\theta}) - z_{1-\alpha/2} SE_B(\hat{\theta}), \hat{\theta} - \widehat{Bias}_B(\hat{\theta}) + z_{1-\alpha/2} SE_B(\hat{\theta})\right] \tag{3.20}$$

where $\widehat{Bias}_B(\hat{\theta}) = \overline{\hat{\theta}}^* - \hat{\theta}$ is an estimated bootstrap bias of $\hat{\theta}$.

The **basic bootstrap** confidence interval is based on the idea that the quantity $\hat{\theta}^* - \hat{\theta}$ has roughly the same distribution as $\hat{\theta} - \theta$. Therefore,

$$\mathbb{P}\left[\hat{\theta}^*_{(B+1)\frac{\alpha}{2}} - \hat{\theta} \leq \hat{\theta}^* - \hat{\theta} \leq \hat{\theta}^*_{(B+1)(1-\frac{\alpha}{2})} - \hat{\theta}\right] \approx 1 - \alpha \tag{3.21}$$

$$\mathbb{P}\left[\hat{\theta}^*_{(B+1)\frac{\alpha}{2}} - \hat{\theta} \leq \hat{\theta} - \theta \leq \hat{\theta}^*_{(B+1)(1-\frac{\alpha}{2})} - \hat{\theta}\right] \approx 1 - \alpha \tag{3.22}$$

$$\mathbb{P}\left[2\hat{\theta} - \hat{\theta}^*_{(B+1)(1-\frac{\alpha}{2})} \leq \theta \leq 2\hat{\theta} - \hat{\theta}^*_{(B+1)\frac{\alpha}{2}}\right] \approx 1 - \alpha \tag{3.23}$$

Then the basic bootstrap confidence interval is given as

$$\left[2\hat{\theta} - \hat{\theta}^*_{(B+1)(1-\frac{\alpha}{2})}, 2\hat{\theta} - \hat{\theta}^*_{(B+1)\frac{\alpha}{2}}\right] \tag{3.24}$$

where $\hat{\theta}^*_{(B+1)\frac{\alpha}{2}}$ is the $(B+1)\frac{\alpha}{2}$th of the B ordered $\hat{\theta}^*$ values. If $(B+1)\frac{\alpha}{2}$ is not an integer, linear interpolation can be used.

The **percentile** confidence interval is based on the quantiles of the B bootstrap replication of $\hat{\theta}$. A $100(1-\alpha)\%$ bootstrap confidence interval for θ is calculated as

$$\left[\hat{\theta}^*_{(B+1)\frac{\alpha}{2}}, \hat{\theta}^*_{(B+1)(1-\frac{\alpha}{2})}\right] \tag{3.25}$$

The **studentized** bootstrap confidence interval is based on estimating the actual distribution of the t statistic from the data. The estimated distribution of T^* is denoted

$$T^* = \frac{\hat{\theta}^* - \hat{\theta}}{\sqrt{\widehat{Var}(\hat{\theta}^*)}} \tag{3.26}$$

where $\hat{\theta}^*$ and $\sqrt{\widehat{Var}(\hat{\theta}^*)}$ are statistics computed from a bootstrap sample. The studentized confidence interval is

$$\left[\hat{\theta} - T^*_{(B+1)(1-\frac{\alpha}{2})}\hat{\sigma}_{\hat{\theta}}, \hat{\theta} - T^*_{(B+1)\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}}\right] \tag{3.27}$$

The final bootstrap interval is the **BCa** confidence interval where the BCa stands for bias-corrected and accelerated. To compute a BCa interval for θ, first compute the bias factor

$$z = \Phi^{-1}\left[\frac{\sum_{i=1}^{B} I\left\{\hat{\theta}^*_i < \hat{\theta}\right\}}{B}\right] \tag{3.28}$$

where Φ^{-1} is the inverse of the standard normal CDF. Next, compute the skewness correction factor:

$$a = \frac{\sum_{i=1}^{n}\left(\overline{\hat{\theta}}_{(-i)} - \hat{\theta}_{(-i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\overline{\hat{\theta}}_{(-i)} - \hat{\theta}_{(-i)}\right)^2\right]^{\frac{3}{2}}} \tag{3.29}$$

where $\hat{\theta}_{(-i)}$ is the value of $\hat{\theta}$ when the ith value is deleted from the sample of n values and $\overline{\hat{\theta}}_{(-i)} = \sum_{i=1}^{n}\frac{\hat{\theta}_{(-i)}}{n}$. Using z and a, compute

$$a_1 = \Phi\left[z + \frac{z + z_{\alpha/2}}{1 - a(z + z_{\alpha/2})}\right] \text{ and } a_2 = \Phi\left[z + \frac{z + z_{1-\alpha/2}}{1 - a(z + z_{1-\alpha/2})}\right] \tag{3.30}$$

The BCa confidence interval is

$$\left[\hat{\theta}^*_{(B+1)a_1}, \hat{\theta}^*_{(B+1)a_2} \right] \tag{3.31}$$

For details, readers are referred to Ugarte, Militino, and Arnholt [68]. All of these five bootstrap confidence intervals can easily be obtained by using the "boot.ci" function in the R package "boot" [13].

3.3 The cross-entropy method for continuous multi-extremal optimization

The main idea for the cross-entropy (CE) method for optimization can be stated as follows: Suppose we wish to maximize some "performance" function $S(\mathbf{x})$ over all elements/states $\mathbf{x}$ in some set $\mathcal{X} \subset \mathbb{R}^n$. Let us denote the maximum by γ^*, thus

$$S(\mathbf{x}^*) = \gamma^* = \max_{\mathbf{x} \in \mathcal{X}} S(\mathbf{x}) \tag{3.32}$$

To proceed with CE, we first randomize our deterministic problem by defining a family of pdfs $\{f(\cdot; \mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ on the set $\mathcal{X}$. Next, we associate with Eq. (3.32) the estimation of

$$\ell(\gamma) = \mathbb{P}_u \left(S(\mathbf{X}) \geq \gamma \right) = \mathbb{E}_u I_{\{S(\mathbf{X}) \geq \gamma\}} \tag{3.33}$$

the so-called associated stochastic problem (ASP). Here, $\mathbf{X}$ is a random vector with pdf $f(\cdot; \mathbf{u})$, for some $\mathbf{u} \in \mathcal{V}$ (for example, $\mathbf{X}$ could be a normal random vector) and γ is a known or unknown parameter. Note that there are in fact two possible estimation problems associated with Eq. (3.33). For a given γ we can estimate ℓ, or alternatively, for a given ℓ we can estimate γ, the root of Eq. (3.33). Let us consider the problem of estimating ℓ for a certain γ close to γ^*. Then, typically $\{S(\mathbf{X}) \geq \gamma\}$ is a rare event, and estimation of ℓ is a non-trivial problem. The CE method solves this efficiently by making adaptive changes to the probability density function according to the Kullback-Leibler CE, thus creating a sequence $f(\cdot; \mathbf{u}), f(\cdot; \mathbf{v}_1), f(\cdot; \mathbf{v}_2), \ldots$ of pdfs that are "steered" in the direction of the theoretically optimal density $f(\cdot; \mathbf{v}^*)$ corresponding to the degenerate density at an optimal point. In fact, the CE method generates a sequence of tuples $\{(\gamma_t, \mathbf{v}_t)\}$, which converges quickly to a small neighborhood of the optimal tuple $(\gamma^*, \mathbf{v}^*)$. More specifically, we initialize by setting $\mathbf{v}_0 = \mathbf{u}$, choosing a not very small quantity ϱ, say $\varrho = 10^{-2}$, and then we proceed as follows:

Adaptive updating of γ_t. For a fixed $\mathbf{v}_{t-1}$, let γ_t be the $(1 - \varrho)$-quantile of $S(\mathbf{X})$ under $\mathbf{v}_{t-1}$. That is, γ_t satisfies

$$\mathbb{P}_{\mathbf{v}_{t-1}}(S(\mathbf{X}) \geq \gamma_t) \geq \varrho \tag{3.34}$$
$$\mathbb{P}_{\mathbf{v}_{t-1}}(S(\mathbf{X}) \leq \gamma_t) \geq 1 - \varrho \tag{3.35}$$

where $\mathbf{X} \sim f(\cdot; \mathbf{v}_{t-1})$

A simple estimator of γ_t, denoted $\hat{\gamma}_t$, can be obtained by drawing a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from $f(\cdot; \mathbf{v}_{t-1})$ and evaluating the sample $(1 - \varrho)$-quantile of the performances as

$$\hat{\gamma}_t = S_{\lceil (1-\varrho)N \rceil} \tag{3.36}$$

Adaptive updating of $\mathbf{v}_t$. For fixed γ_t and $\mathbf{v}_{t-1}$, derive $\mathbf{v}_t$ from the solution of the program

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}} \mathbb{E}_{\mathbf{v}_{t-1}} I_{\{S(\mathbf{X}) \geq \gamma_t\}} \ln f(\mathbf{X}; \mathbf{v}) \tag{3.37}$$

The stochastic counterpart of Eq. (3.37) is as follows: for fixed $\hat{\gamma}_t$ and $\hat{\mathbf{v}}_{t-1}$ (the estimate of $\mathbf{v}_{t-1}$), derive $\hat{\mathbf{v}}_t$ from the following program

$$\max_{\mathbf{v}} \hat{D}(\mathbf{v}) = \max_{\mathbf{v}} \frac{1}{N} \sum_{i=1}^{N} I_{\{S(\mathbf{X}_i) \geq \hat{\gamma}_t\}} \ln f(\mathbf{X}_i; \mathbf{v}) \tag{3.38}$$

Instead of updating the parameter vector $\mathbf{v}$ directly via the solution of Eq. (3.38) we use the following smoothed version

$$\hat{\mathbf{v}}_t = \alpha \tilde{\mathbf{v}}_t + (1 - \alpha)\hat{\mathbf{v}}_{t-1}, \quad i = 1, \dots, n \tag{3.39}$$

where $\tilde{\mathbf{v}}_t$ is the parameter vector obtained from the solution of Eq. (3.38), and α is called the smoothing parameter, with $0.7 < \alpha \leq 1$.

Algorithm Generic CE Algorithm for Optimization

1: Choose some $\hat{\mathbf{v}}_0$. Set $t = 1$.

2: Generate a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from the density $f(\cdot; \hat{\mathbf{v}}_{t-1})$ and compute the sample $(1 - \varrho)$-quantile $\hat{\gamma}_t$ of the performances according to Eq. (3.36).

3: Use the same sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ and solve the stochastic program Eq. (3.38). Denote the solution by $\tilde{\mathbf{v}}_t$.

4: Apply Eq. (3.39) to smooth out the vector $\tilde{\mathbf{v}}_t$. Increase t by 1.

5: Repeat steps 2-4 until a pre-specified stopping criterion is met.

Using normal updating, the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ are sample from an n-dimensional multivariate normal distribution with independent components, $\mathcal{N}(\hat{\boldsymbol{\mu}}_{t-1}, \hat{\boldsymbol{\sigma}}_{t-1}^2)$. While applying CE algorithm, the mean vector $\hat{\boldsymbol{\mu}}_t$ should converge to $\mathbf{x}^*$ and the vector of standard deviations $\hat{\boldsymbol{\sigma}}_t$ converge to the zero vector. In short, we should obtain a degenerated pdf with all mass concentrated in the vicinity of the point $\mathbf{x}^*$. Fig. 3.2 gives an example of the evolution of the sampling pdf for a parameter. More detail for such explanation can be found in [35], [53] or [8], and a short introduction can also be found in [78].
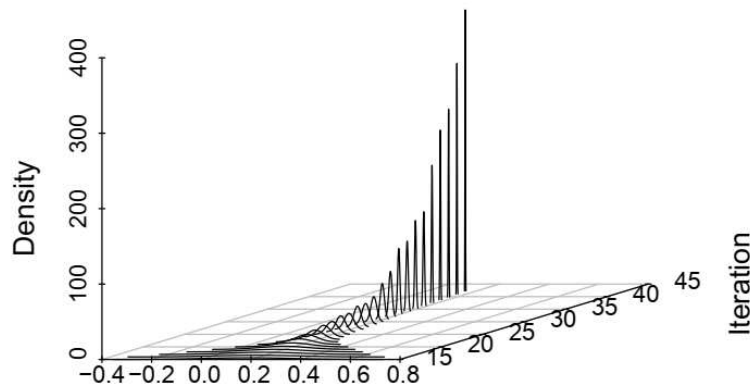
Figure 3.2: The evolution of the sampling pdf for a parameter (source [8])

3.4   Bayesian inference

3.4.1   Bayes' rule

Consider a random variable T that has a probability distribution that depends upon $\boldsymbol{\theta}$ (it can be a scalar or a vector), where $\boldsymbol{\theta}$ is an element of a well-defined set Θ. For example if $\boldsymbol{\theta}$ is the mean of a normal distribution, then $\Theta = \mathbb{R}$. $\boldsymbol{\theta}$ is the unknown parameter, but in Bayesian inference, it is treated as random variables. We denote $\pi(\boldsymbol{\theta})$ as the **prior pdf** of $\boldsymbol{\theta}$. The **prior distribution** of $\boldsymbol{\theta}$ given by the prior pdf $\pi(\boldsymbol{\theta})$ reflects subjective belief of $\boldsymbol{\theta}$ before the sample is drawn. We now denote the pdf of T by $f(t|\boldsymbol{\theta})$ since we think of it as a conditional pdf of T, given $\boldsymbol{\theta}$.

Suppose that $\mathcal{D}: t_1, \ldots, t_n$ is a random sample from the conditional distribution of T given $\boldsymbol{\theta}$ with pdf $f(\mathcal{D}|\boldsymbol{\theta})$. Then the likelihood function, the joint conditional pdf of $\mathcal{D}$ given $\boldsymbol{\theta}$, can be written as

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta}) \tag{3.40}$$

Thus the joint of $\mathcal{D}$ and $\boldsymbol{\theta}$ is

$$\pi(\mathcal{D}, \boldsymbol{\theta}) = L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{3.41}$$

If $\boldsymbol{\theta}$ is a continuous random variable, the joint marginal pdf of $\mathcal{D}$ is given by

$$\pi(\mathcal{D}) = \int_{\Theta} \pi(\mathcal{D}, \boldsymbol{\theta})d\boldsymbol{\theta} = \int_{\Theta} L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{3.42}$$

If $\boldsymbol{\theta}$ is a discrete random variable, integration would be replaced by summation. In either cases, the conditional pdf of $\boldsymbol{\theta}$, given the sample $\mathcal{D}$, is

$$\pi(\boldsymbol{\theta}|\mathcal{D}) = \frac{\pi(\mathcal{D}, \boldsymbol{\theta})}{\pi(\mathcal{D})} = \frac{L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathcal{D})} \tag{3.43}$$

The distribution defined by this conditional pdf is called the **posterior distribution** and (3.43) is called the **posterior pdf**. The prior distribution reflects the subjective belief of $\boldsymbol{\theta}$ before the sample is drawn, while the posterior distribution is the conditional distribution of $\boldsymbol{\theta}$ after the sample is drawn. For more details, see [31].

3.4.2   Prediction

Predictive inferences are the process of making inferences about an unknown observable. Before the data $\mathcal{D}$ are considered, the distribution of the unknown but observable $\tilde{t}$ is

$$\pi(\tilde{t}) = \int_{\Theta} \pi(\tilde{t}, \boldsymbol{\theta})d\boldsymbol{\theta} = \int_{\Theta} f(\tilde{t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{3.44}$$

This is often called the marginal distribution of $\tilde{t}$, but a more informative name is the **prior predictive distribution**: prior because it is not conditional on a previous observation of the process, and predictive because it is the distribution for a quantity that is observable.

After the data $\mathcal{D}$ have been observed, we can predict an unknown observable, $\tilde{t}$, from the same process. The distribution of $\tilde{t}$ is called the **posterior predictive distribution**, posterior because it is conditional on the observed $\mathcal{D}$ and predictive because it is a prediction for an observable $\tilde{t}$:

$$\pi(\tilde{t}|\mathcal{D}) = \int_{\Theta} \pi(\tilde{t}, \boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$

$$= \int_{\Theta} \pi(\tilde{t}|\boldsymbol{\theta}, \mathcal{D})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$

$$= \int_{\Theta} f(\tilde{t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{3.45}$$

The last step follows from the assumed conditional independence of $\mathcal{D}$ and $\tilde{t}$ given θ [22].

3.4.3 Bayesian point estimation

Bayesian point estimation is a process of selecting a decision function $\hat{\theta}$, so that $\hat{\theta}$ is a predicted value of θ. The choice of the decision function should depend upon the loss function $L(\theta, \hat{\theta})$. A Bayes estimate is a decision function $\hat{\theta}$ that minimizes the conditional expectation of the loss

$$\mathbb{E}\left[L(\theta, \hat{\theta})|\mathcal{D}\right] = \int_{\Theta} L(\theta, \hat{\theta})\pi(\theta|\mathcal{D})d\theta \tag{3.46}$$

In estimation problems, it is natural for the loss function to be a function of the distance between the true value of the parameter θ and its estimated value $\hat{\theta}$. The most widely used loss criterion in one parameter estimation problems is squared error loss (SEL) (Fig. 3.3), that is,

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2 \tag{3.47}$$
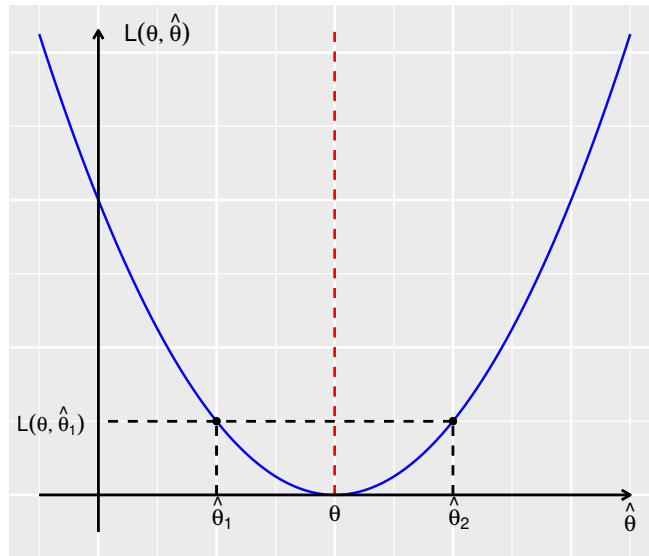
Figure 3.3: Squared error loss function

Squared error loss is a symmetric function that penalizes overestimation and underestimation equally, and takes the value zero when the estimate is right on target [55].

The Bayes estimator of θ under the SEL function is the value $\hat{\theta}$ which minimizes $\mathbb{E}\left[(\hat{\theta} - \theta)^2|\mathcal{D}\right]$. That is

$$\hat{\theta}_{BS} = \mathbb{E}\left[\theta|\mathcal{D}\right] \tag{3.48}$$

where $\mathbb{E}[\cdot|\mathcal{D}]$ denotes the posterior expectation with respect to the posterior density of θ.

The use of symmetric loss function may be inappropriate in the estimation of reliability function as has been recognized by Canfield [12]. Overestimate of reliability function or mean failure time is usually much more serious than underestimate of reliability function or mean failure time. Also, an underestimate of the failure rate results in more serious

consequences than an overestimate of the failure rate. For example, in the disaster of the space shuttle [20] the management underestimated the failure rate and therefore overestimated the reliability of solid-fuel rocket booster [6].

Varian [70] motivated the use of asymmetric loss functions in estimation problems arising in real estate assessment, where the overestimation of a property's value might cause it to remain on the market unsold for an extended period, ultimately costing the seller inordinate and unnecessary expenses. The estimation of peak water flow in the construction of dams or levees clearly has asymmetric consequences; overestimation might lead to increased construction costs while underestimation might lead to the much more serious consequence of subsequent overflows which can seriously threaten lives and property in adjacent communities [55].

There are many forms of asymmetric loss functions. One of the more widely use versions of asymmetric loss is the linear exponential (LINEX) loss function (Fig. 3.4) which can be expressed as

$$L(\theta, \hat{\theta}) = e^{c(\hat{\theta}-\theta)} - c(\hat{\theta}-\theta) - 1, \qquad c \neq 0 \tag{3.49}$$
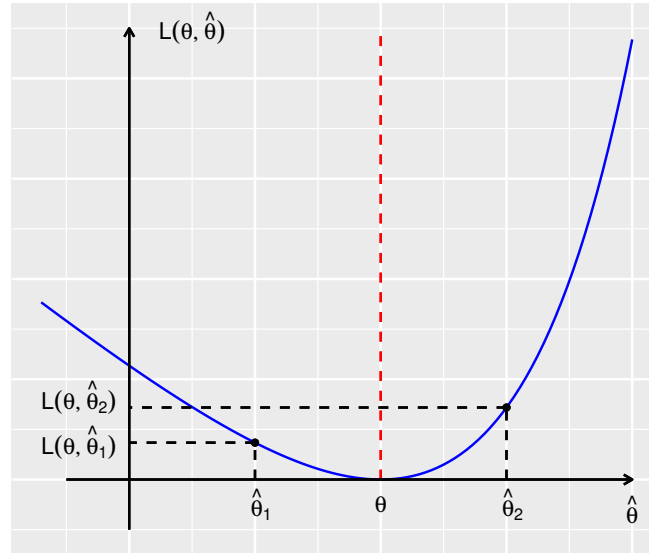
Figure 3.4: Linear exponential loss function ($c > 0$)

This function rises exponentially when $\hat{\theta} > \theta$ and approximately linearly when $\hat{\theta} < \theta$. The sign and magnitude of the parameter c represents the direction and degree of symmetry, respectively. If $c > 0$, the overestimation is more serious than underestimation, and vice versa. For c close to zero, the LINEX loss is approximately SEL and therefore almost symmetric.

The posterior expectation of the LINEX loss function is

$$\mathbb{E}\left[L(\theta, \hat{\theta})|\mathcal{D}\right] = e^{c\hat{\theta}}\mathbb{E}\left[e^{-c\theta}|\mathcal{D}\right] - c\left(\hat{\theta} - \mathbb{E}[\theta|\mathcal{D}]\right) - 1 \tag{3.50}$$

The Bayes estimator of θ under the LINEX loss function is the value $\hat{\theta}$ which minimizes Eq. (3.50). That is

$$\hat{\theta}_{BL} = -\frac{1}{c}\log\left\{\mathbb{E}\left[e^{-c\theta}|\mathcal{D}\right]\right\} \tag{3.51}$$

provided that the expectation $\mathbb{E}\left[e^{-c\theta}|\mathcal{D}\right]$ exists and is finite.

The modified LINEX loss is the general entropy loss (GEL) function, defined as:

$$L(\theta, \hat{\theta}) = \left(\frac{\hat{\theta}}{\theta}\right)^c - c \log\left(\frac{\hat{\theta}}{\theta}\right) - 1 \tag{3.52}$$

The Bayes estimator of θ under the GEL function is given as

$$\hat{\theta}_{BG} = \left(\mathbb{E}\left[\theta^{-c}|\mathcal{D}\right]\right)^{-\frac{1}{c}} \tag{3.53}$$

provided that the expectation $\mathbb{E}\left[\theta^{-c}|\mathcal{D}\right]$ exists and is finite. It can be shown that, when $c = 1$, this loss becomes the entropy loss and the Bayes estimator $\hat{\theta}_{BG}$ coincides with the Bayes estimator under the weighted squared-error loss function. Similarly, when $c = -1$ the Bayes estimator $\hat{\theta}_{BG}$ coincides with the Bayes estimator under squared error loss function [62].

3.4.4 Bayesian interval estimation

In order to obtain an interval estimate of θ, we find two functions $u(\mathcal{D})$ and $v(\mathcal{D})$ so that the conditional probability

$$\mathbb{P}\left[u(\mathcal{D}) < \theta < v(\mathcal{D})|\mathcal{D}\right] = \int_{u(\mathcal{D})}^{v(\mathcal{D})} \pi(\theta|\mathcal{D})d\theta \tag{3.54}$$

is large, for example, 0.95. The the interval $(u(\mathcal{D}), v(\mathcal{D}))$ is an interval estimate of θ in the sense that the conditional probability of θ belonging to that interval is equal to 0.95. To avoid confusing with confidence intervals, these intervals are often called **credible** or **probability intervals** [31].

3.4.5 Bayesian model checking

In parametric Bayesian analysis, we assume a parametric model for the data. Therefore, in order to avoid misleading inferences when the model is poor, checking the model is crucial to statistical analysis. If the model fits, then predictive data generate under the model should look similar to the observed data. In another way, the observed data should look plausible under the posterior predictive distribution. The basic technique for checking the fit of a model to data is to draw simulated values from the posterior predictive distribution and compare these samples to the observed data. Any systematic differences between the simulations and the data indicate potential failings of the model [22].

$$\pi(\tilde{t}|\mathcal{D}) = \int_{\Theta} f(\tilde{t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{3.55}$$

For each draw of the parameters from the posterior distribution, $\boldsymbol{\theta}^{(s)} \sim \pi(\theta|\mathcal{D})$, $s = 1\ldots m$, we draw a vector of n samples $\tilde{\mathbf{t}}^{(s)} = (\tilde{t}_1^{(s)}, \ldots, \tilde{t}_n^{(s)})$ from the posterior predictive distribution by simulating from the data model conditional on parameters $\boldsymbol{\theta}^{(s)}$. Notice that the size of predictive samples is the same as the size of observed data.

Posterior predictive p-values. From a Bayesian context, a posterior p-value or Bayesian p-value is the probability, given the data, that a future observation is more extreme (as measured by some test variable) than the data [21]. We choose a test quantity, or discrepancy measure, $T(\mathcal{D})$ that depends only on data, or $T(\mathcal{D}, \boldsymbol{\theta})$ that depends on data and

model parameter. Then the Bayesian p-value is estimated as [22]:

$$p_B = \mathbb{P}\left(T(\tilde{\mathbf{t}}, \boldsymbol{\theta}) > T(\mathcal{D}, \boldsymbol{\theta})|\mathcal{D}\right) \approx \frac{1}{m}\sum_{s=1}^{m} I_{\left[T(\tilde{\mathbf{t}}^{(s)}, \boldsymbol{\theta}^{(s)}) > T(\mathcal{D}, \boldsymbol{\theta}^{(s)})\right]} \tag{3.56}$$

The closer this value to 0.5, the better the test quantity values calculated from simulated data distribute around the observed test quantity.

Posterior predictive loss criterion. We compute the following quantities [5]: Posterior predictive mean and variance for each observation:

$$\tilde{\mu}_i = \mathbb{E}\left[\tilde{t}_i|\mathcal{D}\right] \approx \frac{1}{m}\sum_{s=1}^{m} \tilde{t}_i^{(s)}, \quad i = 1, \ldots, n \tag{3.57}$$

$$\tilde{\sigma}_i^2 = Var\left[\tilde{t}_i|\mathcal{D}\right] \approx \frac{1}{m}\sum_{s=1}^{m} \left(\tilde{t}_i^{(s)} - \tilde{\mu}_i\right)^2, \quad i = 1, \ldots, n \tag{3.58}$$

Goodness of fit (a model with lower value is better):

$$G = \sum_{i=1}^{n} (t_i - \tilde{\mu}_i)^2 \tag{3.59}$$

Penalize predictive variance (a model with lower value is better):

$$P = \sum_{i=1}^{n} \tilde{\sigma}_i^2 \tag{3.60}$$

Model comparison metric (a model with lower value is better): $D = G + P$

Deviance information criterion (DIC). The DIC is proposed based on the principle DIC = 'goodness of fit'+'complexity' [65]. The DIC is defined as

$$DIC = \overline{D} + p_D \tag{3.61}$$

where $\overline{D}$ is a measure of fit

$$\overline{D} = \mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}}\left[-2\log L(\mathcal{D}|\boldsymbol{\theta})\right] \tag{3.62}$$

and p_D is a measure of "effective number of parameters"

$$p_D = \mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}}\left[-2\log L(\mathcal{D}|\boldsymbol{\theta})\right] + 2\log L(\mathcal{D}|\tilde{\boldsymbol{\theta}}) \tag{3.63}$$

Here $\tilde{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ that can be usually chosen as $\tilde{\boldsymbol{\theta}} = \mathbb{E}\left[\boldsymbol{\theta}|\mathcal{D}\right]$. DIC is easily computed via MCMC methods. A model with smallest value of DIC is considered to be the best approximating model among a set of alternative models (see also [5]).

3.4.6   Empirical Bayes

In traditional Bayesian inferences, we assume that $\boldsymbol{\theta}$ follow a prior distribution with pdf $\pi(\boldsymbol{\theta})$. In fact these prior distributions are also depend upon another parameter γ called hyperparameter with pdf written as $\pi(\boldsymbol{\theta}|\gamma)$. $\boldsymbol{\theta}$ can be a vector of parameters. But in practice, we usually assume that the hyperparameter is known in advance based upon prior knowledge of $\boldsymbol{\theta}$. Then we suppress the prior parameter and write the prior pdf as $\pi(\boldsymbol{\theta})$. In empirical Bayes inferences, we treat the prior parameter as an unknown parameter and

empirical Bayes methodology estimates γ based on the data as follow. Recall that

$$
\begin{aligned}
\pi(\mathcal{D}, \boldsymbol{\theta}|\boldsymbol{\gamma}) &= \frac{\pi(\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\gamma})}{\pi(\boldsymbol{\gamma})} \\
&= \frac{\pi(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})}{\pi(\boldsymbol{\gamma})} \\
&= L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\gamma})
\end{aligned}
\tag{3.64}
$$

Then, consider the marginal likelihood function

$$
m(\mathcal{D}|\boldsymbol{\gamma}) = \int_{\Theta} L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\gamma})d\boldsymbol{\theta}
\tag{3.65}
$$

An empirical Bayes estimate is obtained by maximizing the marginal likelihood:

$$
\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} m(\mathcal{D}|\boldsymbol{\gamma})
\tag{3.66}
$$

The empirical Bayes procedure uses the posterior pdf $\pi(\boldsymbol{\theta}|\mathcal{D}, \hat{\boldsymbol{\gamma}})$ for inference on the parameter $\boldsymbol{\theta}$.

In nonconjugate setting, the high-dimensional integral in (3.65) poses a major difficulty, preventing a direct analytical solution. The Markov chain Monte Carlo empirical Bayes is one of the solutions for this problem. This method employs an EM algorithm to estimate the hyperparameters from MCMC samples. First, write the marginal likelihood as

$$
m(\mathcal{D}|\boldsymbol{\gamma}) = \frac{\pi(\mathcal{D}, \boldsymbol{\theta}|\boldsymbol{\gamma})}{\pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma})}
\tag{3.67}
$$

Then take the expectation of both sides with respect to $\pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma}')$ and switch to the log scale to arrive at

$$
\mathbb{E}_{\boldsymbol{\gamma}'}\left[\log m(\mathcal{D}|\boldsymbol{\gamma})\right] = \mathbb{E}_{\boldsymbol{\gamma}'}\left[\log \pi(\mathcal{D}, \boldsymbol{\theta}|\boldsymbol{\gamma})\right] - \mathbb{E}_{\boldsymbol{\gamma}'}\left[\log \pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma})\right]
\tag{3.68}
$$

for some (current value) $\boldsymbol{\gamma}'$. Expand the last term of the preceding equation:

$$
\mathbb{E}_{\boldsymbol{\gamma}'}\left[\log \pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma})\right] = \int \log \pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma})\pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma}')d\boldsymbol{\theta}
\tag{3.69}
$$

and note that by the Gibbs' inequality this integral is maximized at $\boldsymbol{\gamma} = \boldsymbol{\gamma}'$. Consequently, for every $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}'$, $-\mathbb{E}_{\boldsymbol{\gamma}'}\left[\log \pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma}')\right] < -\mathbb{E}_{\boldsymbol{\gamma}'}\left[\log \pi(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\gamma})\right]$, such that the sequence, which iteratively maximizes the first term in the right-hand side of (3.68)

$$
\boldsymbol{\gamma}^{(k+1)} = \arg\max_{\boldsymbol{\gamma}} \mathbb{E}_{\boldsymbol{\gamma}^{(k)}}\left[\log \pi(\mathcal{D}, \boldsymbol{\theta}|\boldsymbol{\gamma})\right]
\tag{3.70}
$$

is nondecreasing and converges. The expectation in (3.70) will generally not be available in closed form. However, one may approximate it by its Monte Carlo estimate:

$$
\arg\max_{\boldsymbol{\gamma}} \mathbb{E}_{\boldsymbol{\gamma}^{(k)}}\left[\log \pi(\mathcal{D}, \boldsymbol{\theta}|\boldsymbol{\gamma})\right] \approx \arg\max_{\boldsymbol{\gamma}} \frac{1}{M} \sum_{m=1}^{M} \log \pi(\mathcal{D}, \boldsymbol{\theta}^{m,(k)}|\boldsymbol{\gamma})
\tag{3.71}
$$

where $\boldsymbol{\theta}^{m,(k)}$ denotes the mth MCMC sample from the posterior distribution with hyperparameters $\boldsymbol{\gamma}^{(k)}$. For details, readers are referred to reference Wiel, Beest, and Münch [74].

3.5   Accept-Reject sampling method

The accept-reject sampling method allows us to sample from the target density $f(x)$, for which the inverse transform fails to be able to generate the required random variables, by sampling from another easy-to-sample proposal distribution $g(x)$. The proposal distribution $g(x)$ must satisfies the following properties [52]:

- $g(x)$ must have the same support as $f(x)$ (i.e. $g(x) > 0$ when $f(x) > 0$).

- There exists a constant M such that $f(x) \leq Mg(x)$ for all x.

The method only requires us to know the functional form of the density $f(x)$ up to a proportional constant. The accept-reject algorithm for drawing N samples from the target density is described as follows:

For $i = 1$ to N

1. Generate $Y \sim g$, $U \sim U_{(0,1)}$.

2. Accept $X = Y$ if $U \leq \frac{f(Y)}{Mg(Y)}$, otherwise, return to 1.

The accepted Y can be easily show to be drawn from the distribution with density $f(x)$. Indeed,

$$
\begin{aligned}
\mathbb{P}\left(Y \leq x | U \leq \frac{f(Y)}{Mg(Y)}\right) &= \frac{\mathbb{P}\left(Y \leq x, U \leq \frac{f(Y)}{Mg(Y)}\right)}{\mathbb{P}\left(U \leq \frac{f(Y)}{Mg(Y)}\right)} \\
&= \frac{\int_{-\infty}^{x} \int_{0}^{\frac{f(y)}{Mg(y)}} du\, g(y) dy}{\int_{-\infty}^{+\infty} \int_{0}^{\frac{f(y)}{Mg(y)}} du\, g(y) dy} \\
&= \frac{\int_{-\infty}^{x} \frac{f(y)}{Mg(y)} g(y) dy}{\int_{-\infty}^{+\infty} \frac{f(y)}{Mg(y)} g(y) dy} \\
&= \frac{\int_{-\infty}^{x} f(y) dy}{\int_{-\infty}^{+\infty} f(y) dy} \\
&= \int_{-\infty}^{x} f(y) dy = \mathbb{P}(X \leq x) \qquad (3.72)
\end{aligned}
$$

3.6   Markov chain Monte Carlo

3.6.1   The Gibbs sampler

Suppose we would like to simulate samples from a posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$ where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. If samples can be simulated from the full or complete conditional distributions $\{\pi(\theta_i|\boldsymbol{\theta}_{j \neq i}, \mathcal{D}),\ i = 1, \ldots, d\}$ (directly if the full conditional distributions were standard forms or indirectly via rejection sampling methods otherwise), the Gibbs sampler can be used. The Gibbs sampler algorithm for drawing N samples from the target distribution is described as follows:

- Given a set of starting values $\{\theta_2^{(0)}, \ldots, \theta_d^{(0)}\}$.

- For $t = 1, \ldots, N$, repeat:

    1: Draw $\theta_1^{(t)} \sim p\left(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}\right)$

$$\text{2: Draw } \theta_2^{(t)} \sim p\left(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}\right)$$

$$\vdots$$

$$\text{d: Draw } \theta_d^{(t)} \sim p\left(\theta_d | \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{d-1}^{(t)}\right)$$

For t sufficiently large (say, bigger than some n_0), $\{\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \ldots, \theta_d^{(t)}), \, t = n_0+1, \ldots, N\}$ is a (correlated) sample from the posterior $\pi(\boldsymbol{\theta}|\mathcal{D})$ [5].

3.6.2 The Metropolis-Hastings algorithm

The basic idea of the Metropolis-Hastings (MH) algorithm is to generate data from some target density $\pi(\boldsymbol{\theta}|\mathcal{D})$. Given $\boldsymbol{\theta}_n$, we generate a "proposed value" $\boldsymbol{\theta}^*$ from some pre-specified density $q(\boldsymbol{\theta}_n, \cdot)$, and then with probability $\alpha(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*)$ we set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}^*$, otherwise we set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$, where

$$\alpha(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*) = \min\left\{\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}_n)} \frac{q(\boldsymbol{\theta}^*, \boldsymbol{\theta}_n)}{q(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*)}, 1\right\} \tag{3.73}$$

If the proposed distribution is symmetric, then this acceptance probability reduces to

$$\alpha(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*) = \min\left\{\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}_n)}, 1\right\} \tag{3.74}$$

This special case of the MH algorithm is called the Metropolis algorithm [11].

The success (e.g. rapid convergence) of the MH algorithm depends on a good choice of the proposed density q. The usual approach is to use the symmetric random-walk Metropolis algorithm (RWM) in which $\boldsymbol{\theta}^* = \boldsymbol{\theta}_n + \mathbf{Z}$, where the increments $\mathbf{Z}$'s are i.i.d. from some fixed symmetric distribution (e.g. $N(0, \sigma^2 I_d)$). The choice of q then become how to scale the proposal (e.g. how to choose σ): too small and the chain will move to slowly; too large and the proposals will usually be rejected [11].

3.6.3 The adaptive MCMC

The proposal distribution is manually improved through trial and error until we obtain a good sample. However, this can be difficult, especially in high dimensions. An alternative approach is adaptive MCMC, which ask the computer to automatically "learn" better parameter value "on the fly"–that is, while algorithm runs. Suppose $\{P_\gamma\}_{\gamma \in \mathcal{Y}}$ is a family of Markov chain kernels, each having stationary distribution π. (e.g. P_γ corresponds to RWM algorithm with increment distribution $N(0, \gamma^2 I_d)$.) An adaptive MCMC would randomly update the value of γ at each iteration, in an attempt to find the best value [11].

The package "adaptMCMC" [59] is recommended for R users. This package provides an implementation of the generic adaptive Markov chain Monte Carlo sampler proposed by Vihola [71].

3.7 Hamiltonian Monte Carlo

The following brief introduction of Hamiltonian Monte Carlo is based upon what was introduced by Neal [49] and Gelman et al. [22]. MCMC was first introduced in the classic paper of Metropolis et al. [44], where it was used to simulate the distribution of states for a system of idealized molecules. Not long after, another approach to molecular simulation was introduced by Alder and Wainwright [2], in which the motion of the molecules was

deterministic, following Newton's law of motion, which have an elegant formalization as Hamiltonian dynamics.

In 1987, Duane and others merged the MCMC and molecular dynamics together [19]. The method was named "hybrid Monte Carlo," which abbreviates to "HMC,". However the name "Hamiltonian Monte Carlo," retaining the abbreviation, is more specific and descriptive. Duane and others applied HMC not to molecular simulation, but to lattice field theory simulations of quantum chromodynamics. Statistical applications of HMC began with Neal [48] as using it for neural network models. Apparently, HMC seems to be under-appreciated by statisticians, and perhaps also by physicists outside the lattice field theory community [49].

Hamiltonian Monte Carlo (HMC) is a Markov chain Monte Carlo algorithm that avoids the random walk behavior and sensitivity to correlated parameters that plague many MCMC methods by taking a series of steps informed by first-order gradient information. More specifically, it is a generalization of the Metropolis algorithm that includes 'momentum' variables so that each iteration can move further in parameter space, thus allowing faster mixing and moving much more rapidly through the target distribution, especially in high dimensions. These features allow it to converge to high-dimensional target distribution much more quickly than simpler methods such as random walk Metropolis or Gibbs sampling [22].

3.7.1 Hamiltonian dynamics

In physics, Hamiltonian dynamics is used to describe how objects move throughout a system. Hamiltonian dynamics describes an object's motion in terms of its location $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ and momentum $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)$ (object's mass times its velocity) at certain time t. For each location the object takes, there is an associated potential energy $U(\boldsymbol{\theta})$ (the height of the surface at a given position), and for each momentum there is an associated kinetic energy $K(\boldsymbol{\phi})$. The total energy of the system, known as Hamiltonian $H(\boldsymbol{\theta}, \boldsymbol{\phi})$, is constant and defined as the sum of the potential and kinetic energies:

$$H(\boldsymbol{\theta}, \boldsymbol{\phi}) = U(\boldsymbol{\theta}) + K(\boldsymbol{\phi}) \tag{3.75}$$

The partial derivatives of the Hamiltonian determine how $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ change over time t according to Hamiltonian equations:

$$\frac{\partial \theta_i}{\partial t} = \frac{\partial H}{\partial \phi_i} = \frac{\partial K(\boldsymbol{\phi})}{\partial \phi_i}, \quad i = 1, \ldots, d \tag{3.76}$$

$$\frac{\partial \phi_i}{\partial t} = -\frac{\partial H}{\partial \theta_i} = -\frac{\partial U(\boldsymbol{\theta})}{\partial \theta_i}, \quad i = 1, \ldots, d \tag{3.77}$$

Based on the expression of $\frac{\partial U(\boldsymbol{\theta})}{\partial \theta_i}$ and $\frac{\partial K(\boldsymbol{\phi})}{\partial \phi_i}$ and the initial location $\boldsymbol{\theta}_0$ and initial momentum $\boldsymbol{\phi}_0$ of an object at time t_0, we can predict the location and momentum of the object at any time $t = t_0 + T$ by simulating these dynamics for a duration T.

3.7.2 The leapfrog method for simulating Hamiltonian dynamics

The leapfrog method for simulating Hamiltonian dynamics for a duration T is performed by updating the location and momentum variables. A leapfrog step updating the location variable $\boldsymbol{\theta}$ and the momentum variable $\boldsymbol{\phi}$ over ϵ units time, starting at time t, is as follows:

- Take a half time-step to update the momentum variable

$$\phi_i(t + \epsilon/2) = \phi_i(t) - \frac{\epsilon}{2}\frac{\partial U}{\partial \theta_i}(\theta_i(t)) \tag{3.78}$$

- Take a full time-step to update the position variable

$$\theta_i(t + \epsilon) = \theta_i(t) + \epsilon\frac{\partial K}{\partial \phi_i}(\phi_i(t + \epsilon/2)) \tag{3.79}$$

- Take the remaining half time-step to finish updating the momentum variable

$$\phi_i(t + \epsilon) = \phi_i(t + \epsilon/2) - \frac{\epsilon}{2}\frac{\partial U}{\partial \theta_i}(\theta_i(t + \epsilon)) \tag{3.80}$$

We can run L leapfrog steps to simulate Hamiltonian dynamics over ϵL units of time.

3.7.3 Potential energy, kinetic energy and the target distribution

The target distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$ (posterior distribution) that we wish to sample can be related to a potential energy function via the concept of a canonical distribution from statistical mechanics. Given some energy function $E(x)$ for the state x of some physical system, the canonical distribution over states has probability density function

$$\pi(x) = \frac{1}{c}e^{-E(x)/T} \tag{3.81}$$

where T is the temperature of the system and c is the normalizing constant needed for this function to sum or integrate to one. For Hamiltonian Monte Carlo simulation, we choose $T = 1$.

The Hamiltonian $H(\boldsymbol{\theta}, \boldsymbol{\phi})$ is an energy function for the joint state of location $\boldsymbol{\theta}$ and momentum $\boldsymbol{\phi}$. Therefore, following the canonical distribution for energy function, a joint distribution for them is defined as follows:

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\phi}) &\propto e^{-H(\boldsymbol{\theta}, \boldsymbol{\phi})}\\
&= e^{-[U(\boldsymbol{\theta}) + K(\boldsymbol{\phi})]}\\
&= e^{-U(\boldsymbol{\theta})}e^{-K(\boldsymbol{\phi})}\\
&\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})
\end{aligned} \tag{3.82}$$

We see that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are independent and each has canonical distribution, with energy functions $U(\boldsymbol{\theta})$ and $K(\boldsymbol{\phi})$. The potential energy $U(\boldsymbol{\theta})$ will be defined to be minus the log pdf of the target distribution for $\boldsymbol{\theta}$ that we wish to sample. The kinetic energy $K(\boldsymbol{\phi})$ is usually defined as minus of the log pdf of the zero-mean multivariate normal distribution with covariance matrix M. Therefore,

$$U(\boldsymbol{\theta}) = -\log \pi(\boldsymbol{\theta}|\mathcal{D}) \tag{3.83}$$

$$K(\boldsymbol{\phi}) = \frac{\boldsymbol{\phi}^T M^{-1}\boldsymbol{\phi}}{2} \tag{3.84}$$

Here M is a symmetric, positive definite mass matrix, which is typically diagonal, and is often a scalar multiple of the identity matrix. With these forms for U and K, Hamiltonian

equations (3.76) and (3.77) can be rewritten as follows, for $i = 1, \ldots, d$:

$$\frac{\partial \theta_i}{\partial t} = [M^{-1} \boldsymbol{\phi}]_i \tag{3.85}$$

$$\frac{\partial \phi_i}{\partial t} = \frac{\partial \log \pi(\boldsymbol{\theta}|\mathcal{D})}{\partial \theta_i} \tag{3.86}$$

3.7.4 Hamiltonian Monte Carlo algorithm

HMC uses Hamiltonian dynamics rather than a probability distribution as a proposal function to propose future states for Markov chain in order to explore the target distribution more effectively. Starting with the current state $(\boldsymbol{\theta}, \boldsymbol{\phi})$, we simulate Hamiltonian dynamics for a short time using the leapfrog method. Then the final state of the position and momentum variables of the simulation are used as the proposal states $(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*)$ for Markov chain. The proposed state is accepted according to an update rule which is similar to the Metropolis acceptance rule. Specifically if the probability of the proposed state after Hamiltonian dynamics

$$\pi(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*) \propto e^{-U(\boldsymbol{\theta}^*) - K(\boldsymbol{\phi}^*)} \tag{3.87}$$

is greater than probability of the state prior to the Hamiltonian dynamics

$$\pi(\boldsymbol{\theta}, \boldsymbol{\phi}) \propto e^{-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi})} \tag{3.88}$$

then the proposed state is accepted, otherwise, the proposed state is accepted randomly. If the proposed state is rejected, the next state is the same as the current state. The HMC algorithm for drawing N samples from the target distribution is described as follows:

- Given a starting position state $\boldsymbol{\theta}^{(0)}$

- For $i = 0$ to $N - 1$

  1: Draw a momentum variable $\boldsymbol{\phi}^{(i)} \sim \pi(\boldsymbol{\phi})$
  2: Run the leapfrog algorithm starting at $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\phi}^{(i)})$ for L steps with step-size ϵ to obtain proposed state $(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*)$
  3: Calculate the acceptance probability

  $$r = \min\left(1, e^{U(\boldsymbol{\theta}^{(i)}) - U(\boldsymbol{\theta}^*) + K(\boldsymbol{\phi}^{(i)}) - K(\boldsymbol{\phi}^*)}\right) \tag{3.89}$$

  4: Draw $u \sim U(0, 1)$
  5: Set

  $$\boldsymbol{\theta}^{(i+1)} = \begin{cases} \boldsymbol{\theta}^* & \text{if } u \leq r \\ \boldsymbol{\theta}^{(i)} & \text{otherwise} \end{cases} \tag{3.90}$$

There is no need to keep track of $\boldsymbol{\phi}^{(i)}$ after the accept/reject step because we do not care about it in itself, and it immediately get updated at the beginning of the next iteration.

3.7.5 Restricted parameters and areas of zero density

HMC is build to work with all continuous positive target densities. If the algorithm reaches a point of zero density at any point during an iteration, the stepping will be stopped and given up, spending another iteration at the previous value of location $\boldsymbol{\theta}$. This resulting algorithm allows the chain stays in the positive area and preserves detailed balance. An alternative way is to check if the density is positive after each step and, otherwise, change

the sign of the momentum to return to the direction in which it came. Another usual way to handle bounded parameters is to use transformation, e.g. taking the logarithm of a positive parameter or the logit for a parameter restricted to fall between 0 and 1, or more complicated joint transformation [22].

3.7.6 Setting the tuning parameters and the no-U-turn sampler

Beside the choice of the momentum distribution (usually a multivariate normal distribution with mean 0 and covariance set to a prespecified 'mass matrix' M), the efficiency of the HMC depends also on the choice of the scaling factor ϵ of the leapfrog step, and the number of leapfrog steps L per iteration. Recently, a variance of HMC called no-U-turn sampler (NUTS) [30] has been proposed in order to automatically update these parameters during the burn-in (or warm-up) period and then held fixed during the later iteration. For statistical software R, the Rstan package [66] is used to sample from posterior distribution. Rstan is the R interface to Stan [67] which provides full Bayesian inference using NUTS.

Chapter 4

Non-linear failure rate model: A Bayes study using Markov chain Monte Carlo simulation

4.1 Introduction

This chapter comes from my study given in [85]. The exponential distribution is often used in reliability studies as the model for the time until failure of a device. The lack of memory property of the exponential distribution implies that the device does not wear out. The lifetime of a device with failures caused by random shocks might be appropriately modeled as an exponential distribution. However, the lifetime T of a device that suffers slow mechanical wear, such as bearing wear, is better modeled by a distribution such that $P(T < t + \Delta t | T > t)$ increases with t. Distributions such as the Weibull distribution are often used in practice to model the failure time of this type of device [45].

However, the failure rate of Weibull distribution, $h(t) = bt^{k-1}$, equals 0 at time $t = 0$. This model might only be suitable for modeling some physical systems that do not suffer from external random shocks. Some physical systems where from the past experiences the random shocks have been studied, required corrections. For physical systems which suffer from external random shocks and also from wear out failures, this model might be inappropriate. In this regard, the linear failure rate (LFR), $h(t) = a + bt$, provides partly a solution.

The LFR was first introduced by Kodlin [34], and had been studied by Bain [4] as a special case of polynomial failure rate model for type II censoring. It can be considered a generalization of the exponential model ($b = 0$) or the Rayleigh model ($a = 0$). It can also be regarded as a mixture of failure rates of an exponential and a Rayleigh distributions. Notice that the mixture of failure rates (or mixture failure rate) is not necessary to be the convex combination of failure rates as the mixture distribution and the simplest form is just an addition of failure rates.

Bayesian estimation technique of the LFR model was first considered by Pandey, Singh, and Zimmer [50]. The LFR has been mentioned in other reliability publications (e.g. Lawless [40] and Lai and Xie [38]) and has been applied for solving many real world problems. Because of the limitation of the Rayleigh failure rate, as well as the LFR, which is not flexible to capture the most common types of failure rate, new generalizations of LFR must be developed.

In this chapter, a generalized version of the LFR called non-linear failure rate (NLFR) is introduced. It is considered as a mixture of the exponential and Weibull failure rates. This mixture failure rate not only allows for an initial positive failure rate but also takes into account all shapes of Weibull failure rate. The first research work which attempts to solve the NLFR is given by Salem [54] but solutions were hard to obtain due to computational

difficulties. In addition, this generalization differs from Salem [54] in other details. Such model was also introduced by Sarhan [56], Bousquet, Bertholon, and Celeux [10], and Sarhan and Zaindin [58], but with different name, motivation, parameterization, model explanation and purpose. This model can also be considered as a special case of the 4-parameter model introduced by Xie and Lai [75].

In addition to the model, the modern computational methods known as the cross-entropy (CE) [53] and Markov chain Monte Carlo (MCMC) methods are exploited. Mixture models often result in too many parameters. For example, the models in this study and by Salem [54], Sarhan [56], Bousquet, Bertholon, and Celeux [10], and Sarhan and Zaindin [58] have 3 parameters; the model introduced by Xie and Lai [75] has 4 parameters; the models by Almalki and Yuan [3] and He, Cui, and Du [28] have 5 parameters; the model by Wang [72] has 6 parameters. Maximum likelihood estimate (MLE) of such model parameters is based on the log-likelihood function. However, traditional methods of maximization of a log-likelihood function of such mixture models sometimes do not provide the expected result due to multiple optimal points. For mixture modes, we usually rely on the EM algorithm [42]. However, the EM is a local search procedure and therefore there is no guarantee that it converges to the global maximum. As an alternative to EM, the CE algorithm, which most of the time will provide the global maximum, is used. Bayesian estimation is considered under the symmetric and asymmetric loss functions. Bayesian approach to parameter estimation of lifetime distributions has been considered by many authors such as Gupta, Mukherjee, and Upadhyay [26], Soliman et al. [64], and Kundu and Howlader [36]. Likewise, MCMC methods make Bayesian inference for such models be easier and more practical. However, the efficiency of some MCMC algorithms relies on a good choice of the proposal distribution. Here the adaptive MCMC algorithm is used which allows us to sample from posterior distribution without intervening in the proposal distribution but still provides sample better than traditional MCMC algorithms and converges quickly to the target distribution. In comparison with the Hamiltonian Monte Carlo, the adaptive MCMC might explore the target density a bit slower. However, it is easier to implement and can be run in any usual computer.

The benchmark data sets, e.g. the aircraft windshield failure data [9], the lifetime data of male mice exposed to 300 rads of radiation [33], and the U.S.S. Halfbeak Diesel engine data [43] are examples showing that the NLFR model is more suitable than other mixture distributions for modeling, especially when in data files occur failures with more than one failure mode.

The chapter is organized as follows. The intended NLFR model is introduced in Section 4.2 along with the basic characteristics of the corresponding lifetime distribution. Section 4.3 introduces MLEs of the unknown parameters and reliability characteristics of the model on one hand, and the Bayesian estimators under symmetric and asymmetric loss functions on the other hand. Section 4.4 provides a simulation study to compare Bayes estimators under symmetric and asymmetric loss functions with the MLEs. Section 4.5 provides the applications of the NLFR model to some real data sets. Finally, Sections 4.6 and 4.7 bring discussion and conclusions, respectively.

4.2 The model

4.2.1 Non-linear failure rate model

From the beginning and early age of operation, a physical system suffers only from the external random shocks which means that the failure rate evinces a constant course. When the system wears out, due to mechanical wear, it also experiences a wear out failure mode. So let the failure rate function of the system in these two situations be in either of the

following two states: (1) initially, it experiences a constant failure model, and (2) when the system wears out, it also experiences a wear out failure model. That is

$$h(t) = a + bt^{k-1}, \quad a, b, k, t > 0 \tag{4.1}$$

This model allows an initial positive failure rate, $h(0) = a$, whereas $h(0) = 0$ for most other increasing failure rate function. As mentioned by Bain [4] for the LFR, this type of situation would exist if failures result from random as well as wear out or deterioration course. In fact, the sum of two components in Eq. (4.1) results from the assumption of two independent competing risks, i.e. the random shock and wear out. This model is also useful for modeling a series system with two independent components. One component follows the exponential distribution and another component follows the Weibull distribution.

This study is not intending to model shocks themselves. This study models their consequence to the failure rate. That is why, mixtures is considered to increase the model flexibility and keep its sufficiently simple form and also interpretation.

Fig. 4.1 shows various shapes of the NLFR given in Eq. (4.1) which characterizes three of the most common types of failure rate functions. That is increasing, decreasing and constant failure rates.
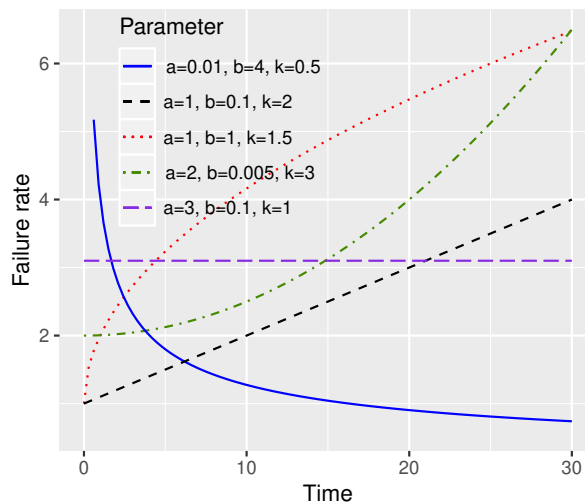
Figure 4.1: NLFR model with decreasing failure rate ($a = 0.01, b = 4, k = 0.5$), linear failure rate ($a = 1, b = 0.1, k = 2$), concave increasing failure rate ($a = 1, b = 1, k = 1.5$), convex increasing failure rate ($a = 2, b = 0.005, k = 3$) and constant failure rate ($a = 3, b = 0.1, k = 1$).

4.2.2 Characteristics of the lifetime distribution

Based on the relationship between failure rate and reliability functions, we have

$$
\begin{aligned}
R(t) &= \exp\left\{ -\int_0^t h(s)ds \right\} \\
&= \exp\left\{ -\int_0^t \left(a + bs^{k-1} \right) ds \right\} \\
&= \exp\left\{ -\left(at + \frac{b}{k}t^k \right) \right\}
\end{aligned}
\tag{4.2}
$$

Then, the probability density function (PDF) is formulated as

$$f(t) = h(t)R(t)$$
$$= \left(a + bt^{k-1}\right) \exp\left\{-\left(at + \frac{b}{k}t^k\right)\right\} \tag{4.3}$$

The cumulative failure rate (CFR) is calculated by

$$H(t) = \int_0^t h(s)ds = at + \frac{b}{k}t^k \tag{4.4}$$

And the MTTF is given by

$$MTTF = \mathbb{E}(T)$$
$$= \int_0^\infty R(t)dt = \int_0^\infty \exp\left\{-\left(at + \frac{b}{k}t^k\right)\right\}dt \tag{4.5}$$

We can use some suitable numerical methods to evaluate this integral.

4.3   Estimation of parameters and reliability characteristics

Let $\mathcal{D}: t_1, ..., t_n$ be the random failure times of n devices under test whose failure time distribution is given as in Eq. (4.3) and $\boldsymbol{\theta} = (a, b, k)$. If there is no censoring, the likelihood function takes the general form

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i|\boldsymbol{\theta})$$
$$= \left[\prod_{i=1}^n \left(a + bt_i^{k-1}\right)\right] \exp\left\{-\sum_{i=1}^n \left(at_i + \frac{b}{k}t_i^k\right)\right\} \tag{4.6}$$

Then, the log-likelihood function can be written as

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^n \log f(t_i|\boldsymbol{\theta})$$
$$= \sum_{i=1}^n \log\left(a + bt_i^{k-1}\right) - \sum_{i=1}^n \left(at_i + \frac{b}{k}t_i^k\right) \tag{4.7}$$

If some observations are censored, we have to make an adjustment to this expression. For an observation of an observed failure, we put in the pdf as above. But for a right-censored observation, we put in the reliability function, indicating that observation is known to exceed a particular value. The likelihood function in general then takes the form

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i|\boldsymbol{\theta})^{\delta_i} R(t_i|\boldsymbol{\theta})^{1-\delta_i}$$
$$= \prod_{i=1}^n h(t_i|\boldsymbol{\theta})^{\delta_i} R(t_i|\boldsymbol{\theta})$$
$$= \left[\prod_{i=1}^n \left(a + bt_i^{k-1}\right)^{\delta_i}\right] \exp\left\{-\sum_{i=1}^n \left(at_i + \frac{b}{k}t_i^k\right)\right\} \tag{4.8}$$

This expression means that when t_i is an observed failure, the censoring indicator is $\delta_i = 1$, and we enter a pdf factor. When t_i is a censored observation, we have $\delta_i = 0$, and we enter a reliability factor [46].

Then, the log-likelihood function for censoring case is

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \log h(t_i|\boldsymbol{\theta}) + \sum_{i=1}^{n} \log R(t_i|\boldsymbol{\theta})$$
$$= \sum_{i=1}^{n} \delta_i \log \left(a + bt_i^{k-1} \right) - \sum_{i=1}^{n} \left(at_i + \frac{b}{k}t_i^k \right) \tag{4.9}$$

4.3.1 Maximum likelihood estimation

In this study, the log-likelihood function Eq. (4.7) in case of non-censored data or Eq. (4.9) in case of censored data is maximized using CE algorithm to produce the maximizer $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{k})$. Then, by using the invariance property of MLE's,

1. The MLE for $R(t)$, say $\hat{R}(t)$, will be

$$\hat{R}(t) = \exp \left\{ -\left(\hat{a}t + \frac{\hat{b}}{\hat{k}}t^{\hat{k}} \right) \right\} \tag{4.10}$$

2. The MLE for $h(t)$, say $\hat{h}(t)$, will be

$$\hat{h}(t) = \hat{a} + \hat{b}t^{\hat{k}-1} \tag{4.11}$$

3. The MLE for $H(t)$, say $\hat{H}(t)$, will be

$$\hat{H}(t) = \hat{a}t + \frac{\hat{b}}{\hat{k}}t^{\hat{k}} \tag{4.12}$$

4. The MLE for MTTF will be

$$\hat{MTTF} = MTTF(\hat{a}, \hat{b}, \hat{k}) \tag{4.13}$$

which can be obtained by installing into formula (4.5) and integrating.

4.3.2 Bayesian estimation

For the mixture model in this study, the Bayesian model is constructed by specifying a prior distribution for $\boldsymbol{\theta} = (a, b, k)$, and then multiplying with the likelihood function to obtain the posterior distribution. Denote the prior distribution of $\boldsymbol{\theta}$ as $\pi(\boldsymbol{\theta})$, the posterior distribution of $\boldsymbol{\theta}$ given $\mathcal{D}: t_1, \ldots, t_n$ is given by

$$\pi(\boldsymbol{\theta}|\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{4.14}$$

Because the denominator in Eq. (4.14) is a normalizing constant, Bayes' theorem is often expressed as:

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{4.15}$$

As mentioned by Jiang, Xie, and Tang [32], the prior distribution is given beforehand, usually based on prior information of the parameters, such as that from historical data, previous experiences, expert suggestions, even wholly subjective suppositions, or simply from the point of mathematical conveniences.

Here, adopted Kundu and Howlader [36], $a, b,$ and k are assumed to have independent gamma(α_1, β_1), gamma(α_2, β_2), and gamma(α_3, β_3) priors respectively, i.e.

$$\pi_1(a) \propto a^{\alpha_1 - 1} \exp\left\{-\beta_1 a\right\}, \, \alpha_1, \beta_1 > 0 \tag{4.16}$$

$$\pi_2(b) \propto b^{\alpha_2 - 1} \exp\left\{-\beta_2 b\right\}, \, \alpha_2, \beta_2 > 0 \tag{4.17}$$

$$\pi_3(k) \propto k^{\alpha_3 - 1} \exp\left\{-\beta_3 k\right\}, \, \alpha_3, \beta_3 > 0 \tag{4.18}$$

If $\alpha_1 = \alpha_2 = \alpha_3 = 1, \beta_1 = \beta_2 = \beta_3 = 0$ we have a generalized uniform distribution on $\mathbb{R}^+$ [32] or a diffuse prior [6], and if $\alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$, we have a non-informative prior. Based on the assumption of independence of individual priors, the joint prior of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}) = \pi_1(a)\pi_2(b)\pi_3(k) \tag{4.19}$$

Since there is no prior information available, the diffuse priors on the model parameters are used in the rest of the chapter.

Then, under the square error loss function, the Bayes estimator of a, b, k, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$a_{BS}^* = \mathbb{E}(a|\mathcal{D}) = \int_{\boldsymbol{\theta}} a\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{4.20}$$

$$b_{BS}^* = \mathbb{E}(b|\mathcal{D}) = \int_{\boldsymbol{\theta}} b\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{4.21}$$

$$k_{BS}^* = \mathbb{E}(k|\mathcal{D}) = \int_{\boldsymbol{\theta}} k\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{4.22}$$

$$h_{BS}^*(t) = \mathbb{E}(h(t; \boldsymbol{\theta})|\mathcal{D}) = \int_{\boldsymbol{\theta}} h(t; \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{4.23}$$

$$R_{BS}^*(t) = \mathbb{E}(R(t; \boldsymbol{\theta})|\mathcal{D}) = \int_{\boldsymbol{\theta}} R(t; \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{4.24}$$

In this study, the adaptive MCMC sampling [59] is used to generate sample $\{\boldsymbol{\theta}_i = (a_i, b_i, k_i), i = 1, \ldots, n\}$ from the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$. We assume that this sample has been taken after burn-in (warm-up) period and reducing autocorrelation. Then, the approximate Bayes estimates of $a_{BS}^*, b_{BS}^*, k_{BS}^*, h_{BS}^*(t)$ and $R_{BS}^*(t)$ are given respectively by

$$a_{BS}^* \approx \frac{1}{n} \sum_{i=1}^{n} a_i \tag{4.25}$$

$$b_{BS}^* \approx \frac{1}{n} \sum_{i=1}^{n} b_i \tag{4.26}$$

$$k_{BS}^* \approx \frac{1}{n} \sum_{i=1}^{n} k_i \tag{4.27}$$

$$h_{BS}^*(t) \approx \frac{1}{n} \sum_{i=1}^{n} h(t; \boldsymbol{\theta}_i) \tag{4.28}$$

$$R_{BS}^*(t) \approx \frac{1}{n} \sum_{i=1}^{n} R(t; \boldsymbol{\theta}_i) \tag{4.29}$$

Under the linear exponential loss function, the Bayes estimator of a, b, k, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$a_{BL}^* = -\frac{1}{c} \log \left(\int_{\boldsymbol{\theta}} e^{-ca} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right) \tag{4.30}$$

$$b_{BL}^* = -\frac{1}{c} \log \left(\int_{\boldsymbol{\theta}} e^{-cb} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right) \tag{4.31}$$

$$k_{BL}^* = -\frac{1}{c} \log \left(\int_{\boldsymbol{\theta}} e^{-ck} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right) \tag{4.32}$$

$$h_{BL}^*(t) = -\frac{1}{c} \log \left(\int_{\boldsymbol{\theta}} e^{-ch(t;\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right) \tag{4.33}$$

$$R_{BL}^*(t) = -\frac{1}{c} \log \left(\int_{\boldsymbol{\theta}} e^{-cR(t;\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right) \tag{4.34}$$

Then, the approximate Bayes estimates of a_{BL}^*, b_{BL}^*, k_{BL}^*, $h_{BL}^*(t)$ and $R_{BL}^*(t)$ are given respectively by

$$a_{BL}^* \approx -\frac{1}{c} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{-ca_i} \right) \tag{4.35}$$

$$b_{BL}^* \approx -\frac{1}{c} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{-cb_i} \right) \tag{4.36}$$

$$k_{BL}^* \approx -\frac{1}{c} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{-ck_i} \right) \tag{4.37}$$

$$h_{BL}^*(t) \approx -\frac{1}{c} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{-ch(t;\boldsymbol{\theta}_i)} \right) \tag{4.38}$$

$$R_{BL}^*(t) \approx -\frac{1}{c} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{-cR(t;\boldsymbol{\theta}_i)} \right) \tag{4.39}$$

Under the general entropy loss function, the Bayes estimator of a, b, k, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$a_{BG}^* = \left(\int_{\boldsymbol{\theta}} a^{-c} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right)^{-\frac{1}{c}} \tag{4.40}$$

$$b_{BG}^* = \left(\int_{\boldsymbol{\theta}} b^{-c} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right)^{-\frac{1}{c}} \tag{4.41}$$

$$k_{BG}^* = \left(\int_{\boldsymbol{\theta}} k^{-c} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right)^{-\frac{1}{c}} \tag{4.42}$$

$$h_{BG}^*(t) = \left(\int_{\boldsymbol{\theta}} h(t;\boldsymbol{\theta})^{-c} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right)^{-\frac{1}{c}} \tag{4.43}$$

$$R_{BG}^*(t) = \left(\int_{\boldsymbol{\theta}} R(t;\boldsymbol{\theta})^{-c} \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \right)^{-\frac{1}{c}} \tag{4.44}$$

Then, the approximate Bayes estimates of a^*_{BG}, b^*_{BG}, k^*_{BG}, $h^*_{BG}(t)$ and $R^*_{BG}(t)$ are given respectively by

$$a^*_{BG} \approx \left(\frac{1}{n} \sum_{i=1}^{n} a_i^{-c} \right)^{-\frac{1}{c}} \tag{4.45}$$

$$b^*_{BG} \approx \left(\frac{1}{n} \sum_{i=1}^{n} b_i^{-c} \right)^{-\frac{1}{c}} \tag{4.46}$$

$$k^*_{BG} \approx \left(\frac{1}{n} \sum_{i=1}^{n} k_i^{-c} \right)^{-\frac{1}{c}} \tag{4.47}$$

$$h^*_{BG}(t) \approx \left(\frac{1}{n} \sum_{i=1}^{n} h(t; \boldsymbol{\theta}_i)^{-c} \right)^{-\frac{1}{c}} \tag{4.48}$$

$$R^*_{BG}(t) \approx \left(\frac{1}{n} \sum_{i=1}^{n} R(t; \boldsymbol{\theta}_i)^{-c} \right)^{-\frac{1}{c}} \tag{4.49}$$

4.4   Monte Carlo simulations

A Monte Carlo simulation study is conducted to compare the performance of CE and MCMC estimators for the parameters of the non-linear failure rate. For each of the following choice of parameters, 1000 data sets are simulated with sample size $n = 25, 50, 100$ and 200, respectively, and based on each data set the CE and MCMC estimators for the model parameters are computed. The data sets are generated from the failure distribution Eq.(4.3) using the accept-reject method (Chapter 3). In order to obtain MCMC estimators, the adaptive MCMC algorithm is implemented to construct a Markov chain of length 20,000 with burn-in of 1000 and reduced autocorrelation by retaining only every 5 iterations of the chain. The Bayes estimators are obtained under the three loss functions, i.e. SEL, LINEX loss and GEL. The parameters of the asymmetric loss functions are selected as $c = -1.5$ (c_1), -0.5 (c_2), 0.5 (c_3) and 1.5 (c_4). Note here that when $k = 1$, the studying model coincides with the exponential model with constant failure rate $\lambda = a + b$.

1. $a = 0.03, b = 0.07$, and $k = 0.5$

2. $a = 0.03, b = 0.07$, and $k = 2$

3. $a = 0.03, b = 0.07$, and $k = 3$

Tables 4.1-4.3 list the results of the simulation study. The mean square error (MSE) is calculated as the mean of the squared differences between 1000 estimators and the true value.

It is clear from Tables 4.1-4.3 that the MSEs of a and b are relatively small. The most interesting thing here is the MSEs of k. In the three selected cases of k:

- The MSEs of Bayes estimators under LINEX loss function when $c = -1.5, -0.5, 0.5$ and 1.5 decrease respectively (see Figs. 4.2, 4.5 and 4.8). And this fact is also true under GEL (see Figs. 4.3, 4.6 and 4.9).

- In cases $k = 2$ or 3, the MSEs of Bayes estimators under LINEX (when $c > 0$) seem to be always smaller than the MSEs of Bayes estimators under GEL in comparison to each value of c (see Figs. 4.7 and 4.10). This fact is opposite when $k = 0.5$ (see Fig. 4.4).

Table 4.1: MSEs of MLEs (namely CE) and Bayes estimators under SEL, LINEX, and GEL when $k = 0.5$

θ	n	CE	SEL	LINEX (LEL)				GEL			
				$c_1 = -1.5$	$c_2 = -0.5$	$c_3 = 0.5$	$c_4 = 1.5$	$c_1 = -1.5$	$c_2 = -0.5$	$c_3 = 0.5$	$c_4 = 1.5$
a	25	0.0003	0.0002	0.0002	0.0002	0.0002	0.0001	0.0002	0.0001	0.0002	0.0005
a	50	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0004
a	100	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0004
a	200	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003
b	25	0.0016	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0009	0.0015	0.0028
b	50	0.0009	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0008	0.0014
b	100	0.0006	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0006
b	200	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0002	0.0003	0.0003
k	25	0.0485	0.0241	0.0424	0.0285	0.0204	0.0151	0.0294	0.0195	0.0124	0.0103
k	50	0.0253	0.0118	0.0162	0.0130	0.0108	0.0092	0.0133	0.0105	0.0086	0.0081
k	100	0.0134	0.0077	0.0088	0.0081	0.0075	0.0070	0.0082	0.0074	0.0069	0.0068
k	200	0.0072	0.0056	0.0060	0.0057	0.0055	0.0053	0.0058	0.0054	0.0052	0.0050

Table 4.2: MSEs of MLEs (namely CE) and Bayes estimators under SEL, LINEX, and GEL when $k = 2$

θ	n	CE	SEL	LINEX (LEL)				GEL			
				$c_1 = -1.5$	$c_2 = -0.5$	$c_3 = 0.5$	$c_4 = 1.5$	$c_1 = -1.5$	$c_2 = -0.5$	$c_3 = 0.5$	$c_4 = 1.5$
a	25	0.0013	0.0028	0.0030	0.0028	0.0027	0.0025	0.0037	0.0019	0.0005	0.0005
a	50	0.0009	0.0012	0.0013	0.0013	0.0012	0.0011	0.0016	0.0009	0.0004	0.0006
a	100	0.0006	0.0007	0.0007	0.0007	0.0007	0.0006	0.0008	0.0005	0.0004	0.0006
a	200	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0004	0.0003	0.0003	0.0005
b	25	0.0013	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0006	0.0015	0.0032
b	50	0.0009	0.0005	0.0005	0.0005	0.0005	0.0005	0.0004	0.0005	0.0010	0.0019
b	100	0.0006	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0004	0.0007	0.0011
b	200	0.0004	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0004	0.0006
k	25	0.0669	0.1549	0.5043	0.2135	0.1217	0.0967	0.1710	0.1419	0.1256	0.1284
k	50	0.0415	0.1079	0.2494	0.1378	0.0874	0.0633	0.1173	0.0995	0.0858	0.0757
k	100	0.0274	0.0693	0.1205	0.0814	0.0600	0.0467	0.0736	0.0654	0.0585	0.0527
k	200	0.0168	0.0371	0.0517	0.0408	0.0340	0.0290	0.0386	0.0358	0.0333	0.0312

Table 4.3: MSEs of MLEs (namely CE) and Bayes estimators under SEL, LINEX, and GEL when $k = 3$

θ	n	CE	SEL	LINEX (LEL)				GEL			
				$c_1 = -1.5$	$c_2 = -0.5$	$c_3 = 0.5$	$c_4 = 1.5$	$c_1 = -1.5$	$c_2 = -0.5$	$c_3 = 0.5$	$c_4 = 1.5$
a	25	0.0012	0.0023	0.0025	0.0024	0.0022	0.0021	0.0031	0.0016	0.0006	0.0007
a	50	0.0007	0.0011	0.0011	0.0011	0.0011	0.0010	0.0014	0.0008	0.0005	0.0007
a	100	0.0004	0.0005	0.0005	0.0005	0.0005	0.0005	0.0006	0.0004	0.0004	0.0006
a	200	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0003	0.0002	0.0003	0.0004
b	25	0.0017	0.0012	0.0013	0.0012	0.0011	0.0011	0.0014	0.0010	0.0011	0.0020
b	50	0.0010	0.0007	0.0007	0.0007	0.0007	0.0007	0.0008	0.0007	0.0008	0.0012
b	100	0.0007	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0006	0.0007
b	200	0.0006	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
k	25	0.1211	0.2774	0.9352	0.3904	0.2181	0.1845	0.2972	0.2606	0.2351	0.2202
k	50	0.0795	0.1817	0.4082	0.2293	0.1509	0.1192	0.1915	0.1729	0.1579	0.1462
k	100	0.0672	0.1062	0.1607	0.1199	0.0957	0.0818	0.1094	0.1032	0.0979	0.0935
k	200	0.0552	0.0491	0.0596	0.0520	0.0468	0.0437	0.0499	0.0485	0.0473	0.0463
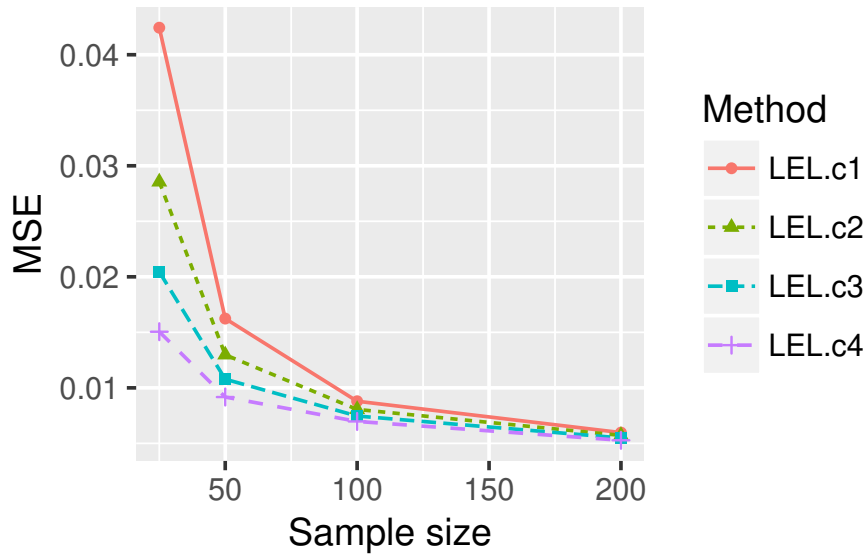
Figure 4.2: MSEs of Bayes estimators under LINEX when $k = 0.5$
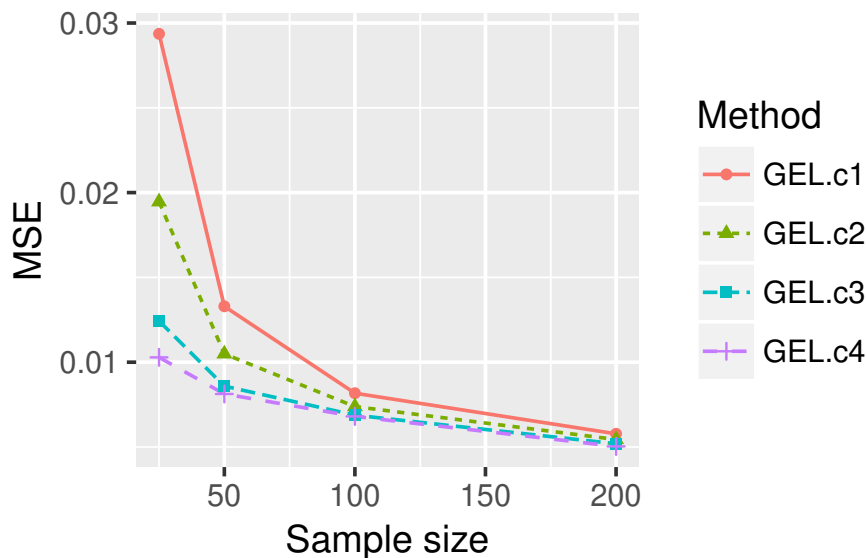
Figure 4.3: MSEs of Bayes estimators under GEL when $k = 0.5$

- The MSEs of Bayes estimators under asymmetric loss function (when $c > 0$) seem to be always smaller than the MSEs of Bayes estimators under SEL (see Figs. 4.4, 4.7 and 4.10).

- When $k = 0.5$, the MSEs of Bayes estimators under asymmetric loss functions seem to be always smaller than the MSEs of CE estimators (see Fig. 4.4), and this fact is opposite in cases $k = 2$ or 3 (see Figs. 4.7 and 4.10).

- The MSEs of Bayes estimators under GEL (when $c < 0$) seem to be always smaller than the MSEs of Bayes estimators under LINEX loss in all cases (see Figs. 4.11, 4.12 and 4.13).
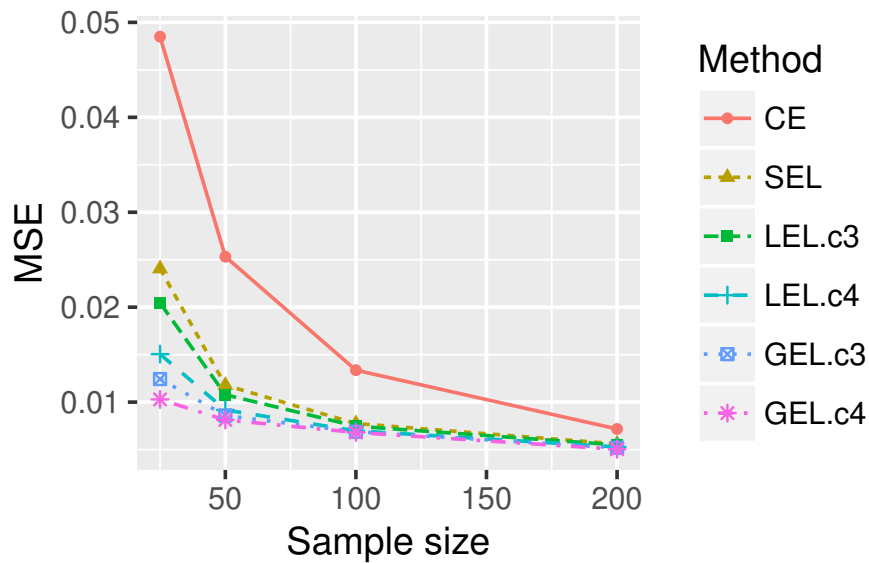
Figure 4.4: MSEs of CE estimate and Bayes estimators under SEL, LINEX and GEL when $k = 0.5$
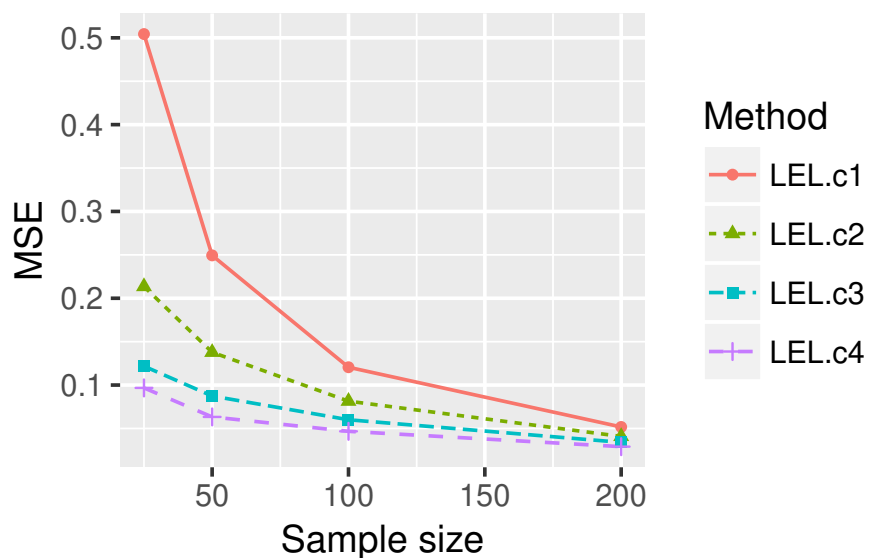
Figure 4.5: MSEs of Bayes estimators under LINEX when $k = 2$

4.5 Illustrative examples

In this Section, three examples are presented to illustrate the estimate procedures discussed in this paper, and only the Bayes estimators under SEL function is shown.

4.5.1 The aircraft windshield failure data

Table 4.4 contains the failure data of aircraft windshields [9]. Among 153 observations, there are 88 failed observed windshields and 65 censored observations. The unit for measurement of this data is 1000 hours.

Bayes estimates via MCMC and MLEs via CE for the parameters and reliability characteristics are provided. Table 4.5 shows Bayes estimates obtained by using adaptive MCMC
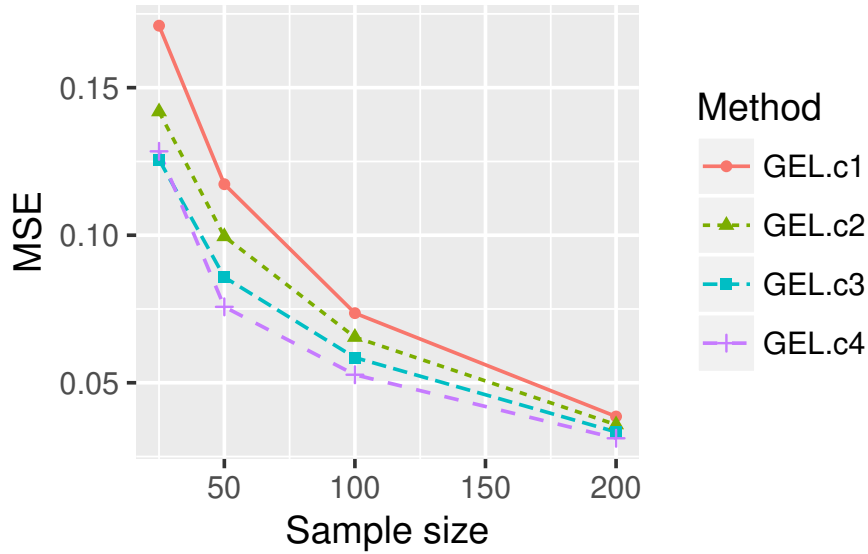
Figure 4.6: MSEs of Bayes estimators under GEL when $k = 2$
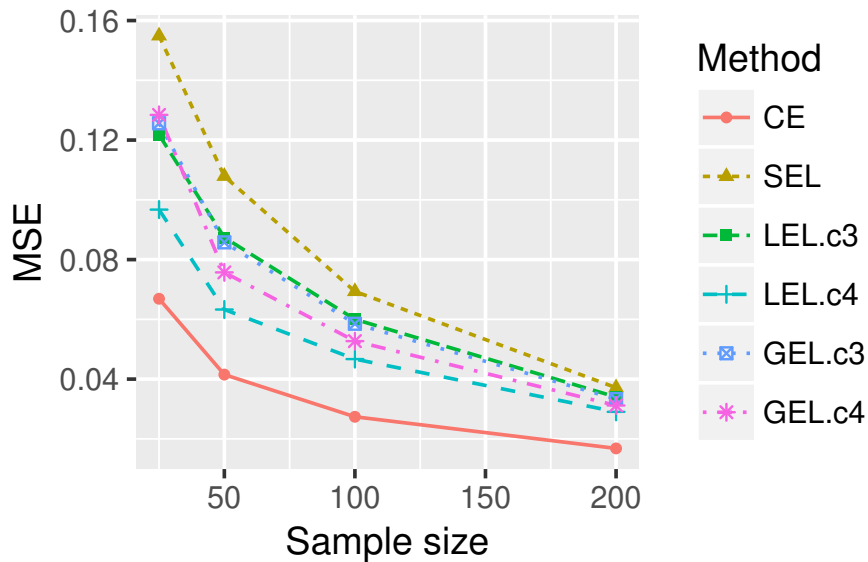
Figure 4.7: MSEs of CE estimate and Bayes estimators under SEL, LINEX and GEL when $k = 2$

along with highest posterior density (HPD) intervals for the parameters and MTTF, and Table 4.6 shows MLEs obtained by using CE along with bootstrap percentile confident (BPC) intervals for the parameters and MTTF. The bootstrap CIs are obtained by using the "censboot" function in the "boot" package [13]. The nonparametric estimation of the failure rate function called 'step function' is performed by divided the time domain into bins of equal width and then estimate the failure rate in each bin as the numbers of events in that bin divided by the number of items at risk in that bin. This estimate is done by using the "pehaz" function in the "muhaz" package [23]. Figs. 4.14-4.15 show trace plots and histograms of a, b and k obtained by adaptive MCMC, and time courses of $R(t), h(t)$ and $H(t)$ are displayed in Figs. 4.16-4.18. From these results, we see that both methods provide almost the same results.
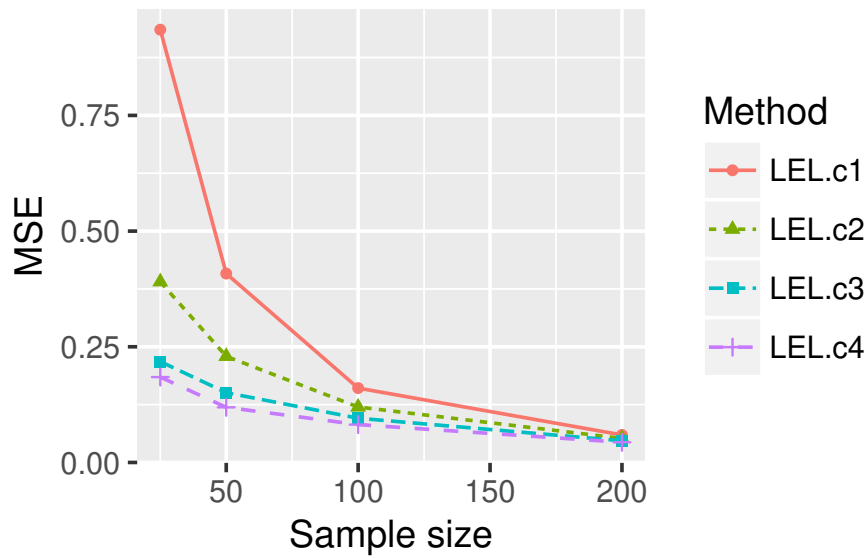
Figure 4.8: MSEs of Bayes estimators under LINEX when $k = 3$
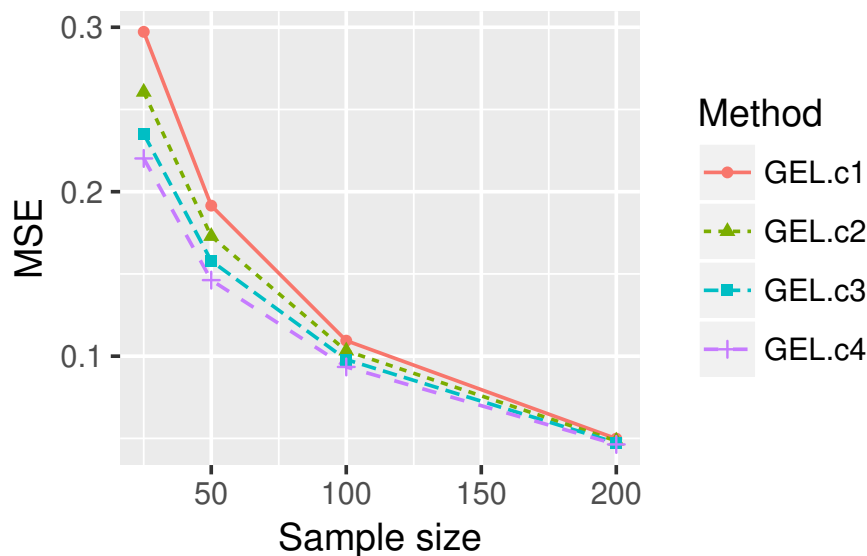
Figure 4.9: MSEs of Bayes estimators under GEL when $k = 3$

Tables 4.7-4.8 represent the Akaike Information Criterion (AIC) values and MLEs of parameters for mixture models, respectively. The Akaike Information Criterion (AIC) values for the first seven models have been provided by Blischke, Karim, and Murthy [9]. From Table 4.7 we see that NLFR model has smallest AIC value. Therefore, it is considered to be the best approximating model among the given models.

4.5.2 Male mice exposed to 300 rads data

Data in Table 4.9 represent the days until death for male mice exposed to 300 rads of radiation. The unit for measurement is 1000 days. Here only the group maintained in a germ-free environment is considered and the causes of death is due to the effect of other causes. The new feature of male mice data is that more than one failure mode occurs [33].
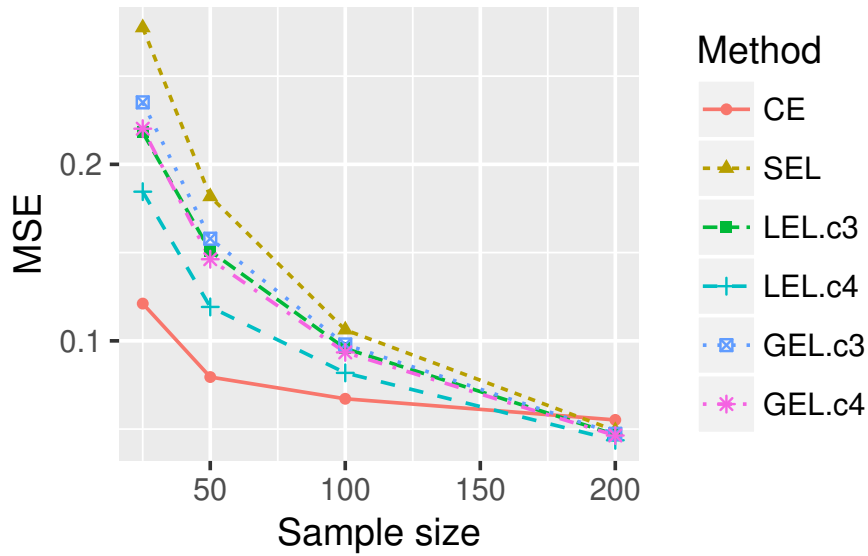
Figure 4.10: MSEs of CE estimate and Bayes estimators under SEL, LINEX and GEL when $k = 3$
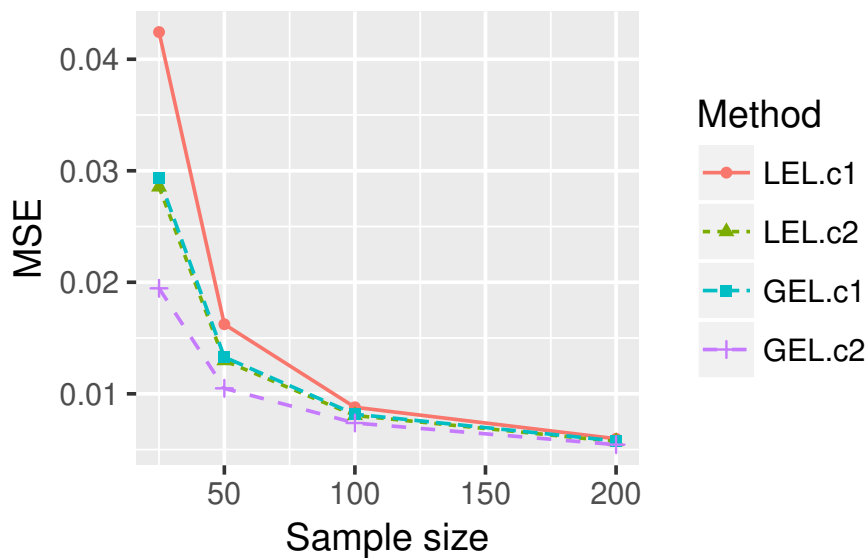
Figure 4.11: MSEs of Bayes estimators under LINEX and GEL when $k = 0.5$

Here the same procedure described in Subsection 4.5.1 is used for MCMC. Table 4.10 shows MCMC point estimates and two-sided 90% and 95% HPD intervals for a, b, k and MTTF. Table 4.11 shows CE point estimates and two-sided 90% and 95% BCa (bias corrected and accelerated) bootstrap confident intervals for a, b, k and MTTF. The bootstrap CIs are obtained by using the "boot" function in the "boot" package [13].

Figs. 4.19-4.20 show posterior distributions and trace plots of each parameter of the Bayesian model obtained by MCMC algorithm, and Figs. 4.21-4.23 show time courses of all relevant functions obtained by both CE and MCMC methods, i.e. reliability, failure rate and cumulative failure rate functions.

From Figs. 4.21-4.23, we see, that although the sample size is small, our fitting models achieved by both CE and MCMC methods are very close to the nonparametric estimate of the reliability characteristics. In case of small datasets, the CE method seems to be a bit
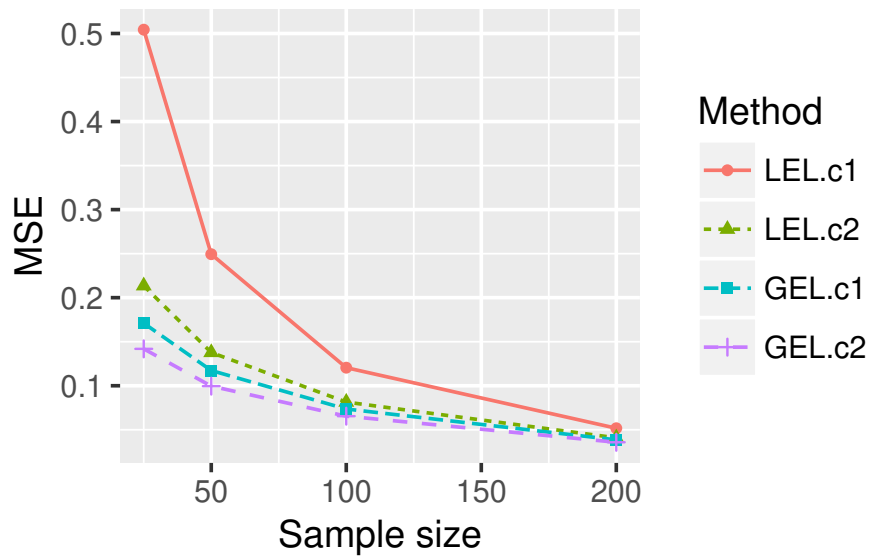
Figure 4.12: MSEs of Bayes estimators under LINEX and GEL when $k = 2$
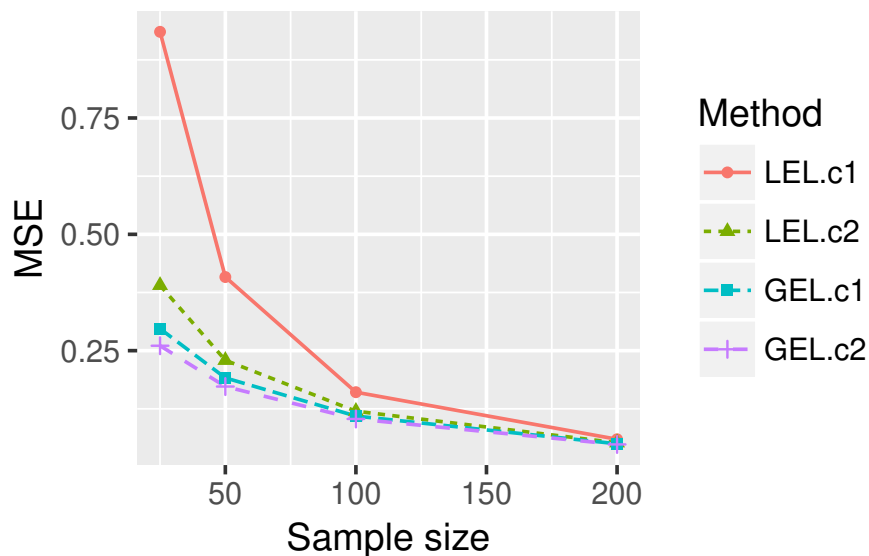
Figure 4.13: MSEs of Bayes estimators under LINEX and GEL when $k = 3$

better than MCMC method. This is due to the stochastic search of CE algorithm. We run the algorithm through trial and error until we get the optimizer point which corresponds to the largest optimum value. The AIC value of NLFR model obtained in this case is -18.813.

4.5.3 U.S.S. Halfbeak Diesel Engine data

Table 4.12 gives times of unscheduled maintenance actions for the U.S.S. Halfbeak number 4 main propulsion diesel engine over $25,518$ operating hours [43]. The unit for measurement is 10,000 hours. The data are times of recurrent events on one machine, hence not typical lifetime data. Here, the time from one maintenance to the following one is assumed as lifetime data in order to demonstrate the proposed model for uncensored data.

The same procedure described in Subsection 4.5.1 is used for MCMC. Table 4.13 shows MCMC point estimators and two-sided 90% and 95% HPD intervals for a, b, k and MTTF.

Table 4.4: Aircraft windshield failure data

Failure Times				Service Times		
0.040	1.866	2.385	3.443	0.046	1.436	2.592
0.301	1.876	2.481	3.467	0.140	1.492	2.600
0.309	1.899	2.610	3.478	0.150	1.580	2.670
0.557	1.911	2.625	3.578	0.248	1.719	2.717
0.943	1.912	2.632	3.595	0.280	1.794	2.819
1.070	1.914	2.646	3.699	0.313	1.915	2.820
1.124	1.981	2.661	3.779	0.389	1.920	2.878
1.248	2.010	2.688	3.924	0.487	1.963	2.950
1.281	2.038	2.823	4.035	0.622	1.978	3.003
1.281	2.085	2.890	4.121	0.900	2.053	3.102
1.303	2.089	2.902	4.167	0.952	2.065	3.304
1.432	2.097	2.934	4.240	0.996	2.117	3.483
1.480	2.135	2.962	4.255	1.003	2.137	3.500
1.505	2.154	2.964	4.278	1.010	2.141	3.622
1.506	2.190	3.000	4.305	1.085	2.163	3.665
1.568	2.194	3.103	4.376	1.092	2.183	3.695
1.615	2.223	3.114	4.449	1.152	2.240	4.015
1.619	2.224	3.117	4.485	1.183	2.341	4.628
1.652	2.229	3.166	4.570	1.244	2.435	4.806
1.652	2.300	3.344	4.602	1.249	2.464	4.881
1.757	2.324	3.376	4.663	1.262	2.543	5.140
1.795	2.349	3.385	4.694	1.360	2.560	
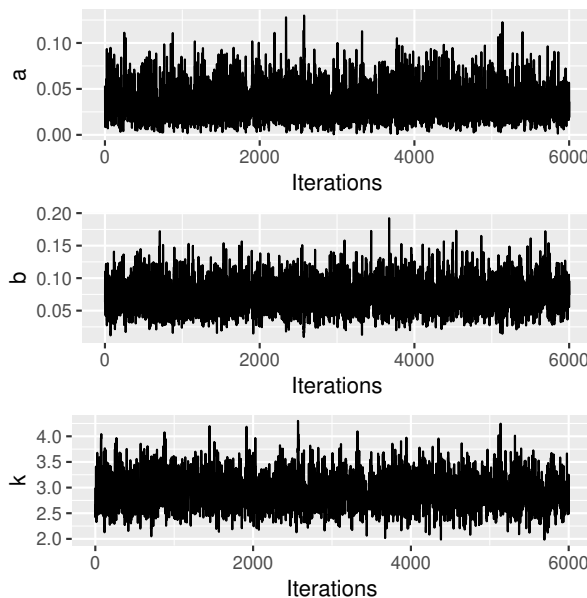
Figure 4.14: Trace plots of a, b and k produced by adaptive MCMC.

Table 4.14 shows CE point estimators and two-sided 90% and 95% BCa bootstrap confident intervals for a, b, k and MTTF.

Figs. 4.24-4.25 show posterior distributions and trace plots of each parameter of the Bayesian model obtained by MCMC algorithm, and Figs. 4.26-4.28 show time courses of all relevant functions obtained by both CE and MCMC methods, i.e. reliability, failure rate
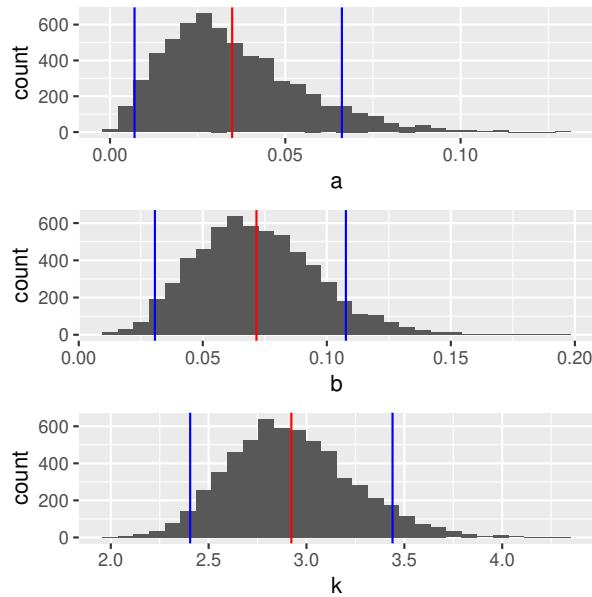
Figure 4.15: Histograms of a, b and k produced by adaptive MCMC.

Table 4.5: Bayes estimates via MCMC and HPD intervals for the parameters and MTTF.

	MCMC	90% HPD Interval	95% HPD Interval
a	0.0348	$[0.0070, 0.0661]$	$[0.0025, 0.0722]$
b	0.0716	$[0.0307, 0.1077]$	$[0.0274, 0.1201]$
k	2.9227	$[2.4056, 3.4397]$	$[2.3223, 3.5605]$
$MTTF$	3.0351	$[2.8177, 3.2302]$	$[2.7918, 3.2895]$

Table 4.6: MLEs via CE and bootstrap percentile confident intervals for the parameters and MTTF.

	CE	90% BPC Interval	95% BPC Interval
a	0.0269	$[0.0000, 0.0569]$	$[0.0000, 0.0626]$
b	0.0692	$[0.0375, 0.1127]$	$[0.0321, 0.1244]$
k	2.9284	$[2.5112, 3.4443]$	$[2.4331, 3.5554]$
$MTTF$	3.0535	$[2.8478, 3.2726]$	$[2.8105, 3.2996]$

and cumulative failure rate functions.

Figs. 4.26-4.28 imply that although the dataset is quite small, the fitting models achieved by both CE and MCMC methods are very close to the nonparametric estimate of the reliability characteristics. In case of moderate datasets, CE method also seems to be slightly better than MCMC method. The AIC value obtained here is 67.348 with respect to the MLEs represented in Table 4.14.
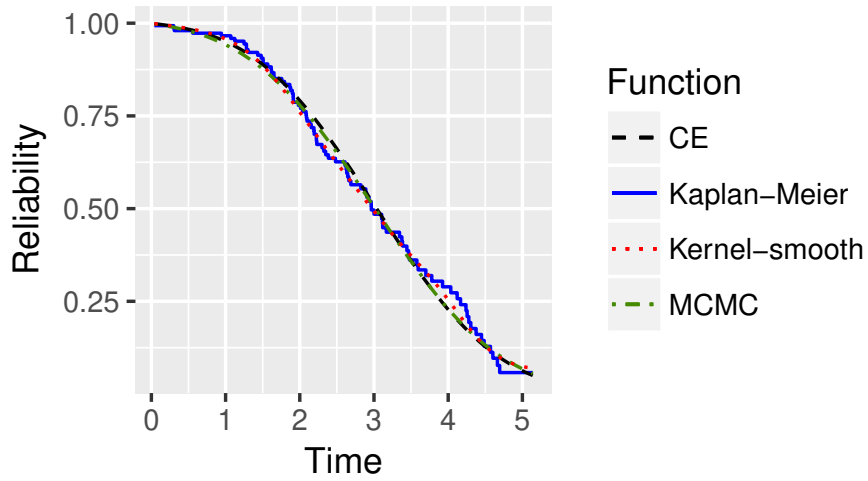
Figure 4.16: Time courses of $R(t)$ when fitting to aircraft windshield failure data.
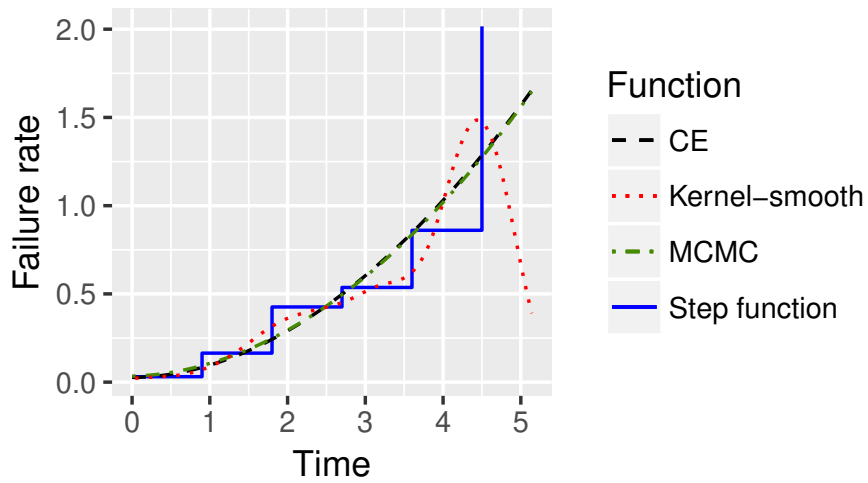
Figure 4.17: Time courses of $h(t)$ when fitting to aircraft windshield failure data.
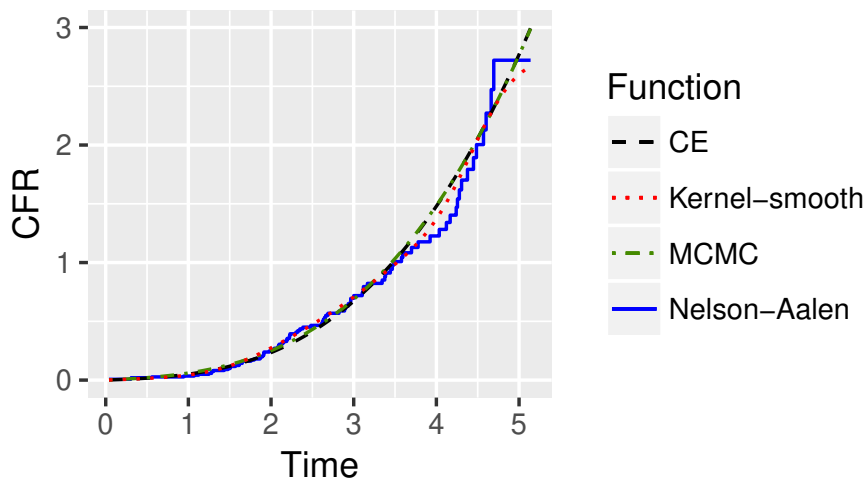
Figure 4.18: Time courses of $H(t)$ when fitting to aircraft windshield failure data.

Table 4.7: Estimated AIC for nine mixture models for aircraft windshield failure data

Mixture models	Models forms and parameters	AIC
1. Weibull-Weibull	$p \times \text{Weib}(\beta_1, \alpha_1) + (1-p) \times \text{Weib}(\beta_2, \alpha_2)$	350.159
2. Weibull-Exponential	$p \times \text{Weib}(\beta_1, \alpha_1) + (1-p) \times \text{Exp}(\lambda)$	348.260
3. Weibull-Normal	$p \times \text{Weib}(\beta_1, \alpha_1) + (1-p) \times \text{Nor}(\mu, \sigma)$	349.710
4. Weibull-Lognormal	$p \times \text{Weib}(\beta_1, \alpha_1) + (1-p) \times \text{Lnor}(\mu, \sigma)$	351.235
5. Normal-Exponential	$p \times \text{Nor}(\mu, \sigma) + (1-p) \times \text{Exp}(\lambda)$	351.513
6. Normal-Lognormal	$p \times \text{Nor}(\mu, \sigma) + (1-p) \times \text{Lnor}(\mu_2, \sigma_2)$	351.579
7. Lognormal-Exponential	$p \times \text{Lnor}(\mu, \sigma) + (1-p) \times \text{Exp}(\lambda)$	349.301
8. Linear failure rate	$h(t) = a + bt$	359.106
9. Non-linear failure rate	$h(t) = a + bt^{k-1}$	347.371

Table 4.8: MLEs of parameters for nine mixture models

Mixture models	MLEs of parameters
1. Weibull-Weibull	$\hat{\beta}_1 = 1.249, \hat{\alpha}_1 = 0.245, \hat{\beta}_2 = 2.777, \hat{\alpha}_2 = 3.485, \hat{p} = 0.017$
2. Weibull-Exponential	$\hat{\beta} = 2.768, \hat{\alpha} = 3.484, \hat{\lambda} = 4.052, \hat{p} = 0.983$
3. Weibull-Normal	$\hat{\beta} = 7.359, \hat{\alpha} = 4.481, \hat{\mu} = 2.303, \hat{\sigma} = 0.868, \hat{p} = 0.387$
4. Weibull-Lognormal	$\hat{\beta} = 1.246, \hat{\alpha} = 0.395, \hat{\mu} = 1.075, \hat{\sigma} = 0.440, \hat{p} = 0.029$
5. Normal-Exponential	$\hat{\mu} = 3.053, \hat{\sigma} = 1.220, \hat{\lambda} = 24.003, \hat{p} = 0.995$
6. Normal-Lognormal	$\hat{\mu}_1 = 0.302, \hat{\sigma}_1 = 0.182, \hat{\mu}_2 = 1.073, \hat{\sigma}_2 = 0.443, \hat{p} = 0.027$
7. Lognormal-Exponential	$\hat{\mu} = 1.080, \hat{\sigma} = 0.435, \hat{\lambda} = 1.729, \hat{p} = 0.965$
8. Linear failure rate	$\hat{a} = 0.004, \hat{b} = 0.161$
9. Non-linear failure rate	$\hat{a} = 0.027, \hat{b} = 0.069, \hat{k} = 2.928$

Table 4.9: Male mice exposed to 300 rads of radiation (other causes in germ-free group)

0.136	0.246	0.255	0.376	0.421	0.565	0.616
0.617	0.652	0.655	0.658	0.660	0.662	0.675
0.681	0.734	0.736	0.737	0.757	0.769	0.777
0.800	0.807	0.825	0.855	0.857	0.864	0.868
0.870	0.870	0.873	0.882	0.895	0.910	0.934
0.943	1.015	1.019				

4.6 Discussion

The results of illustrative examples show that some physical systems that likely suffer from random shock always have positive initial failure rate, i.e. $a > 0$. Like the aircraft windshield failure data and the U.S.S. Halfbeak Diesel engine data, failures actually suffer from random shocks which are unable to avoid during their operation. These results also
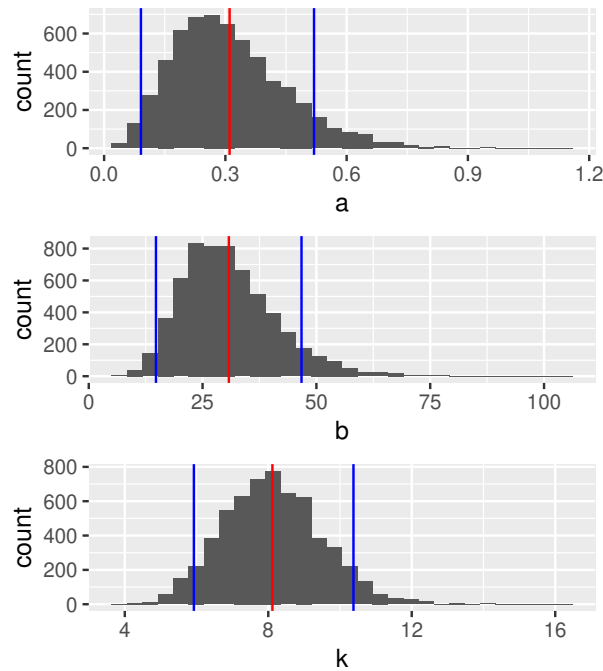
Figure 4.19: Posterior density of each parameter of the Bayesian model.
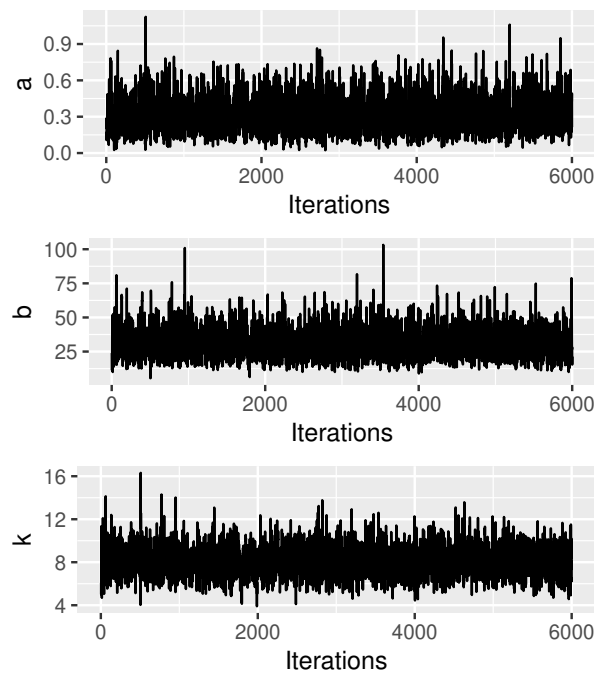
Figure 4.20: Trace plots of each parameter of the Bayesian model.

show that even life data, here used referenced data of animals, might also be originated from more than one failure mode.

The illustrative examples present only the results of Bayes estimation based on SEL function. In case one has good knowledge of the consequences of overestimation and underestimation of a particular problem which one is concerned, one is able to choose the corresponding loss function.

In case of small datasets like the lifetime data of male mice exposed to 300 rads, the

Table 4.10: Point estimates and two-sided 90% and 95% HPD intervals for a, b, k and MTTF

	MCMC	90% HPD	95% HPD
a	0.3105	$[0.0911, 0.5193]$	$[0.0746, 0.5933]$
b	30.774	$[14.761, 46.729]$	$[13.695, 52.843]$
k	8.1159	$[5.9259, 10.376]$	$[5.4736, 10.780]$
$MTTF$	0.7082	$[0.6499, 0.7655]$	$[0.6353, 0.7735]$

Table 4.11: Point estimates and two-sided 90% and 95% BCa bootstrap confident intervals for a, b, k and MTTF

	CE	90% BCa	95% BCa
a	0.2422	$[0.0788, 0.5248]$	$[0.0526, 0.6211]$
b	25.925	$[16.510, 35.880]$	$[15.270, 38.020]$
k	7.4382	$[6.0710, 9.3660]$	$[5.7440, 9.7500]$
$MTTF$	0.7202	$[0.6560, 0.7712]$	$[0.6449, 0.7799]$
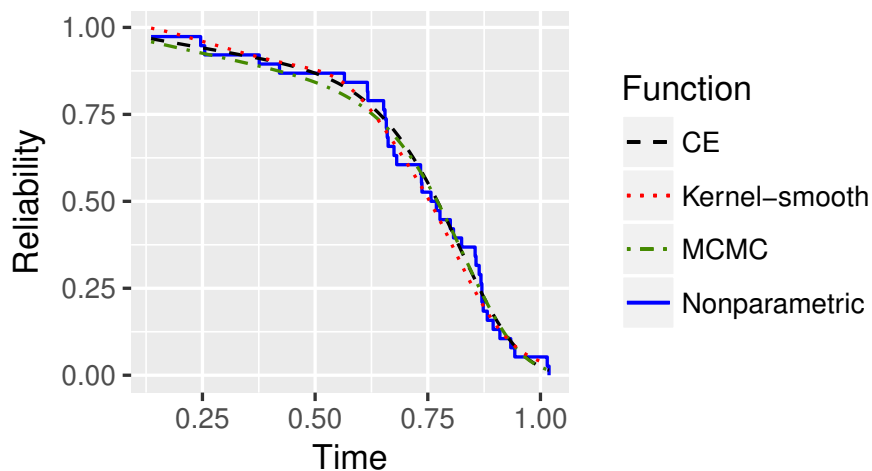
Figure 4.21: The time courses of the reliability functions.

CE method seems to be slightly better in comparison with the MCMC method. This fact is due to the stochastic search of CE method. To improve the result of CE algorithm, one can choose the initial value(s) of (vector) standard deviation $\hat{\boldsymbol{\sigma}}_0$ (sometimes changing the initial value(s) of (vector) mean $\hat{\boldsymbol{\mu}}_0$ is needed) through trial and error until we get the smallest optimum value which yields the optimal estimator(s). The CE method is not so sensitive to the choice of the initial values of the algorithm in comparison with other alternative optimization methods. Both kinds of estimators were completed by confidence intervals (90% and 95%): Bayes credible intervals versus bootstrap confidence intervals. The interval analysis shows that most of obtained 90% and 95% bootstrap confidence intervals are narrower than corresponding HPD intervals, especially for shape parameter k, what shows on less uncertainty connected with CE estimators, and greater variance related to MCMC estimators.
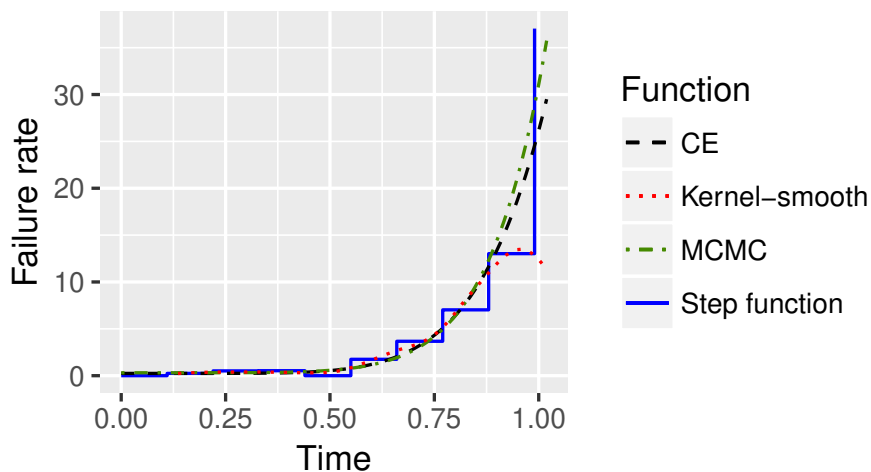
Figure 4.22: The time courses of the failure rate functions.
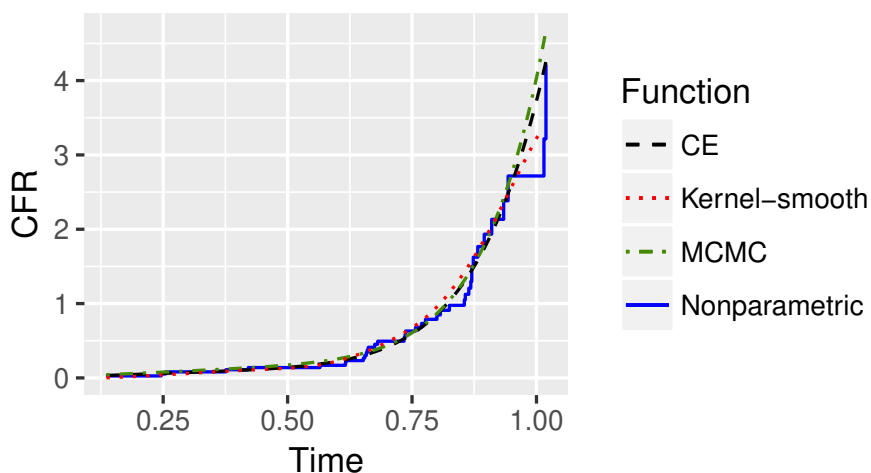
Figure 4.23: The time courses of the cumulative failure rate functions.

Table 4.12: U.S.S. Halfbeak Diesel Engine Data

0.1382	0.2990	0.4124	0.6827	0.7472	0.7567
0.8845	0.9450	0.9794	1.0848	1.1993	1.2300
1.5413	1.6497	1.7352	1.7632	1.8122	1.9067
1.9172	1.9299	1.9360	1.9686	1.9940	1.9944
2.0121	2.0132	2.0431	2.0525	2.1057	2.1061
2.1309	2.1310	2.1378	2.1391	2.1456	2.1461
2.1603	2.1658	2.1688	2.1750	2.1815	2.1820
2.1822	2.1888	2.1930	2.1943	2.1946	2.2181
2.2311	2.2634	2.2635	2.2669	2.2691	2.2846
2.2947	2.3149	2.3305	2.3491	2.3526	2.3774
2.3791	2.3822	2.4006	2.4286	2.5000	2.5010
2.5048	2.5268	2.5400	2.5500	2.5518	
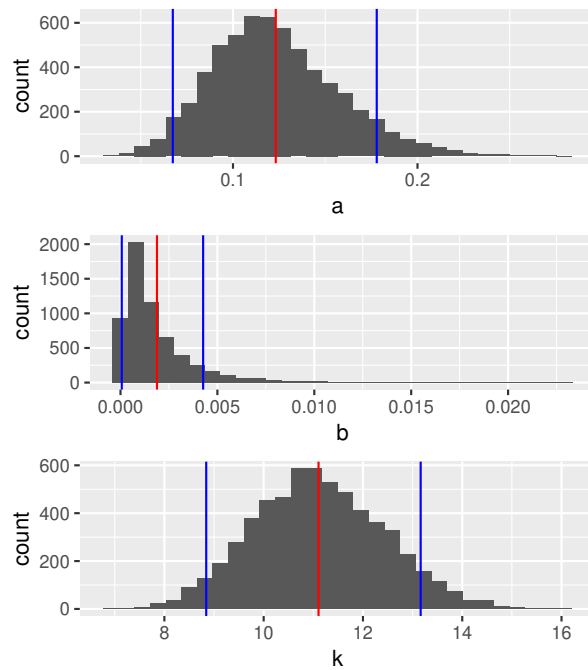
Figure 4.24: Posterior density of each parameter of the Bayesian model.
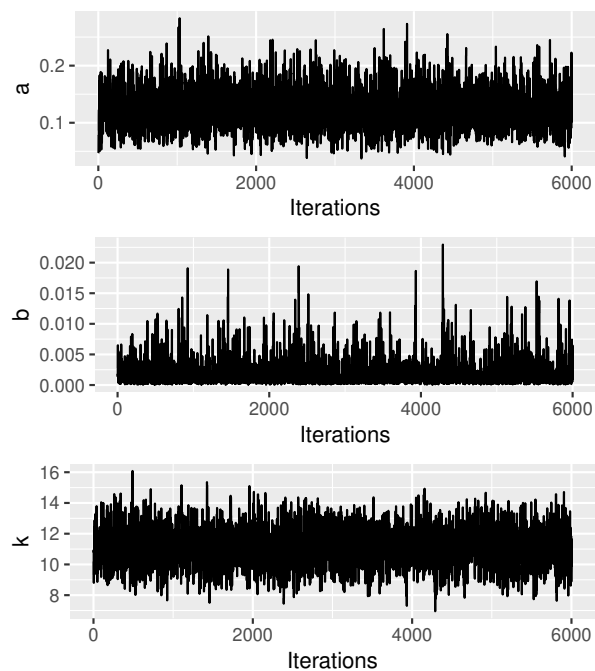
Figure 4.25: Trace plots of each parameter of the Bayesian model.

4.7 Conclusions

The benchmark datasets, especially the aircraft windshield failure data, show that the innovative non-linear failure rate might be more appropriate than some mixtures of distributions for modeling the data which might be originated from more than one failure mode.

The CE and MCMC methods are recommended for using as the mutual add-in tools for parameter estimation. In case of small datasets, the CE method seems to be slightly

Table 4.13: Point estimates and two-sided 90% and 95% HPD intervals for a, b, k and MTTF

	MCMC	90% HPD	95% HPD
a	0.1232	$[0.0674, 0.1779]$	$[0.0596, 0.1928]$
b	0.0019	$[0.0001, 0.0043]$	$[0.0000, 0.0059]$
k	11.105	$[8.8428, 13.163]$	$[8.5930, 13.667]$
$MTTF$	1.9095	$[1.7993, 2.0229]$	$[1.7744, 2.0389]$

Table 4.14: Point estimates and two-sided 90% and 95% BCa bootstrap confident intervals for a, b, k and MTTF

	CE	90% BCa	95% BCa
a	0.1182	$[0.0710, 0.1819]$	$[0.0624, 0.1949]$
b	4.669×10^{-4}	$[0.0002, 0.0006]$	$[0.0002, 0.0006]$
k	12.269	$[12.070, 13.120]$	$[12.060, 13.300]$
$MTTF$	1.9336	$[1.8100, 2.0410]$	$[1.7870, 2.0570]$
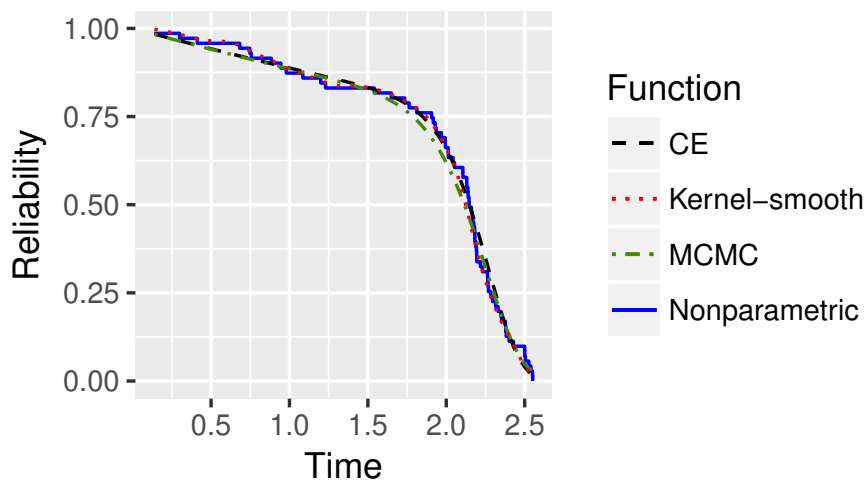
Figure 4.26: The time courses of the reliability functions.

more favored in comparison with the MCMC method. However, Bayes credible intervals are easily obtained by using MCMC method.

Resulting from the simulation study, the asymmetric loss functions are recommended for the non-linear failure rate model in such reliability situations when overestimation is more serious than underestimation.

In this study, only the sum of two components in the non-linear failure rate which results from the assumption of two independent competing risks (i.e. the random shock and the wear out) is considered. The case of dependent competing risks will be considered next. And only the diffuse priors on the parameters for Bayesian estimation is used, so this study can also be continued with other alternative priors. And there are still other loss functions can be considered (e.g. precaution loss, entropy loss, etc.).
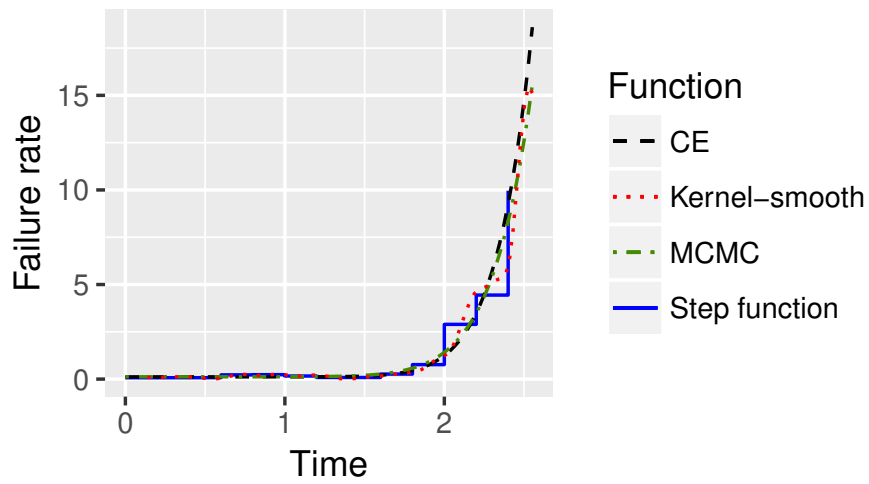
Figure 4.27: The time courses of the failure rate functions.
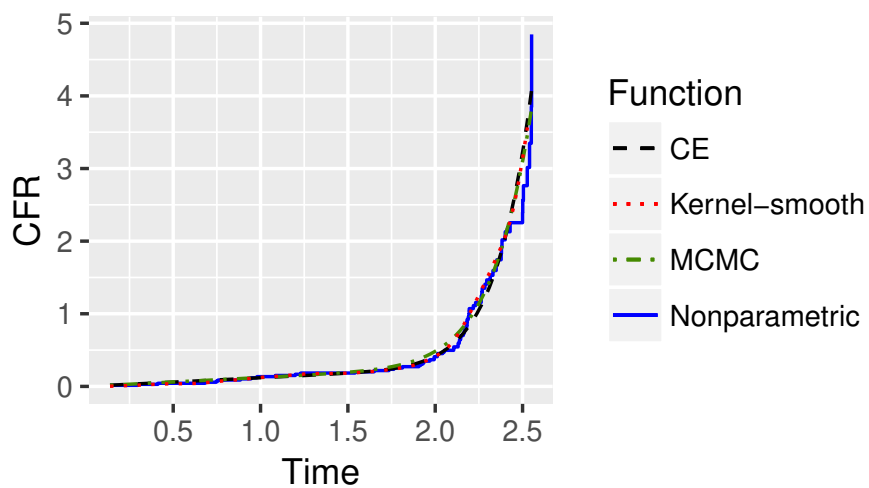
Figure 4.28: The time courses of the cumulative failure rate functions.

Chapter 5

Reparameterized Weibull model: A Bayes study using Hamiltonian Monte Carlo simulation

5.1 Introdoction

This chapter comes from my study given in [86]. The Weibull distribution [73] which was named after Swedish mathematician Waloddi Weibull, who described it in detail in 1951, is the widely used distribution not only in reliability but also in many other fields [51]. It's cumulative distribution function (CDF) is defined as

$$F(t) = 1 - e^{-\left(\frac{t}{\beta}\right)^k}, \quad t > 0 \tag{5.1}$$

where β and k are positive, with β being scale parameter and k being shape parameter. In this chapter, it is named as standard Weibull (SW) distribution. Notice that, if $X \sim SW(\beta, k)$, then $X/\beta \sim SW(1, k)$. As mentioned by Lai [37], for some applications, one may find it be more convenient to reparameterize the standard Weibull distribution as

$$F(t) = 1 - e^{-bt^k} \tag{5.2}$$

The shape parameter is still the same as above, but the scale parameter is $b = 1/\beta^k$. Here it is named as reparameterized Weibull (RW) distribution. Indeed, this reparameterized form has been used in medical statistics [16] and many modified Weibull distributions was based upon this parameterization. The RW model might be convenient and simple. However, as it will be demonstrated in later sections, it leads to an undesirable problem which produce a very high correlated parameters, especially when the values of parameters are large. As a result, Bayesian inference via Markov chain Monte Carlo (MCMC) methods has been heavily affected when using vague prior.

In fact, there is another frequently used of parameterization of the Weibull distribution where the CDF is given as follows

$$F(t) = 1 - e^{-(at)^k} \tag{5.3}$$

This parameterization was used by Lawless [40] and Xie and Lai [75]. However, this parameterization is skipped in this chapter.

This chapter is organized as follows. Section 5.2 gives the examinations for contour plots of likelihood functions of the two Weibull forms. Section 5.3 brings the methods of parameter estimation. Section 5.4 provides a Monte Carlo simulation study for some special selected cases. Section 5.5 brings a special illustrative example on a real data set. Finally, Section 5.6 gives conclusions.

5.2 Contour plots of likelihood functions

In this section, the likelihood functions of the two Weibull forms are examined in an intuitive way via their contour plots for some special selected parameter values in order to understand the correlated parameters for the two forms. Let $\mathcal{D}: t_1, ..., t_n$ be a random sample from the Weibull distribution. The likelihood function is defined as

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta}) \tag{5.4}$$

Figure 5.1 shows contour plots of the likelihood functions of the SW and RW forms respectively in case data generated from the SW model with $\beta = 250$ and $k = 0.5$. The contour plot of the RW shows a medium correlated parameters. In case $\beta = 250$ and $k = 3$, i.e when $k > 1$, Figure 5.2 reveals that the contour plot of the SW shows almost no correlated parameters whereas the RW form shows a very high correlation. For the RW with large value of scale parameter, the larger the value of the shape parameter the higher the correlated parameters. And, in case $\beta = 1$ and $k = 10$, i.e scale parameter being 1 and large shape parameter, the correlated parameters of the two Weibull forms are very small and are all most the same, see Figures 5.3. The contour plots of the SW likelihood function show that the correlated parameters are very small in any case. These analyses suggest that

- In case $k = 0.5$, i.e. decreasing failure rate, the two Weibull forms will have no effects on the Bayesian inference via MCMC methods. The RW might be even better than the SW in producing the accuracy of parameter estimate.

- In case $k = 3$, i.e. increasing failure rate, the SW is more appropriate than the RW for using MCMC methods. However, in case data have a small scale as if it (approximately) comes from the SW with scale parameter $\beta = 1$, both forms are appropriate. In this case, the RW is recommended due to its simple form.

These conclusions will also be reinforced through the simulation study in a later section.

(a)   (b)
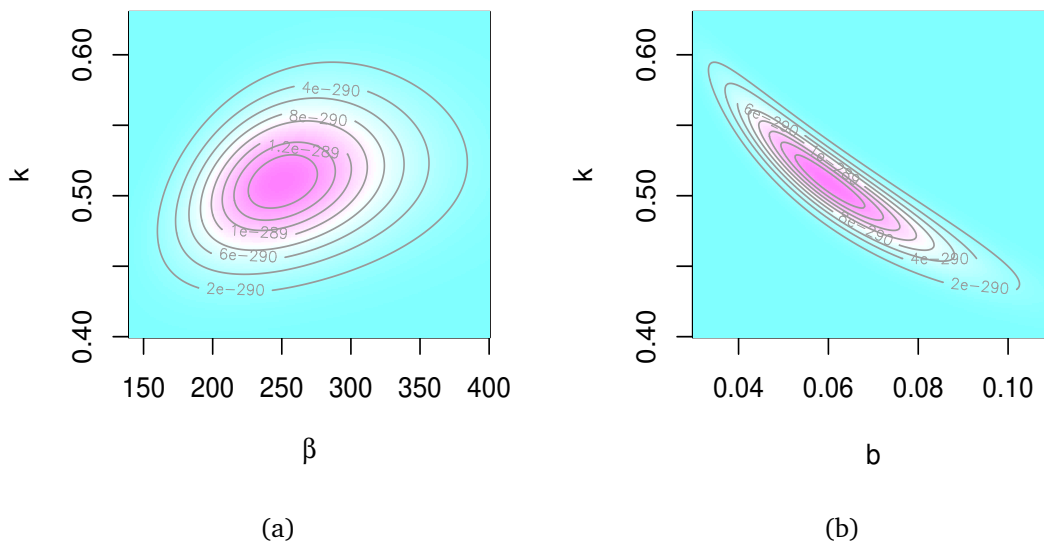
Figure 5.1: Contour plots of (a) SW and (b) RW likelihood functions in case $\beta = 250$ and $k = 0.5$.
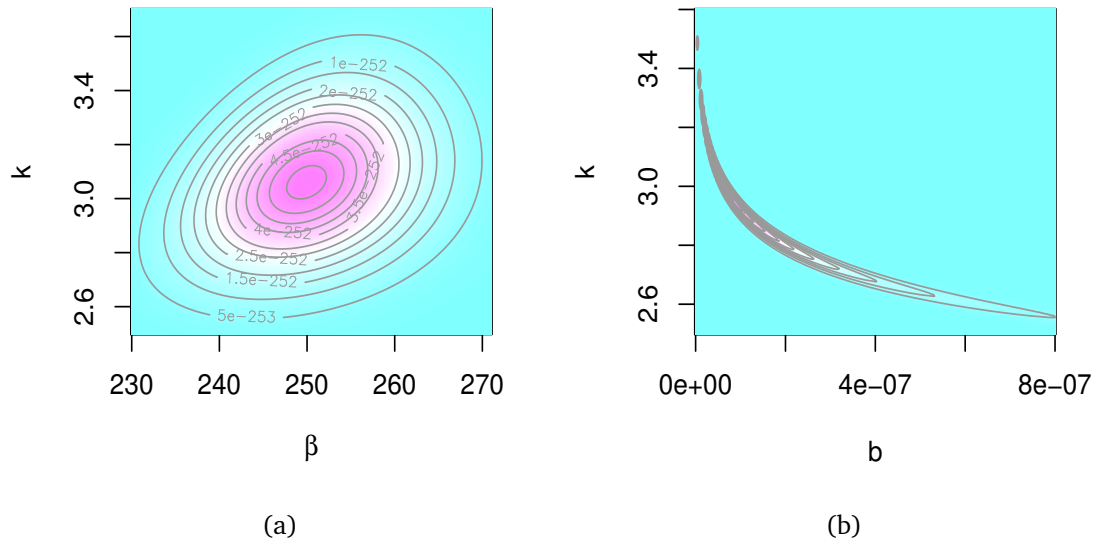
Figure 5.2: Contour plots of (a) SW and (b) RW likelihood functions in case $\beta = 250$ and $k = 3$.
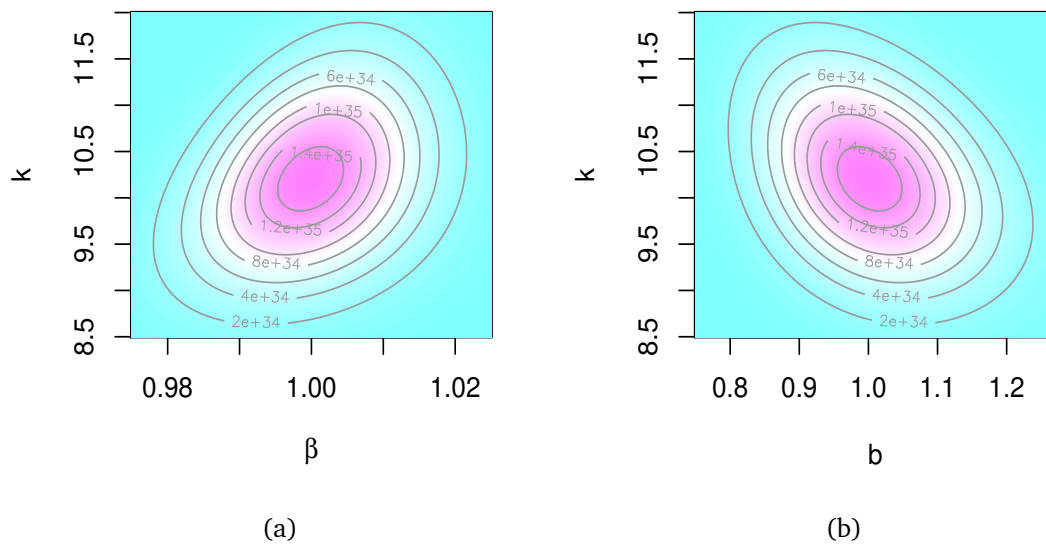
Figure 5.3: Contour plots of (a) SW and (b) RW likelihood functions in case $\beta = 1$ and $k = 10$.

5.3 Parameter estimation methods

5.3.1 Maximum likelihood estimation

The log-likelihood function can be derived from Eq. (5.4) as

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(t_i|\boldsymbol{\theta}) \tag{5.5}$$

The MLEs of the model's parameters are obtained by solving the following equations

$$\frac{\partial \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_1} = 0, \quad \frac{\partial \log L(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_2} = 0 \tag{5.6}$$

These equations can not be solved analytically and need to be solved by some suitable numerical methods. Here the CE method is used to optimize the log-likelihood function given in Eq. (5.5).

5.3.2   Bayesian estimation

Denote the prior distribution of $\boldsymbol{\theta}$ as $\pi(\boldsymbol{\theta})$, the posterior distribution of $\boldsymbol{\theta}$ given $\mathcal{D}: t_1, \ldots, t_n$ is given by

$$\pi(\boldsymbol{\theta}|\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{5.7}$$

Here θ_1 and θ_2 are assumed to be independent and have gamma(α_1, β_1) and gamma(α_2, β_2) priors respectively, i.e

$$\pi_1(\theta_1) \propto \theta_1^{\alpha_1-1} e^{-\beta_1 \theta_1}, \; \alpha_1, \beta_1 > 0 \tag{5.8}$$

$$\pi_2(\theta_2) \propto \theta_2^{\alpha_2-1} e^{-\beta_2 \theta_2}, \; \alpha_2, \beta_2 > 0 \tag{5.9}$$

If $\alpha_1 = \alpha_2 = 1, \beta_1 = \beta_2 = 0$ we have diffuse priors, and if $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$, we have non-informative priors. Since there is no prior information available, the diffuse priors are used in later sections.

Then under the square error loss function, the Bayes estimators of the parameters are given by

$$\theta_1^* = \mathbb{E}(\theta_1|\mathcal{D}) \tag{5.10}$$

$$\theta_2^* = \mathbb{E}(\theta_2|\mathcal{D}) \tag{5.11}$$

$$h^*(t) = \mathbb{E}(h(t;\boldsymbol{\theta})|\mathcal{D}) \tag{5.12}$$

$$R^*(t) = \mathbb{E}(R(t;\boldsymbol{\theta})|\mathcal{D}) \tag{5.13}$$

Suppose that the sample $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}), i = 1, \ldots, N$ is simulated from the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$. Then when i is sufficiently large (say, bigger than n_0), $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}), i = n_0 + 1, \ldots, N$ is a (correlated) sample from the true posterior. In practice, we usually run m parallel chains (say, $m = 3, 4$ or 5), instead of only 1, for assessing sampler convergence. Then, the approximate Bayes estimates of θ_1^* and θ_2^* by calculating the means:

$$\theta_1^* \approx \frac{1}{m(N-n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \theta_1^{(ij)} \tag{5.14}$$

$$\theta_2^* \approx \frac{1}{m(N-n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \theta_2^{(ij)} \tag{5.15}$$

$$h^*(t) \approx \frac{1}{m(N-n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}^{(ij)}) \tag{5.16}$$

$$R^*(t) \approx \frac{1}{m(N-n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} R(t;\boldsymbol{\theta}^{(ij)}) \tag{5.17}$$

5.4 Simulation study

A Monte Carlo simulation study is conducted to compare the HMC estimators as well as the CE estimators for the parameters of the two Weibull forms in term of their mean squared errors (MSE). The data sets were simulated from the SW distribution with different selected parameter values as follows:

- $\beta = 250$ and $k = 0.5$

- $\beta = 250$ and $k = 3$

- $\beta = 250$ and $k = 10$

- $\beta = 1$ and $k = 10$

For each of the above choice of parameters, 1000 data sets were simulated for each sample size $n = 25, 50, 100$ and 200, and based on each data set the CE and HMC estimators for the parameters of the Weibull forms were computed. In order to obtain HMC estimators, samples are simulated from the posterior distribution by using the HMC algorithm to construct Markov chains of length 2000 with burn-in (warm-up) of 1000. The MSE is calculated as the average squared difference between estimated values and the true value. Tables 5.1 - 5.4 show the MSEs of the parameters of the Weibull forms.

Table 5.1: MSEs of the parameters of SW and RW models in case $\beta = 250$ and $k = 0.5$.

n	Method	SW		RW	
		β	k	b	k
25	CE	5046.6502	0.0078	0.0012	0.0089
	HMC	Inf	0.0099	0.0019	0.0060
50	CE	3373.0718	0.0033	0.0006	0.0036
	HMC	10753.7600	0.0037	0.0008	0.0030
100	CE	2087.5945	0.0017	0.0003	0.0018
	HMC	3768.5854	0.0018	0.0004	0.0016
200	CE	1207.1086	0.0008	0.0001	0.0008
	HMC	2026.4754	0.0008	0.0002	0.0008

Since the two Weibull forms have the same shape parameter k, only the comparison of the MSEs of k is provided. Resulting from the simulation study, we see that

- In case $\beta = 250$ and $k = 0.5$ (decreasing failure rate), the RW form has medium correlated parameters as demonstrate in Section 5.2. Therefore, it does not much affect HMC estimates. The MSEs of HMC estimators for the parameter k of the RW form are smallest compared to the SW form and other CE estimators (see Figure 5.4).

- In case $\beta = 250$ and $k = 3$ (increasing failure rate), the MSEs of HMC estimators for the parameter k of the RW form are largest whereas their MSEs of CE estimators for the parameter k are smallest (see Figure 5.5).

- In case $\beta = 250$ and $k = 10$ (increasing failure rate, but with larger shape parameter), the MSEs of HMC estimators for the parameter k of the RW form are also

Table 5.2: MSEs of the parameters of SW and RW models in case $\beta = 250$ and $k = 3$.

n	Method	SW		RW	
		β	k	b	k
25	CE	297.3406	0.3019	1.0344×10^{-14}	0.1297
	HMC	304.3224	0.2974	7.1049×10^{-9}	0.7190
50	CE	163.8943	0.1273	8.5746×10^{-15}	0.0764
	HMC	165.7818	0.1252	7.8539×10^{-11}	0.2607
100	CE	72.3507	0.0570	6.7516×10^{-15}	0.0435
	HMC	72.4627	0.0568	1.4444×10^{-12}	0.0989
200	CE	37.9752	0.0297	4.0948×10^{-15}	0.0248
	HMC	37.9923	0.0298	1.0274×10^{-13}	0.0397

Table 5.3: MSEs of the parameters of SW and RW models in case $\beta = 250$ and $k = 10$.

n	Method	SW		RW	
		β	k	b	k
25	CE	28.5926	4.0114	1.1818×10^{-48}	0.1781
	HMC	28.3067	3.9204	6.6065×10^{-14}	37.5833
50	CE	13.8543	1.3341	1.0774×10^{-48}	0.1241
	HMC	13.7233	1.3193	1.3393×10^{-21}	18.3233
100	CE	6.9635	0.6270	9.3148×10^{-49}	0.0927
	HMC	6.9416	0.6255	1.1950×10^{-29}	6.8448
200	CE	3.5088	0.3155	8.1162×10^{-49}	0.0639
	HMC	3.5236	0.3149	1.0723×10^{-35}	2.1954

Table 5.4: MSEs of the parameters of SW and RW models in case $\beta = 1$ and $k = 10$.

n	Method	SW		RW	
		β	k	b	k
25	CE	0.0004	3.2768	0.0546	3.2861
	HMC	0.0004	3.1911	0.0598	3.3909
50	CE	0.0002	1.5857	0.0248	1.5948
	HMC	0.0002	1.5668	0.0259	1.6240
100	CE	0.0001	0.6344	0.0118	0.6376
	HMC	0.0001	0.6332	0.0120	0.6440
200	CE	0.0001	0.3204	0.0056	0.3204
	HMC	0.0001	0.3189	0.0056	0.3206

largest whereas their MSEs of CE estimators for the parameter k are smallest (see Figure 5.6).
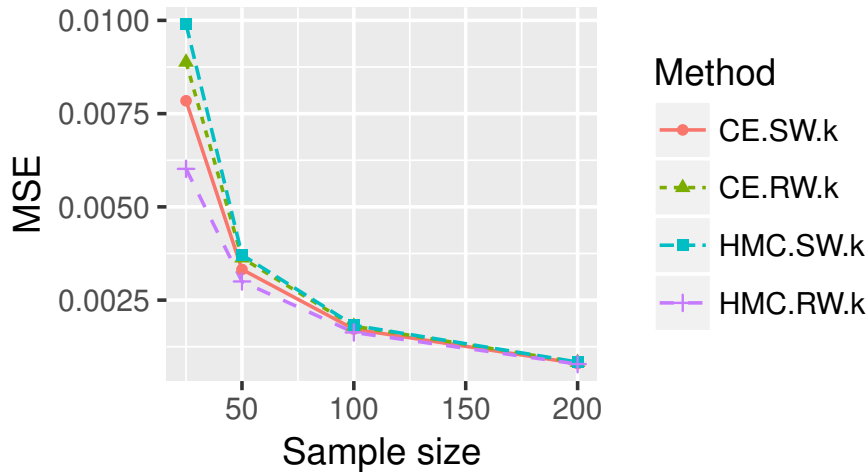
Figure 5.4: MSEs of the parameter k of SW and RW models in case $\beta = 250$ and $k = 0.5$
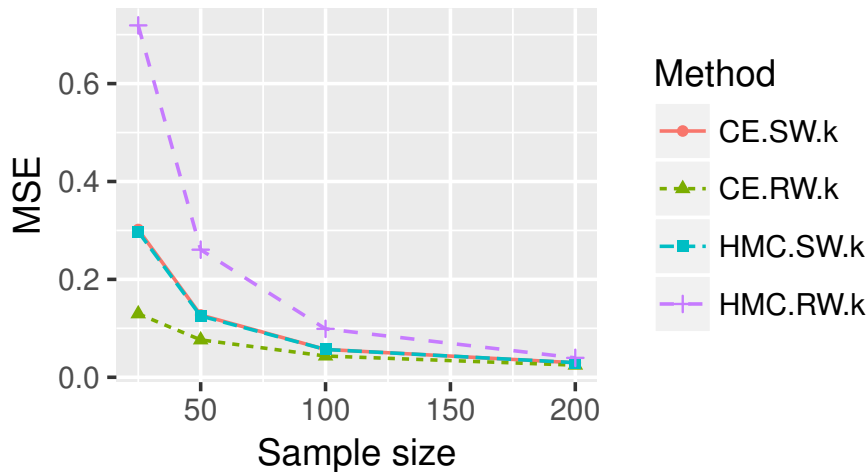
Figure 5.5: MSEs of the parameter k of SW and RW models in case $\beta = 250$ and $k = 3$

- In case $\beta = 1$ and $k = 10$ (also increasing failure rate, but with scale being 1), the MSEs of estimators for the parameter k of the two Weibull forms obtained by both methods are almost the same (see Figure 5.7).

- In case increasing failure rate with large scale parameter, the MSEs of CE estimators of RW form are smallest (see Figures 5.5-5.6).

These results have been obtained because Bayes estimators (reliability characteristics also) under squared error loss function are based on the posterior mean. Since in case of high correlation, the marginal distributions of parameters are not symmetric, this fact will affect the posterior mean.

Figure 5.8 shows the plot of the Weibull PDFs with different selected shape parameter values. For $k < 3$, we see the clear differences between these functions, but for $k > 3$, there is not much difference between them. So only a large sample can detect the differences. This is the reason why small sample sizes result in high MSE in the simulation study.
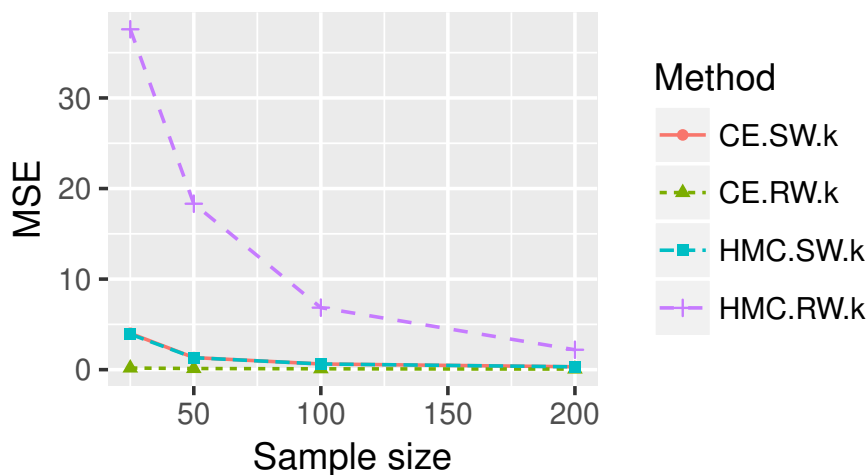
Figure 5.6: MSEs of the parameter k of SW and RW models in case $\beta = 250$ and $k = 10$
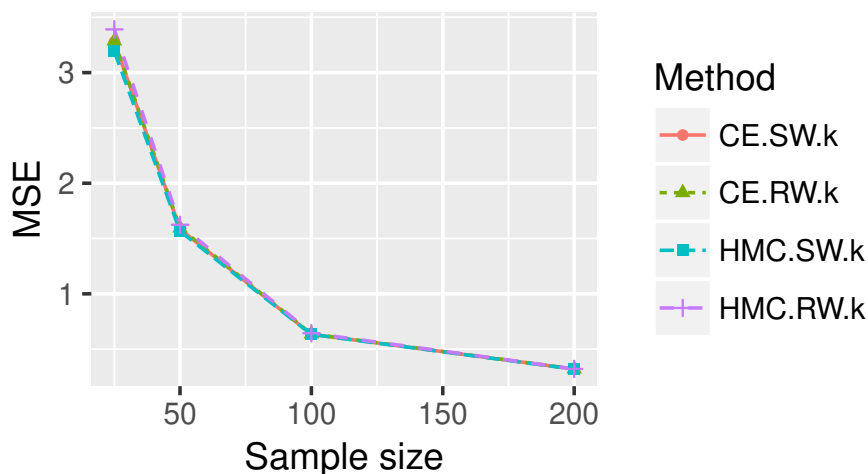
Figure 5.7: MSEs of the parameter k of SW and RW models in case $\beta = 1$ and $k = 10$

5.5 Illustrative example

5.5.1 The Weibull distribution

Data in Table 5.5 are used to demonstrate the consequence of the reparameterization on the Bayes estimators via HMC. The data represent lifetimes of diesel engines [18]. The unit for measurement is hour.

Table 5.5: Time to failure of diesel engine.

1276	720	1135	1854	1687	2570	2440
2547	1100	2117	1876	1633	2646	1556
2470	1250	1895	2607	896	401	

In order to produce the HMC samples, 4 parallel chains are constructed, each with length of 2000 iterations and burn-in (warm-up) period of 1000. Figure 5.9 shows the trace plots and density estimates of HMC output for the SW parameters. The trace plots
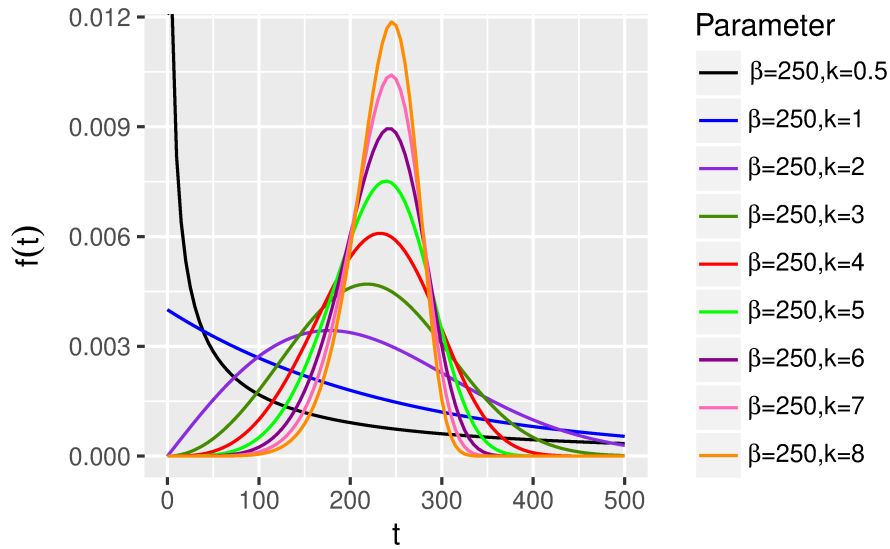
Figure 5.8: Weibull pdf with variate values of shape parameter

show that the 4 parallel chains converge to the same stationary distribution for each parameter and the density estimates are distributed approximately symmetrically which mean that they will produce good estimates by using posterior mean function. Since the diffuse priors are used, the posterior distribution is nothing but the likelihood function. Figure 5.10 represents the contour of the SW likelihood function superimposed by 500 sample points from a single chain. Likewise, Figure 5.11 shows the trace plots and density estimates of HMC output for the RW parameters. The trace plots show that the 4 parallel chains also converge to the same stationary distribution for each parameter, but the posterior distribution for b is not symmetric. Even the posterior distribution for k is symmetric, it does not provide the expected estimate for k. The contour plot of RW likelihood function superimposed by HMC sample points in Figure 5.12 shows a very high correlation between the RW parameters. Notice that even in case of very high correlation, HMC still provides the good representative sample.

(a)   (b)
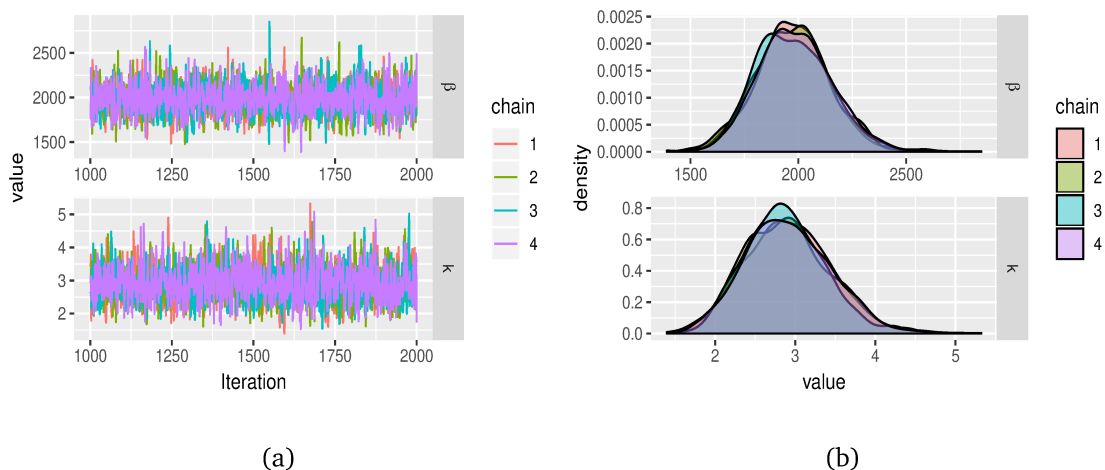
Figure 5.9: (a) Trace plots and (b) density estimates of HMC output when fitting the SW to the data.

Table 5.6 displays CE and HMC point estimates and Bayes highest posterior density (HPD) intervals for the two Weibull forms. From this table we can see that for the SW
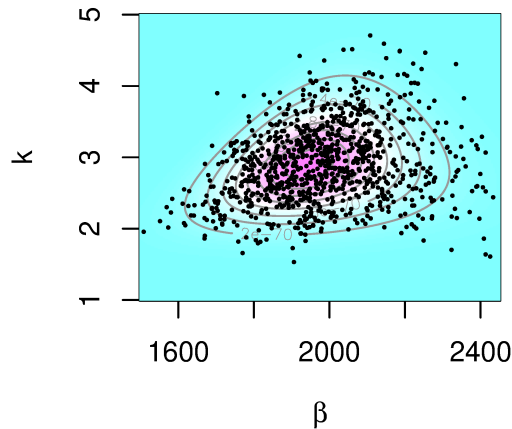
Figure 5.10: Contour plot of SW likelihood function superimposed by an HMC sample.
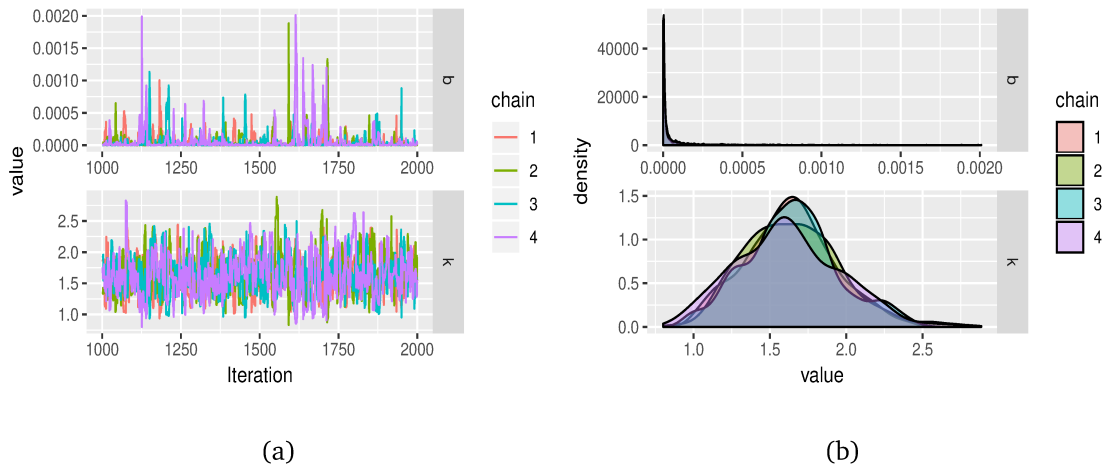
Figure 5.11: (a) Trace plots and (b) density estimates of HMC output when fitting the RW to the data.

form, the CE and HMC estimates are almost the same whereas HMC estimates for the RW form are much more different from the CE estimates. The estimated reliability and failure rate functions of the SW and RW forms also provide in Figure 5.13. From this figure it is easy to observe that the RW form provides poor HMC estimates for the reliability and failure rate functions.

Another way that help HMC working well with the two Weibull forms is also provided. As mentioned before, if $X \sim SW(\beta, k)$, then $X/\beta \sim SW(1, k)$. Since the estimate for the scale parameter β is 1946.4814 (see Table 5.6). Therefore if we divide the data in Table 5.5 by 2000, we will obtain the new data in Table 5.7 as if it comes from (approximately) the SW with $\beta = 1$. Now the unit of measurement is 2000 hours. In this case, Figures 5.14-5.15 show that the chains converge to the same distribution and the density estimates are approximately symmetric for each parameter of both Weibull forms. And Figure 5.16 shows a small correlated parameters for both forms. Table 5.8 shows that HMC estimates for parameters of each Weibull form are very close to the CE estimates. The CE and HMC estimates of reliability and failure rate functions are also displayed in Figures 5.17. Both Weibull forms provide good CE and HMC estimates for these functions in this case. And
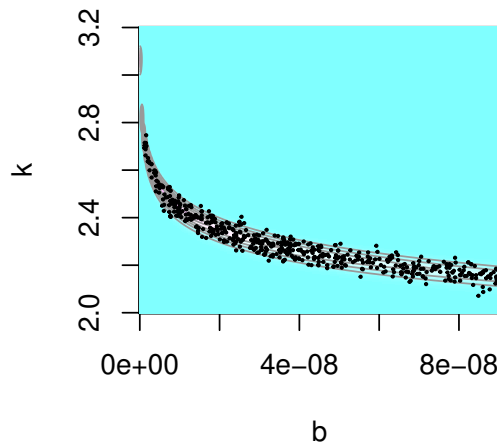
Figure 5.12: Contour plot of RW likelihood function with superimposed by HMC sample.

Table 5.6: CE and HMC point estimates and HPD intervals for parameters of the two forms

		CE	HMC	90% HPD	95% HPD
SW	β	1946.4814	1986.5335	$[1713.8150, 2254.7640]$	$[1649.3379, 2298.1889]$
	k	2.8984	2.8976	$[2.0225, 3.7430]$	$[1.91296, 3.960445]$
RW	b	1.6316×10^{-10}	0.0003×10^{-1}	$[2.0772 \times 10^{-11}, 9.3965 \times 10^{-5}]$	$[2.0772 \times 10^{-11}, 0.0002]$
	k	2.9742	1.6494	$[1.1052, 2.1285]$	$[1.0596, 2.2777]$

(a)

(b)
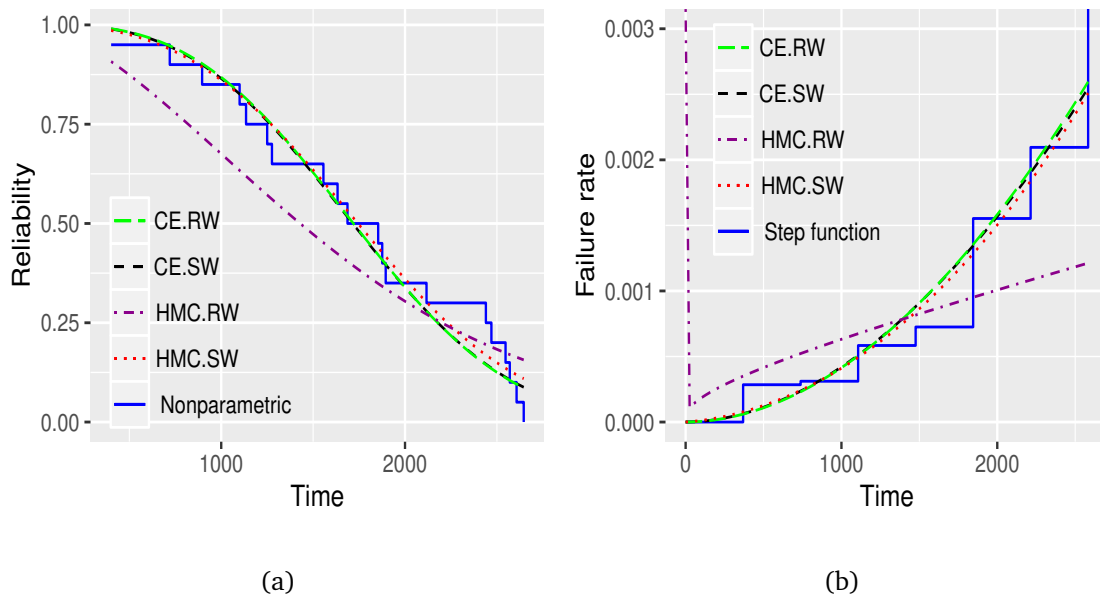
Figure 5.13: Estimated (a) reliability and (b) failure rate functions of the SW and RW forms when fitting to the data using both CE and HMC methods.

in comparison with the results obtained by Dolas, Jaybhaye, and Deshmukh [18], the CE and HMC methods provide better estimators of parameters than the method used in their study.

Table 5.7: Time to failure of cooling system of Diesel engine: data are divided by 2000

0.6380	0.3600	0.5675	0.9270	0.8435	1.2850	1.2200
1.2735	0.5500	1.0585	0.9380	0.8165	1.3230	0.7780
1.2350	0.6250	0.9475	1.3035	0.4480	0.2005	

(a)   (b)
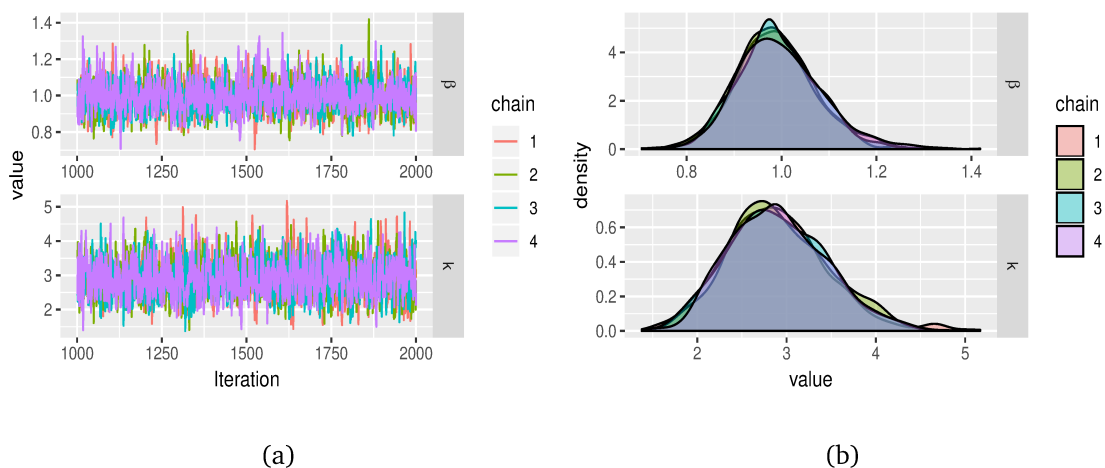
Figure 5.14: (a) Trace plots and (b) density estimates of HMC output when fitting the SW to the scaling data.

(a)   (b)
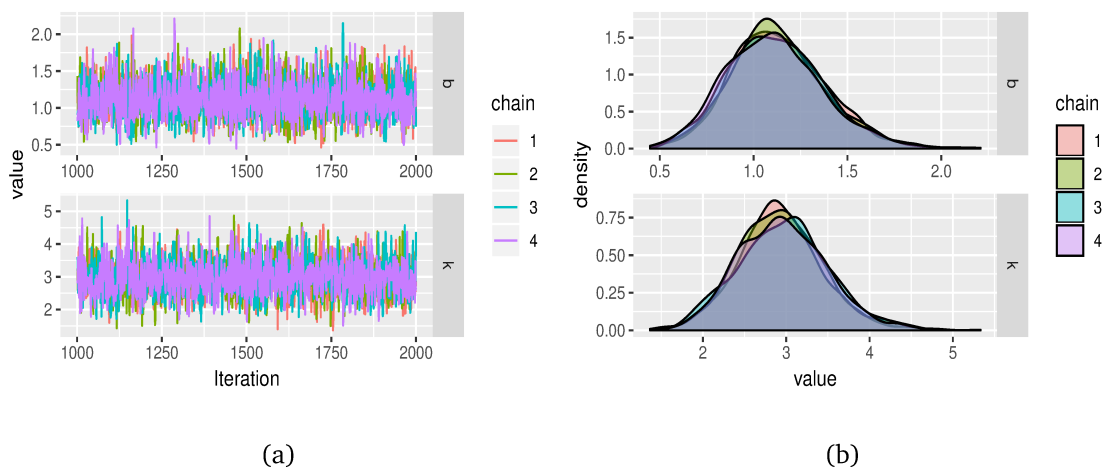
Figure 5.15: (a) Trace plots and (b) density estimates of HMC output when fitting the RW to the scaling data.

5.5.2   The non-linear failure rate distribution

The NLFR model (see Chapter 4) is used to demonstrate how to apply successfully HMC algorithm for Bayesian posterior analysis when fitting this model to a real dataset. The

(a) (b)
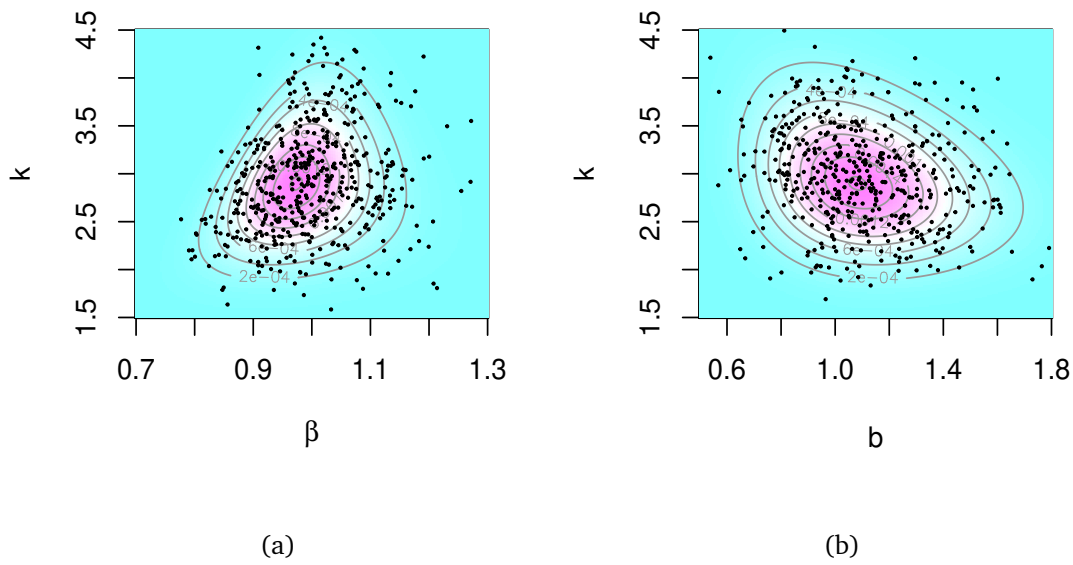
Figure 5.16: Contour plots of (a) the SW and (b) RW likelihood functions superimposed by HMC sample points for scaling data.

Table 5.8: CE and HMC point estimates and HPD intervals for parameters of the two forms; scaling data

Model	Parameter	CE	HMC	90% HPD	95% HPD
SW	β	0.9732	0.9951	$[0.8502, 1.1227]$	$[0.8341, 1.1719]$
	k	2.8984	2.9040	$[2.0109, 3.7267]$	$[1.8883, 3.9162]$
RW	b	1.0818	1.1228	$[0.8502, 1.1227]$	$[0.8341, 1.1719]$
	k	2.8984	2.9473	$[2.0109, 3.7267]$	$[1.8883, 3.9162]$

(a) (b)
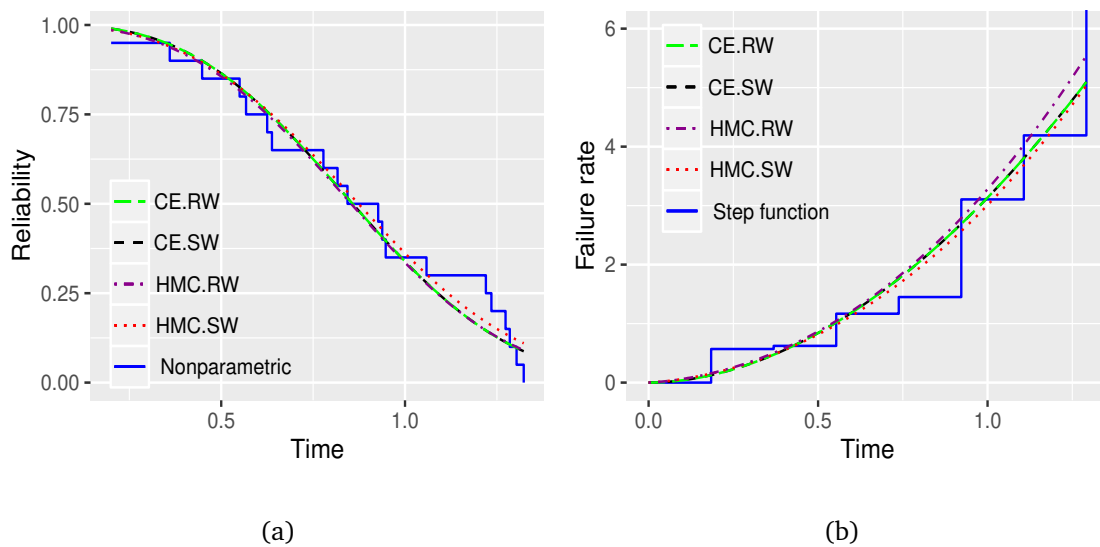
Figure 5.17: Estimated (a) reliability and (b) failure rate functions of the SW and RW forms when fitting to the data using both CE and HMC methods.

NLFR model, with the failure rate function

$$h(t) = a + bt^{k-1} \tag{5.18}$$

induced from the Weibull model, was designed for modeling data sets in which failures result from both random shock and wearout. For example, data in Table 5.9 represent the days until death for male mice exposed to 300 rads of radiation. The unit for measurement is day. Here only the group maintained in a germ-free environment is considered and the causes of death is due to the effect of other causes. The feature of the dataset is that more than one failure mode occurs ([33]). Fig. 5.18 shows that NLFR model is probably more appropriate than Weibull model for this data set. A reparameterized form of the NLFR model, named "NLFR1" which is given as follows

$$h(t) = a + \frac{k}{b}\left(\frac{t}{b}\right)^{k-1} \tag{5.19}$$

is also given to observe which form provides good Bayes estimates via HMC. Fig. 5.19 show the Bayes estimates of reliability and failure rate functions of the two forms along with the MLEs of these functions. From this figure, it is easy to observe that the NLFR1 form provides better result than does the NLFR form. Once again, we observe that the MLE method is not affected by any forms. In contract, it works quite well with the NLFR form which provides even better result than the NLFR1 form and this fact has already been demonstrated in the simulation study for the Weibull forms.

Table 5.9: Male mice exposed to 300 rads of radiation (other causes in germ-free group)

136	246	255	376	421	565	616	617	652	655
658	660	662	675	681	734	736	737	757	769
777	800	807	825	855	857	864	868	870	870
873	882	895	910	934	943	1015	1019		
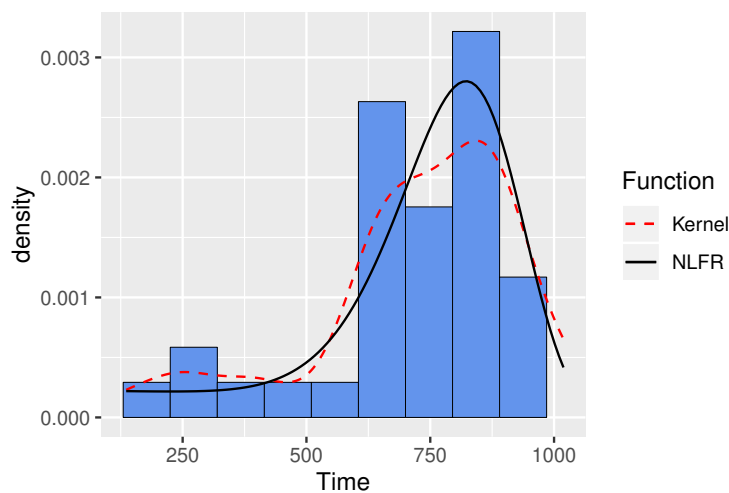
Figure 5.18: Histogram along with density plots of male mice data

The way of changing the data scale is provided also in order to obtain good Bayes estimates. Table 5.10 show the male mice dataset with different scale. In this table, the data is divided by 1000 which changed the unit of measurement into 1000 days. From
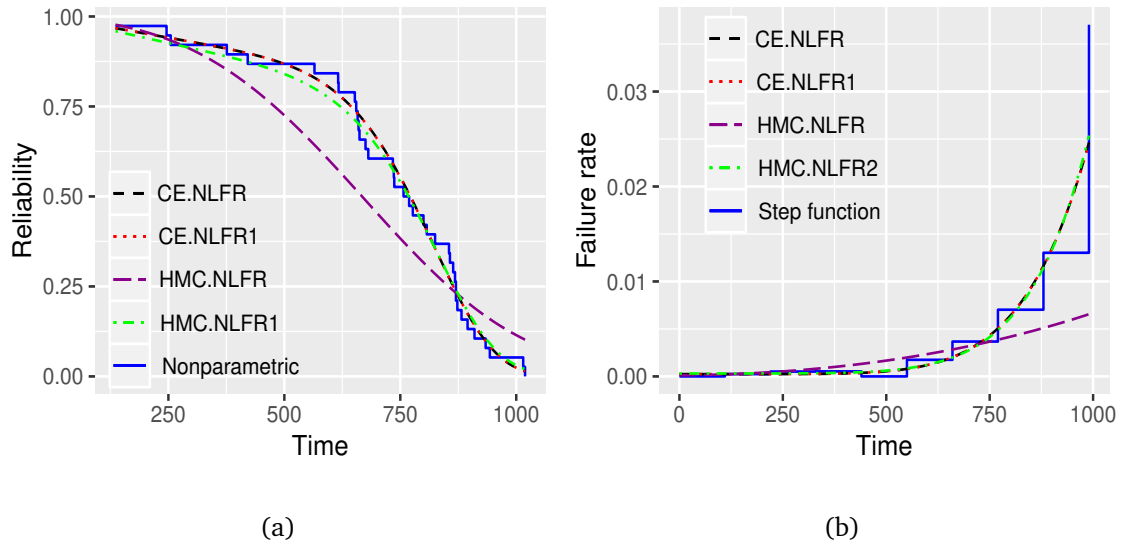
Figure 5.19: Estimated (a) reliability and (b) failure rate functions in case unit measurement of data is day.

Figures 5.20 we see that in this scale, any forms of the NLFR model provide good Bayes estimates.

Table 5.10: Male mice exposed to 300 rads of radiation (measurement unit: 1000 days)

0.136	0.246	0.255	0.376	0.421	0.565	0.616	0.617
0.652	0.655	0.658	0.660	0.662	0.675	0.681	0.734
0.736	0.737	0.757	0.769	0.777	0.800	0.807	0.825
0.855	0.857	0.864	0.868	0.870	0.870	0.873	0.882
0.895	0.910	0.934	0.943	1.015	1.019		
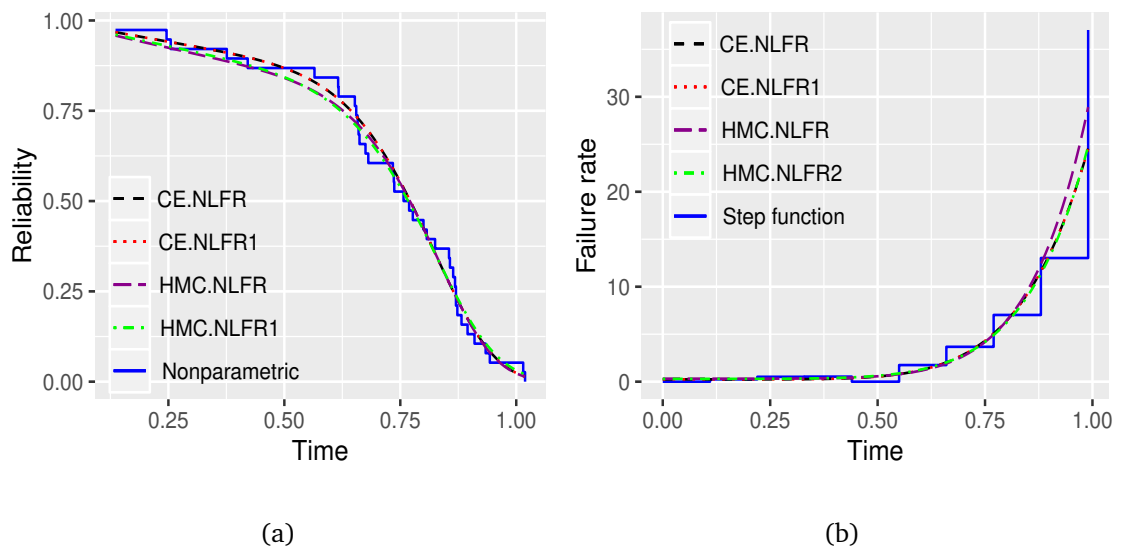
Figure 5.20: Estimated (a) reliability and (b) failure rate functions in case unit measurement of data is 1000 days.

5.6 Conclusions

This study shows how the reparameterization and changing data scale can provide good Bayes estimates via HMC method for the Weibull model as well as the models resulting from the Weibull model. These ways also work well for Bayesian inference using other MCMC methods provided that they produce good samples as HMC does. For frequentist approach, the RW form is recommended due to its simpler form and less MSE than the SW form. For Bayesian inference via MCMC methods, the SW form is recommended. However, in case of decreasing failure rate the RW form is recommended. If a dataset is rescaled as if it (approximately) comes from the SW model with scale parameter 1, then any Weibull forms can be used.

Chapter 6

An additive Chen-Weibull model: A Bayes study using Hamiltonian Monte Carlo simulation

6.1 Introduction

This chapter comes from my study given in [89]. The Weibull distribution [73], is the most widely used distribution in reliability data analysis. However, the failure rate function of the Weibull distribution can only be increasing, decreasing or constant. It fails to capture a lifetime data with bathtub-shaped failure rate such as human mortality, failure rate of newly lunched product, etc., which can involve high initial failure rates (infant mortality, design defects, production errors, inexperienced maintenance errors), then approximately low constant failure rate for a period of time (useful life, random failure) and eventual high failure rates due to aging and wearout, indicating a bathtub-shaped failure rate. Therefore, many generalizations, extensions and modifications of the Weibull distribution have been developed to meet the requirements. For example Mann, Schafer, and Singpurwalla [41] proposed mixtures of Weibull distributions. Hjorth [29] studied a 3-parameter family obtained by generalizing the Rayleigh distribution which with increasing, decreasing, and bathtub failure rates. Mudholkar and Srivastava [47] introduced an exponentiated Weibull (EW) distribution for analyzing bathtub failure rate data. Xie and Lai [75] proposed an additive Weibull (AddW) distribution by combining two Weibull distributions with cumulative distribution function (CDF)

$$F(x) = 1 - e^{-\alpha x^\theta - \beta x^\gamma}, \ x \geq 0; \ \alpha, \beta \geq 0, \ \theta > 1, \ 0 < \gamma < 1 \tag{6.1}$$

Chen [15] introduced a new two-parameter lifetime distribution with bathtub-shape or increasing failure rate function. The CDF of the Chen distribution is given by

$$F(x) = 1 - e^{\lambda(1 - e^{x^\beta})}, \ x \geq 0; \ \lambda, \beta > 0 \tag{6.2}$$

Xie, Tang, and Goh [76] presented a modified Weibull extension (MWE) distribution by adding a scale parameter to the Chen distribution with CDF

$$F(x) = 1 - e^{\alpha \lambda \left(1 - e^{(x/\alpha)^\beta}\right)}, \ x \geq 0; \ \alpha, \beta, \lambda > 0$$

Lai, Xie, and Murthy [39] proposed a modified Weibull (MW) distribution with CDF

$$F(x) = 1 - e^{-\alpha x^\theta e^{\gamma x}}, \ x \geq 0; \ \alpha, \theta > 0, \ \gamma \geq 0 \tag{6.3}$$

which can have an increasing or a bathtub-shaped failure rate function. Bebbington, Lai, and Zitikis [7] introduced a flexible Weibull extension which is able to model various ageing classes of life distributions including increasing and bathtub-shaped failure rates. Dimitrakopoulou, Adamidis, and Loukas [17] proposed a three-parameter lifetime distribution with increasing, decreasing, bathtub, and upside down bathtub shaped failure rates. Carrasco, Ortega, and Cordeiro [14] studied a generalized modified Weibull distribution which has ability to model monotone as well as non-monotone failure rates. Silva, Ortega, and Cordeiro [61] proposed a beta modified Weibull distribution which accommodates monotone, unimodal and bathtub-shaped failure rates. Almalki and Yuan [3] introduced a new modified Weibull (NMW) distribution by combining the Weibull distribution and the MW distributions with CDF

$$F(x) = 1 - e^{-\left(\alpha x^\theta + \beta x^\gamma e^{\lambda x}\right)}, \, x \geq 0; \, \alpha, \beta, \gamma, \theta > 0, \, \lambda \geq 0$$

Sarhan and Apaloo [57] presented an exponentiated modified Weibull extension (EMWE) distribution with CDF

$$F(x) = \left[1 - e^{\alpha\lambda\left(1 - e^{(x/\alpha)^\beta}\right)}\right]^\gamma, \, x \geq 0; \, \alpha, \beta, \gamma, \lambda > 0$$

He, Cui, and Du [28] proposed an additive modified Weibull (AMW) distribution by combining the MW distribution and the Gompertz distribution [25] with CDF

$$F(x) = 1 - e^{-\left(\alpha x^\theta e^{\gamma x} + e^{\lambda x - \beta} - e^{-\beta}\right)}, \, x \geq 0; \, \alpha, \beta, \theta > 0, \, \gamma, \lambda \geq 0$$

Zeng, Lan, and Chen [77] presented five and four-parameter lifetime distributions for bathtub-shaped failure rate using Perks mortality equation. More recent, Shakhatreh, Lemonteb, and Moreno–Arenas [60] introduced a log-normal modified Weibull distribution by combining the MW distribution and the log-normal distribution.

In this chapter, a new continuous lifetime distribution, called the additive Chen-Weibull distribution, is proposed by combining the Weibull distribution and the Chen distribution in a series system with two independent components. One component follows the Chen distribution and the other follows the Weibull distribution. The new distribution provides the flexibility and diversity of shapes of failure rate function. The usefulness of the new distribution is illustrated by fitting to two well-know data sets with bathtub-shaped failure rate and the proposed distribution is demonstrated to be better than many other existing distributions when fitting to these data sets.

The rest of the chapter is organized as follows. Section 6.2 introduces the new lifetime distribution. Some properties of the new distribution are studied in Section 6.3. Section 6.4 deals with the estimation of the parameters. Applications to two real data sets are given Section 6.5. And finally, Section 6.6 brings conclusions.

6.2 The new lifetime distribution

The CDF of the additive Chen-Weibull (ACW) distribution with four parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \lambda)$ is defined by

$$F(x) = 1 - e^{\lambda(1 - e^{x^\gamma}) - (\alpha x)^\beta}, \quad x \geq 0; \, \alpha, \beta, \gamma > 0, \lambda \geq 0 \tag{6.4}$$

The probability density function (PDF) of the ACW distribution has the following form

$$f(x) = \left(\lambda\gamma x^{\gamma-1} e^{x^\gamma} + \alpha\beta(\alpha x)^{\beta-1}\right) e^{\lambda(1 - e^{x^\gamma}) - (\alpha x)^\beta}, \quad x \geq 0 \tag{6.5}$$

And the failure rate and reliability/survival functions are, respectively

$$h(x) = \lambda\gamma x^{\gamma-1}e^{x^\gamma} + \alpha\beta(\alpha x)^{\beta-1} \tag{6.6}$$

and

$$R(x) = e^{\lambda(1-e^{x^\gamma})-(\alpha x)^\beta} \tag{6.7}$$

The reliability function can be written as

$$R(x) = e^{-H(x)} \tag{6.8}$$

where

$$H(x) = \lambda(e^{x^\gamma} - 1) + (\alpha x)^\beta \tag{6.9}$$

is called the cumulative failure rate function.

The new model is useful for modeling a series system with two independent components. One component follows Chen distribution and the other follows Weibull distribution. It can also be used for modeling lifetime data in which failure might be originated from more than one failure mode.

6.3 Properties of the model

6.3.1 The failure rate function

The failure rate function given in Eq. (6.6) is increasing when $\beta, \gamma \geq 1$ and bathtub shaped otherwise. If we choose $\lambda = 0$ in Eq. (6.6), we have the Weibull failure rate which is increasing, decreasing or constant. If we choose $\alpha = 0$, we have the Chen failure rate which is increasing or bathtub-shaped. The plot of the PDFs and the corresponding failure rate functions of the ACW distribution with different values of parameters are displayed in Fig. 6.1. As we can see that the proposed model provides a variety of shapes of the distribution and failure rate for modeling complicated lifetime data. Various shapes of the bathtub-shaped failure rate function of the ACW distribution with long useful lifetime are shown in Fig. 6.2.

(a)   (b)
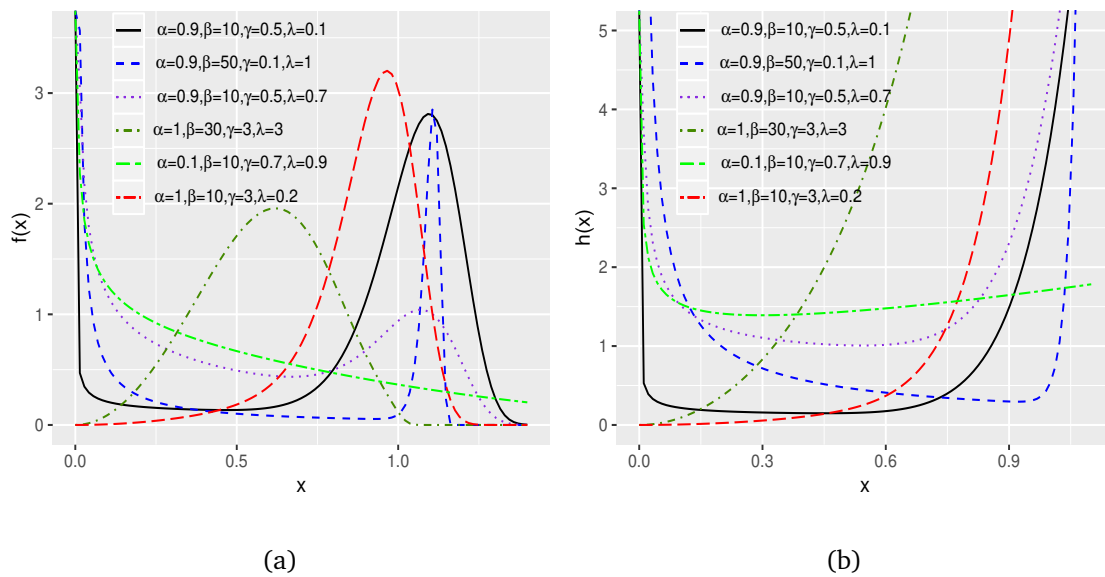
Figure 6.1: (a) Probability density functions and (b) the corresponding failure rate functions of the ACW.
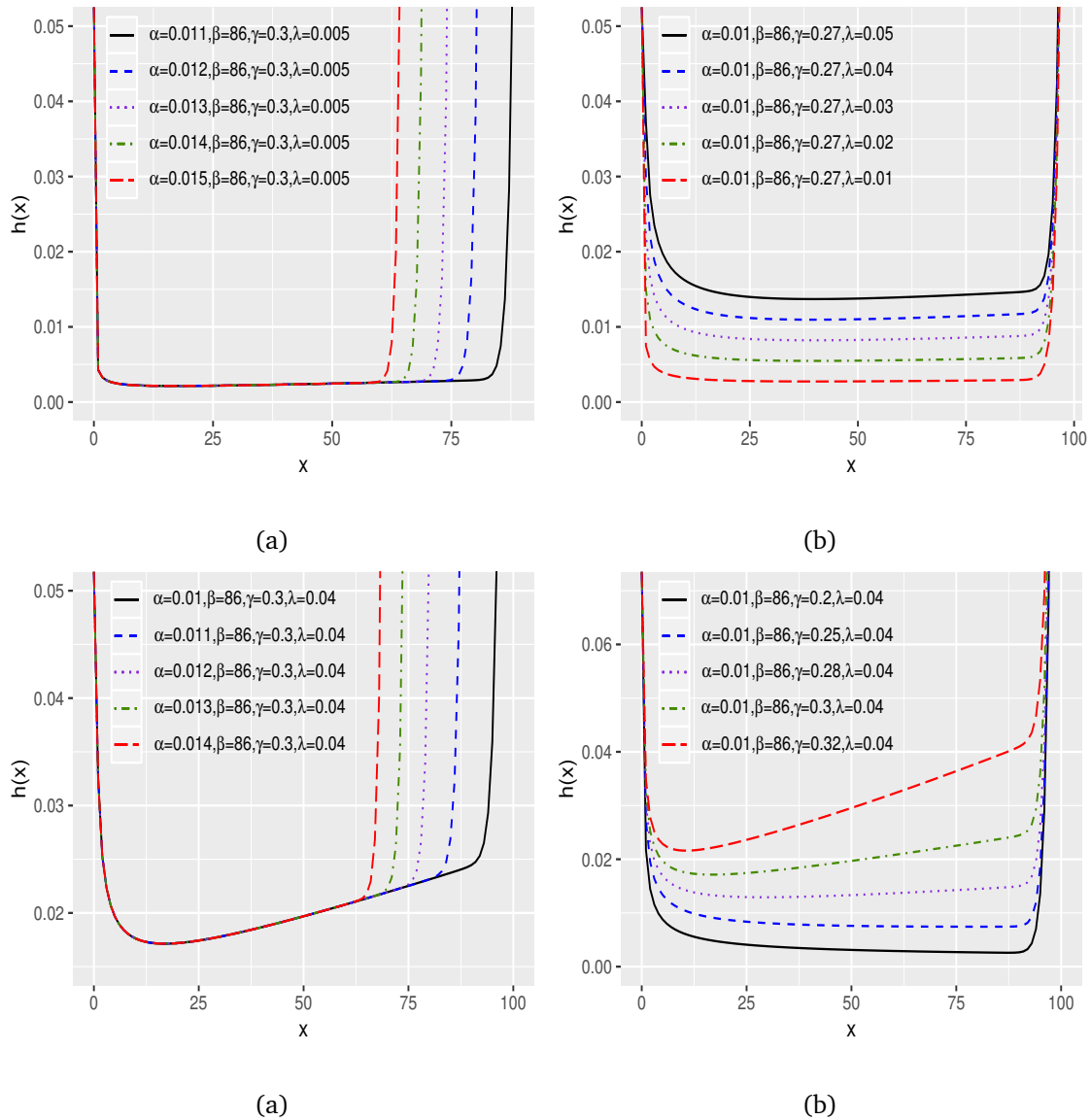
Figure 6.2: Bathtub-shaped failure rate with long useful lifetime of the ACW distribution with different values of parameters.

6.3.2 The moments

The rth non-central moment or the rth moment about the origin of the ACW distribution can be derived as follows by using the Taylor expression of e^x

$$
\begin{aligned}
\mu'_r &= \int_0^{+\infty} x^r dF(x) \\
&= -\int_0^{+\infty} x^r d e^{\lambda(e^{x^\gamma}-1)-(\alpha x)^\beta} \\
&= \int_0^{+\infty} r x^{r-1} e^{\lambda(e^{x^\gamma}-1)-(\alpha x)^\beta} dx \\
&= r e^{-\lambda} \int_0^{+\infty} x^{r-1} e^{\lambda e^{x^\gamma}-(\alpha x)^\beta} dx
\end{aligned}
$$

$$= re^{-\lambda} \sum_{n=0}^{+\infty} \sum_{m=0}^{+\infty} \frac{\lambda^n n^m}{n!\,m!} \int_0^{+\infty} x^{r-1} x^{m\gamma} e^{-(\alpha x)^\beta}\, dx$$

$$= \frac{re^{-\lambda}}{\beta} \sum_{n=0}^{+\infty} \sum_{m=0}^{+\infty} \frac{\lambda^n n^m}{n!\,m!} \alpha^{-(m\gamma+r)} \Gamma\left(\frac{m\gamma+r}{\beta}\right) \tag{6.10}$$

for $r = 1, 2, \ldots$, where $\Gamma(\cdot)$ is the gamma function.

6.3.3 Order statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample from the ACW distribution and $X_{k:n}$ is the kth order statistic of the sample, then the PDF of $X_{k:n}$ is given by

$$f_{k:n}(x) = \frac{1}{B(k, n-k+1)} F(x)^{k-1} (1 - F(x))^{n-k} f(x) \tag{6.11}$$

where $B(\cdot, \cdot)$ is the beta function.

Since we have

$$(1 - F(x))^{n-k} = e^{-(n-k)H(x)}$$

and

$$F(x)^{k-1} = (1 - e^{-H(x)})^{k-1} = \sum_{l=0}^{k-1} \binom{k-1}{l} (-1)^l e^{-lH(x)}$$

Therefore,

$$f_{k:n}(x) = \frac{1}{B(k, n-k+1)} \sum_{l=0}^{k-1} \binom{k-1}{l} (-1)^l h(x) e^{-(n+l+1-k)H(x)}$$

$$= n \binom{n-1}{k-1} \sum_{l=0}^{k-1} \binom{k-1}{l} (-1)^l \left(\lambda\gamma x^{\gamma-1} e^{x^\gamma} + \alpha\beta(\alpha x)^{\beta-1}\right) e^{-(n+l+1-k)\left(\lambda(e^{x^\gamma}-1)+(\alpha x)^\beta\right)}$$

$$= n \binom{n-1}{k-1} \sum_{l=0}^{k-1} \binom{k-1}{l} \frac{(-1)^l}{n+l+1-k} f(x; \alpha', \beta, \gamma, \lambda')$$

where $\alpha' = \alpha \sqrt[\beta]{n+l+1-k}$ and $\lambda' = (n+l+1-k)\lambda$

Using (6.10), the rth non-central moment of the kth order statistics $X_{k:n}$ is

$$\mu_r'^{(k:n)} = \frac{nre^{-\lambda}}{\beta} \binom{n-1}{k-1} \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \sum_{l=0}^{k-1} \binom{k-1}{l} \frac{(-1)^l \lambda^i i^j \alpha^{-(j\gamma+r)}}{(n+l+1-k)^{(j\gamma+r)/\beta-i+1} i!\,j!} \Gamma\left(\frac{j\gamma+r}{\beta}\right)$$

6.4 Parameter estimation

6.4.1 Maximum likelihood estimation

Let $\mathcal{D}: x_1, \ldots, x_n$ be an observed sample from the ACW distribution with unknown parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \lambda)$. Then the log-likelihood function is derived as

$$\mathcal{L} = \sum_{i=1}^n \log\left(\lambda\gamma x_i^{\gamma-1} e^{x_i^\gamma} + \alpha\beta(\alpha x_i)^{\beta-1}\right) - \sum_{i=1}^n \left(\lambda(1 - e^{x_i^\gamma}) + (\alpha x_i)^\beta\right) \tag{6.12}$$

To obtain the MLE of θ, we first calculate the first partial derivatives of $\mathcal{L}$ with respect to α, β, γ and λ.

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} \frac{\beta^2 (\alpha x_i)^{\beta-1}}{\lambda \gamma x_i^{\gamma-1} e^{x_i^{\gamma}} + \alpha \beta (\alpha x_i)^{\beta-1}} - \sum_{i=1}^{n} \beta \alpha^{\beta-1} x_i^{\beta} \tag{6.13}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^{n} \frac{\alpha (\alpha x_i)^{\beta-1} + \alpha \beta (\alpha x_i)^{\beta-1} \log(\alpha x_i)}{\lambda \gamma x_i^{\gamma-1} e^{x_i^{\gamma}} + \alpha \beta (\alpha x_i)^{\beta-1}} - \sum_{i=1}^{n} (\alpha x_i)^{\beta} \log(\alpha x_i) \tag{6.14}$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \sum_{i=1}^{n} \frac{\lambda x_i^{\gamma-1} e^{x_i^{\gamma}} \left(1 + \gamma \log x_i + \gamma x_i^{\gamma} \log x_i \right)}{\lambda \gamma x_i^{\gamma-1} e^{x_i^{\gamma}} + \alpha \beta (\alpha x_i)^{\beta-1}} + \sum_{i=1}^{n} \lambda x_i^{\gamma} e^{x_i^{\gamma}} \log x_i \tag{6.15}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{n} \frac{\gamma x_i^{\gamma-1} e^{x_i^{\gamma}}}{\lambda \gamma x_i^{\gamma-1} e^{x_i^{\gamma}} + \alpha \beta (\alpha x_i)^{\beta-1}} + \sum_{i=1}^{n} (e^{x_i^{\gamma}} - 1) \tag{6.16}$$

Then setting these expressions to zero and solving them simultaneously gives the MLE $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\lambda})$. These equations can not be solved analytically. Therefore, a numerical method should be employed as, for example, the Newton-Raphson algorithm. However, in the study the CE method (Chapter 3) is used to optimize the log-likelihood function given in Eq. (6.12).

6.4.2   Bayesian estimation

The Bayesian model is constructed by specifying a prior distribution $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \lambda)$, and then multiplying with the likelihood function to obtain the posterior distribution. The posterior distribution of $\boldsymbol{\theta}$ given $\mathcal{D}: t_1, \ldots, t_n$ is given by

$$\pi(\boldsymbol{\theta}|\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{6.17}$$

Since the denominator in Eq. (6.17) is a normalizing constant and not necessary for Bayesian inference using MCMC methods, the posterior distribution is often expressed as:

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{6.18}$$

Here also adopted Kundu and Howlader [36], the prior distributions of $\alpha, \beta, \gamma, \theta$ and λ are assumed to be independent and each parameter follows gamma distribution, i.e.

$$\pi_1(\alpha) \propto \alpha^{a_1-1} \exp\{-b_1 \alpha\}, \, a_1, b_1 > 0 \tag{6.19}$$

$$\pi_2(\beta) \propto \beta^{a_2-1} \exp\{-b_2 \beta\}, \, a_2, b_2 > 0 \tag{6.20}$$

$$\pi_3(\gamma) \propto \gamma^{a_3-1} \exp\{-b_3 \gamma\}, \, a_3, b_3 > 0 \tag{6.21}$$

$$\pi_4(\lambda) \propto \lambda^{a_4-1} \exp\{-b_4 \lambda\}, \, a_4, b_4 > 0 \tag{6.22}$$

If $a_1 = a_2 = a_3 = a_4 = 1$ and $b_1 = b_2 = b_3 = b_4 = 0$ we have generalized uniform distributions on $\mathbb{R}^+$ or diffuse priors, and if $a_1 = a_2 = a_3 = a_4 = b_1 = b_2 = b_3 = b_4 = 0$, we have non-informative priors. Then, under the square error loss function, the Bayes estimator of $\alpha, \beta, \gamma, \lambda$, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$\alpha^* = \mathbb{E}(\alpha|\mathcal{D}) = \int_{\boldsymbol{\theta}} \alpha \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \tag{6.23}$$

$$\beta^* = \mathbb{E}(\beta|\mathcal{D}) = \int_{\boldsymbol{\theta}} \beta \pi(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \tag{6.24}$$

$$\gamma^* = \mathbb{E}(\gamma|\mathcal{D}) = \int_{\boldsymbol{\theta}} \gamma \pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{6.25}$$

$$\lambda^* = \mathbb{E}(\lambda|\mathcal{D}) = \int_{\boldsymbol{\theta}} \lambda \pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{6.26}$$

$$h^*(t) = \mathbb{E}(h(t;\boldsymbol{\theta})|\mathcal{D}) = \int_{\boldsymbol{\theta}} h(t;\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{6.27}$$

$$R^*(t) = \mathbb{E}(R(t;\boldsymbol{\theta})|\mathcal{D}) = \int_{\boldsymbol{\theta}} R(t;\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{6.28}$$

In this study, HMC (see Subsection 3.7) is used to simulate samples from posterior distribution. Suppose that $\{\boldsymbol{\theta}_i, i = 1, \ldots, N\}$ is generated from the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$. Then when i is sufficiently large (say, bigger than n_0), $\{\boldsymbol{\theta}_i, i = n_0 + 1, \ldots, N\}$ is a (correlated) sample from the true posterior. Then, the approximate Bayes estimate of α^*, β^*, γ^*, λ^*, $h^*(t)$ and $R^*(t)$ by calculating the means:

$$\alpha^* \approx \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} \alpha_i \tag{6.29}$$

$$\beta^* \approx \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} \beta_i \tag{6.30}$$

$$\gamma^* \approx \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} \gamma_i \tag{6.31}$$

$$\lambda^* \approx \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} \lambda_i \tag{6.32}$$

$$h^*(t) \approx \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}_i) \tag{6.33}$$

$$R^*(t) \approx \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} R(t;\boldsymbol{\theta}_i) \tag{6.34}$$

In practice, some experts suggest to run m parallel chains (say, $m = 3, 4$ or 5), instead of only 1, for assessing sampler convergence. Then the posterior means are obtained as follows

$$\alpha^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \alpha_{i,j} \tag{6.35}$$

$$\beta^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \beta_{i,j} \tag{6.36}$$

$$\gamma^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \gamma_{i,j} \tag{6.37}$$

$$\lambda^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \lambda_{i,j} \tag{6.38}$$

$$h^*(t) \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}_{i,j}) \tag{6.39}$$

Table 6.1: Aarset data

0.1	0.2	1.0	1.0	1.0	1.0	1.0	2.0	3.0	6.0
7.0	11.0	12.0	18.0	18.0	18.0	18.0	18.0	21.0	32.0
36.0	40.0	45.0	46.0	47.0	50.0	55.0	60.0	63.0	63.0
67.0	67.0	67.0	67.0	72.0	75.0	79.0	82.0	82.0	83.0
84.0	84.0	84.0	85.0	85.0	85.0	85.0	85.0	86.0	86.0

Table 6.2: The MLEs of parameters for fitting ACW distribution along with other modified Weibull distributions to Aarset data.

Models	CDF	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$
MW	$1 - e^{-\alpha x^\theta e^{\gamma x}}$	0.0626		0.0230	0.3595	
MWE	$1 - e^{\alpha\lambda\left(1 - e^{(x/\alpha)^\beta}\right)}$	13.7467	0.5877			0.0088
ACW	$1 - e^{\lambda(1 - e^{x^\gamma}) - (\alpha x)^\beta}$	0.0118	86.2313	0.2782		0.0420
AddW	$1 - e^{-\alpha x^\theta - \beta x^\gamma}$	1.1095×10^{-8}	0.0857	0.4765	4.2190	
EMWE	$\left[1 - e^{\alpha\lambda\left(1 - e^{(x/\alpha)^\beta}\right)}\right]^\gamma$	49.1958	3.1645	0.1445		7.0213×10^{-5}
AMW	$1 - e^{-\left(\alpha x^\theta e^{\gamma x} + e^{\lambda x - \beta} - e^{-\beta}\right)}$	0.0763	90.1357	0.0104	0.4579	1.0604
NMW	$1 - e^{-\left(\alpha x^\theta + \beta x^\gamma e^{\lambda x}\right)}$	0.0710	7.0150×10^{-8}	0.0160	0.5950	0.1970

$$R^*(t) \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} R(t; \boldsymbol{\theta}_{i,j}) \tag{6.40}$$

6.5 Applications

In this section, the proposed model is applied to two well-known data sets and a comparison with other typical models based on Kolmogorov-Smirnov (K-S) statistic, Akaike information criterion (AIC), Bayesian information criterion (BIC) and bias-corrected Akaike information criterion (AICc) is provided.

6.5.1 Aarset data

Data in Table 6.1 represent the lifetimes of 50 devices [1]. This failure data has been analysed by numerous authors, see Almalki and Yuan [3] and He, Cui, and Du [28] and references therein. It is known to have a bathtub-shaped failure rate function as indicated by the scaled TTT-transform plot which is first convex and then concave (Fig. 6.3).

Table 6.2 gives the MLEs of parameters of the ACW as well as the MW, MWE, AddW, EMWE, AMW and NMW models when fitting to Aarset data and the measure of fit values are given in Table 6.3. From Table 6.3 we find that the five-parameter model AMW provides the best fit to this data and the proposed four-parameter model ACW fits to the data almost as good as the AMW model. Fig 6.4 shows the reliability functions and the failure rate functions of the ACW, MW, MWE, AddW, EMWE, AMW and NMW models when fitting to the data set. From these plots we can see that the ACW is as close as the AMW to the nonparametric estimates of these functions.

For Bayesian inference, the prior information needs to be specified. Since we have no prior information available, the diffuse priors are used as the prior information for
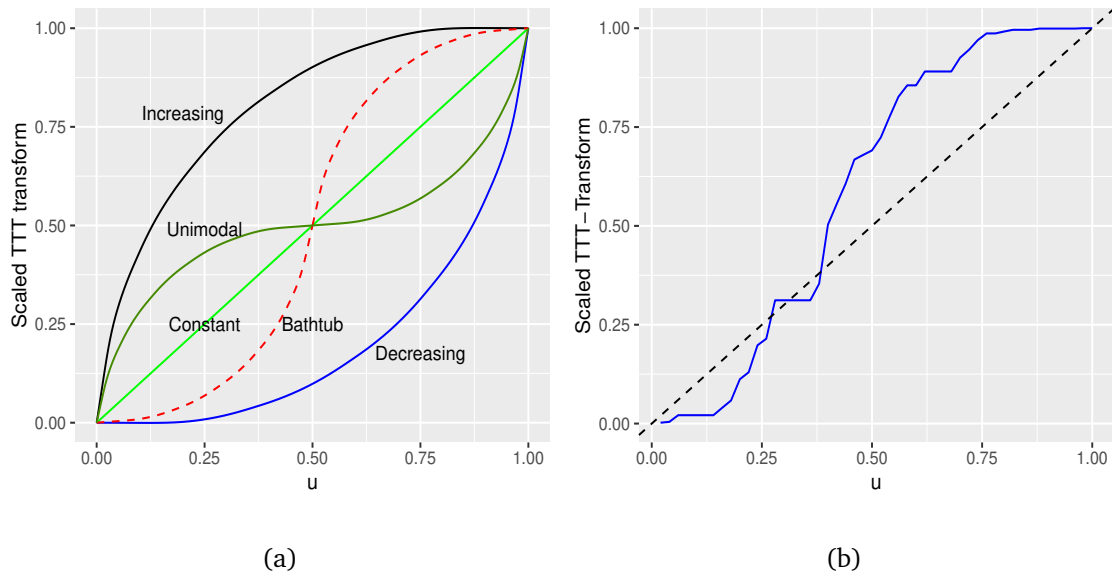
Figure 6.3: (a) Shapes of the scaled TTT-transform plot with corresponding types of failure rate and (b) the empirical scaled TTT-transform plot for Aarset data.
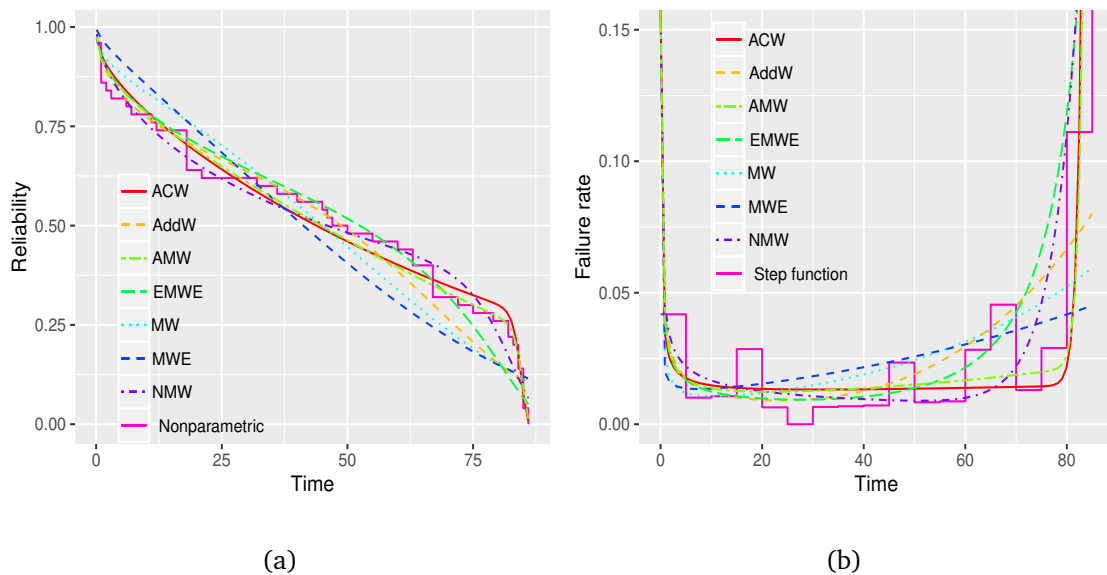
Figure 6.4: The estimated (a) reliability functions and (b) failure rate functions obtained by fitting ACW distribution and other modified Weibull distributions to Aarset data.

the model parameters. Fig. 6.5 shows the trace plots and density estimates of the parameters obtained by HMC algorithm. The trace plots show that the 4 parallel chains for each parameter produced by HMC algorithm converge quickly to the same target distribution. The densities are distributed approximately symmetrically around the central values which means that they provide good Bayes estimates under square error loss function. The scatter plot matrix of HMC output shows the posterior correlations between the parameters (Fig. 6.7). In the graph, most pairs of parameters have small posterior correlations whereas (γ, λ) appear to have higher posterior correlations. This is due to the parameterization of the Chen distribution. These higher correlations, however, have little effect on the Bayes estimators in this situation.

Table 6.4 shows the HMC point estimates and two-sided 90% and 95% HPD (highest

Table 6.3: Log-likelihood, K-S statistic, AIC, BIC and AICc for fitting ACW distribution along with other modified Weibull distribution to Aarset data.

Model	Log-lik	K-S (p-value)	AIC	BIC	AICc
3 parameters					
MW	-227.16	$0.134\,(0.328)$	460.32	466.06	460.84
MWE	-231.65	$0.161\,(0.150)$	469.29	475.03	469.81
4 parameters					
ACW	-205.35	$0.070\,(0.965)$	418.71	426.36	419.60
AddW	-221.49	$0.127\,(0.397)$	450.98	458.63	451.87
EMWE	-213.86	$0.139\,(0.287)$	435.72	443.37	436.61
5 parameters					
AMW	-203.57	$0.068\,(0.976)$	417.14	426.70	418.50
NMW	-212.90	$0.088\,(0.838)$	435.80	445.36	437.16

posterior density) intervals for $\alpha, \beta, \gamma, \lambda$ and MTTF. Fig. 6.6 displays the estimated reliability and failure rate functions obtained by MLE and Bayesian methods when fitting ACW to the data. It is easy to see that both methods provide comparable results.
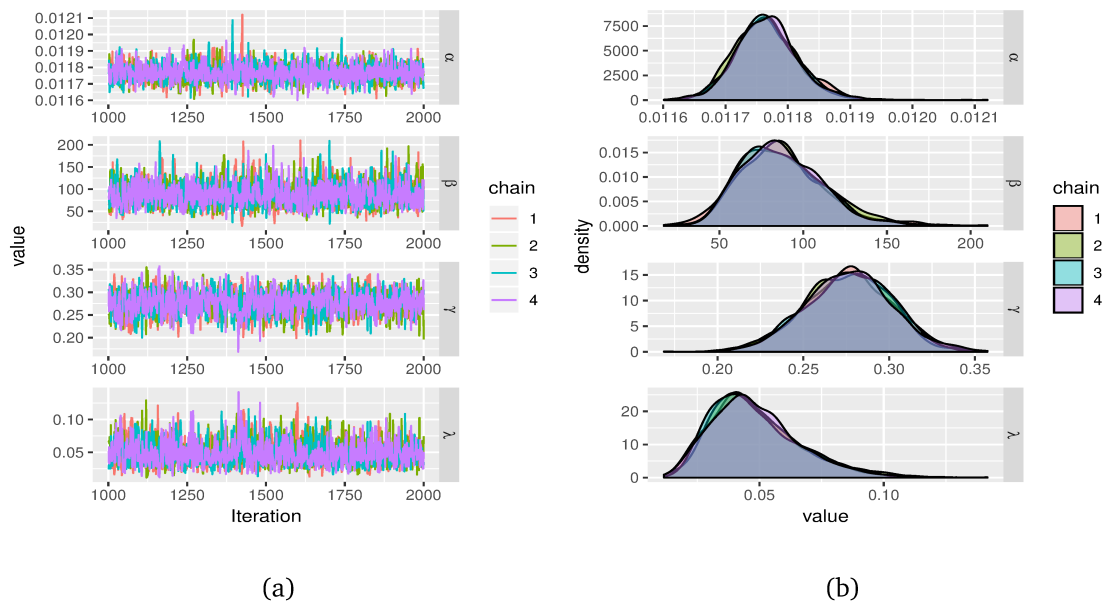
Figure 6.5: (a) Trace plots and (b) density estimates of the parameters produced by HMC sampling with 4 parallel chains obtained by fitting ACW to Aarset data.

6.5.2   Meeker-Escobar data

Data in Table 6.5 represent failure and running times of 30 devices [43, p. 383]. Two failure modes were observed for this dataset. This data set has been studied extensively by numerous authors, and the most recent studies are given by Almalki and Yuan [3] and He, Cui, and Du [28]. As indicated by the scaled TTT-transform plot given in Fig. 6.8, the data is shown to have a bathtub-shaped failure rate.

Table 6.6 gives the MLEs of parameters of the ACW as well as the MW, MWE, AddW, EMWE, AMW and NMW models when fitting to Meeker-Escobar data and the measure

Table 6.4: Bayes estimates via HMC and HPD intervals for the parameters and MTTF for fitting ACW to Aarset data

Parameter	Point estimate	90% HPD	95% HPD
α	0.0118	$[0.0117, 0.0118]$	$[0.0117, 0.0119]$
β	87.5177	$[48.7484, 126.2103]$	$[42.3028, 136.8757]$
γ	0.2761	$[0.2339, 0.3153]$	$[0.2272, 0.3247]$
λ	0.0484	$[0.0199, 0.0749]$	$[0.0199, 0.0864]$
$MTTF$	43.7496	$[36.7722, 51.0436]$	$[35.0335, 52.0174]$

(a)

(b)

Figure 6.6: MLE, Bayes and nonparametric estimates of the (a) reliability functions and (b) failure rate functions obtained by fitting ACW distribution to Aarset data.

of fit values are given in Table 6.7. From Table 6.7 we find that in this case the proposed four-parameter model, ACW, provides the best fit to Meeker-Escobar data. Fig 6.9 shows the estimated reliability functions and the failure rate functions of the ACW, MW, MWE, AddW, EMWE, AMW and NMW models when fitting to Meeker-Escobar data. From these plots we can see that the ACW is very close to the nonparametric estimates of these functions.

For analyzing Meeker-Escobar data, we use the same procedures given in Subsection 6.5.1 for Bayesian estimation. However, for this data set, the informative prior is used since we have very little data points. For the prior distributions given in Eqs. 6.20-6.22, the hyper-parameters have been chosen such that the prior means approximate the MLEs of the parameters. Since the MLEs of parameters for fitting ACW to Meeker-Escobar

Table 6.5: Meeker-Escobar data: running times of 30 devices

2	10	13	23	23	28	30	65	80	88
106	143	147	173	181	212	245	247	261	266
275	293	300	300	300	300	300	300	300	300

Table 6.6: The MLEs of parameters for fitting ACW distribution along with other modified Weibull distributions to Meeker-Escobar data.

Models	CDF	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$
MW	$1 - e^{-\alpha x^{\theta} e^{\gamma x}}$	0.0181		0.0071	0.4538	
MWE	$1 - e^{\alpha\lambda\left(1 - e^{(x/\alpha)^{\beta}}\right)}$	85.4922	0.8020			0.0016
ACW	$1 - e^{\lambda(1 - e^{x^{\gamma}}) - (\alpha x)^{\beta}}$	0.00333	259.42759	0.26068		0.01518
AddW	$1 - e^{-(\alpha x)^{\beta} - (\theta x)^{\gamma}}$	1.3109×10^{-7}	0.0187	0.6024	2.8358	
EMWE	$\left[1 - e^{\alpha\lambda\left(1 - e^{(x/\alpha)^{\beta}}\right)}\right]^{\gamma}$	197.2165	4.4955	0.1289		5.4673×10^{-6}
AMW	$1 - e^{-\left(\alpha x^{\theta} e^{\gamma x} + e^{\lambda x - \beta} - e^{-\beta}\right)}$	0.0142	116.9665	0.0019	0.6788	0.3902
NMW	$1 - e^{-\left(\alpha x^{\theta} + \beta x^{\gamma} e^{\lambda x}\right)}$	0.024	5.991×10^{-8}	0.012	0.629	0.056

Table 6.7: Log-likelihood, K-S statistic, AIC, BIC and AICc for fitting ACW distribution along with other modified Weibull distributions to Meeker-Escobar data.

Model	Log-lik	K-S (p-value)	AIC	BIC	AICc
3 parameters					
MW	-178.06	$0.180\,(0.282)$	362.12	366.32	363.04
MWE	-179.20	$0.184\,(0.261)$	364.41	368.61	365.33
4 parameters					
ACW	-151.34	$0.134\,(0.652)$	310.67	316.28	312.27
AddW	-178.10	$0.193\,(0.214)$	364.20	369.80	365.80
EMWE	-166.34	$0.191\,(0.223)$	340.68	346.28	342.28
5 parameters					
AMW	-155.58	$0.167\,(0.374)$	321.16	328.17	323.66
NMW	-166.18	$0.149\,(0.522)$	342.36	349.37	344.86
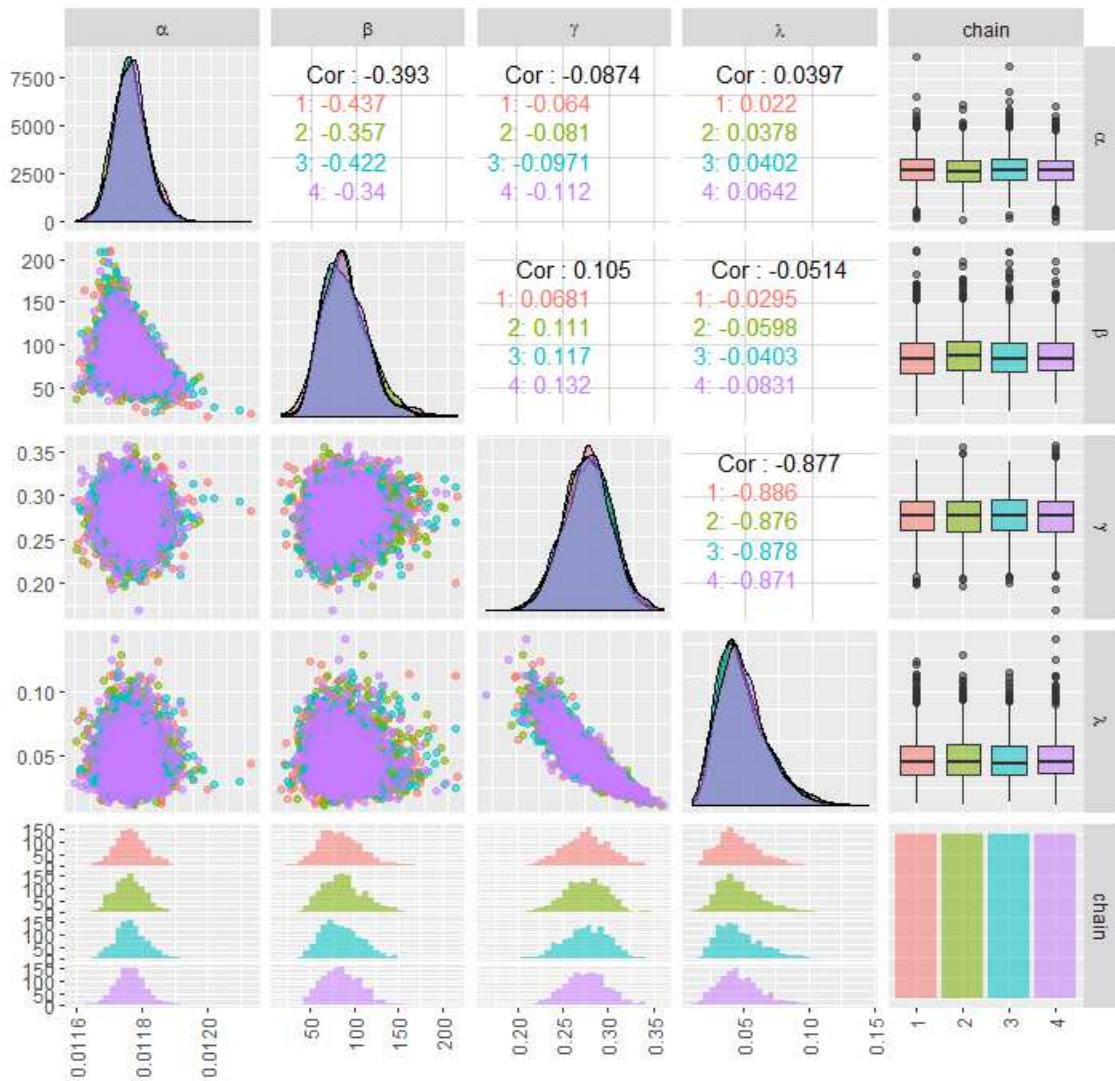
Figure 6.7: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting ACW to Aarset data.

data are $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\lambda}) = (0.00333, 259.42759, 0.26068, 0.01518)$, the hyper-parameters have been chosen as $(a_1 = 50, b_1 = 50/0.00333), (a_2 = 50, b_2 = 50/259.42759), (a_3 = 50, b_3 = 50/0.26068)$, and $(a_4 = 50, b_4 = 50/0.01518)$.

Fig. 6.10 shows the trace plots and density estimates of the parameters obtained by HMC algorithm. The trace plots show that the 4 parallel chains for each parameter produced by HMC algorithm also converge quickly to the same target distribution. The estimated densities are distributed almost symmetrically around the central values which means that they provide good Bayes estimates under square error loss function, thanks to the informative prior. The scatter plot matrix of HMC output shows the estimated posterior correlations between the parameters (Fig. 6.11). Most pairs of parameters have a very small correlation whereas the pair (β, λ) appears to have a little higher correlation.

Table 6.8 shows the HMC point estimates and two-sided 90% and 95% HPD (highest posterior density) intervals for $\alpha, \beta, \gamma, \lambda$ and MTTF. Fig. 6.12 displays the estimated reliability and failure rate functions obtained by MLE and Bayesian methods when fitting ACW to the data. It is easy to see that both methods also give comparable results.
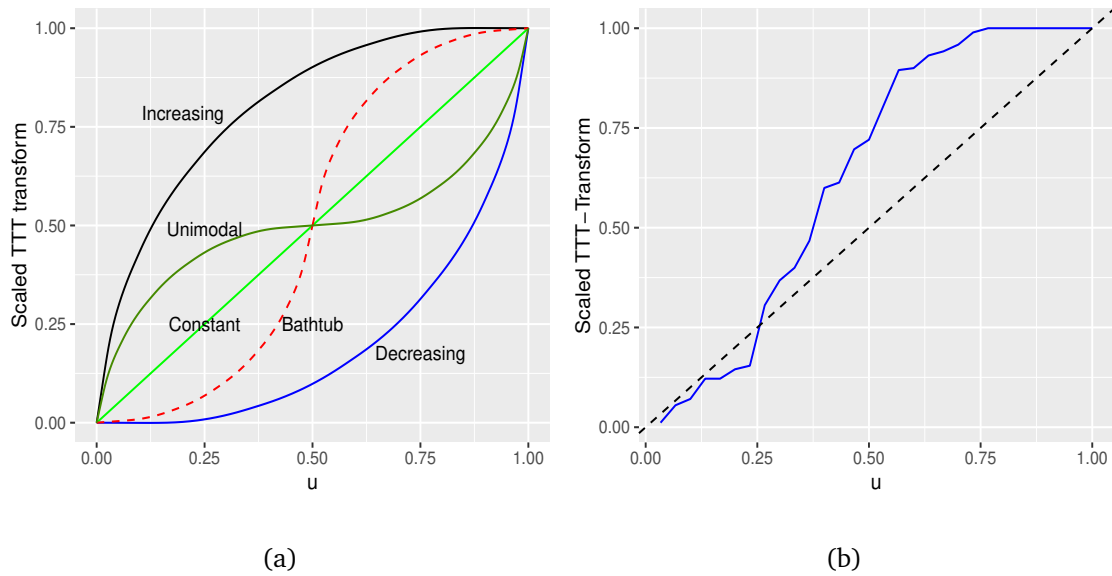
Figure 6.8: (a) Shapes of the scaled TTT-transform plot with corresponding types of failure rate and (b) the empirical scaled TTT-transform plot for Meeker-Escobar data.
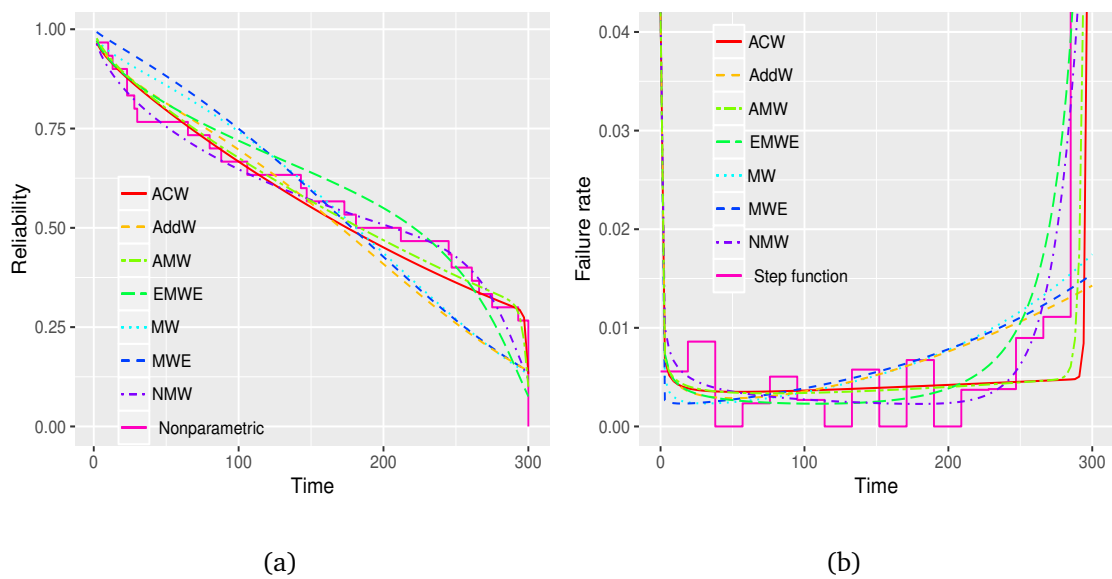
Figure 6.9: The estimated (a) reliability functions and (b) failure rate functions obtained by fitting ACW distribution and other modified Weibull distributions to Meeker-Escobar data.

6.6   Conclusions

The ACW distribution has been developed and has been demonstrated to be better than many existing models for modeling the two well-known data sets. The new distribution has a simple form and flexible failure rate function which might appropriate for many real data sets. Both classical and Bayesian inferences for its parameters have been investigated. With the power of modern computations as, for example, the cross-entropy method for optimization, MCMC simulation methods such as Hamiltonian Monte Carlo method, resampling methods like bootstrapping, researchers can develop new statistical models with many parameters that can be flexible enough to meet the realistic demand.
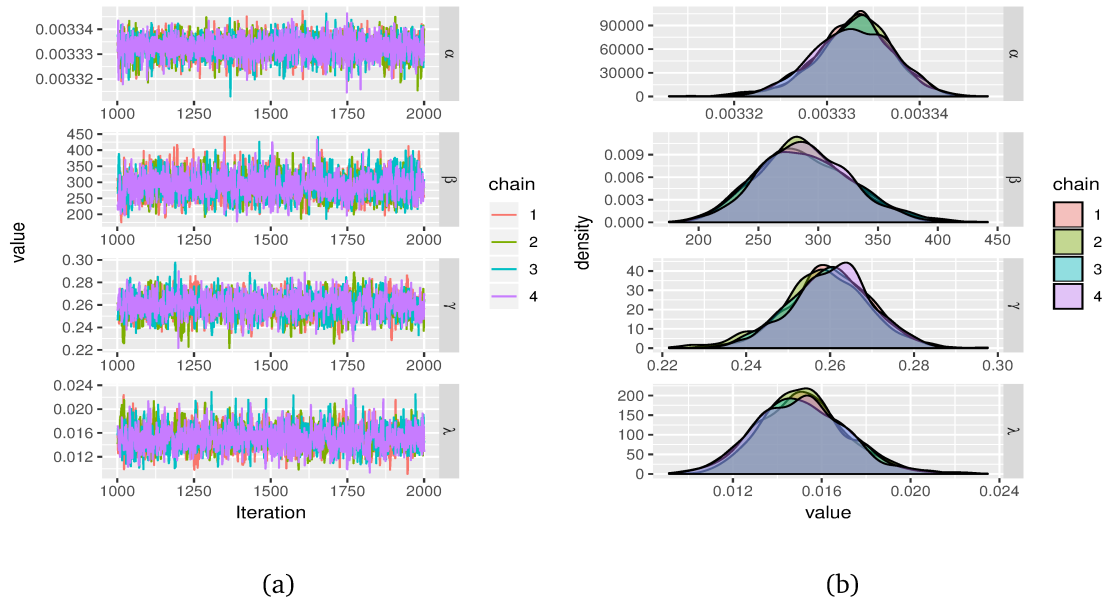
Figure 6.10: (a) Trace plots and (b) density estimates of the parameters produced by HMC sampling with 4 parallel chains obtained by fitting ACW to Meeker-Escobar data.

Table 6.8: Bayes estimates via HMC and HPD intervals for the parameters and MTTF for fitting ACW to Meeker-Escobar data

Parameter	Point estimate	90% HPD	95% HPD
α	0.0033	$[0.0033, 0.0033]$	$[0.0033, 0.0033]$
β	288.2150	$[220.9576, 348.8948]$	$[214.2431, 365.3712]$
γ	0.2596	$[0.2432, 0.2758]$	$[0.2392, 0.2787]$
λ	0.0151	$[0.0117, 0.0182]$	$[0.0114, 0.0192]$
$MTTF$	174.1098	$[150.0000, 198.0324]$	$[146.3728, 202.9088]$
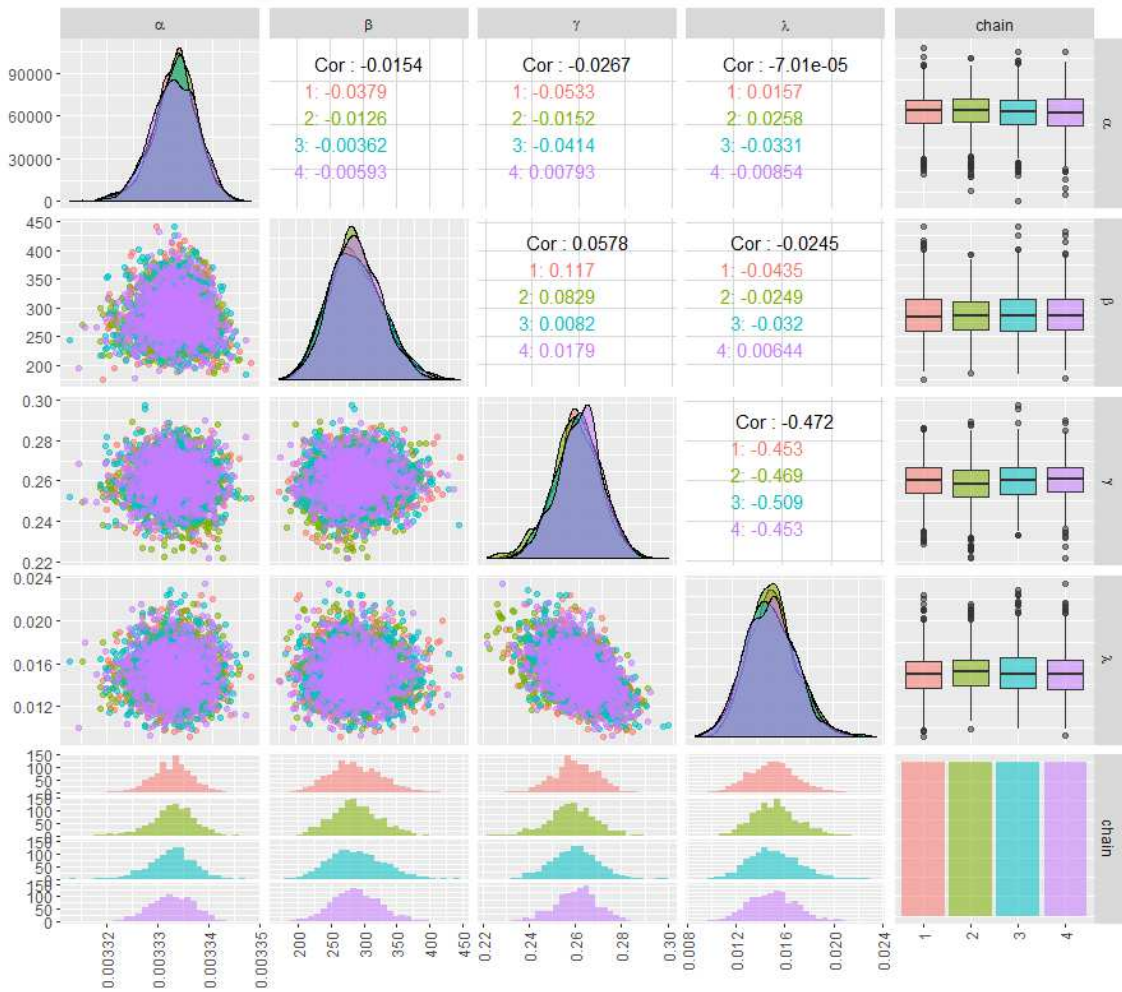
Figure 6.11: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting ACW to Meeker-Escobar data.

(a)   (b)
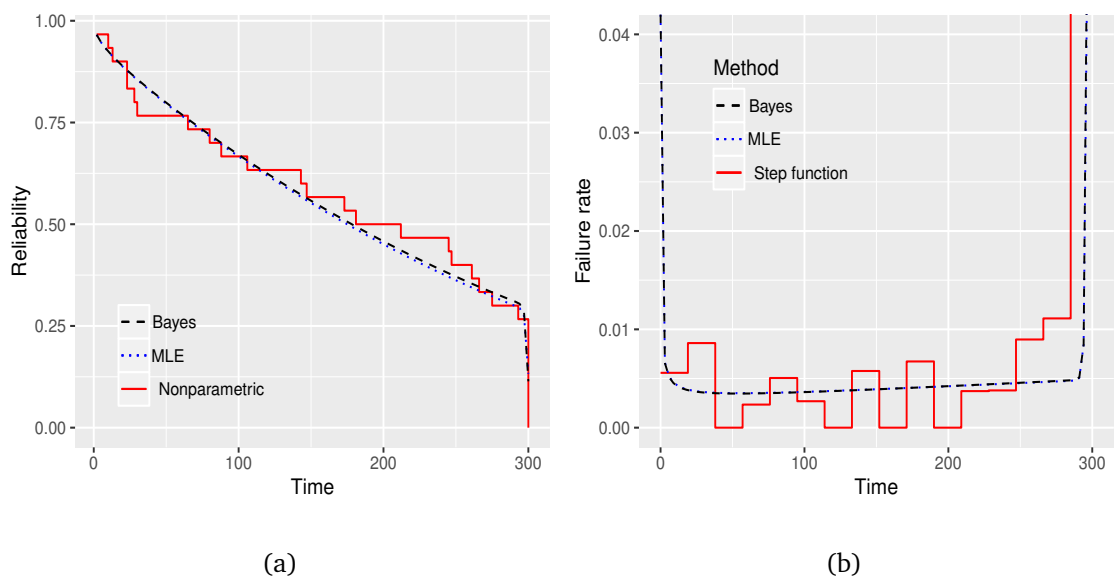
Figure 6.12: MLE, Bayes and nonparametric estimates of the (a) reliability functions and (b) failure rate functions obtained by fitting ACW distribution to Meeker-Escobar data.

Chapter 7

Improving new modified Weibull model: A Bayes study using Hamiltonian Monte Carlo simulation

7.1 Introduction

This chapter comes from my study given in [87]. As we know the bathtub-shaped failure rate function is the most popular non-monotonic failure rate function which can be used for modeling of human mortality, failure rate of newly launched products, etc. In 2013, a new modified Weibull (NMW) distribution has been published in a journal of engineering [3]. This new model provides best fits of specific data sets, that evince bathtub-shaped failure rate, comparing to all other existing models. (Here we skip the discussion of earlier developments of bathtub-shaped failure rate models since it was already provided by Almalki and Yuan [3].) Therein, authors introduce a new lifetime distribution by considering a series system with one component following the Weibull distribution and another following the modified Weibull distribution [39]. The NMW distribution has been defined by the following CDF:

$$F(t) = 1 - \exp\left\{-\left(\alpha t^\theta + \beta t^\gamma e^{\lambda t}\right)\right\}, \quad t \geq 0 \tag{7.1}$$

where $\alpha, \beta, \theta, \gamma$ and λ are non-negative, with θ and γ being shape parameters and α and β being scale parameters and λ acceleration parameter. Its corresponding failure rate function is given as

$$h(t) = \alpha\theta t^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}, \quad t \geq 0 \tag{7.2}$$

This is the mixture failure rate of the Weibull and modified Weibull failure rates. The important property of preceding failure rate function is that it enables a more flexible bathtub shaped failure rate function comparing to other existing alternative failure time models.

Short time later, a new additive modified Weibull (AMW) distribution has been released and published in the same journal as the NMW model [28]. The new model was demonstrated to be even better than the NMW and other existing models for fitting to the data sets. Therein, the authors also considered a series system of two components. But instead, one component is still following the modified Weibull distribution and another is following a special case of Gompertz's distribution [25]. The AMW distribution has been defined by the following CDF:

$$F(t) = 1 - \exp\left\{-\left(\alpha t^\theta e^{\gamma t} + e^{\lambda t - \beta} - e^{-\beta}\right)\right\}, \quad t \geq 0 \tag{7.3}$$

where $\alpha > 0, \beta > 0, \theta > 0, \gamma \geq 0$ and $\lambda \geq 0$. Its corresponding failure rate function is given as

$$h(t) = \alpha(\theta + \gamma t)t^{\theta-1}e^{\gamma t} + \lambda e^{\lambda t - \beta}, \quad t \geq 0 \tag{7.4}$$

This chapter shows that the failure rate function given in Eq. (7.2) can be adjusted into a slightly different form which fits to bathtub-shaped failure data at least as good as the model introduced by He, Cui, and Du [28]. Then a full Bayesian analysis of the model is provided. Bayes estimators under square error loss function are obtained by using Hamiltonian Monte Carlo (HMC), a Markov chain Monte Carlo (MCMC) method, for posterior simulations. Furthermore, MLEs of model parameters are obtained by using the cross-entropy (CE) method to optimize the log-likelihood function and using the invariance property of MLE to provide the MLEs of reliability characteristics.

As mentioned by Gupta, Mukherjee, and Upadhyay [26], the idea of combining two or more models will result in too many parameters of the proposed model and, as such, the resulting inferences may be difficult to obtain, especially in the reliability studies with too little amount of data. Indeed, we have rarely seen or perhaps have never seen so far a publication, that provides a full Bayesian analysis of failure time distributions with more than three parameters. In this regard, the conventional MCMC methods are hard to implement to find a good posterior sample. Therefore, this study shows the advantage of the HMC algorithm which allow us to successfully simulate posterior samples from such models, even in case of high correlated parameters or in high dimensions. Here, the model under study has five parameters and this study will demonstrate how to apply successfully HMC algorithm for posterior simulation of the model. Because recently there are only few publications that provide Bayesian model checking, this chapter provides another routine to continue search in this issue.

The remainder of this chapter is organized as follows. The improved model is introduced in Section 7.2 along with the basic characteristics of the corresponding lifetime distribution. Section 7.3 provides MLEs of unknown parameters as well as reliability characteristics of the model on one hand, and Bayes estimators on the other hand. Section 7.4 demonstrates the proposed innovative model to two well-known data sets adopted from references. Finally, Section 7.5 brings conclusions.

7.2 The model and its characteristics

7.2.1 Improving NMW model (INMW)

Based on the results obtained by Almalki and Yuan [3], this study point out an important property of the failure rate function in Eq. (7.2). As we can see, the first component of the failure rate function in Eq. (7.2) is the Weibull failure rate function which can be increasing, decreasing or constant, and the second component is the modified Weibull failure rate function which can be either increasing or bathtub-shaped [39] and as pointed out by Upadhyay and Gupta [69] this failure rate function does not provide a good fit for bathtub-shaped failure rate data with sharp change in the wear-out phase like Aarset data or Meeker-Escobar data, for example. When Almalki and Yuan [3] applied their model for fitting to the data sets, they obtained the estimates of γ and θ both less than unity in both cases. This means that the first component is decreasing and has been absorbed into the decreasing part of the second component which has a bathtub curve. This means that their model is not as flexible as the model introduced by He, Cui, and Du [28]. To improve the model, the first component of the failure rate function in Eq. (7.2) is coerced to be always increasing (by restricting $\theta > 1$) which gives the following new formula of

the mixture failure rate:

$$h(t) = \alpha\theta(\alpha t)^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}, \ \alpha, \beta, \gamma, \lambda \geq 0, \theta > 1 \tag{7.5}$$

In the preceding function, the first component on the right side of (7.2) is also reparameterized in order to reduce correlated parameters so that it can work well with MCMC methods. In addition, this new modified model has another good property that when reduced into a sub-model by setting $\lambda = 0$, it produces a four-parameter model which avoids the non-identifiability problem:

$$h(t) = \alpha\theta(\alpha t)^{\theta-1} + \beta\gamma t^{\gamma-1} \tag{7.6}$$

whereas the sub-model of the failure rate function in Eq. (7.2) when $\lambda = 0$ is

$$h(t) = \alpha\theta t^{\theta-1} + \beta\gamma t^{\gamma-1} \tag{7.7}$$

that commits the non-identifiability problem which means that two or more parameter sets result in the same model, i.e. such model would be ambiguous. In fact, the failure rate models in Eqs. (7.6) and (7.7) are just the failure rate model defined by Xie and Lai [75], but with slightly different parametrization. Notice that the failure rate model introduced by Xie and Lai [75] does not commit the non-identifiability problem due to the fact that the authors restricted the parameters.

7.2.2 Characteristics of lifetime distribution

Here the characteristics of the lifetime distribution of the INMW model is derived. Using the relationship between reliability and failure rate functions, the reliability/survival function is

$$R(t) = \exp\left\{-\int_0^t h(s)ds\right\} = \exp\left\{-(\alpha t)^\theta - \beta t^\gamma e^{\lambda t}\right\} \tag{7.8}$$

Then, the probability density (PDF) function is given as

$$f(t) = h(t)R(t) = \left(\alpha\theta(\alpha t)^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}\right)\exp\left\{-(\alpha t)^\theta - \beta t^\gamma e^{\lambda t}\right\} \tag{7.9}$$

The cumulative failure rate (CFR) function is given by

$$H(t) = -\log(R(t)) = (\alpha t)^\theta + \beta t^\gamma e^{\lambda t} \tag{7.10}$$

And the mean time to failure (MTTF) is given by

$$MTTF = \mathbb{E}(T)$$
$$= \int_0^\infty R(t)dt = \int_0^\infty \exp\left\{-(\alpha t)^\theta - \beta t^\gamma e^{\lambda t}\right\}dt \tag{7.11}$$

This integral can be obtained by using some suitable numerical methods.

7.3 Estimation of parameters and reliability characteristics

Let $\mathcal{D}: t_1, \ldots, t_n$ be a random sample from the INMW distribution with parameter $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \theta, \lambda)$. Then the likelihood function is given as

$$L(\mathcal{D}|\boldsymbol{\theta}) = \left[\prod_{i=1}^{n} \left(\alpha\theta(\alpha t_i)^{\theta-1} + \beta(\gamma + \lambda t_i)t_i^{\gamma-1}e^{\lambda t_i} \right) \right] \times \exp \left\{ -\sum_{i=1}^{n} \left((\alpha t_i)^{\theta} + \beta t_i^{\gamma}e^{\lambda t_i} \right) \right\}$$
(7.12)

and the log-likelihood function is derived as

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \left(\alpha\theta(\alpha t_i)^{\theta-1} + \beta(\gamma + \lambda t_i)t_i^{\gamma-1}e^{\lambda t_i} \right) - \sum_{i=1}^{n} \left((\alpha t_i)^{\theta} + \beta t_i^{\gamma}e^{\lambda t_i} \right) \quad (7.13)$$

7.3.1 Maximum likelihood estimation

In this study, the log-likelihood function in Eq. (7.13) is maximized by using CE algorithm to produce the maximizer $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\lambda})$. And using the invariance property of MLE,

1. The MLE for $R(t)$, say $\hat{R}(t)$, will be

$$\hat{R}(t) = \exp \left\{ -(\hat{\alpha}t)^{\hat{\theta}} - \hat{\beta}t^{\hat{\gamma}}e^{\hat{\lambda}t} \right\} \tag{7.14}$$

2. The MLE for $h(t)$, say $\hat{h}(t)$, will be

$$\hat{h}(t) = \hat{\alpha}\hat{\theta}(\hat{\alpha}t)^{\hat{\theta}-1} + \hat{\beta}(\hat{\gamma} + \hat{\lambda}t)t^{\hat{\gamma}-1}e^{\hat{\lambda}t} \tag{7.15}$$

3. The MLE for $H(t)$, say $\hat{H}(t)$, will be

$$\hat{H}(t) = (\hat{\alpha}t)^{\hat{\theta}} + \hat{\beta}t^{\hat{\gamma}}e^{\hat{\lambda}t} \tag{7.16}$$

4. The MLE for MTTF will be

$$M\hat{T}TF = MTTF(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\lambda}) \tag{7.17}$$

which can be obtained by installing into formula (7.11) and integrating.

7.3.2 Bayesian estimation

The Bayesian model is constructed by specifying a prior distribution $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \theta, \lambda)$, and then multiplying with the likelihood function to obtain the posterior distribution. The posterior distribution of $\boldsymbol{\theta}$ given $\mathcal{D}: t_1, \ldots, t_n$ is given by

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{7.18}$$

The prior distributions of $\alpha, \beta, \gamma, \theta$ and λ are assumed to be independent and each parameter follows gamma distribution, i.e.

$$\pi_1(\alpha) \propto \alpha^{a_1-1} \exp \left\{ -b_1\alpha \right\}, \ a_1, b_1 > 0 \tag{7.19}$$
$$\pi_2(\beta) \propto \beta^{a_2-1} \exp \left\{ -b_2\beta \right\}, \ a_2, b_2 > 0 \tag{7.20}$$
$$\pi_3(\gamma) \propto \gamma^{a_3-1} \exp \left\{ -b_3\gamma \right\}, \ a_3, b_3 > 0 \tag{7.21}$$
$$\pi_4(\theta) \propto \theta^{a_4-1} \exp \left\{ -b_4\theta \right\}, \ a_4, b_4 > 0 \tag{7.22}$$

$$\pi_5(\lambda) \propto \lambda^{a_5-1} \exp\left\{-b_5\lambda\right\}, \, a_5, b_5 > 0 \tag{7.23}$$

If $a_1 = a_2 = a_3 = a_4 = a_5 = 1$ and $b_1 = b_2 = b_3 = b_4 = b_5 = 0$ we have generalized uniform distributions on $\mathbb{R}^+$ or diffuse priors, and if $a_1 = a_2 = a_3 = a_4 = a_5 = b_1 = b_2 = b_3 = b_4 = b_5 = 0$, we have non-informative priors. Then, under the square error loss function, the Bayes estimator of $\alpha, \beta, \gamma, \theta, \lambda$, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$\alpha^* = \mathbb{E}(\alpha|\mathcal{D}) = \int_{\boldsymbol{\theta}} \alpha\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.24}$$

$$\beta^* = \mathbb{E}(\beta|\mathcal{D}) = \int_{\boldsymbol{\theta}} \beta\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.25}$$

$$\gamma^* = \mathbb{E}(\gamma|\mathcal{D}) = \int_{\boldsymbol{\theta}} \gamma\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.26}$$

$$\theta^* = \mathbb{E}(\theta|\mathcal{D}) = \int_{\boldsymbol{\theta}} \theta\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.27}$$

$$\lambda^* = \mathbb{E}(\lambda|\mathcal{D}) = \int_{\boldsymbol{\theta}} \lambda\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.28}$$

$$h^*(t) = \mathbb{E}(h(t;\boldsymbol{\theta})|\mathcal{D}) = \int_{\boldsymbol{\theta}} h(t;\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.29}$$

$$R^*(t) = \mathbb{E}(R(t;\boldsymbol{\theta})|\mathcal{D}) = \int_{\boldsymbol{\theta}} R(t;\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta} \tag{7.30}$$

In this study, HMC is also used to simulate samples from posterior distribution. Suppose that $\{\boldsymbol{\theta}_i, i = 1, \ldots, N\}$ is generated from the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$. Then when i is sufficiently large (say, bigger than n_0), $\{\boldsymbol{\theta}_i, i = n_0 + 1, \ldots, N\}$ is a (correlated) sample from the true posterior. Then, the approximate Bayes estimate of α^*, β^*, γ^*, θ^*, λ^*, $h^*(t)$ and $R^*(t)$ by calculating the means:

$$\alpha^* \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} \alpha_i \tag{7.31}$$

$$\beta^* \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} \beta_i \tag{7.32}$$

$$\gamma^* \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} \gamma_i \tag{7.33}$$

$$\theta^* \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} \theta_i \tag{7.34}$$

$$\lambda^* \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} \lambda_i \tag{7.35}$$

$$h^*(t) \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}_i) \tag{7.36}$$

$$R^*(t) \approx \frac{1}{N-n_0} \sum_{i=n_0+1}^{N} R(t;\boldsymbol{\theta}_i) \tag{7.37}$$

Here again, m parallel chains are run (say, $m = 3, 4$ or 5), instead of only 1, for assessing sampler convergence. Then the posterior means are obtained as follows

$$\alpha^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \alpha_{i,j} \tag{7.38}$$

$$\beta^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \beta_{i,j} \tag{7.39}$$

$$\gamma^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \gamma_{i,j} \tag{7.40}$$

$$\theta^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \theta_{i,j} \tag{7.41}$$

$$\lambda^* \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} \lambda_{i,j} \tag{7.42}$$

$$h^*(t) \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} h(t; \boldsymbol{\theta}_{i,j}) \tag{7.43}$$

$$R^*(t) \approx \frac{1}{m(N - n_0)} \sum_{j=1}^{m} \sum_{i=n_0+1}^{N} R(t; \boldsymbol{\theta}_{i,j}) \tag{7.44}$$

7.4 Application

7.4.1 Aarset data

This section provides an application of the new model to the Aarset data given in Chapter 6. It is known to have a bathtub-shaped failure rate function. In order to obtain the Bayes estimates of the parameters and reliability characteristics, the HMC algorithm is implemented in order to simulate samples from the posterior distribution by constructing 4 parallel Markov chains, each of length 2000, with burn-in (warm-up) of 1000 and final posterior sample of size 1000 for each chain is obtained. For the HMC algorithm, there is no need to reduce the autocorrelation, i.e. setting "lag", in the samples. The diffuse priors are used as prior information for parameters.

Fig. 7.1 shows the trace plots and density estimates of the parameters obtained by HMC algorithm. The trace plots show that the 4 parallel chains for each parameter produced by HMC algorithm converge quickly to the same target distribution. The densities are distributed approximately symmetrically around the central values which means that they provide good Bayes estimates under square error loss function. The scatter plot matrix of HMC output shows the posterior correlations between the parameters (Fig. 7.2). In the graph, most pairs of parameters have small posterior correlations whereas the two pairs (β, γ) and (γ, λ) appear to have higher posterior correlations. This is due to the parameterization of the modified Weibull distribution [39]. These higher correlations, however, have little effect on the Bayes estimators in this situation.

Table 7.1 shows the HMC point estimates and two-sided 90% and 95% HPD (highest posterior density) intervals for $\alpha, \beta, \gamma, \theta, \lambda$ and MTTF. Fig. 7.3 displays the time courses of the estimated reliability and failure rate functions obtained by CE and HMC methods when fitting INMW and NMW to the data. It is easy to see that INMW fits to the data set much better than its original NMW.

Table 7.2 provides the measure of fit values, i.e. the log-likelihood (Log-lik), Kolmogorov-Smirnov (K-S) test, Akaike information criterion (AIC), Baysian information criterion (BIC), and corrected Akaike information criterion (AICc), resulting from fitting the INMW, NMW and AMW models to the data set. The corresponding MLEs of parameters are given in Table 7.3. From Table 7.2 we can observe that the INMW model fits to Aarset data at least as good as the AMW model. These results can also be seen in Fig. 7.4. Here, only the comparison with the two newly best models is provided. For an exhausted comparison with other existing models, the readers are referred to Almalki and Yuan [3] and He, Cui, and Du [28].

Table 7.4 provides the Bayesian model checking when fitting INMW, NMW and AMW to the data. Adopted Gupta, Mukherjee, and Upadhyay [26] and Upadhyay and Gupta [69], here the smallest and largest observed values, $T_{\min}$ and $T_{\max}$, are chosen as the test statistics for calculating Bayesian p-value since discrepancy, if any, between the model and data, generally occurs in the tails whereas most reliability models provide good fit to the data in the central region. Fig. 7.5 displays the density estimates of the choosing test statistics based on replicated future data sets, where the observed test values are also displayed by means of vertical lines. From the results we see, that the INMW and AMW provide good fits to the data, whereas the NMW does not provide a good fit at the upper tail. The results in Table 7.4 also show that the INMW model is slightly better than the AMW model. The DIC for the NMW model in Table 7.4 cannot be obtained since the p_D value is negative. This is due to the fact that the high correlated parameters inside the model make the posterior distributions of the parameters to be extremely skewed. As a result of this, the posterior means are very poor estimators of the parameters. This is also the reason why the model is needed to reparameterize in early section.
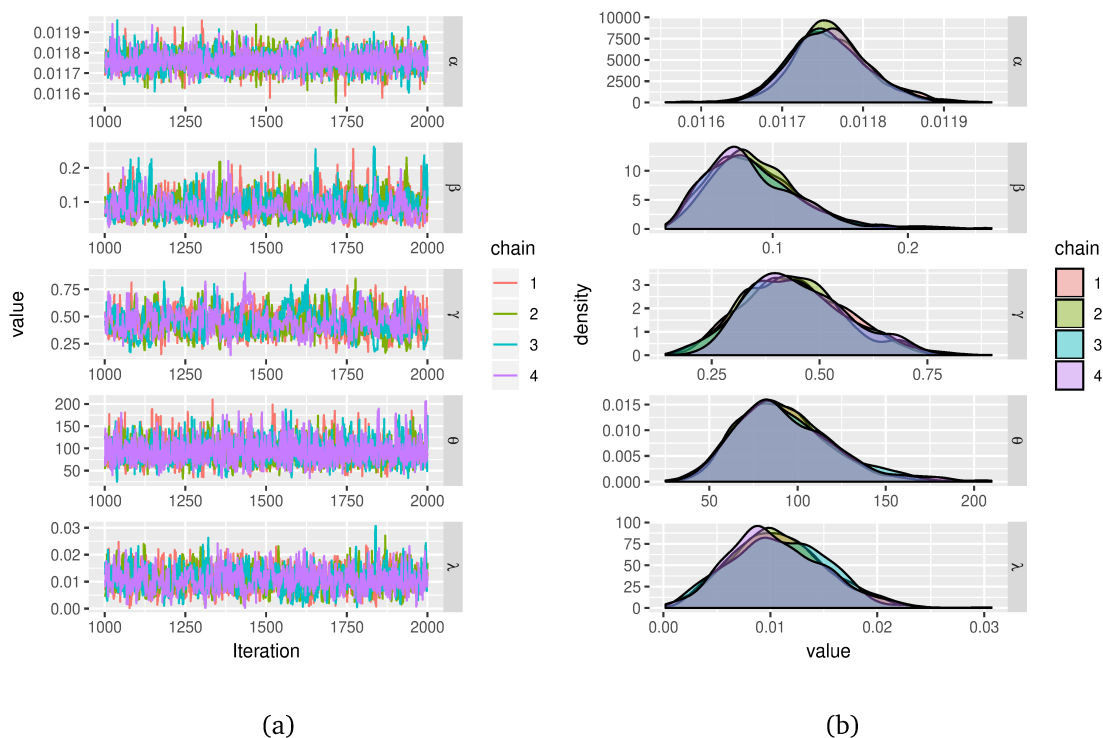

(a)   (b)

Figure 7.1: (a) Trace plots and (b) density estimates of the parameters produced by HMC sampling with 4 parallel chains obtained by fitting INMW to Aarset data.
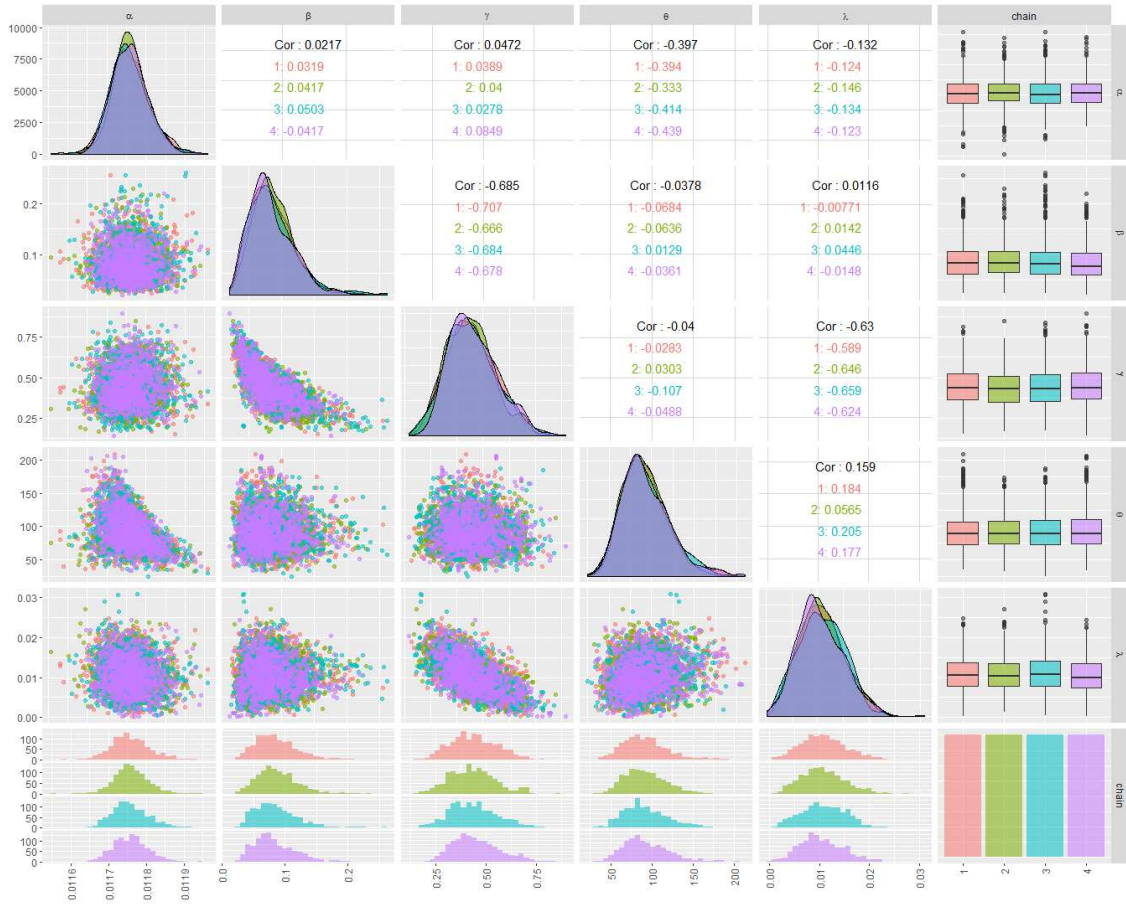
Figure 7.2: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting INMW to Aarset data.

Table 7.1: Bayes estimates via HMC and HPD intervals for the parameters and MTTF obtained by fitting INMW to Aarset data.

Parameter	Point estimate	90% HPD interval	95% HPD interval
α	0.0118	$[0.0117, 0.0118]$	$[0.0117, 0.0119]$
β	0.0860	$[0.0335, 0.1340]$	$[0.0314, 0.1510]$
γ	0.4432	$[0.2488, 0.6335]$	$[0.2323, 0.6817]$
θ	92.6352	$[49.5442, 134.1876]$	$[44.9898, 148.7824]$
λ	0.0107	$[0.0036, 0.0180]$	$[0.0025, 0.0195]$
$MTTF$	44.8617	$[37.9911, 52.2731]$	$[36.9839, 53.6662]$

7.4.2 Meeker-Escobar data

Again the Meeker-Escobar data given in Chapter 6 is used for demonstrating the superior of the proposed model. As mentioned earlier, the data possesses a bathtub-shaped failure rate. For analyzing Meeker-Escobar data, the same procedure given in Subsection 7.4.1 is used for Bayesian estimation. However, for this data set, the informative prior is used since very little data points are shown in the data set. For the prior distributions given in Eqs. 7.20-7.23, the hyper-parameters are chosen such that the prior means approximate

Figure 7.3: The MLEs via CE and Bayes estimates via HMC of (a) reliability and (b) failure rate functions obtained by fitting INMW and NMW to Aarset data.
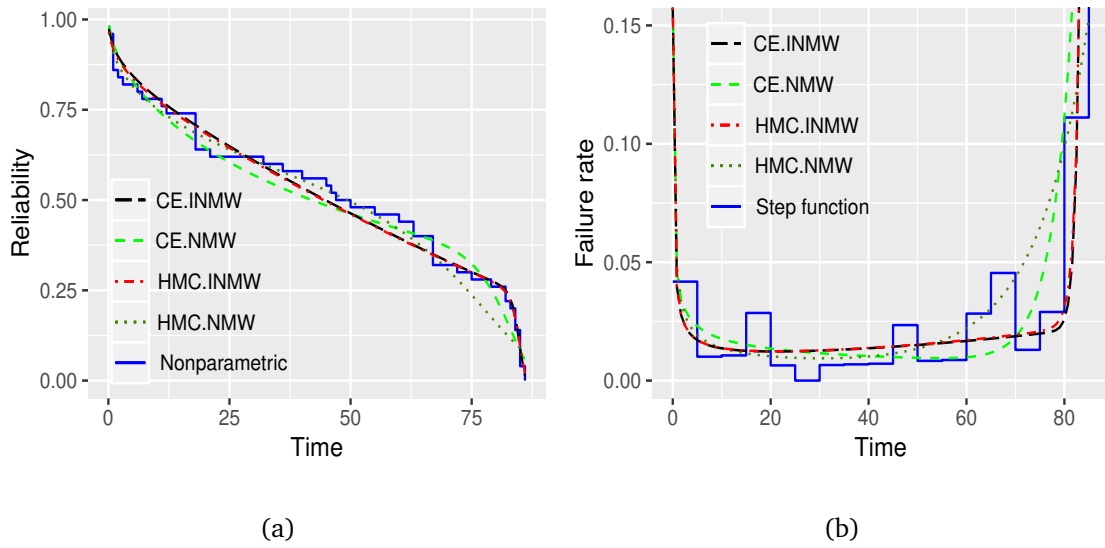
Table 7.2: Log-likelihood, K-S statistic, AIC, BIC and AICc obtained by fitting INMW, AMW and NMW to Aarset data.

Models	Log-lik	K-S (p-value)	AIC	BIC	AICc
INMW	-203.58	$0.067\,(0.977)$	417.16	426.72	418.52
AMW	-203.57	$0.068\,(0.963)$	417.14	426.70	418.50
NMW	-212.88	$0.075\,(0.921)$	435.76	445.32	437.12

Table 7.3: The MLEs of parameters obtained by fitting INMW, AMW and NMW to Aarset data.

Models	Failure rate function	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$
INMW	$(\alpha t)^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}$	0.0118	0.0771	0.4544	90.0578	0.0105
AMW	$\lambda e^{\lambda t - \beta} + \alpha(\theta + \gamma t)t^{\theta-1}e^{\gamma t}$	0.0763	90.1357	0.0104	0.4579	1.0604
NMW	$\alpha\theta t^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}$	0.0709	6.9952×10^{-8}	0.0168	0.6008	0.1976

Table 7.4: Bayesian p-value and DIC obtained by fitting INMW, NMW and AMW to Aarset data.

Model	Bayesian p-value		Deviance information criterion		
	$T_{\min}$	$T_{\max}$	DIC	$\overline{D}$	p_D
INMW	0.282	0.502	416.420	411.934	4.486
AMW	0.284	0.484	416.655	412.201	4.454
NMW	0.220	0.952	$--$	436.726	-130.017

the MLEs of the parameters. Since the MLEs of parameters for fitting INMW to Meeker-Escobar data are $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\lambda}) = (0.0033, 0.0198, 0.5942, 154.3077, 0.0025)$, we have chosen $(a_1 = 50, b_1 = 16666.67), (a_2 = 50, b_2 = 2500), (a_3 = 50, b_3 = 83.33), (a_4 = 50, b_4 = 0.32)$, and $(a_5 = 50, b_5 = 20000)$.
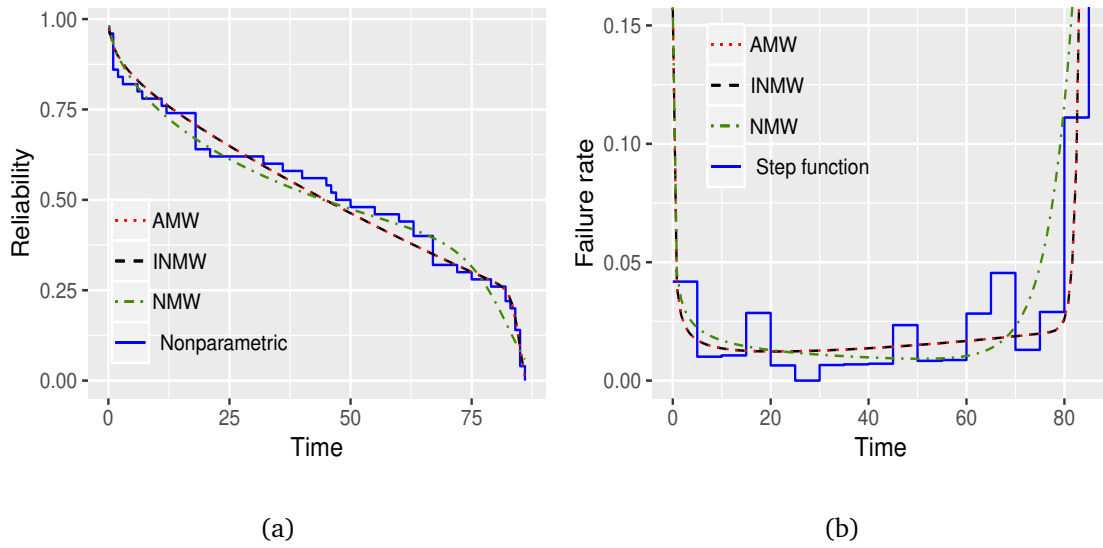
Figure 7.4: The MLEs of (a) reliability and (b) failure rate functions obtained by fitting INMW, AMW and NMW to Aarset data.
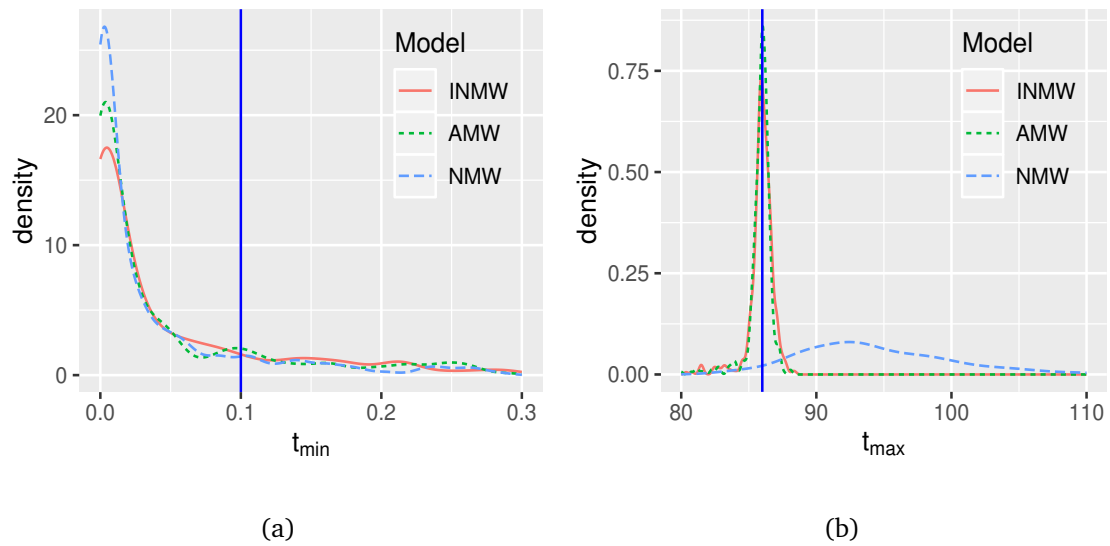

Figure 7.5: Density estimates of (a) smallest ordered future observations and (b) largest ordered future observations for INMW, NMW, and AMW, vertical lines represent corresponding observed values for Aarset data.

Fig. 7.6 shows the trace plots and density estimates of the parameters obtained by HMC algorithm. The trace plots show that the 4 parallel chains for each parameter produced by HMC algorithm also converge quickly to the same target distribution. However, it took more time for this small data set. The densities are distributed almost symmetrically around the central values which means that they provide good Bayes estimates under square error loss function, thanks to the informative priors. The scatter plot matrix of HMC output shows the estimated posterior correlations between the parameters (Fig. 7.7). Most pairs of parameters have a very small correlation whereas the pair (β, λ) appears to have higher posterior correlation.

Table 7.5 shows the HMC point estimates and two-sided 90% and 95% HPD (highest posterior density) intervals for $\alpha, \beta, \gamma, \theta, \lambda$ and MTTF. Fig. 7.8 displays the time courses of the estimated reliability and failure rate functions obtained by MLE and Bayesian methods

when fitting INMW and NMW to the data. It is easy to see that INMW fits to the data set much better than its original NMW.

Table 7.6 provides the measure of fit values resulting from fitting the INMW, NMW and AMW models to the data set. The corresponding MLEs of parameters are given in Table 7.7. From Table 7.6 we can observe that the INMW model fits to the data even better than the AMW model. These results can also be seen in Fig. 7.9.

Table 7.4 provides the Bayesian model checking when fitting INMW, NMW and AMW to the data. Fig. 7.5 displays the density estimates of the choosing test statistics, $T_{\min}$ and $T_{\max}$ based on replicated future data sets, where the observed test values are also displayed by means of vertical lines. From the results we see, that the INMW and AMW provide good fits to the data, whereas the NMW does not provide a good fit at the upper tail. The DIC values in Table 7.4 also show that the INMW model is better than the other models.
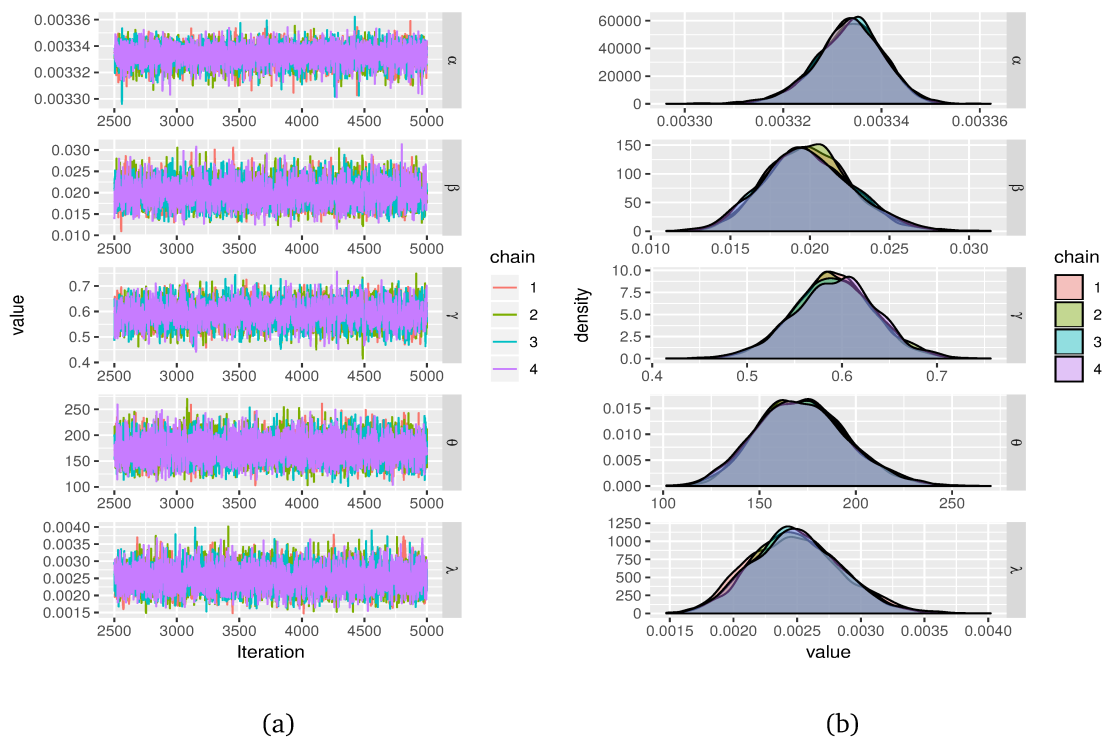
Figure 7.6: (a) Trace plots and (b) density estimates of the parameters produced by HMC sampling with 4 parallel chains obtained by fitting INMW to Meeker-Escobar data.

7.5 Conclusions

This chapter have provided a full Bayesian analysis of the INMW model, an improvement of the NMW model, using the Hamiltonian Monte Carlo simulation. The innovative model has provided a good flexibility for modeling two real data sets that exhibit bathtub-shaped failure rate. The study has demonstrated that the HMC algorithm works well with the proposed model, particularly in context with producing the estimates of parameters and reliability characteristics. The CE algorithm is also applied to optimize the log-likelihood function which results from the MLE method. The study demonstrated that the INMW model is much better than its original NMW model, and to some extent even better than the AMW model. The INMW model is generally preferable to the AMW model because the
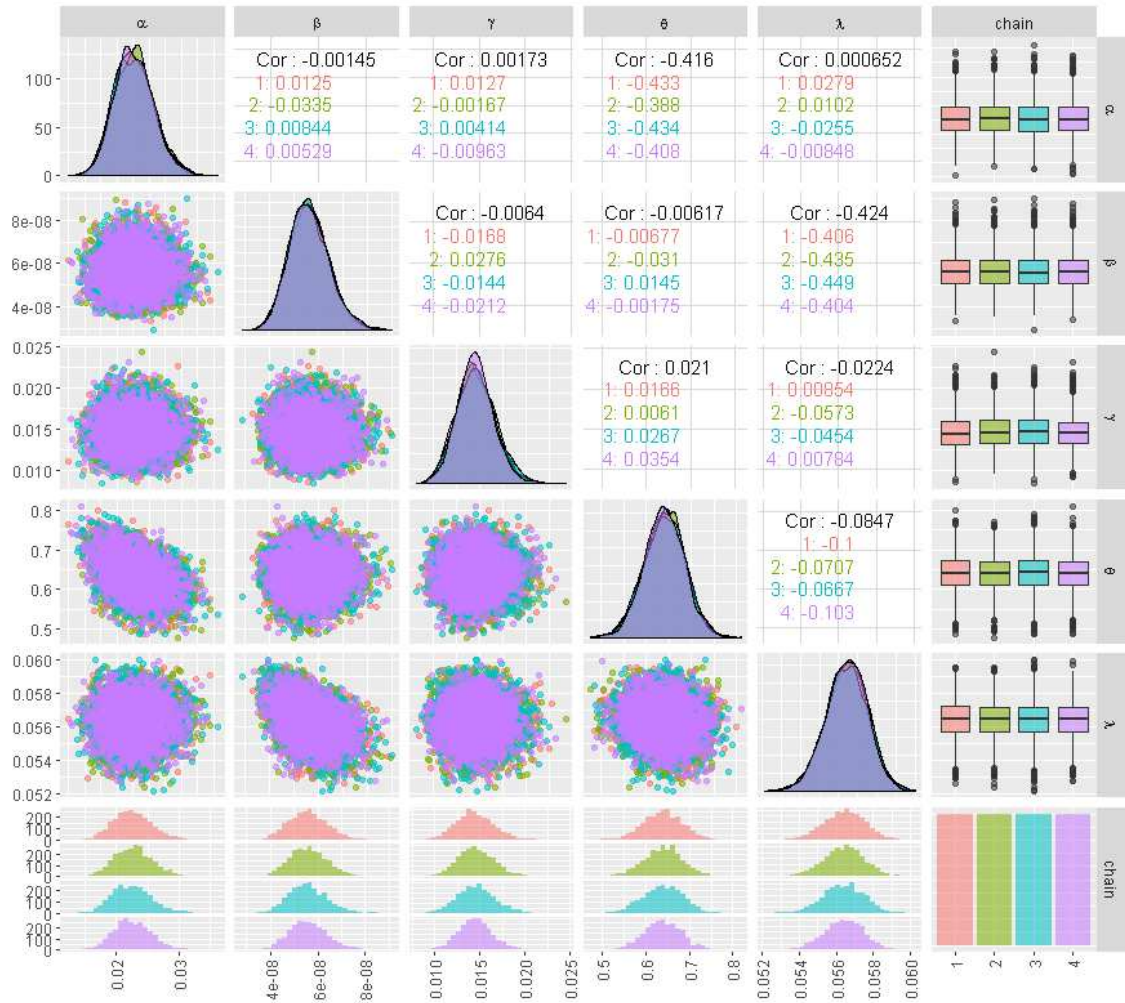
Figure 7.7: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting INMW to Meeker-Escobar data.

Table 7.5: Bayes estimates via HMC and HPD intervals for the parameters and MTTF obtained by fitting INMW to Meeker-Escobar data.

Parameter	Point estimate	90% HPD interval	95% HPD interval
α	0.0033	$[0.0033, 0.0033]$	$[0.0033, 0.0033]$
β	0.0199	$[0.0151, 0.0241]$	$[0.0147, 0.0252]$
γ	0.5915	$[0.5214, 0.6620]$	$[0.5008, 0.6718]$
θ	172.3540	$[134.6809, 211.2877]$	$[127.3370, 217.8504]$
λ	0.0025	$[0.0019, 0.0031]$	$[0.0018, 0.0032]$
$MTTF$	175.6658	$[149.9354, 200.0784]$	$[145.68, 205.0139]$

HMC sampling for the INMW model always converges quickly to the target distribution, whereas the HMC sampling for the AMW model might need many trials to obtain a good convergence. This will be demonstrated in future work.

Figure 7.8: The MLEs and Bayes estimates of (a) the reliability and (b) failure rate functions obtained by fitting INMW and NMW to Meeker-Escobar data.

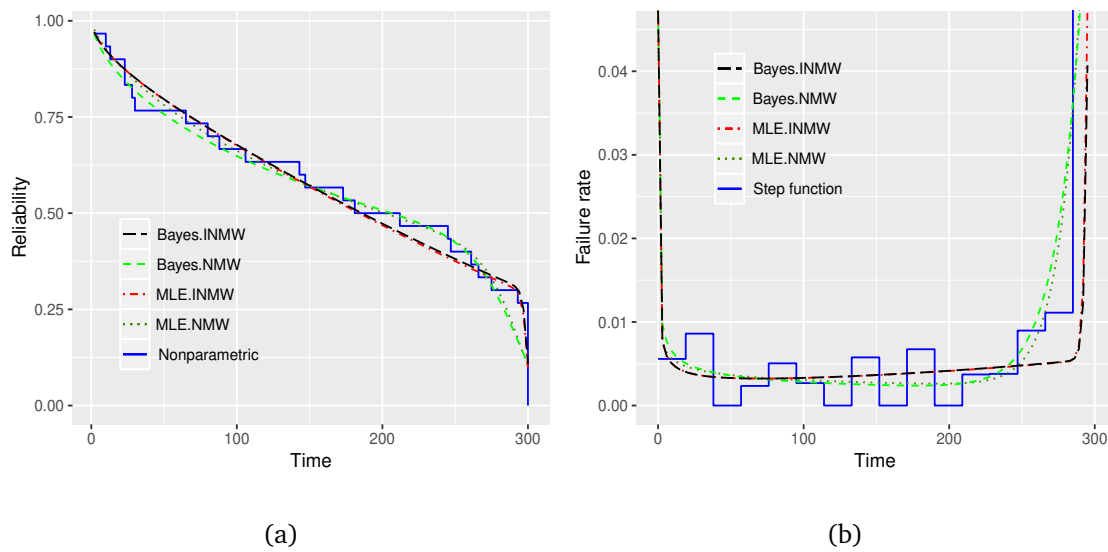**Table 7.6:** Log-likelihood, K-S statistic, AIC, and BIC obtained by fitting INMW, AMW and NMW to Meeker-Escobar data.

Models	Log-lik	K-S (p-value)	AIC	BIC	AICc
INMW	-153.90	$0.099(0.557)$	317.80	324.80	320.30
AMW	-155.58	$0.100(0.896)$	321.16	328.17	323.66
NMW	-165.02	$0.102(0.538)$	340.04	347.05	342.54

Table 7.7: The MLEs of parameters obtained by fitting INMW, AMW and NMW to Aarset data.

Models	Failure rate function	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$
INMW	$(\alpha t)^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}$	0.0033	0.0198	0.5942	154.3077	0.0025
AMW	$\lambda e^{\lambda t - \beta} + \alpha(\theta + \gamma t)t^{\theta-1}e^{\gamma t}$	0.0142	116.9665	0.0019	0.6788	0.3902
NMW	$\alpha\theta t^{\theta-1} + \beta(\gamma + \lambda t)t^{\gamma-1}e^{\lambda t}$	0.0137	6.1255×10^{-10}	0.1308	0.7391	0.0692

Table 7.8: Bayesian p-value and DIC obtained by fitting INMW, NMW and AMW to Meeker-Escobar data.

Model	Bayesian p-value		Deviance information criterion		
	$T_{\min}$	$T_{\max}$	DIC	D	p_D
INMW	0.226	0.922	310.472	308.490	1.982
AMW	0.316	0.936	314.499	312.505	1.993
NMW	0.158	0.972	335.791	333.999	1.793

(a) (b)

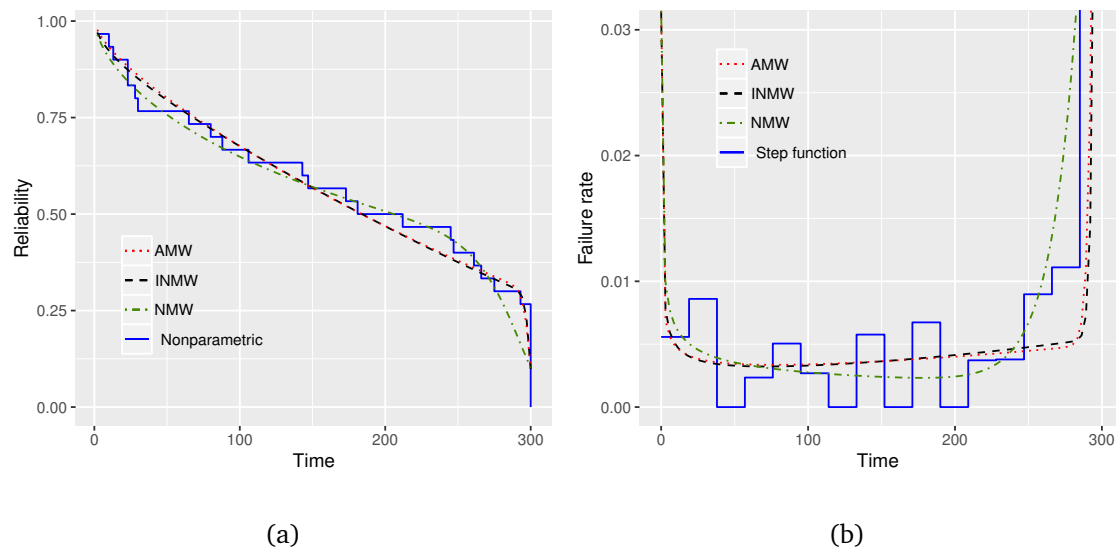**Figure 7.9:** The MLEs of (a) the reliability and (b) failure rate functions obtained by fitting INMW, AMW and NMW to Meeker-Escobar data.
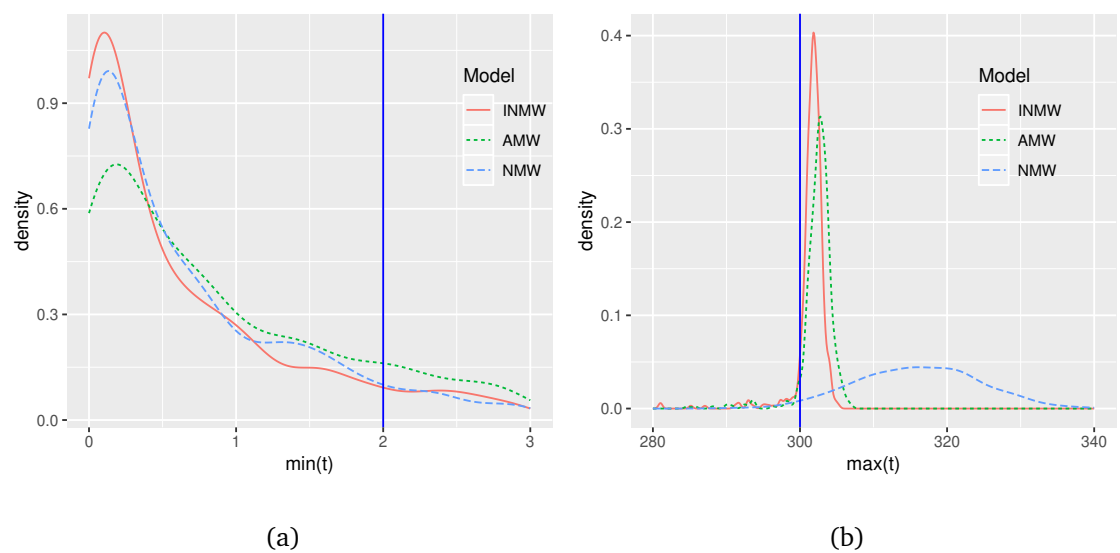

(a) (b)

Figure 7.10: Density estimates of (a) smallest ordered future observations and (b) largest ordered future observations for INMW, NMW, and AMW, vertical lines represent corresponding observed values for Meeker-Escobar data.

Chapter 8

Concluding remarks

The thesis have developed statistical models for modeling failure time data and studied the convenient statistical computation methods which allow practitioners and researchers to adopt and apply Bayesian analysis for analyzing such complicated failure time models with several parameters. The non-linear failure rate model, the additive Chen-Weibull model and the improvement of the new modified Weibull failure rate model have been developed and studied in detail. Many well-known lifetime data sets have been analyzed successfully by using the proposed models. The statistical methods, for example the cross-entropy method, Markov chain Monte Carlo methods, maximum likelihood, Bayesian inference, bootstrapping, have been applied successfully. Obviously, the main goal of the thesis involving individual sub-goals has been met. To be successful in this, indispensable programming codes (in R setting) were created. The successful analyses of the proposed models by using modern statistical methods will allow researchers to develop new model with many parameters. For future research, there are a few possibilities as follows:

- Consider a mixture failure rate with more than two components.

- Mixtures of other failure rate models, instead of Weibull and modified Weibull models, are also worth to consider and whether it would be possible to select the distribution type in advance, based on certain data features.

- Another possibility is to consider failure rates with change points or a simpler case, an incrementally constructed failure rate.

- In the thesis, all the mixture failure rate models are considered in case of independent competing risks. Thus the mixture failure rate models can also be considered in case of dependent competing risks.

Bibliography

[1] M. V. Aarset. "How to identify bathtub hazard rate". In: *IEEE Transactions on Reliability* 36.1 (1987), 106–8.

[2] B. Alder and T. Wainwright. "Studies in molecular dynamics I. General method". In: *Journal of Chemical Physics* 31.2 (1959), pp. 459–466.

[3] S. J. Almalki and J. Yuan. "A new modified Weibull distribution". In: *Reliability Engineering & System Safety* 111 (2013), 164–70.

[4] L. J. Bain. "Analysis for the linear failure-rate life-testing distribution". In: *Technometrics* 16 (1974), pp. 551–559.

[5] S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. 2nd edition. Chapman and Hall/CRC, 2014.

[6] A. P. Basu and N. Ebrahimi. "Bayesian approach to life testing and reliability estimation using asymmetric loss function". In: *Journal of Statistical Planning and Inference* 29 (1991), pp. 21–31.

[7] M. Bebbington, C. D. Lai, and R. Zitikis. "A flexible Weibull extension". In: *Reliability Engineering & System Safety* 92 (2007), 719–26.

[8] T. Benham et al. "CEoptim: Cross-Entropy R Package for Optimization". In: *Journal of Statistical Software* 76.8 (2017), pp. 1–29.

[9] W. R. Blischke, M. R. Karim, and D. P. Murthy. *Warranty data collection and analysis*. Springer, 2011.

[10] N. Bousquet, H. Bertholon, and G. Celeux. "An alternative competing risk model to the Weibull distribution for modeling aging in lifetime data analysis". In: *Lifetime Data Analysis* 12 (2006), pp. 481–504.

[11] S. Brooks et al. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[12] R. Canfield. "A Bayesian approach to reliability estimation using a loss function". In: *IEEE Transactions on Reliability* 19 (1970), pp. 13–16.

[13] A. Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-20. 2017.

[14] M. Carrasco, E. M. Ortega, and G. M. Cordeiro. "A generalized modified Weibull distribution for lifetime modeling". In: *Computational Statistics & Data Analysis* 53 (2008), 450–62.

[15] Z. Chen. "A new two parameter lifetime distribution with bathtub shaped or increasing failure rate function". In: *Statistics & Probability Letters* 49 (2000), pp. 155–61.

[16] D Collett. *Modelling survival data in medical research*. 2nd. Boca Raton: CRC press, 2015.

[17] T. Dimitrakopoulou, K. Adamidis, and S. Loukas. "A lifetime distribution with an upside-down bathtub-shaped hazard function". In: *IEEE Transactions on Reliability* 56.2 (2007), 308–11.

[18]   D. R. Dolas, M. D. Jaybhaye, and S. D. Deshmukh. "Estimation the System Reliability using Weibull Distribution". In: *International Proceedings of Economics Development and Research* 75 (2014), pp. 144–148.

[19]   S. Duane et al. "Hybrid Monte Carlo". In: *Physics Letters B* 195.2 (1987), pp. 216–222.

[20]   R. P. Feynman. "Mr. Feynman goes to Washington". In: *Engineering and Science, California Institute of Technology, Pasadena, CA* (1987), pp. 6–22.

[21]   A. Gelman. "Comment: Fuzzy and Bayesian p-values and u-values". In: *Statistical Science* 20 (2005), pp. 380–381.

[22]   A. Gelman et al. *Bayesian Data Analysis*. 3rd edition. Chapman and Hall/CRC, 2014.

[23]   R. Gentleman. *muhaz: A package for producing a smooth estimate of the hazard function for censored data*. R package version 1.2.6. 2015.

[24]   C. J. Geyer. *Stat 5102 Notes: Fisher Information and Confidence Intervals Using Maximum Likelihood*. 2007.

[25]   B Gompertz. "On the nature of the function expressive of the law of human mortality and on the new model of determining the value of life contingencies". In: *Philosophical Transactions of the Royal Society of London* 115 (1825), 513–85.

[26]   A. Gupta, B. Mukherjee, and S. K. Upadhyay. "Weibull extension model: A Bayes study using Markov chain Monte Carlo simulation". In: *Reliability Engineering & System Safety* 93.10 (2008), pp. 1434–1443.

[27]   M. S. Hamada et al. *Bayesian reliability*. Springer, 2008.

[28]   B. He, W. Cui, and X. Du. "An additive modified Weibull distribution". In: *Reliability Engineering & System Safety* 145 (2016), pp. 28–37.

[29]   U. Hjorth. "A reliability distribution with increasing, decreasing, and bathtub-shaped failure rates". In: *Technometrics* 22 (1980), pp. 99–107.

[30]   M. D. Hoffman and A. Gelman. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research* 15 (2014), 1593–1623.

[31]   R. V. Hogg, J. W. McKean, and A. T. Craig. *Introduction to mathematical statistics*. 8th edition. Pearson, 2019.

[32]   H. Jiang, M. Xie, and L. C. Tang. "Markov chain Monte Carlo methods for parameter estimation of the modified weibull distribution". In: *Journal of Applied Statistics* 35.6 (2008), pp. 647–658.

[33]   J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*. Wiley, New York, 2002.

[34]   D. Kodlin. "A new response time distribution". In: *Biometrics* 2 (1967), pp. 227–239.

[35]   D. P. Kroese, S. Porotsky, and R. Y. Rubinstein. "The cross-entropy method for continuous multi-extremal optimization". In: *Methodology and Computing in Applied Probability* 8 (2006), pp. 383–407.

[36]   D. Kundu and H. Howlader. "Bayesian inference and prediction of the inverse Weibull distribution for type-II censored data". In: *Computational Statistics and Data Analysis* 54.6 (2010), pp. 1547–1558.

[37]   C. D. Lai. *Generalized Weibull Distributions*. Springer, 2013.

[38]   C. D. Lai and M. Xie. *Stochastic ageing and dependence for reliability*. Springer, 2006.

[39] C. D. Lai, M. Xie, and D. N. P. Murthy. "A modified Weibull distribution". In: *IEEE Transactions on Reliability* 52 (2003), pp. 33–7.

[40] J. F. Lawless. *Statistical models and methods for lifetime data*. Wiley-Interscience, 2002.

[41] N. R. Mann, R. E. Schafer, and N. D. Singpurwalla. *Methods for Statistical Analysis of Reliability and Life Data*. John Wiley & Sons, 1974.

[42] G. J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. 2nd. Wiley, 2008.

[43] W. Q. Meeker and L. A. Escobar. *Statistical methods for reliability data*. Wiley, 1998.

[44] N. Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21 (1953), pp. 1087–1092.

[45] D. C. Montgomery and G. C. Runger. *Applied statistics and probability for engineers*. 7th ed. Wiley, 2019.

[46] D. F. Moore. *Applied survival analysis using R*. Springer, 2016.

[47] G. S. Mudholkar and D. K. Srivastava. "Exponentiated Weibull family for analysing bathtub failure rate data". In: *IEEE Transactions on Reliability* 42 (1993), 299–302.

[48] R. M. Neal. "Bayesian Learning for Neural Networks, Lecture Notes in Statistics". In: *Springer, New York* 118 (1996).

[49] R. M. Neal. *Handbook of Markov chain Monte Carlo*. Ed. by S. Brooks et al. CRC Press, 2011. Chap. 5: MCMC using Hamilton dynamics.

[50] A. Pandey, A. Singh, and W. J. Zimmer. "Bayes estimation of the linear hazard-rate model". In: *IEEE Transactions on Reliability* 42 (1993), pp. 636–640.

[51] H. Rinne. *The Weibull Distribution: A Handbook*. CRC Press, 2008.

[52] C. P. Robert and G. Casella. *Introducing Monte Carlo methods with R*. Springer-Verlag, New York, 2010.

[53] R. Y. Rubinstein and D. P. Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte Carlo simulation and machine learning*. Springer-Verlag, New York, 2004.

[54] S. A. Salem. "Bayesian estimation of a non-linear failure rate from censored samples type II". In: *Microelectronics and Reliability* 32 (1992), pp. 1385–1388.

[55] F. J. Samaniego. *A comparison of the Bayesian and Frequentist approaches to estimation*. Springer, 2010.

[56] A. Sarhan. "Bayes estimation of the general hazard rate model". In: *Reliability Engineering & System Safety* 66.1 (1999), pp. 85–91.

[57] A. M. Sarhan and J. Apaloo. "Exponentiated modified Weibull extension distribution". In: *Reliability Engineering & System Safety* 112 (2013), 137–44.

[58] A. M. Sarhan and M. Zaindin. "Modified Weibull distribution". In: *Applied Sciences* 11.1 (2009), 123–136.

[59] A. Scheidegger. *adaptMCMC: Implementation of a Generic Adaptive Monte Carlo Markov Chain Sampler*. R package version 1.3. 2018.

[60] M. K. Shakhatreh, A. J. Lemonteb, and G. Moreno–Arenas. "The log-normal modified Weibull distribution and its reliability implications". In: *Reliability Engineering & System Safety* 188 (2019), pp. 6–22.

[61]   G. O. Silva, E. M. Ortega, and G. M. Cordeiro. "The beta modified Weibull distribution". In: *Lifetime Data Analysis* 16 (2010), 409–30.

[62]   P. K. Singh, S. K. Singh, and U. Singh. "Bayesian estimator of inverse Gaussian parameters under general entropy loss function using Lindley's approximation". In: *Communication in Statistics-Simulation and Computation* 37.9 (2010), pp. 1750–1762.

[63]   P. J. Smith. *Analysis of Failure and Survival Data*. Chapman and Hall/CRC, 2002.

[64]   A. A. Soliman et al. "Modified Weibull model: A Bayes study using MCMC approach based on progressive censoring data". In: *Reliability Engineering & System Safety* 100 (2012), pp. 48–57.

[65]   D. J. Spiegelhalter et al. "The deviance information criterion: 12 years on." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.3 (2014), 485–493.

[66]   Stan Development Team. *RStan: the R interface to Stan*. R package version 2.17.3. 2018. URL: http://mc-stan.org/rstan/.

[67]   Stan Development Team. *Stan Modeling Language User's Guide and Reference Manual*. 2018. URL: http://mc-stan.org/.

[68]   M. D. Ugarte, A. F. Militino, and A. T. Arnholt. *Probability and Statistics with R*. 2th ed. CRC press, 2015.

[69]   S.K. Upadhyay and A. Gupta. "A Bayesian analysis of modified Weibull distribution using Markov Chain Monte Carlo simulation". In: *Journal of Statistical Computation and Simulation* 80.3 (2010), 241–254.

[70]   H. R. Varian. "A bayesian approach to real estate assessment. In: L. J. Savage, S. E. Feinberg and A. Zellner, Eds., Studies in Bayesian Econometrics and Statistics: In Honor of L. J. Savage". In: *North-Holland Pub. Co., Amsterdam* (1975), pp. 195–208.

[71]   M. Vihola. "Robust adaptive metropolis algorithm with coerced acceptance rate". In: *Statistics and Computing* 22.5 (2012), pp. 997–1008.

[72]   F. K. Wang. "A new model with bathtub-shaped failure rate using an additive Burr XII distribution". In: *Reliability Engineering & System Safety* 70.3 (2000), pp. 305–312.

[73]   W. A. Weibull. "Statistical distribution function of wide applicability". In: *Journal of Applied Mechanics* 18 (1951), 293–96.

[74]   M. A. van de Wiel, D. E. T. Beest, and M. M. Münch. "Learning from a lot: Empirical Bayes for high-dimensional model-based prediction". In: *Scandinavian Journal of Statistics* 100 (2018), pp. 1–24.

[75]   M. Xie and C. D. Lai. "Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function". In: *Reliability Engineering & System Safety* 52 (1995), 87–93.

[76]   M. Xie, Y. Tang, and T. N. Goh. "A modified Weibull extension with bathtub-shaped failure rate function". In: *Reliability Engineering & System Safety* 76 (2002), 279–85.

[77]   H. Zeng, T. Lan, and Q. Chen. "Five and four-parameter lifetime distributions for bathtub-shaped failure rate using perks mortality equation". In: *Reliability Engineering & System Safety* 152 (2016), 307–15.

[78]   E. Zio. *The Monte Carlo simulation method for system reliability and risk analysis*. Springer, 2013.

Author's publications

[79] R. Bris and T. T. Thach. "Bayesian approach to estimate the mixture of failure rate model". *(2016) Applied Mathematics in Engineering and Reliability - Proceedings of the 1st International Conference on Applied Mathematics in Engineering and Reliability, ICAMER 2016*, pp. 9-18. ISBN: 978-113802928-6. (SCOPUS, major contribution)

[80] T. T. Thach and R. Bris (2016). "Bayes estimators of the mixture of failure rate model". In: *Proceedings of the 14th annual workshop WOFEX*, p. 402-407. ISBN: 978-80-248-3961-5.

[81] T. T. Thach and R. Bris (2017). "Mixture failure rate: A study based on product of spacings". In: *Proceedings of the 15th annual workshop WOFEX*, p. 371-376. ISBN:978-80-248-4056-7.

[82] T. T. Thach, R. Bris and F. P. A. Coolen (2017). "Mixture failure rate: A study based on cross-entropy and mcmc method". In: *2017 International Conference on Information and Digital Technologies (IDT)*, Zilina, 2017, pp. 373-382. Doi: 10.1109/DT.2017.8024325, Electronic ISBN: 978-1-5090-5689-7, ©2017 IEEE. (SCOPUS, major contribution)

[83] T. T. Thach and R. Bris: MLE versus MCMC estimators of the mixture of failure rate model, Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds), © 2018 Taylor & Francis Group, London, pg.937-944, ISBN: 978-0-8153-8682-7. (SCOPUS, major contribution)

[84] T. T. Thach and R. Bris: Non-linear failure rate: A comparison of the Bayesian and frequentist approaches to estimation, *Proceedings of International Conference on Mathematics ICM 2018*, ITM Web of Conferences 20, paper 03001 (2018) `https://doi.org/10.1051/itmconf/20182003001`, Open Access, ISBN: 978-7598-9058-3, 16 pg. Edited by: Radim Bris, Gyu Whan Chang, Chu Duc Khanh, Mohsen Razzaghi, Krzysztof Stempak, Phan Thanh Toan. (SCOPUS, major contribution)

[85] T. T. Thach and R. Bris, P. Volf and F. P. A. Coolen (2019). "Non-linear failure rate: a study using Markov chain Monte Carlo and crossentropy methods". Submitted to the Journal of Approximate Reasoning. (IF 1.982(Q2), Manuscript ID: IJA_2019_61, Under review)

[86] T. T. Thach and R. Bris (2019). "Reparameterized Weibull distribution: A Bayes study using Hamiltonian Monte Carlo". Submitted to ESREL 2019: 29th European Safety and Reliability Conference, Hannover, Germany. (accepted)

[87] T. T. Thach and R. Bris (2019). "New modified Weibull model: a Bayes study using Hamiltonian Monte Carlo". Submitted to the Journal of Risk and Reliability. (IF 1.313(Q3), Manuscript ID: JRR-19-0103, Under review)

[88] T. T. Thach and R. Bris: Hamiltonian Monte Carlo method for parameter estimation of the additive Weibull distribution, Proceedings of the International Conference on

Information and Digital Technologies 2019, June 25-27, pg. 503-509, Zilina, Slovakia, DOI: 10.1109/DT.2019.8813441, Electronic ISBN: 978-1-7281-1401-9, Electronic ISSN: 2575-677X, ©2019 IEEE.

[89] T. T. Thach and R. Bris: An additive Chen-Weibull distribution and its applications in lifetime data analysis. Submitted to the Lifetime Data Analysis. (IF 0.948(Q3), Manuscript ID: LIDA-D-19-00096, Under review)