2016-01-01

Bayesian Parameter Estimation for the Birnbaum-Saunders Distribution and its Extension

Tun Lee Ng
University of Texas at El Paso, tng@miners.utep.edu

BAYESIAN PARAMETER ESTIMATION FOR THE

BIRNBAUM-SAUNDERS DISTRIBUTION

AND ITS EXTENSION

TUN LEE NG

Master's Program in Mathematical Sciences

APPROVED:

Naijun Sha, Ph.D., Chair

Jonghyun Yun, Ph.D.

Yirong Lin, Ph.D.

Charles Ambler, Ph.D.
Dean of the Graduate School

to my

MOTHER and FATHER

with love

BAYESIAN PARAMETER ESTIMATION FOR THE BIRNBAUM-SAUNDERS DISTRIBUTION AND ITS EXTENSION

by

TUN LEE NG

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Master's Program in Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

May 2016

Acknowledgements

I would like to express my deep-felt gratitude to my advisor, Dr. Naijun Sha of Department of Mathematical Sciences at The University of Texas at El Paso, for his advice, encouragement, enduring patience and constant support.

I also wish to thank the other members of my committee, Dr. Jonghyun Yun of Department of Mathematical Sciences and Dr. Yirong Lin of Department of Mechanical Engineering, both at The University of Texas at El Paso. Their suggestions, comments and additional guidance were invaluable to the completion of this work.

NOTE: This thesis was submitted to my Supervising Committee on the May 16, 2016.

Abstract

We utilize the Bayesian approach to estimate the parameters of the Birnbaum-Saunders (BS) distribution devised by Birnbaum and Saunders (1969a), as well as the Generalized Birnbaum-Saunders (GBS) distribution obtained by Owen (2006), in the presence of random right censored data. We also derive the classical MLE expressions for the observed Information matrix of the GBS distribution, in order to illustrate the fact that no closed form expressions are available for the MLE, and numerical approximations are required to obtain the point estimates and asymptotic confidence intervals. Where Bayesian approach is concerned, new sets of priors are considered based on the model assumptions adopted by Birnbaum and Saunders (1969a) and Owen (2006). To handle the presence of random right censored observations, we utilize the data augmentation technique introduced by Tanner and Wong (1987), to circumvent the arduous expressions involving the censored data in obtaining posterior inferences. Simulation studies were carried out to assess performance of these methods under different parameter values, with small and large sample sizes, as well as various degrees of censoring. Two illustrative examples and some concluding remarks were finally presented.

Table of Contents

Page

Acknowledgements . . . v

Abstract . . . vi

Table of Contents . . . vii

List of Tables . . . ix

List of Figures . . . x

Chapter

1 Introduction . . . 1

1.1 The BS Distribution . . . 1

1.2 The Generalized BS (GBS) Distribution . . . 3

2 Literature Review . . . 8

2.1 Frequentist/Classical Parameter Estimation . . . 9

2.2 Bayesian Parameter Estimation . . . 10

2.3 Motivation for our Research . . . 10

3 Methodology . . . 12

3.1 The BS Distribution . . . 13

3.1.1 Likelihood Principle . . . 13

3.1.2 Augmented Likelihood . . . 13

3.1.3 Prior Specification . . . 15

3.1.4 Posterior Inference . . . 17

3.1.5 Sampling Algorithm . . . 18

3.2 The GBS Distribution . . . 23

3.2.1 Likelihood Principle . . . 24

3.2.2 Augmented Likelihood . . . 24

3.2.3 Prior Specification . . . 25

3.2.4 Posterior Inference 26
3.2.5 Sampling Algorithm 27
4 Simulation Study 31
4.1 The BS Distribution 31
4.2 The Generalized BS Distribution 42
5 Real Data Analyses 56
5.1 Data Set 1 56
5.2 Data Set 2 60
6 Concluding Remarks 69
References 71
Appendix
Appendix 76
Curriculum Vitae 81

List of Tables

4.1 Simulation Results for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$) 33

4.2 Simulation Results for Parameter Setting 2 ($\alpha = 1$, $\beta = 1$) 36

4.3 Simulation Results for Parameter Setting 3 ($\alpha = 2$, $\beta = 1$) 39

4.4 Simulation Results for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$) . . . 44

4.5 Simulation Results for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$) 48

4.6 Simulation Results for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$) 52

5.1 Lifetimes (in months) of 20 cancer patients receiving a new treatment . . . 56

5.2 Point and Interval Estimates for Data Set 1 58

5.3 Fatigue life of 6061-T6 aluminium coupons exerted with maximum stress per cycle of 21,000 psi 60

5.4 Point and Interval Estimates for Data Set 2 63

List of Figures

1.1 Hazard function of the Extended BS distribution under different values of α, β and κ 5

4.1 Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$) 34

4.2 Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$) 35

4.3 Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 2 ($\alpha = 1$, $\beta = 1$) 37

4.4 Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 2 ($\alpha = 1$, $\beta = 1$) 38

4.5 Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 3 ($\alpha = 2$, $\beta = 1$) 40

4.6 Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 3 ($\alpha = 2$, $\beta = 1$) 41

4.7 Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$) 45

4.8 Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$) 46

4.9 Average Estimates, SRMSE and Average Length of 95% CI of κ for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$) 47

4.10 Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$) 49

4.11 Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$) 50

4.12 Average Estimates, SRMSE and Average Length of 95% CI of κ for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$) 51
4.13 Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$) 53
4.14 Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$) 54
4.15 Average Estimates, SRMSE and Average Length of 95% CI of κ for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$) 55

5.1 Histograms of the simulated samples for the parameters and the reliability plots 59
5.2 Histograms of the simulated samples for the parameters and the reliability plots for $n = 101$ 64
5.3 Histograms of the simulated samples for the parameters and the reliability plots for $n = 90$ 65
5.4 Histograms of the simulated samples for the parameters and the reliability plots for $n = 80$ 66
5.5 Histograms of the simulated samples for the parameters and the reliability plots for $n = 70$ 67
5.6 Histograms of the simulated samples for the parameters and the reliability plots for $n = 60$ 68

Chapter 1

Introduction

Birnbaum and Saunders (1969a) developed a family of two-parameter distributions to model failure time due to fatigue under cyclic loading of "stress". This distribution is known as the Birnbaum-Saunders (henceforth abbreviated as BS) distribution. It is worth noting that the term "stress" is not only restricted to mechanical stress, but also encompasses other engineering response processes such as temperature, voltage and many others. The BS distribution, and its generalizations, have seen vast areas of practical applications, including engineering, business, environmental, medical and many other applications (Fierro et al., 2012). In order to facilitate our subsequent discussion, we will first provide an overview on the derivation of this distribution by Birnbaum and Saunders (1969a).

1.1 The BS Distribution

Consider a specimen (be it a product, a machine or even a biological subject) that is exerted with cyclical stress. The sources and types of stress may vary depending on the types of experiments or processes. Fatigue or crack builds up in the specimen gradually, and the specimen "breaks down" entirely once the internal crack breaches the breakdown threshold of the specimen.

Let W_j be the crack extension during the jth unit time interval. Assume that these crack extensions have contant mean μ and constant variance σ^2, and they are independent of one another. Then, by Central Limit Theorem (CLT), the total crack after time

$T = \sum_{j=1}^{T} W_j$ is approximately distributed by $\mathcal{N}(T\mu, T\sigma^2)$.

Denote the specimen's breakdown threshold with ω. Then, the probability of the specimen reaching its lifetime ($T \leq t$) is equivalent to the probability of total crack exceeding the specimen's breakdown threshold:

$$\begin{aligned} \mathrm{P}(T \leq t) &= \mathrm{P}\left[\sum_{j=1}^{t} W_j \geq \omega\right] \\ &\approx \Phi\left[\frac{\mu\sqrt{t}}{\sigma} - \frac{\omega}{\sigma\sqrt{t}}\right] \\ &= \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right], \end{aligned} \tag{1.1}$$

where $\Phi\{.\}$ is the cdf (cumulative distribution function) of the standard normal distribution, $\alpha = \frac{\sigma}{\sqrt{\omega\mu}} > 0$, and $\beta = \frac{\omega}{\mu} > 0$. Equation (1.1) is the cdf of the BS distribution with parameters α and β, with its corresponding pdf (probability density function) $f_T(t)$ given by

$$f_T(t) = \frac{t+\beta}{2\sqrt{2\pi}t^{3/2}\alpha\sqrt{\beta}} \exp\left\{-\frac{1}{2\alpha^2}\left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right\}. \tag{1.2}$$

From its cdf given in (1.1), it could be easily shown that the relationship between the BS distribution $T \sim BS(\alpha, \beta)$ and the standard normal distribution Z is given by the following identity:

$$Z = \frac{1}{\alpha}\left\{\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{t}}\right\} \sim \mathcal{N}(0, 1). \tag{1.3}$$

In addition, α is a shape parameter and β is a scale parameter. In fact, β is also the median of the BS distribution since $F_T(0) = \Phi\{0\} = 0.5$.

The identity given in Equation (1.3) also comes in handy for the purpose of random number generation and derivation of integer moments. The mean $E(T)$ and variance $Var(T)$

of the BS distribution T are given by

$$E(T) = \beta\left(1+\frac{\alpha^2}{2}\right) \tag{1.4}$$

$$Var(T) = (\alpha\beta)^2\left(1+\frac{5\alpha^2}{4}\right). \tag{1.5}$$

Another interesting fact is that, if $T \sim BS(\alpha,\beta)$, then $T^{-1} \sim BS(\alpha,\beta^{-1})$. Therefore, the mean $E(T^{-1})$ and variance $Var(T^{-1})$ of the BS distribution T^{-1} are given by

$$E(T^{-1}) = \beta^{-1}\left(1+\frac{\alpha^2}{2}\right) \tag{1.6}$$

$$Var(T^{-1}) = \left(\frac{\alpha}{\beta}\right)^2\left(1+\frac{5\alpha^2}{4}\right). \tag{1.7}$$

1.2 The Generalized BS (GBS) Distribution

Owen (2006) extended the BS distribution by relaxing the independence assumption among the crack extensions. He argued that the crack extension at the current time interval depends on the existing cracks in the specimen, which were built up from previous crack extensions. It is probable that the specimen wears out faster once there are existing internal cracks that constitute the "weak point" in the body of the specimen. This may lead to the phenomenon where crack extension tends to be smaller at the beginning of the experiment and gradually becomes larger as the specimen is getting close to failure. In other words, the old cracks may be influencing the formation of new crack extension. This validates the idea of modeling the sequence of crack extensions $W_1,\ldots,W_T$ as a stochastic process with non-zero correlation

$$\rho(i,j) = \frac{E[(W_i-\mu)(W_j-\mu)]}{\sigma^2} \neq 0 \text{ for } i \neq j.$$

By utilizing the statistical result proven by Beran (1994), if the sequence of crack extensions is a stationary and self-similar stochastic process (meaning that the process is invariant in distribution under a scaling of time), then we obtain the variance of the sample mean of crack extensions as follows:

$$Var(\bar{W}) = \sigma^2 T^{2\kappa - 2}, \tag{1.8}$$

where $0 < \kappa < 1$. Here, the additional parameter κ is the self-similarity parameter introduced by Beran (1994), and can be viewed as the rate of decay of $Var(\bar{W})$. When $\kappa = 0.5$, this corresponds to the classical CLT result of $Var(\bar{W}) = \sigma^2/T$ under the independence assumption which gives us the original BS distribution. When $\kappa > 0.5$, this corresponds to a long memory process whereby the rate of decay of variance is slower. When $\kappa < 0.5$, the rate of decay of variance is faster and leads to short memory process.

We now explain the role of κ from a practical point of view. If the crack extensions are a short memory process (i.e. $\kappa < 0.5$), the formation of new crack extension is more influenced by recent cracks. As experiment continues, crack extensions tend to get larger, and the formation of new crack extension is more influenced by more recent and larger cracks, rather than being influenced by older, smaller cracks. This correlation with large cracks may cause the new crack extension to be large as well, and in turn speed up the instanteneous failure rate or hazard rate of the specimen as experiment carries on.

On the other hand, if the crack extensions are a long memory process (i.e. $\kappa > 0.5$), the formation of new crack extension is also heavily influenced by older cracks. Even though the size of crack extensions may still increase with time, the formation of new crack extension is also heavily correlated with older, smaller cracks. The new crack extension may not be as large or as fast as its counterpart in a short memory process. Figure 1.1 adequately captures the fact that, as experiment time increases, a short memory process ($\kappa < 0.5$) always has a higher hazard rate, followed by the independent process ($\kappa = 0.5$) and finally the long memory process ($\kappa > 0.5$), under different values of α and β.

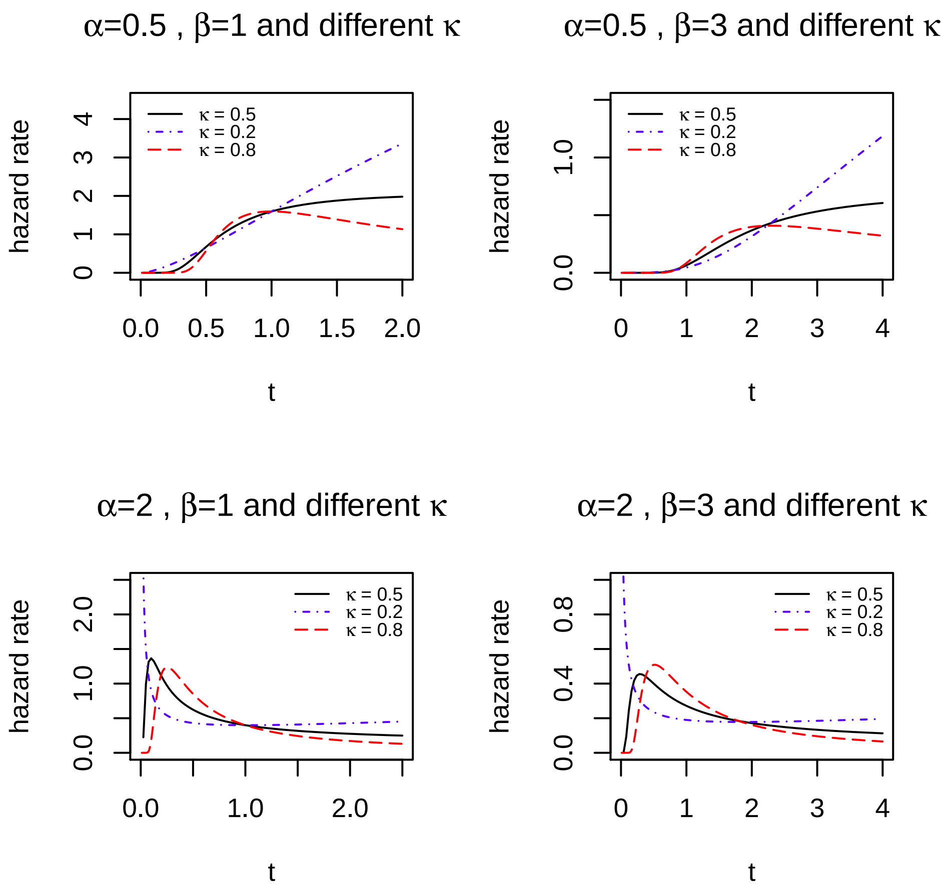

Figure 1.1: Hazard function of the Extended BS distribution under different values of α, β and κ

Due to CLT, $\bar{W}$ is still approximately distributed by $\mathcal{N}(\mu, \sigma^2 T^{2\kappa-2})$. By modeling these crack extensions as a stationary and self-similar stochastic process, Owen (2006) derived (and named) the *generalized three-parameter* BS distribution as follows:

$$\begin{aligned} P(T \le t) &= P\left[\sum_{j=1}^{T} W_j \ge \omega\right] \\ &\approx \Phi\left[\frac{\mu T^{1-\kappa}}{\sigma} - \frac{\omega}{\sigma T^{\kappa}}\right] \\ &= \Phi\left[\frac{1}{\alpha}\left(\frac{T^{1-\kappa}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{T^{\kappa}}\right)\right], \end{aligned}$$

where, again, $\alpha = \dfrac{\sigma}{\sqrt{\omega\mu}} > 0$, and $\beta = \dfrac{\omega}{\mu} > 0$. Therefore, when $T \sim GBS(\alpha, \beta, \kappa)$, its cdf $F_T(t)$ is given by

$$F_T(t) = \Phi\left\{\frac{1}{\alpha}\left(\frac{t^{1-\kappa}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{t^{\kappa}}\right)\right\}, \tag{1.9}$$

with its corresponding pdf $f_T(t)$ given by:

$$f_T(t) = \frac{1-\kappa+\frac{\beta\kappa}{t}}{\sqrt{2\pi}\alpha\sqrt{\beta}t^{\kappa}} \exp\left\{-\frac{(t-\beta)^2}{2\alpha^2\beta t^{2\kappa}}\right\}. \tag{1.10}$$

Obviously, $GBS(\alpha, \beta, \frac{1}{2}) = BS(\alpha, \beta)$. From the cdf given by (1.9), it is clear that the relationship between the GBS distribution $T \sim GBS(\alpha, \beta, \kappa)$ and the standard normal distribution Z is given by the following identity::

$$Z = \frac{1}{\alpha}\left\{\frac{T^{1-\kappa}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{T^{\kappa}}\right\} \sim \mathcal{N}(0, 1). \tag{1.11}$$

Furthermore, β is still the median of the GBS distribution but it is no longer a scale parameter. There are also no closed form expressions for the moments of T. In addition, the reciprocal property is preserved as follows: if $T \sim GBS(\alpha, \beta, \kappa)$, then $T^{-1} \sim GBS(\alpha, \beta^{-1}, 1-\kappa)$. In other words, if T is a long memory process, then T^{-1} is the short memory analog.

This thesis unfolds as follows: We will first provide an overview about the existing research done on the BS distribution in Section 2, followed by a detailed explanation on our methodology to estimate parameters for the BS and GBS distibutions in Section 3. Subsequently, we will carry out simulation studies for the BS and GBS distributions in Section 4. For illustrative purpose, real data sets will be analyzed in Section 5, followed by some concluding remarks in Section 6.

Chapter 2

Literature Review

Many authors have contributed extensively to the development of the BS distribution. For instance, Desmond (1985) provided a more general and robust proof for the BS distribution in a biological random environment context. Subsequently, Desmond (1986) proved that the BS distribution is in fact a mixture of the inverse normal distribution and its reciprocal with a mixing probability of $\frac{1}{2}$. Later, Rieck (1999) derived the moment generating function of the sinh-normal distribution in order to obtain both integer and non-integer moments of the BS distribution. Kundu et al. (2008) established that the hazard function of the BS distribution is always concave for all values of its shape parameter α, and developed several estimators for the maximum point of the hazard function.

Different generalizations and applications of the BS distributions have been introduced by various authors over the years. The GBS distribution constructed by Owen (2006), as discussed in previous chapter, is a form of generalization obtained by relaxing the independence assumption of the crack extensions. Other generalizations are constructed by introducing different kernels or additional parameters, or by using extreme value and non-centrality arguments. See, for example, Diaz-Garcia and Leiva-Sanchez (2005), Leiva et al. (2008), Sanhueza et al. (2008), Gomez et al. (2009), Fierro et al. (2012), and many others. In the context of accelerated life testing, Rieck and Nedelman (1991) constructed a log-linear model for the BS distribution. Owen and Padgett (2000) developed the BS inverse power law accelerated life model.

In accordance with our research focus, we dedicate the next two subsections below for a more in-depth overview of past researches on parameter estimation for the BS distribution via the frequentist and bayesian approaches respectively.

2.1 Frequentist/Classical Parameter Estimation

In the context of fully-observed data, the maximum likelihood estimators (MLEs) were originally obtained by Birnbaum and Saunders (1969b) with their asymptotic distributions derived by Engelhardt et al. (1981). Dupuis and Mills (1998) utilized the OBRE (optimal bias-robust estimator) to estimate the parameters and quantiles of the BS distribution. Later, Ng et al. (2003) proposed modified moment estimators (MMEs) with two methods of bias correction, one via inspection of bias pattern based on extensive Monte-Carlo simulation studies, and another by the Jackknife method. Subsequently, Wu and Wong (2004) improved interval estimation of the parameters using the signed log-likelihood ratio method introduced by Barndorff-Nielsen (1991). Lemonte et al. (2007) derived modified MLEs (which are bias-free to second order), compared several interval estimation methods and obtained a Bartlett correction which improves finite-sample performance of the likelihood ratio test (LRT). The expression for the Fisher information matrix provided by Lemonte et al. (2007) only involves numerical evaluation of the standard normal distribution function, which is an improvement from the expression given by Engelhardt et al. (1981) that requires numerical integration. Recently, Balakrishnan and Zhu (2014) obtained another form of estimator (with three methods of interval estimations, namely Jacknifing, Bootstrap and Asymptotic) which has the same mean squared error (MSE) but smaller bias compared to the MLE and MME proposed by Ng et al. (2003).

Where Type-II censored samples are present, Ng et al. (2006) discussed MLE point and interval estimation and proposed a simple bias correction method which originated from Hirose (1999). Meanwhile, Steven and Li (2006) introduced four families of estimators based on fully observed data, and two of them can be extended to the case of right-censored samples. Under random censoring, Wang et al. (2006) constructed the modified censored moment estimators together with their asymtotic distributions and bias-reduced versions based on counting process theory and martingale techniques.

2.2 Bayesian Parameter Estimation

Bayesian parameter estimation for the BS distribution is a relatively novel approach in the literature. In the context of fully observed data, Achcar (1993) first provided the expressions for marginal posterior distribution of the parameters based on the Jeffrey's prior and the reference prior, which was first introduced by Bernardo (1979). More specifically, Achcar (1993) used Laplace method for integral approximation in his derivations. Xu and Tang (2010) obtained posterior estimation using Lindley's approximation (Lindley, 1980) as well as Gibbs sampling based on the reference priors.

However, Xu and Tang (2011) proved that those reference priors lead to improper posterior distributions, and improvised their Bayesian inference using the reference priors with partial information, which was first introduced by Sun and Berger (1998). To simplify sampling issue in the presence of Type-II right censored data, they adopted a slice sampler by introducing an independent auxiliary variable. In addition, Achcar and Moala (2010) obtained a Bayesian inference based on independent inverse gamma priors for the case of censored data and covariates using WinBugs software. Recently, Wang et al. (2015) critiqued the slice sampler used by Xu and Tang (2011) and proposed a sampling algorithm in the context of fully observed data by using the generalized ratio of uniforms method.

2.3 Motivation for our Research

Owen (2006) provides a compelling and practical argument to relax the independence assumption among crack extensions. However, thus far, no other research was done to further investigate parameter estimation for this GBS distribution, other than the classical Maximum Likelihood (ML) estimation obtained by Owen (2006) for the case of fully-observed data. This motivates us to extend the MLE results to the case where right-censored data is involved. On top of that, our primary research focus in this paper would be the use of Bayesian methods to estimate the parameters due to the following reasons. Firstly, the Bayesian approach allows researchers' prior knowledge to be incorporated in obtaining inferences. Secondly, no closed form expressions are available for the MLE and numerical optimization to obtain the MLE may be unstable due to the data-dependent nature of the MLE. The arduous MLE expressions with respect to censored data could also be circumvented with the use of the Bayesian methods (which will be explained in detail in subsequent chapters), thus rendering the Bayesian approach to be relatively more appealing. In addition, large sample size is needed to construct asymptotic confidence interval for the estimators, whereas using the Bayesian methods, credible intervals for the parameters can still be obtained even with small sample size.

Under our Bayesian approach, we will consider a new set of priors as well as the data augmentation technique proposed by Tanner and Wong (1987) to handle the presence of right-censored data in estimating the parameters for the BS and GBS distributions.

Chapter 3

Methodology

We will split our discussion about parameter estimation into two main sections, the first for the BS distribution, and the second for the GBS distribution.

Suppose we have n observed samples and m censored samples from an experiment. In accordance to statistical convention, denote the random variables representing the n observed data by $T_1, \ldots, T_n$, and the random variables representing the m right censored data by $T_{n+1}, \ldots, T_{n+m}$. In addition, the values of the observed samples are denoted by $t_{n+1}, \ldots, t_{n+m}$, whereas the values of the censored samples are denoted by $c_1, \ldots, c_m$ respectively, with $T_{n+j} > c_j$ for $j = 1, \ldots, m$. Note that the subscripts here do not refer to order statistics, but only serve the purpose of simplifying notation.

Throughout our discussion, we assume independent censoring of the observations, which means that the censoring mechanism is independent of the event process (Lawless, 2003). Thus, censoring would not affect parameter estimation, so we do not consider the distribution of censoring when we construct the likelihood functions for parameter estimation. More specifically, we consider random right censoring in our simulation study. It can be easily shown that our inferences on parameter estimation can be readily applied to the case of Type-II right censored data under the independent censoring assumption. Whilst we would not consider Type-II right censoring in our simulation, we would illustrate the use of our algorithms to estimate the parameters under Type-II right censoring with a real data set in Chapter 5.

3.1 The BS Distribution

3.1.1 Likelihood Principle

We explain briefly about the classical maximum likelihood estimation for the BS distribution. To simplify notation, define the following functions:

$$z(t) = \frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right) \tag{3.1}$$

$$S(t) = 1 - \Phi\{z(t)\}, \tag{3.2}$$

where $\Phi\{.\}$ denotes the cdf of a standard normal distribution.

Let $\boldsymbol{\theta} = (\alpha, \beta)^T$ be the parameter vector, and let $\boldsymbol{D}$ be the collected data which consists of the observed samples and the censored samples. Then the likelihood function $L(\boldsymbol{\theta}|\boldsymbol{D})$ and the loglikelihood function $l(\boldsymbol{\theta}|\boldsymbol{D})$ for the BS distribution are given by the following:

$$L(\boldsymbol{\theta}|\boldsymbol{D}) \propto \prod_{i=1}^{n}\left[\frac{t_i + \beta}{\alpha\sqrt{\beta}}\exp\left\{-\frac{z^2(t_i)}{2}\right\}\right] \times \prod_{j=1}^{m} S(c_j), \tag{3.3}$$

$$l(\boldsymbol{\theta}|\boldsymbol{D}) = \sum_{i=1}^{n}\left[\log\left(\frac{t_i + \beta}{\alpha\sqrt{\beta}}\right) - \frac{z^2(t_i)}{2}\right] + \sum_{j=1}^{m}\log S(c_j) + \mathcal{C}, \tag{3.4}$$

for some constant $\mathcal{C}$. The inferential procedures to obtain ML estimates and asymptotic confidence intervals are explained in the Appendix section.

3.1.2 Augmented Likelihood

Now we dedicate our focus towards Bayesian parameter estimation for the BS distribution using the Markov Chain Monte Carlo (MCMC) methods, such as the Gibbs Sampling algorithm (see, for example, Casella and George (1992)) and the Metropolis-Hastings algorithm (see, for example, Chib and Greenberg (1995)).

Before we begin discussion on our Bayesian approach, first note that the likelihood function in Equation (3.3) could be simplified into

$$L(\boldsymbol{\theta}|\boldsymbol{D}) \propto (\alpha\sqrt{\beta})^{-n} \exp\left\{-\frac{1}{2\alpha^2}\sum_{i=1}^{n}\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}-2\right)\right\}\prod_{i=1}^{n}(t_i+\beta)$$

in the absence of censored data. This allows the parameter α^2 to have an inverse gamma conjugate prior. Furthermore, as we have shown in the Appendix section, closed form expressions for the Fisher information of the parameters are also available without censored data. This allows us to properly specify the variance-covariance of our proposal distributions in our MCMC procedures.

Therefore, in our Bayesian approach, we adopt the data augmentation technique which was first popularized by Tanner and Wong (1987). Under this method, we consider the censored data as latent variables which are to be sampled together with the parameters in our MCMC procedures. These sampled latent variable values are denoted by $(t_{n+1}, \ldots, t_{n+m})$. Together with the values of observed data $(t_1, \ldots, t_n)$, they constitute the augmented data, denoted by $\boldsymbol{t}$. By sampling these latent variables, we continue our inference procedures as if we finally manage to "observe" all data. In other words, by working with the *augmented likelihood* instead of the original likelihood function, we manage to overcome the limitations brought about by the likelihood expressions involving the censored data.

To simplify notations, we define the following functions:

$$\varphi_1(\boldsymbol{t}) = \sum_{i=1}^{n} t_i + \sum_{j=1}^{m} t_{n+j}, \tag{3.5}$$

$$\varphi_2(\boldsymbol{t}) = \sum_{i=1}^{n} t_i^{-1} + \sum_{j=1}^{m} t_{n+j}^{-1}, \tag{3.6}$$

$$\varphi_3(\boldsymbol{t}) = \frac{\varphi_1(\boldsymbol{t})}{\beta} + \beta\varphi_2(\boldsymbol{t}) - 2(n+m). \tag{3.7}$$

Then the augmented likelihood is given by:

$$L(\boldsymbol{\theta}|\boldsymbol{t}) \propto (\alpha^2\beta)^{-\frac{n+m}{2}} \exp\left\{-\frac{\varphi_3(\boldsymbol{t})}{2\alpha^2}\right\} \prod_{i=1}^{n}(t_i+\beta)\prod_{j=1}^{m}(t_{n+j}+\beta). \tag{3.8}$$

The augmented loglikelihood becomes:

$$l(\boldsymbol{\theta}|\boldsymbol{t}) = -\frac{n+m}{2}\log(\alpha^2\beta) - \frac{\varphi_3(\boldsymbol{t})}{2\alpha^2} + \sum_{i=1}^{n}\log(t_i+\beta) + \sum_{j=1}^{m}\log(t_{n+j}+\beta) + \mathcal{C} \tag{3.9}$$

for some constant $\mathcal{C}$. We refrain from lumping together the notations into $t_1, \ldots, t_{n+m}$ in order to distinguish between the latent variable values $t_{n+1}, \ldots, t_{n+m}$ (which are to be sampled), and the observed samples $t_1, \ldots, t_n$ (which are fixed). Again, the subscripts here do not refer to order statistics.

Note that the augmented likelihood $L(\boldsymbol{\theta}|\boldsymbol{t})$ has the same functional form as the original likelihood function $L(\boldsymbol{\theta}|\boldsymbol{D})$ when $\boldsymbol{D}$ does not contain censored data. This shows that, even if we have censored samples in our data, the augmented likelihood allows us to maintain an inverse gamma conjugate prior for α^2. We could also utilize the closed form expressions for the Fisher information of the augmented data in specifying the variance-covariance of the proposal distributions in our MCMC procedures, as we will show in subsequent subsections.

3.1.3 Prior Specification

We first revisit the original assumption made by Birnbaum and Saunders (1969a) in the derivation of the BS distribution, whereby $\beta = \frac{\omega}{\mu}$, and $\alpha = \frac{\sigma}{\sqrt{\omega\mu}}$. Consequently,

$$\alpha^2 = \frac{\sigma^2}{\omega\mu} = \frac{\sigma^2}{\omega^2} \times \frac{\omega}{\mu} = \frac{\sigma^2}{\omega^2}\beta \propto \beta.$$

From above, we obtain the idea of constructing a joint prior $\pi(\alpha^2, \beta) = \pi(\alpha^2|\beta) \times \pi(\beta)$, such that the conditional prior mean $E(\alpha^2|\beta) \propto \beta$.

Again, from (3.8), it is easily discernible that the inverse gamma distribution is a conjugate prior for conditional distribution $(\alpha^2|\beta)$. Therefore, we specify the conditional prior distribution of α^2 given β to be

$$\alpha^2|\beta \sim \mathcal{IG}\left(\frac{a_0}{2}, \frac{a_0\beta}{2a_1}\right), \tag{3.10}$$

with $a_1 > 0$ and $a_0 > 4$ to ensure existence of $Var(\alpha^2|\beta)$. It follows that the conditional prior density of α^2 given β is given by

$$\pi\left(\alpha^2|\beta\right) \propto \left(\alpha^2\right)^{-\left(\frac{a_0}{2}+1\right)} \times \exp\left\{-\frac{a_0}{2a_1}\left(\frac{\beta}{\alpha^2}\right)\right\} \times \beta^{\frac{a_0}{2}}. \tag{3.11}$$

From (3.8), it is clear that no standard distribution works as the conjugate prior for β. However, the conditional augmented likelihood for $(\beta|\alpha, \boldsymbol{t})$ is given by

$$L(\beta|\alpha, \boldsymbol{t}) \propto \beta^{-\frac{n+m}{2}} \exp\left\{-\frac{\varphi_3(\boldsymbol{t})}{2\alpha^2}\right\} \prod_{i=1}^{n}(t_i + \beta) \prod_{j=1}^{m}(t_{n+j} + \beta). \tag{3.12}$$

Hence, we may consider a prior distribution for β which has a similar functional form as its conditional augmented likelihood function. In this case, we pick the marginal prior of β to be

$$\beta \sim \mathcal{IG}\left(\frac{b_0}{2}, \frac{b_0}{2b_1}\right), \tag{3.13}$$

with $b_1 > 0$ and $b_0 > 4$ to ensure existence of $Var(\beta)$. Then the prior density of β can be written as

$$\pi(\beta) \propto \beta^{-\left(\frac{b_0}{2}+1\right)} e^{-\frac{b_0}{2b_1}\left(\frac{1}{\beta}\right)}. \tag{3.14}$$

To specify the hyperparameters a_0, a_1, b_0 and b_1, first note that the prior mean and prior variance of β are given by:

$$E[\beta] = \frac{b_0}{b_1(b_0 - 2)} \tag{3.15}$$

$$Var[\beta] = \frac{2b_0^2}{b_1^2(b_0 - 2)^2(b_0 - 4)}, \tag{3.16}$$

whereas the the conditional prior mean and conditional prior variance of α^2 given β are given by:

$$E[\alpha^2|\beta] = \frac{a_0\beta}{a_1(a_0-2)} \tag{3.17}$$

$$Var[\alpha^2|\beta] = \frac{2a_0^2\beta^2}{a_1^2(a_0-2)^2(a_0-4)}. \tag{3.18}$$

Then, we may consider the following factors:

- β is the median of the BS distribution. We can refer to the sample median when we attempt to specify b_0 and b_1.
- The shape parameter α determines the shape of the hazard function. Hence, we can incorporate prior knowledge or cross-industry expertise about the hazard function in order to specify a_0 and a_1.

3.1.4 Posterior Inference

To simplify notations, let $\nu_0 = \frac{a_0+n+m}{2}$, $\tau_0 = \frac{b_0-a_0+n+m}{2}$, and

$$\varphi_4(\boldsymbol{t}) = -\frac{1}{2\alpha^2}\left[\varphi_3(\boldsymbol{t}) + \frac{a_0\beta}{a_1}\right] - \frac{b_0}{2b_1\beta}.$$

From (3.8), (3.11) and (3.14), the joint posterior distribution of the parameters is given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{t}) \propto \left(\alpha^2\right)^{-(\nu_0+1)} \beta^{-(\tau_0+1)} e^{\varphi_4(\boldsymbol{t})} \prod_{i=1}^{n}(t_i+\beta) \prod_{j=1}^{m}(t_{n+j}+\beta). \tag{3.19}$$

It follows that the conditional posterior of $(\alpha^2|\beta, \boldsymbol{t})$ has an inverse gamma distribution as follows

$$\left(\alpha^2|\beta, \boldsymbol{t}\right) \sim \mathcal{IG}\left(\nu_0, \frac{1}{2}\left[\varphi_3(\boldsymbol{t}) + \frac{a_0\beta}{a_1}\right]\right). \tag{3.20}$$

In other words, the conditional posterior density of $(\alpha^2|\beta, \boldsymbol{t})$ can be written as

$$\pi(\alpha^2|\beta, \boldsymbol{t}) \propto \left(\alpha^2\right)^{-(\nu_0+1)} \times \exp\left\{-\frac{1}{2\alpha^2}\left[\varphi_3(\boldsymbol{t}) + \frac{a_0\beta}{a_1}\right]\right\}. \tag{3.21}$$

The conditional posterior of $(\beta|\alpha^2, \boldsymbol{t})$ is not a standard distribution, which is given by the following:

$$\pi(\beta|\alpha^2, \boldsymbol{t}) \propto \beta^{-(\tau_0+1)} e^{\varphi_4(\boldsymbol{t})} \prod_{i=1}^{n}(t_i + \beta) \prod_{j=1}^{m}(t_{n+j} + \beta). \tag{3.22}$$

3.1.5 Sampling Algorithm

We will implement two slightly different Gibbs sampling procedures for our Bayesian parameter estimation. The first algorithm involves sampling the parameters individually (henceforth known as conditional sampling), and the second involves sampling the parameters jointly (henceforth known as joint sampling).

Sampling Scheme 1 - Conditional Sampling

First, set initial values for the parameters α and β. Then at iteration step $t+1$,

1. Sample the latent variables $T_{n+1}, \ldots, T_{n+m}$ to obtain the updated augmented data $\boldsymbol{t}$.
2. Draw β_{t+1} from $\pi(\beta|\alpha_t^2, \boldsymbol{t})$ via a Metropolis-Hastings procedure, where $\boldsymbol{t}$ is the augmented data updated in Step 1, and α_t^2 represents the value of α^2 drawn from the previous step t.
3. Draw $\left(\alpha_{t+1}^2|\beta_{t+1}, \boldsymbol{t}\right) \sim \mathcal{IG}\left(\nu_0, \frac{1}{2}\left[\varphi_3(\boldsymbol{t}) + \frac{a_0}{a_1}\beta_{t+1}\right]\right)$ using the updated augmented data $\boldsymbol{t}$ and β_{t+1} from Steps 1 and 2.

After sampling from the conditional posteriors, we can obtain point estimates and credible intervals of the parameters α and β.

We will now discuss Steps 1 and 2 in detail. In Step 1, we update the latent variables based on the values of the parameters drawn from the previous step (denoted by $\boldsymbol{\theta}_t$) before we proceed to sample the parameters in Steps 2 and 3. Since we assume an underlying

BS distribution for the data, sampling these latent variables is equivalent to drawing truncated Birnbaum-Saunders random variates $(T_{n+j}|T_{n+j} > c_j, \boldsymbol{\theta}_t)$ for $j = 1, \ldots, m$ at every iteration step $t + 1$.

From the relationship between the BS distribution T and the standard normal distribution Z given in (1.3), it can be easily verified that one variable is an increasing function of another. This provides us a convenient way to sample the truncated BS random variates. At step $t + 1$, we could first draw truncated normal random variates $(Z_{n+j}|Z_{n+j} > d_j)$, where

$$d_j = \frac{1}{\alpha_t}\left[\sqrt{\frac{c_j}{\beta_t}} - \sqrt{\frac{\beta_t}{c_j}}\right]. \tag{3.23}$$

Then for $j = 1, \ldots, m$, we could convert the sampled Z_{n+j} to T_{n+j} using the following relationship:

$$T_{n+j} = \beta_t\left(1 + \frac{\alpha_t^2 Z_{n+j}^2}{2} + \alpha_t Z_{n+j}\sqrt{1 + \frac{\alpha_t^2 Z_{n+j}^2}{4}}\right). \tag{3.24}$$

Subsequently in Step 2, we propose β from a lognormal distribution using a Metropolis-Hastings step. In other words, propose $\log\beta_p \sim N\left(\log\beta_t, \gamma_\beta^2\sigma_t^2\right)$. The subscript p denotes proposal, $\gamma_\beta > 0$ is a tuning parameter for the proposal distribution of β, and

$$\begin{aligned}\sigma_t^2 &= \left(-E\left\{\frac{\partial^2\left(\log\pi(\beta|\alpha_t^2, \boldsymbol{t}) + \log|J|\right)}{\partial(\log\beta)^2}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}\right)^{-1} \\ &= \left(\frac{(n+m)[\alpha_t(2\pi)^{-1/2}h(\alpha_t) + 1]}{\alpha_t^2} + \frac{b_0}{2b_1\beta_t} + \frac{a_0\beta_t}{2a_1\alpha_t^2}\right)^{-1},\end{aligned} \tag{3.25}$$

where $h(\alpha_t) = \alpha_t\sqrt{(\pi/2)} - \pi e^{2/\alpha_t^2}[1 - \Phi(2/\alpha_t)]$, and the Jacobian term here is $J = \beta$. For each data set, the tuning parameter γ_β is calibrated such that the acceptance rate of the Metropolis-Hastings step lies within the preferable range of around 40% (Gelman et al., 2004). Via our simulation study, we find that the rule of thumb $\gamma_\beta = 2.4$ as suggested by Gelman et al. (2004) is adequate in ensuring the acceptance rates to hover around 41% -

47%.

For the purpose of sampling efficiency, we intend to specify a proposal distribution which closely resembles the conditional posterior distribution (Gelman et al., 2004). This consideration prompts us to specify the variance term of our normal proposal distribution based on the posterior likelihood, which is essentially the conditional posterior distribution of $\log \beta$. This explains why we take the second derivative of $(\log \pi(\beta|\alpha_t^2, \boldsymbol{t}) + \log |J|)$ with respect to $\log \beta$. The Jacobian term J is needed here as we make a log transformation on β.

We also highlight the fact that the existing closed-form expression of the Fisher information can be easily modified and implemented to obtain the first term of (3.25) because we are now working with the augmented data which "removes" the cumbersome censored variables. The second and third term of (3.25) are the resultant derivative terms from the prior component in the conditional posterior distribution.

After obtaining $\log \beta_p$, exponentiate it to get β_p. Next, denote the lognormal proposal density of β with $q_\beta(.)$, and take

$$\beta_{t+1} = \begin{cases} \beta_p & \text{with probability } \lambda_\beta \\ \beta_t & \text{with probability } 1 - \lambda_\beta \end{cases},$$

where

$$\lambda_\beta = \min \left\{ 1, \frac{\pi(\beta_p|\alpha_t^2, \boldsymbol{t}) \times q_\beta(\beta_t|\beta_p)}{\pi(\beta_t|\alpha_t^2, \boldsymbol{t}) \times q_\beta(\beta_p|\beta_t)} \right\}, \tag{3.26}$$

with

$$\frac{q_\beta(\beta_t|\beta_p)}{q_\beta(\beta_p|\beta_t)} = \frac{(\sigma_p \beta_t)^{-1} \exp\left\{ -\frac{(\log \beta_t - \log \beta_p)^2}{2\gamma_\beta^2 \sigma_p^2} \right\}}{(\sigma_t \beta_p)^{-1} \exp\left\{ -\frac{(\log \beta_p - \log \beta_t)^2}{2\gamma_\beta^2 \sigma_t^2} \right\}}, \tag{3.27}$$

and

$$\sigma_p^2 = \left(\frac{(n+m)[\alpha_t (2\pi)^{-1/2} h(\alpha_t) + 1]}{\alpha_t^2} + \frac{b_0}{2b_1 \beta_p} + \frac{a_0 \beta_p}{2a_1 \alpha_t^2} \right)^{-1}.$$

A possible modification of Step 2 would be to first integrate out α from the joint posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{t})$ to obtain the marginal posterior distribution of β. Then, employ a Metropolis-Hastings step to sample β from its marginal posterior distribution. This explains why we chose to first sample β before α^2. This method works well too, but, by drawing from the conditional posterior of β given α, we could easily modify the existing formula for the Fisher information (which involves α) to specify the variance term of the proposal distribution of β, as we have shown earlier.

Sampling Scheme 2 - Joint Sampling

To sample the parameters jointly, first set initial values for the parameters α and β. Then at iteration step $t+1$,

1. Sample the latent variables $T_{n+1}, \ldots, T_{n+m}$ to update the augmented data $\boldsymbol{t}$ using the method explained in Step 1 of Sampling Scheme 1.

2. Draw the two parameters $\boldsymbol{\theta} = (\alpha, \beta)^T$ from the joint posterior density of the parameters $\pi(\boldsymbol{\theta}|\boldsymbol{t})$, where $\boldsymbol{t}$ includes the set of observed data $(t_1, \ldots, t_n)$ and the latent variable values $(t_{n+1}, \ldots, t_{n+m})$ sampled from Step 1. This step requires a Metropolis-Hastings procedure as explained below.

After sampling from the joint posterior, we can obtain point estimates and credible intervals of the parameters α and β.

We will propose the two parameters jointly from a bivariate lognormal distribution. In other words, denote $\log \boldsymbol{\theta} = (\log \alpha, \log \beta)^T$ and propose

$$\log \boldsymbol{\theta}_p \sim N_2\left(\log \boldsymbol{\theta}_t, \gamma_{\boldsymbol{\theta}}^2 \Sigma_t\right), \tag{3.28}$$

for some tuning parameter $\gamma_{\boldsymbol{\theta}} > 0$. Here, the subscript p refers to proposal, and $\log \boldsymbol{\theta}_t$

denotes $(\log \alpha_t, \log \beta_t)^T$. By defining

$$h(\alpha_t) = \alpha_t \sqrt{(\pi/2)} - \pi e^{2/\alpha_t^2}[1 - \Phi(2/\alpha_t)],$$
$$g(\alpha_t) = \frac{(n+m)[\alpha_t (2\pi)^{-\frac{1}{2}} h(\alpha_t) + 1]}{\alpha_t^2},$$

we have

$$\begin{aligned}\Sigma_t &= \left[-E \left\{ \frac{\partial^2 \left(\log \pi(\boldsymbol{\theta}|\boldsymbol{t}) + \log |J| \right)}{\partial(\log \boldsymbol{\theta}) \partial(\log \boldsymbol{\theta})^T} \right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_t} \right]^{-1} \\ &= \begin{bmatrix} 2(n+m) + \dfrac{2a_0\beta_t}{a_1\alpha_t^2} & -\dfrac{a_0\beta_t}{a_1\alpha_t^2} \\ -\dfrac{a_0\beta_t}{a_1\alpha_t^2} & g(\alpha_t) + \dfrac{b_0}{2b_1\beta_t} + \dfrac{a_0\beta_t}{2a_1\alpha_t^2} \end{bmatrix}^{-1}, \end{aligned} \tag{3.29}$$

where the Jacobian term above is $J = \alpha\beta$. For each data set, the tuning parameter $\gamma_{\boldsymbol{\theta}}$ is calibrated such that the acceptance rate of the Metropolis-Hastings step lies within the preferable range of around 30% (Gelman et al., 2004). Via our simulation study, we find that the rule of thumb $\gamma_{\boldsymbol{\theta}} = \frac{2.4}{\sqrt{2}}$ as suggested by Gelman et al. (2004) is adequate in ensuring the acceptance rate to hover around 31% - 36%.

Again, the covariance term in our normal proposal distribution is specified based on posterior likelihood for the purpose of efficient sampling as discussed previously in Sampling Scheme 1. Note that, even though α and β are asymptotically independent in the absence of censored data, the non-diagonal terms in the covariance matrix of the proposal distribution are non-zero because they are the resultant derivative terms from the conditional priors $\pi(\alpha^2, \beta) = \pi(\alpha^2|\beta) \times \pi(\beta)$ that we specified earlier. Again, by working with the augmented data, the existing closed-form expressions of the Fisher information can be easily modified and implemented to obtain the covariance term above.

Next, exponentiate $\log \boldsymbol{\theta}_p$ to obtain $\boldsymbol{\theta}_p$. Denote $q_{\boldsymbol{\theta}}(.)$ as the bivariate lognormal proposal distribution of $\boldsymbol{\theta}$, and define

$$h(\alpha_p) = \alpha_p\sqrt{(\pi/2)} - \pi e^{2/\alpha_p^2}[1 - \Phi(2/\alpha_p)],$$
$$g(\alpha_p) = \frac{(n+m)[\alpha_p(2\pi)^{-\frac{1}{2}}h(\alpha_p) + 1]}{\alpha_p^2}.$$

Then, take

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}_p & \text{with probability } \lambda_{\boldsymbol{\theta}} \\ \boldsymbol{\theta}_t & \text{with probability } 1 - \lambda_{\boldsymbol{\theta}} \end{cases},$$

where

$$\lambda_{\boldsymbol{\theta}} = \min\left\{1, \frac{\pi(\boldsymbol{\theta}_p|\boldsymbol{t}) \times q_{\boldsymbol{\theta}}(\boldsymbol{\theta_t}|\boldsymbol{\theta_p})}{\pi(\boldsymbol{\theta}_t|\boldsymbol{t}) \times q_{\boldsymbol{\theta}}(\boldsymbol{\theta_p}|\boldsymbol{\theta_t})}\right\}, \tag{3.30}$$

with

$$\frac{q_{\boldsymbol{\theta}}(\boldsymbol{\theta_t}|\boldsymbol{\theta_p})}{q_{\boldsymbol{\theta}}(\boldsymbol{\theta_p}|\boldsymbol{\theta_t})} = \frac{(\alpha_t\beta_t)^{-1}|\Sigma_p|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\log\boldsymbol{\theta}_t - \log\boldsymbol{\theta}_p)^T(\gamma_{\boldsymbol{\theta}}^2\Sigma_p)^{-1}(\log\boldsymbol{\theta}_t - \log\boldsymbol{\theta}_p)\}}{(\alpha_p\beta_p)^{-1}|\Sigma_t|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\log\boldsymbol{\theta}_p - \log\boldsymbol{\theta}_t)^T(\gamma_{\boldsymbol{\theta}}^2\Sigma_t)^{-1}(\log\boldsymbol{\theta}_p - \log\boldsymbol{\theta}_t)\}}, \tag{3.31}$$

and

$$\Sigma_p = \begin{bmatrix} 2(n+m) + \dfrac{2a_0\beta_p}{a_1\alpha_p^2} & -\dfrac{a_0\beta_p}{a_1\alpha_p^2} \\ -\dfrac{a_0\beta_p}{a_1\alpha_p^2} & g(\alpha_p) + \dfrac{b_0}{2b_1\beta_p} + \dfrac{a_0\beta_p}{2a_1\alpha_p^2} \end{bmatrix}^{-1}.$$

3.2 The GBS Distribution

We will now discuss our methodology for estimating the three parameters α, β and κ in the GBS distribution.

3.2.1 Likelihood Principle

We explain briefly about the classical maximum likelihood estimation for the GBS distribution. To simplify notation, define the following functions:

$$z(t) = \frac{1}{\alpha}\left(\frac{t^{1-\kappa}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{t^{\kappa}}\right), \tag{3.32}$$

$$S(t) = 1 - \Phi\{z(t)\}, \tag{3.33}$$

$$\eta(t) = t(1-\kappa) + \beta\kappa, \tag{3.34}$$

where $\Phi\{.\}$ denotes the cdf of a standard normal distribution.

Let $\boldsymbol{\theta} = (\alpha, \beta, \kappa)^T$ be the parameter vector, and, again, let $\boldsymbol{D}$ be the collected data which consists of the observed samples and the censored samples. Then the likelihood function $L(\boldsymbol{\theta}|\boldsymbol{D})$ and the loglikelihood function $l(\boldsymbol{\theta}|\boldsymbol{D})$ for the GBS distribution are given by the following:

$$L(\boldsymbol{\theta}|\boldsymbol{D}) \propto \prod_{i=1}^{n}\left[\frac{t_i^{-(\kappa+1)}\eta(t_i)}{\alpha\sqrt{\beta}}\exp\left\{-\frac{z^2(t_i)}{2}\right\}\right] \times \prod_{j=1}^{m} S(c_j), \tag{3.35}$$

$$l(\boldsymbol{\theta}|\boldsymbol{D}) = \sum_{i=1}^{n}\left[\log\left(\frac{t_i^{-(\kappa+1)}\eta(t_i)}{\alpha\sqrt{\beta}}\right) - \frac{z^2(t_i)}{2}\right] + \sum_{j=1}^{m}\log S(c_j) + \mathcal{C}, \tag{3.36}$$

for some constant $\mathcal{C}$. The inferential procedures to obtain ML estimates and asymptotic confidence intervals are explained in the Appendix section.

3.2.2 Augmented Likelihood

Again, we will utilize the same strategy to circumvent the arduous expressions of the likelihood function involving the censored data (as shown in the Appendix section) by sampling these latent variables and working with augmented data $\boldsymbol{t}$ instead (Tanner and

Wong, 1987). To simplify expressions, we will first introduce the following notations:

$$\varphi_1(\boldsymbol{t}) = \sum_{i=1}^{n} t_i^{2-2\kappa} + \sum_{j=1}^{m} t_{n+j}^{2-2\kappa} \tag{3.37}$$

$$\varphi_2(\boldsymbol{t}) = \sum_{i=1}^{n} t_i^{-2\kappa} + \sum_{j=1}^{m} t_{n+j}^{-2\kappa} \tag{3.38}$$

$$\varphi_3(\boldsymbol{t}) = \sum_{i=1}^{n} t_i^{1-2\kappa} + \sum_{j=1}^{m} t_{n+j}^{1-2\kappa} \tag{3.39}$$

$$\varphi_4(\boldsymbol{t}) = \frac{\varphi_1(\boldsymbol{t})}{\beta} + \beta\varphi_2(\boldsymbol{t}) - 2\varphi_3(\boldsymbol{t}), \tag{3.40}$$

$$\varphi_5(\boldsymbol{t}) = t^{-(\kappa+1)}\eta(t). \tag{3.41}$$

Then, the augmented likelihood is given by:

$$L(\boldsymbol{\theta}|\boldsymbol{t}) \propto (\alpha^2\beta)^{-\frac{n+m}{2}} \exp\left\{-\frac{\varphi_4(\boldsymbol{t})}{2\alpha^2}\right\} \prod_{i=1}^{n} \varphi_5(t_i) \prod_{j=1}^{m} \varphi_5(t_{n+j}). \tag{3.42}$$

For some constant $\mathcal{C}$, the augmented loglikelihood becomes:

$$l(\boldsymbol{\theta}|\boldsymbol{t}) = -\frac{n+m}{2}\log(\alpha^2\beta) - \frac{\varphi_4(\boldsymbol{t})}{2\alpha^2} + \sum_{i=1}^{n} \log[\varphi_5(t_i)] + \sum_{j=1}^{m} \log[\varphi_5(t_{n+j})] + \mathcal{C}. \tag{3.43}$$

3.2.3 Prior Specification

From the augmented likelihood in (3.42), $(\alpha^2|\beta)$ maintains an inverse gamma conjugate prior. In addition, based on the model assumption of the GBS distribution by Owen (2006), we propose a joint prior $\pi(\alpha^2, \beta, \kappa) = \pi(\alpha^2|\beta) \times \pi(\beta) \times \pi(\kappa)$, whereby

$$\alpha^2|\beta \sim \mathcal{IG}\left(\frac{a_0}{2}, \frac{a_0\beta}{2a_1}\right), \tag{3.44}$$

with $a_1 > 0$ and $a_0 > 4$ to ensure existence of $Var(\alpha^2|\beta)$, and

$$\beta \sim \mathcal{IG}\left(\frac{b_0}{2}, \frac{b_0}{2b_1}\right), \tag{3.45}$$

with $b_1 > 0$ and $b_0 > 4$ to ensure existence of $Var(\beta)$. The conditional prior density of $(\alpha^2|\beta)$ as wells as the marginal prior density of β are given by (3.11) and (3.14) respectively.

Since $0 < \kappa < 1$, we specify the Beta prior distribution for κ:

$$\kappa \sim Beta(d_0, d_1), \tag{3.46}$$

such that

$$\pi(\kappa) \propto \kappa^{d_0-1}(1-\kappa)^{d_1-1}. \tag{3.47}$$

If no prior information or idea is available about κ, then pick $d_0 = d_1 = 1$, which is equivalent to picking a standard uniform prior for κ.

3.2.4 Posterior Inference

Let $\nu_0 = \frac{a_0+n+m}{2}$, $\tau_0 = \frac{b_0-a_0+n+m}{2}$, and

$$\varphi_6(\boldsymbol{t}) = -\frac{\left[\varphi_4(\boldsymbol{t}) + \frac{a_0\beta}{a_1}\right]}{2\alpha^2} - \frac{b_0}{2b_1\beta}. \tag{3.48}$$

From (3.42), (3.11), (3.14), and (3.47), the joint posterior distribution of the parameters is given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{t}) \propto \left(\alpha^2\right)^{-(\nu_0+1)} \beta^{-(\tau_0+1)} e^{\varphi_6(\boldsymbol{t})} \kappa^{d_0-1}(1-\kappa)^{d_1-1} \prod_{i=1}^{n} \varphi_5(t_i) \prod_{j=1}^{m} \varphi_5(t_{n+j}). \tag{3.49}$$

It follows that the conditional posterior of $(\alpha^2|\beta, \kappa, \boldsymbol{t})$ has an inverse gamma distribution with parameters

$$\left(\alpha^2|\beta, \kappa, \boldsymbol{t}\right) \sim \mathcal{IG}\left(\nu_0, \frac{1}{2}\left[\varphi_4(\boldsymbol{t}) + \frac{a_0\beta}{a_1}\right]\right). \tag{3.50}$$

In other words,

$$\pi(\alpha^2|\beta, \kappa, \boldsymbol{t}) \propto \left(\alpha^2\right)^{-(\nu_0+1)} \times \exp\left\{-\frac{1}{2\alpha^2}\left[\varphi_4(\boldsymbol{t}) + \frac{a_0\beta}{a_1}\right]\right\}. \tag{3.51}$$

The conditional posterior of $(\beta|\alpha^2, \kappa, \boldsymbol{t})$ is then given by

$$\pi(\beta|\alpha^2, \kappa, \boldsymbol{t}) \propto \beta^{-(\tau_0+1)} e^{\varphi_6(\boldsymbol{t})} \prod_{i=1}^{n} \eta(t_i) \prod_{j=1}^{m} \eta(t_{n+j}) \tag{3.52}$$

and the conditional posterior of $(\kappa|\alpha^2, \beta, \boldsymbol{t})$ becomes

$$\pi(\kappa|\alpha^2, \beta, \boldsymbol{t}) \propto \kappa^{d_0-1}(1-\kappa)^{d_1-1} e^{\left\{-\frac{\varphi_4(\boldsymbol{t})}{2\alpha^2}\right\}} \prod_{i=1}^{n} \varphi_5(t_i) \prod_{j=1}^{m} \varphi_5(t_{n+j}). \tag{3.53}$$

3.2.5 Sampling Algorithm

We will implement a Gibbs sampling algorithm to sample the parameters individually/conditionally. Again, at every iteration step, we will first draw the latent variables to update our augmented data before we proceed to sample our parameters.

Sampling Scheme - Conditional Sampling

First, set initial values for the parameters α, β and κ. Then at iteration step $t+1$,

1. Sample the latent variables $T_{n+1}, \ldots, T_{n+m}$.

2. Draw κ from $\pi(\kappa|\alpha_t^2, \beta_t, \boldsymbol{t})$ using a Metropolis-Hastings procedure, where augmented data $\boldsymbol{t}$ includes the set of observed data $(t_1, \ldots, t_n)$ and the latent variable values $t_{n+1}, \ldots, t_{n+m}$ sampled from Step 1, while α_t^2 and β_t represent the values of α^2 and β sampled from the previous step t.

3. Draw β from $\pi(\beta|\alpha_t^2, \kappa_{t+1}, \boldsymbol{t})$ using a Random Walk (RW) Metropolis procedure, where κ_{t+1} represents the updated value of κ from Step 2, and $\boldsymbol{t}$ represents the updated augmented data from Step 1.

4. Draw $\alpha^2 \sim \mathcal{IG}\left(\nu_0, \frac{1}{2}\left[\varphi_4(\boldsymbol{t}) + \frac{a_0}{a_1}\beta_{t+1}\right]\right)$ using the updated augmented data $\boldsymbol{t}$ from Step 1 and the updated values of β and κ from Steps 2 and 3.

After sampling from their respective conditional posterior distributions, we can obtain point estimates and credible intervals of the parameters α, β and κ.

We will now explain Steps 1-3 in detail. For Step 1, the attempt to sample the latent variables is equivalent to drawing truncated GBS random variates $(T_{n+j}|T_{n+j} > c_j, \boldsymbol{\theta}_t)$ for $j = 1, \ldots, m$. Recall that the relationship between the standard normal random variable Z and the GBS distribution T is given by $Z = \frac{1}{\alpha}\left[\frac{T^{1-\kappa}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{T^{\kappa}}\right]$. It can be easily verified that one variable is an increasing function of another. This provides us a convenient way to sample the truncated GBS random variates. At step $t+1$, we could first draw truncated standard normal random variates $(Z_{n+j}|Z_{n+j} > d_j)$, where

$$d_j = \frac{1}{\alpha_t}\left[\frac{c_j^{1-\kappa_t}}{\sqrt{\beta_t}} - \frac{\sqrt{\beta_t}}{c_j^{\kappa_t}}\right]. \tag{3.54}$$

Then for $j = 1, \dots, m$, we could convert the sampled Z_{n+j} to T_{n+j} by solving the following non-linear equation for T_{n+j}:

$$\alpha_t\sqrt{\beta_t}Z_{n+j}T_{n+j}^{\kappa_t} - T_{n+j} + \beta_t = 0. \tag{3.55}$$

Since T is an increasing function of Z (and vice versa), there is only one root for the non-linear equation above.

In Step 2, propose $\kappa_p \sim Beta\left(\gamma_\kappa\kappa_t\ ,\ \gamma_\kappa(1-\kappa_t)\right)$, where the subscript p denotes proposal and $\gamma_\kappa > 0$ is a tuning parameter to be calibrated for each data set to achieve reasonable acceptance rates for this Metropolis-Hastings procedure in Step 2.

Denote the Beta proposal density of κ with $q_\kappa(.)$, and take

$$\kappa_{t+1} = \begin{cases} \kappa_p & \text{with probability } \lambda_\kappa \\ \kappa_t & \text{with probability } 1-\lambda_\kappa \end{cases},$$

where

$$\lambda_\kappa = \min\left\{1, \frac{\pi(\kappa_p|\alpha_t^2, \beta_t, \boldsymbol{t}) \times q_\kappa(\kappa_t|\kappa_p)}{\pi(\kappa_t|\alpha_t^2, \beta_t, \boldsymbol{t}) \times q_\kappa(\kappa_p|\kappa_t)}\right\}, \tag{3.56}$$

and

$$\frac{q_\kappa(\kappa_t|\kappa_p)}{q_\kappa(\kappa_p|\kappa_t)} = \frac{\Gamma(\gamma_\kappa\kappa_t)\Gamma[\gamma_\kappa(1-\kappa_t)]}{\Gamma(\gamma_\kappa\kappa_p)\Gamma[\gamma_\kappa(1-\kappa_p)]} \times \frac{\kappa_t^{\gamma_\kappa\kappa_p-1}(1-\kappa_t)^{\gamma_\kappa(1-\kappa_p)-1}}{\kappa_p^{\gamma_\kappa\kappa_t-1}(1-\kappa_p)^{\gamma_\kappa(1-\kappa_t)-1}}. \tag{3.57}$$

In Step 3, draw β from a lognormal proposal density with a Random Walk (RW) Metropolis step. In other words, propose $\log \beta_p \sim \mathcal{N}(\log \beta_t, \gamma_\beta^2 \widehat{\sigma}^2)$, where the subscript p denotes proposal, $\gamma_\beta > 0$ is a tuning parameter to be calibrated for each data set to achieve reasonable acceptance rates, and

$$\widehat{\sigma}^2 = \left(-\beta^2 \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial \beta^2} + \frac{b_0}{2b_1\beta} + \frac{a_0\beta}{2a_1\alpha^2} \right)^{-1}_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{3.58}$$

From our simulation study, the rule of thumb $\gamma_\beta = 2.4$ as suggested by Gelman et al. (2004) provides adequate tuning to achieve reasonable acceptance rate for this RW Metropolis step. Exponentiate $\log \beta_p$ to obtain β_p. Denote the proposal density of β with $q_\beta(.)$, and take

$$\beta_{t+1} = \begin{cases} \beta_p & \text{with probability } \lambda_\beta \\ \beta_t & \text{with probability } 1 - \lambda_\beta \end{cases},$$

where

$$\lambda_\beta = \min \left\{ 1, \frac{\pi(\beta_p|\alpha_t^2, \kappa_{t+1}, \boldsymbol{t}) \times q_\beta(\beta_t|\beta_p)}{\pi(\beta_t|\alpha_t^2, \kappa_{t+1}, \boldsymbol{t}) \times q_\beta(\beta_p|\beta_t)} \right\}, \tag{3.59}$$

with

$$\frac{q_\beta(\beta_t|\beta_p)}{q_\beta(\beta_p|\beta_t)} = \frac{\beta_p}{\beta_t}. \tag{3.60}$$

We will now explain the reason behind the use of a RW Metropolis procedure instead of a Metropolis-Hastings procedure in Step 3. Again, the variance term above is specified based on the posterior likelihood

$$\left[-\widehat{E} \left\{ \frac{\partial^2 \left(\log \pi(\beta|\alpha_t^2, \boldsymbol{t}) + \log |J| \right)}{\partial (\log \beta)^2} \right\} \right]^{-1} = \left[-\beta^2 \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{t})}{\partial \beta^2} + \frac{b_0}{2b_1\beta} + \frac{a_0\beta}{2a_1\alpha^2} \right]^{-1}, \tag{3.61}$$

where the Jacobian term is $J = \beta$.

Note that, under the case of GBS distribution, there is no closed form expression for the Fisher information $-E\left\{ \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{t})}{\partial \beta^2} \right\}$, so instead, we have to rely on the observed Fisher information

$$-\widehat{E}\left\{ \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{t})}{\partial \beta^2} \right\} = -\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{t})}{\partial \beta^2}$$

in specifying the variance term for our normal proposal density, as shown in (3.61). However, by using the observed Fisher information (which is dependent on data), this "approximate" variance term may not be positive-definite for the entire parameter space $\{\alpha \in (0,\infty), \beta \in (0,\infty), \kappa \in (0,1)\}$. A Metropolis-Hastings algorithm will produce error when the the parameters assume values which belong to the subset of parameter space where the "approximate" variance term is not positive-definite.

To overcome this problem, we will replace the Metropolis-Hastings step with a RW Metropolis step by fixing the variance term for each iteration step. Therefore, we estimate the observed Fisher information $-\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{t})}{\partial \beta^2}$ for the augmented data (which is updated in every iteration step) with the observed MLE Fisher information $\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial \beta^2}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ which is a fixed value. All parameters in the "approximate" variance term would also assume their corresponding ML estimates, and we obtain the fixed variance term $\widehat{\sigma}^2$ as given by (3.58) for our normal proposal density.

Chapter 4

Simulation Study

4.1 The BS Distribution

We perform several simulation examples to assess performance of the ML and Bayesian estimators. Rieck and Nedelman (1991) pointed out that, in practice, if the fatigue lifetime of a metal specimen subjected to cyclical stress loading has a BS distribution, then the shape parameter α would usually not exceed one. Therefore, in our simulation examples, we take the shape parameter $\alpha = 0.5, 1$ and 2. Without loss of generaliy, the scale parameter β is fixed at 1. In addition, we apply 4 random right censoring percentages (CEP) at 10%, 20%, 30% and 40% for two sample sizes, $n = 20$ and 50.

For each of these 24 parameter settings, we generate 10,000 data sets, and apply the MLE, conditional sampling and joint sampling algorithms, as explained in Chapter 3, to obtain the point estimates and interval estimates of the parameters α and β for these simulated data sets. Then, we compute the average point estimates, average length (AL) of the 95% credible intervals (for conditional sampling and joint sampling) and average length (AL) of the 95% confidence intervals (for MLE), coverage probability (CP), and square root of mean squared error (SRMSE) of the parameters.

Where Gibbs sampling is concerned, we adopt the priors as explained in Chapter 3 and pick these hyperparameters $a_0 = a_1 = b_0 = b_1 = 5$, such that the prior distributions are rather "flat" or "less informative", in order to illustrate the case where we have little prior knowledge about the parameters. For conditional sampling, the Gibbs sampler has 8,000 iterations with a burn-in period of 2,000. The tuning parameter γ_β for the proposal of β is fixed at 2.4, as suggested by Gelman et al. (2004), to ensure acceptance rates of the Metropolis-Hastings step to remain at 40% to 47%. For joint sampling, the Gibbs sampler has 6,000 iterations with a burn-in period of 1,500. The tuning parameter $\gamma_{\boldsymbol{\theta}}$ for the joint proposal is fixed at $2.4/\sqrt{2}$ as suggested by Gelman et al. (2004), to ensure acceptance rates of the Metropolis-Hastings step to remain at 30% to 37%.

We separate the simulation results into three tables, namely Table 4.1 for parameter setting $\alpha = 0.5$, Table 4.2 for parameter setting $\alpha = 1$, and Table 4.3 for parameter setting $\alpha = 2$. Each table contains results for different CEP and sample sizes. To aid visualization, we also plot the average estimates, AL of 95% CI and SRMSE for both α and β under different CEP and sample sizes for each of the α parameter settings.

From these tables and plots, some features can be summarized as follows:

1. Bias, SRMSE and average length (AL) of 95% CI decrease with sample size n but increase with censoring percentage (CEP).

2. Compared to conditional sampling, joint sampling obtains slightly smaller bias for α but much higher bias and wider 95% CI for β. Overall, conditional sampling appears to outperform joint sampling. We do not consider joint sampling for the case of GBS distribution.

3. With larger sample size, the performances of the three algorithms (in terms of bias, SRMSE and 95% CI width) are similar to one another.

4. The effect of priors in posterior inference is more prevalent with smaller sample size. Due to the choice of relatively "flat" or "uninformative" priors, conditional sampling and joint sampling produce smaller bias and SRMSE as compared to MLE, but slightly wider 95% credible intervals than the 95% confidence intervals.

Table 4.1: Simulation Results for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$)

Parameter			α				β			
n	Method	CEP(%)	Mean	AL	CP(%)	SRMSE	Mean	AL	CP(%)	SRMSE
20	MLE	10	0.4888	0.3127	90.87	0.0838	1.0466	0.4374	93.20	0.1243
		20	0.4992	0.3345	91.79	0.0893	1.0955	0.4739	90.25	0.1556
		30	0.5111	0.3650	93.14	0.0974	1.1544	0.5215	84.18	0.2028
		40	0.5251	0.4069	93.92	0.1102	1.2293	0.5858	73.44	0.2701
	*Cond	10	0.5111	0.3181	97.02	0.0710	1.0353	0.4679	95.18	0.1166
		20	0.5237	0.3446	97.09	0.0779	1.0849	0.5223	92.62	0.1461
		30	0.5383	0.3778	96.66	0.0875	1.1450	0.5952	86.58	0.1932
		40	0.5558	0.4205	95.68	0.1015	1.2220	0.7002	75.11	0.2619
	Joint	10	0.5010	0.3064	96.97	0.0693	1.0605	0.4746	93.30	0.1297
		20	0.5134	0.3329	97.07	0.0747	1.1134	0.5343	88.09	0.1677
		30	0.5284	0.3677	96.90	0.0836	1.1784	0.6178	79.06	0.2229
		40	0.5469	0.4148	96.09	0.0978	1.2634	0.7452	63.84	0.3020
50	MLE	10	0.5048	0.2030	93.65	0.0540	1.0462	0.2844	92.10	0.0859
		20	0.5193	0.2181	93.73	0.0613	1.0996	0.3106	79.75	0.1259
		30	0.5361	0.2395	93.01	0.0732	1.1656	0.3456	55.95	0.1853
		40	0.5572	0.2703	91.12	0.0918	1.2497	0.3937	25.00	0.2656
	*Cond	10	0.5133	0.2081	95.29	0.0520	1.0417	0.2959	93.14	0.0826
		20	0.5285	0.2274	93.66	0.0610	1.0953	0.3317	80.90	0.1220
		30	0.5461	0.2517	90.26	0.0746	1.1618	0.3803	57.62	0.1814
		40	0.5682	0.2842	85.25	0.0944	1.2466	0.4520	26.55	0.2625
	Joint	10	0.5085	0.2041	95.32	0.0506	1.0519	0.2967	90.72	0.0889
		20	0.5236	0.2233	94.16	0.0586	1.1070	0.3336	75.72	0.1319
		30	0.5412	0.2476	91.13	0.0715	1.1754	0.3845	50.48	0.1944
		40	0.5635	0.2805	86.25	0.0912	1.2633	0.4605	21.14	0.2790

*Cond = Conditional Sampling

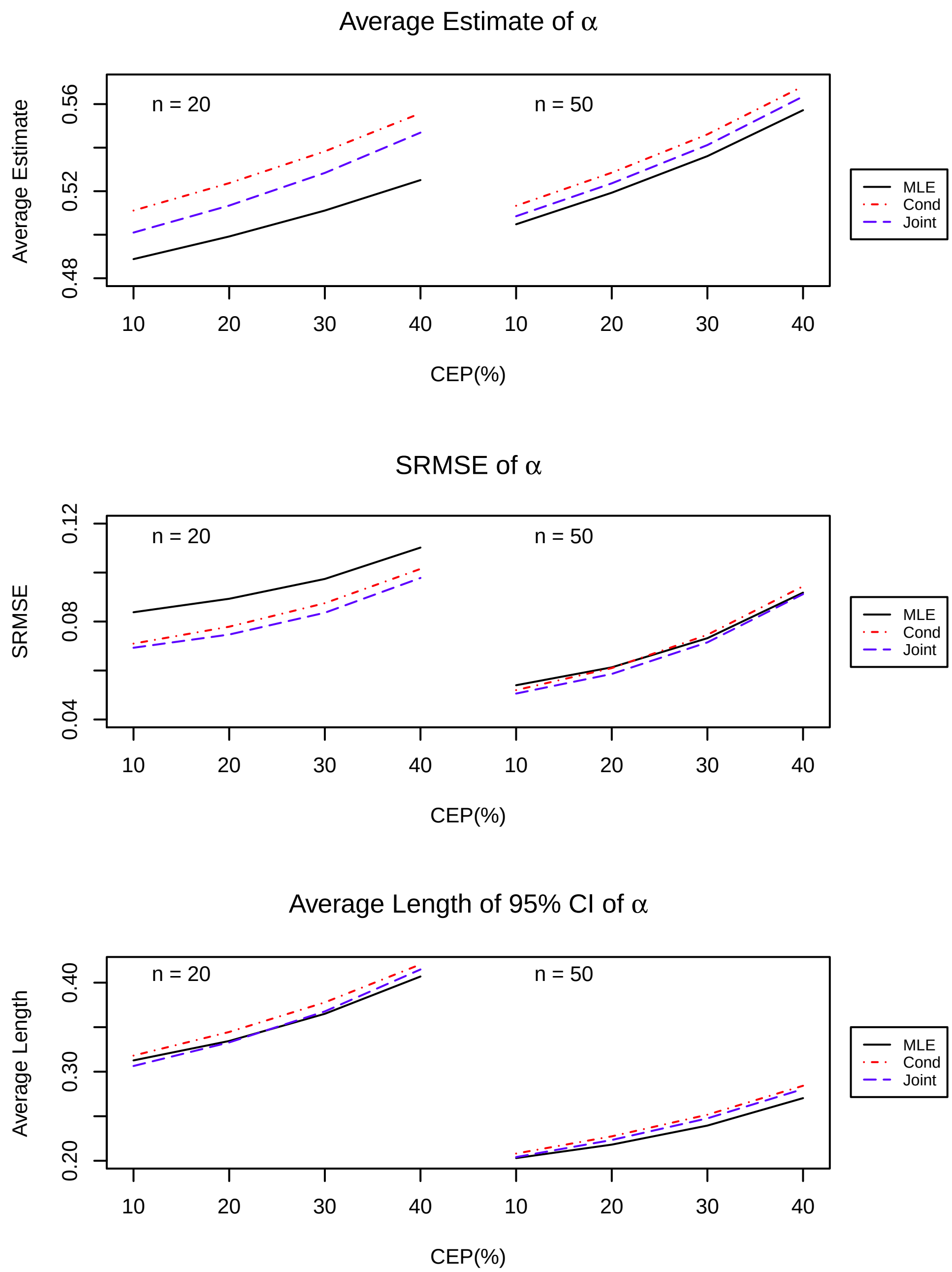

Figure 4.1: Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$)

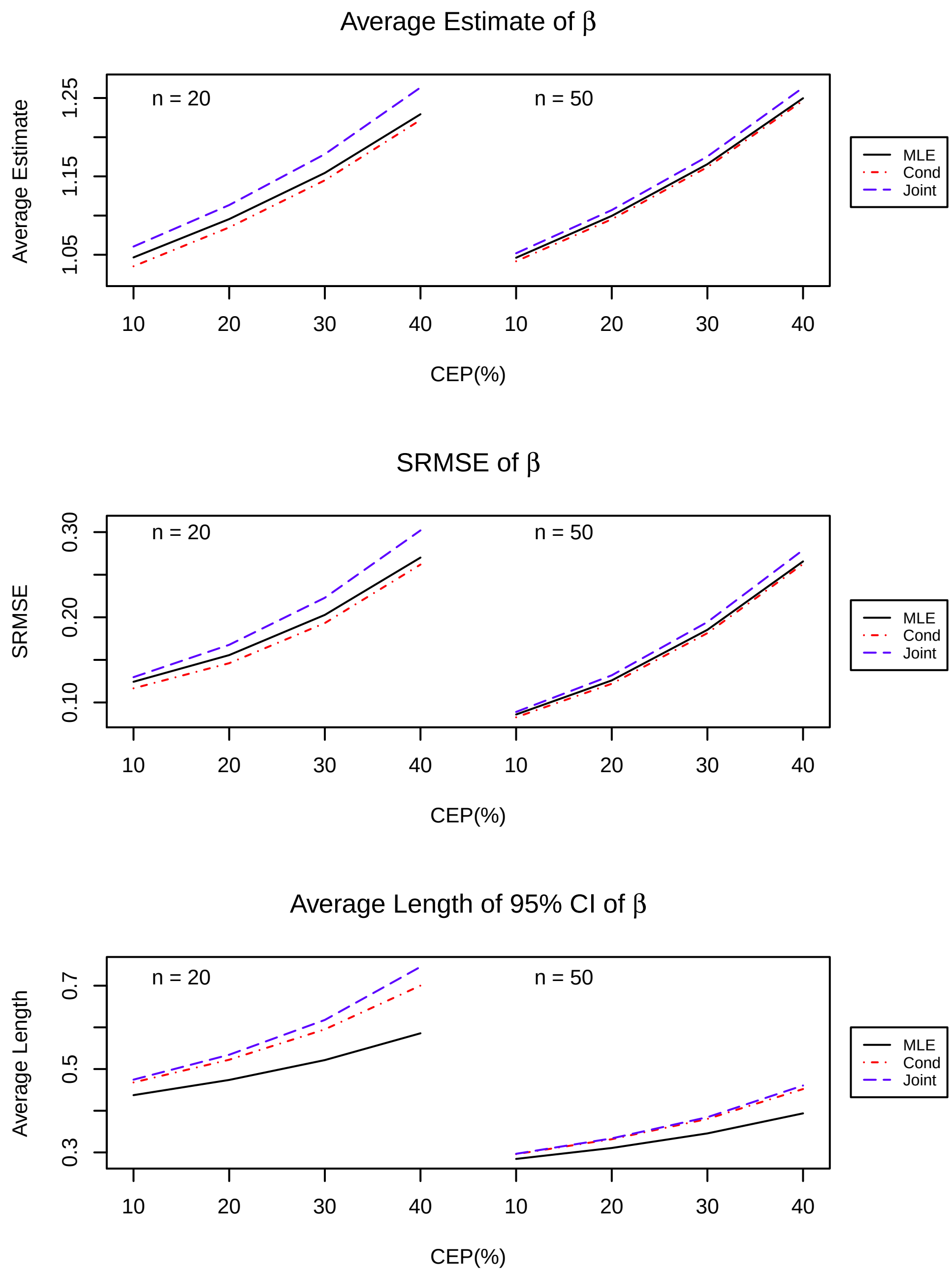

Figure 4.2: Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$)

Table 4.2: Simulation Results for Parameter Setting 2 ($\alpha = 1$, $\beta = 1$)

Parameter			α				β			
n	Method	CEP(%)	Mean	AL	CP(%)	SRMSE	Mean	AL	CP(%)	SRMSE
20	MLE	10	0.9762	0.6245	90.42	0.1701	1.0942	0.8191	93.73	0.2441
		20	0.9975	0.6680	91.56	0.1806	1.1860	0.9060	92.28	0.3099
		30	1.0248	0.7291	92.47	0.1999	1.3051	1.0288	87.55	0.4115
		40	1.0609	0.8185	93.35	0.2323	1.4634	1.2081	78.99	0.5594
	*Cond	10	0.9464	0.5925	91.82	0.1632	1.1059	0.8420	91.91	0.2437
		20	0.9638	0.6403	93.12	0.1674	1.1945	0.9674	87.94	0.3096
		30	0.9854	0.7025	94.23	0.1769	1.3077	1.1451	80.11	0.4075
		40	1.0126	0.7854	94.23	0.1944	1.4552	1.4080	68.14	0.5460
	Joint	10	0.9329	0.5829	90.07	0.1681	1.1850	0.9155	86.47	0.3012
		20	0.9535	0.6401	92.05	0.1718	1.2897	1.0823	78.56	0.3922
		30	0.9806	0.7216	93.33	0.1833	1.4292	1.3431	66.98	0.5236
		40	1.0190	0.8497	93.67	0.2092	1.6273	1.8238	51.22	0.7418
50	MLE	10	1.0077	0.4052	93.84	0.1075	1.0892	0.5276	93.07	0.1646
		20	1.0373	0.4352	94.34	0.1211	1.1912	0.5900	81.78	0.2446
		30	1.0751	0.4792	92.94	0.1487	1.3233	0.6793	57.64	0.3656
		40	1.1234	0.5430	89.57	0.1909	1.5038	0.8119	26.25	0.5415
	*Cond	10	0.9939	0.4044	94.36	0.1037	1.0953	0.5485	90.60	0.1670
		20	1.0214	0.4425	94.83	0.1127	1.1961	0.6395	77.67	0.2475
		30	1.0559	0.4935	93.16	0.1347	1.3258	0.7711	53.50	0.3670
		40	1.0991	0.5641	89.45	0.1694	1.5010	0.9739	25.13	0.5378
	Joint	10	0.9863	0.3989	93.81	0.1041	1.1293	0.5643	85.85	0.1918
		20	1.0150	0.4393	94.75	0.1118	1.2372	0.6653	68.31	0.2843
		30	1.0518	0.4945	93.26	0.1342	1.3783	0.8161	41.73	0.4175
		40	1.0995	0.5747	89.49	0.1723	1.5727	1.0570	15.77	0.6090

*Cond = Conditional Sampling

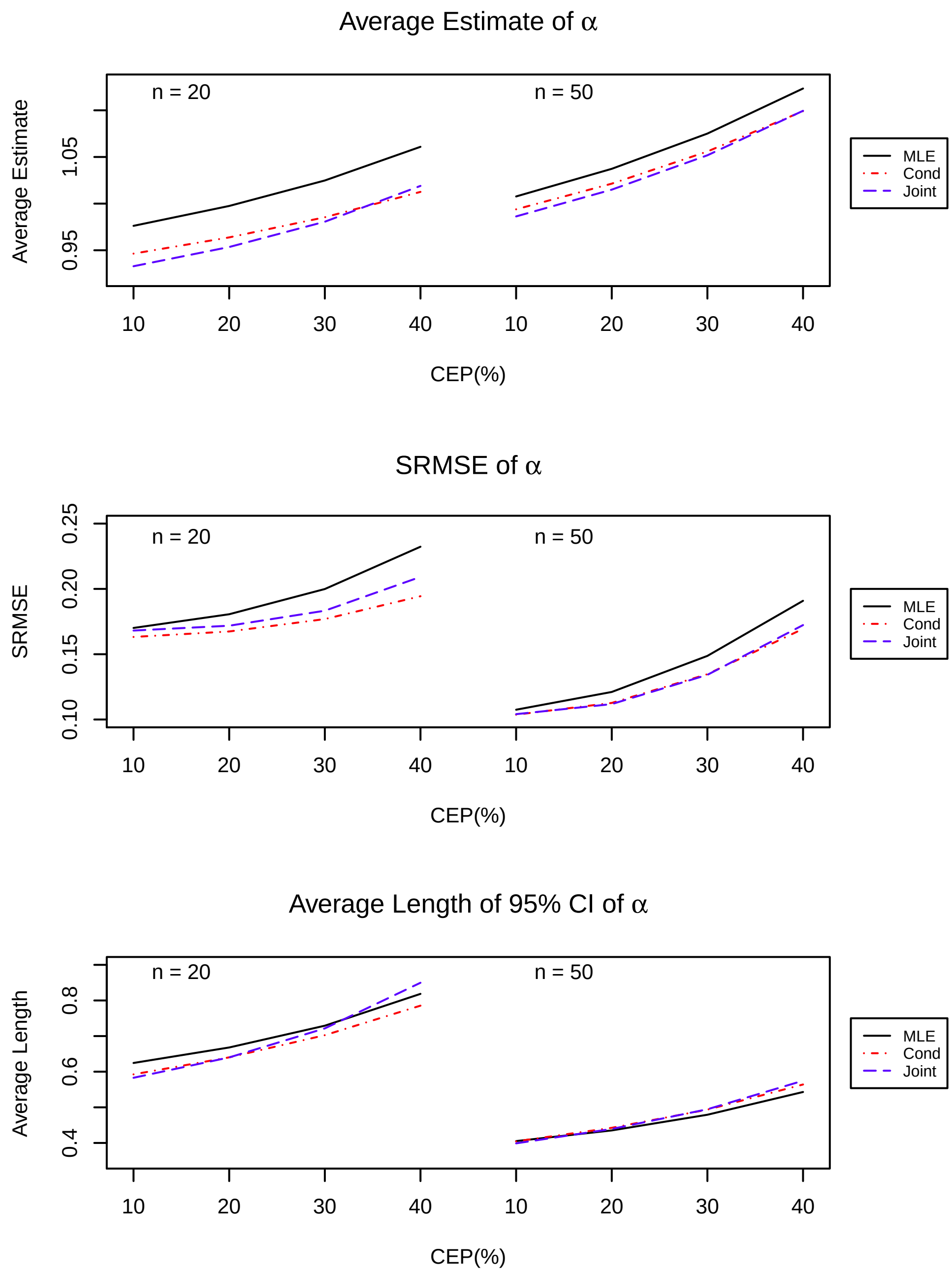

Figure 4.3: Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 2 ($\alpha = 1$, $\beta = 1$)

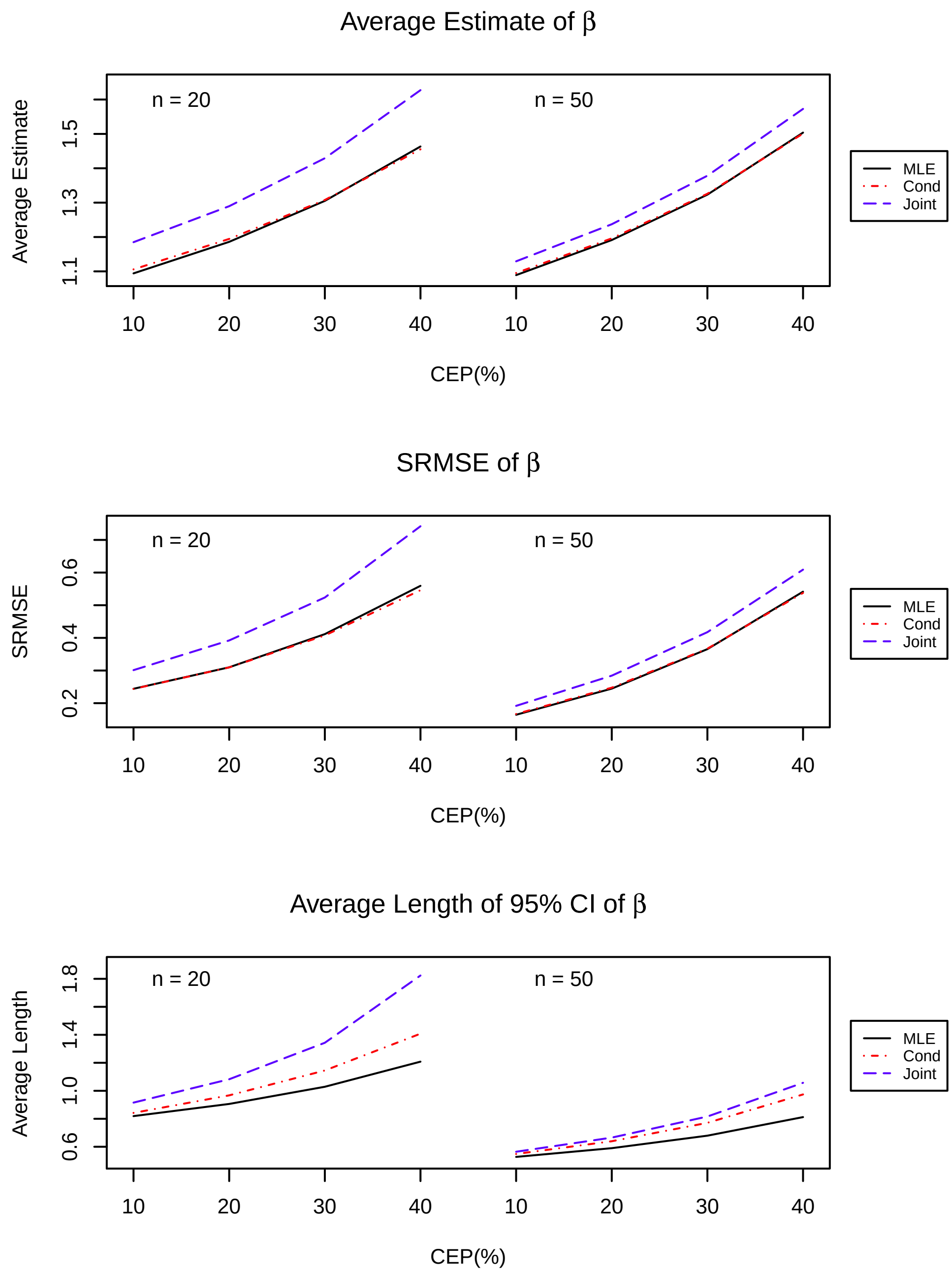

Figure 4.4: Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 2 ($\alpha = 1$, $\beta = 1$)

Table 4.3: Simulation Results for Parameter Setting 3 ($\alpha = 2$, $\beta = 1$)

Parameter			α				β			
n	Method	CEP(%)	Mean	AL	CP(%)	SRMSE	Mean	AL	CP(%)	SRMSE
20	MLE	10	1.9464	1.2440	90.19	0.3410	1.1570	1.2808	93.44	0.4041
		20	1.9909	1.3266	91.10	0.3646	1.2979	1.4457	93.43	0.5208
		30	2.0519	1.4480	92.07	0.4071	1.4859	1.6895	91.51	0.6976
		40	2.1294	1.6212	92.38	0.4771	1.7664	2.0803	85.65	0.9846
	*Cond	10	1.8488	1.1678	89.35	0.3487	1.2176	1.3525	89.80	0.4313
		20	1.8801	1.2654	90.64	0.3573	1.3553	1.6059	85.82	0.5520
		30	1.9205	1.3919	91.85	0.3753	1.5331	1.9608	78.57	0.7226
		40	1.9652	1.5580	92.29	0.4058	1.7867	2.5154	67.96	0.9827
	Joint	10	1.8393	1.1915	88.95	0.3559	1.4126	1.6506	80.50	0.5978
		20	1.8860	1.3387	91.08	0.3648	1.6058	2.0774	71.45	0.7865
		30	1.9529	1.5689	92.40	0.3930	1.8808	2.8299	59.95	1.0713
		40	2.0467	1.9856	93.33	0.4695	2.3726	4.9964	44.87	2.6583
50	MLE	10	2.0156	0.8095	93.78	0.2147	1.1331	0.8028	93.95	0.2547
		20	2.0785	0.8685	94.12	0.2457	1.2830	0.9147	85.33	0.3786
		30	2.1569	0.9537	91.85	0.3029	1.4885	1.0854	64.78	0.5721
		40	2.2651	1.0803	87.35	0.3982	1.7875	1.3556	34.92	0.8705
	*Cond	10	1.9734	0.8073	93.91	0.2100	1.1613	0.8597	90.04	0.2712
		20	2.0300	0.8879	94.58	0.2277	1.3110	1.0394	77.86	0.3999
		30	2.0984	0.9954	93.22	0.2685	1.5131	1.3074	57.24	0.5922
		40	2.1900	1.1535	89.98	0.3405	1.8023	1.7446	30.77	0.8814
	Joint	10	1.9632	0.8044	93.51	0.2116	1.2419	0.9262	82.38	0.3360
		20	2.0259	0.8961	94.48	0.2287	1.4136	1.1423	64.48	0.4945
		30	2.1049	1.0250	93.07	0.2754	1.6527	1.4854	40.09	0.7284
		40	2.2166	1.2351	89.37	0.3667	2.0102	2.1022	16.80	1.0901

*Cond = Conditional Sampling

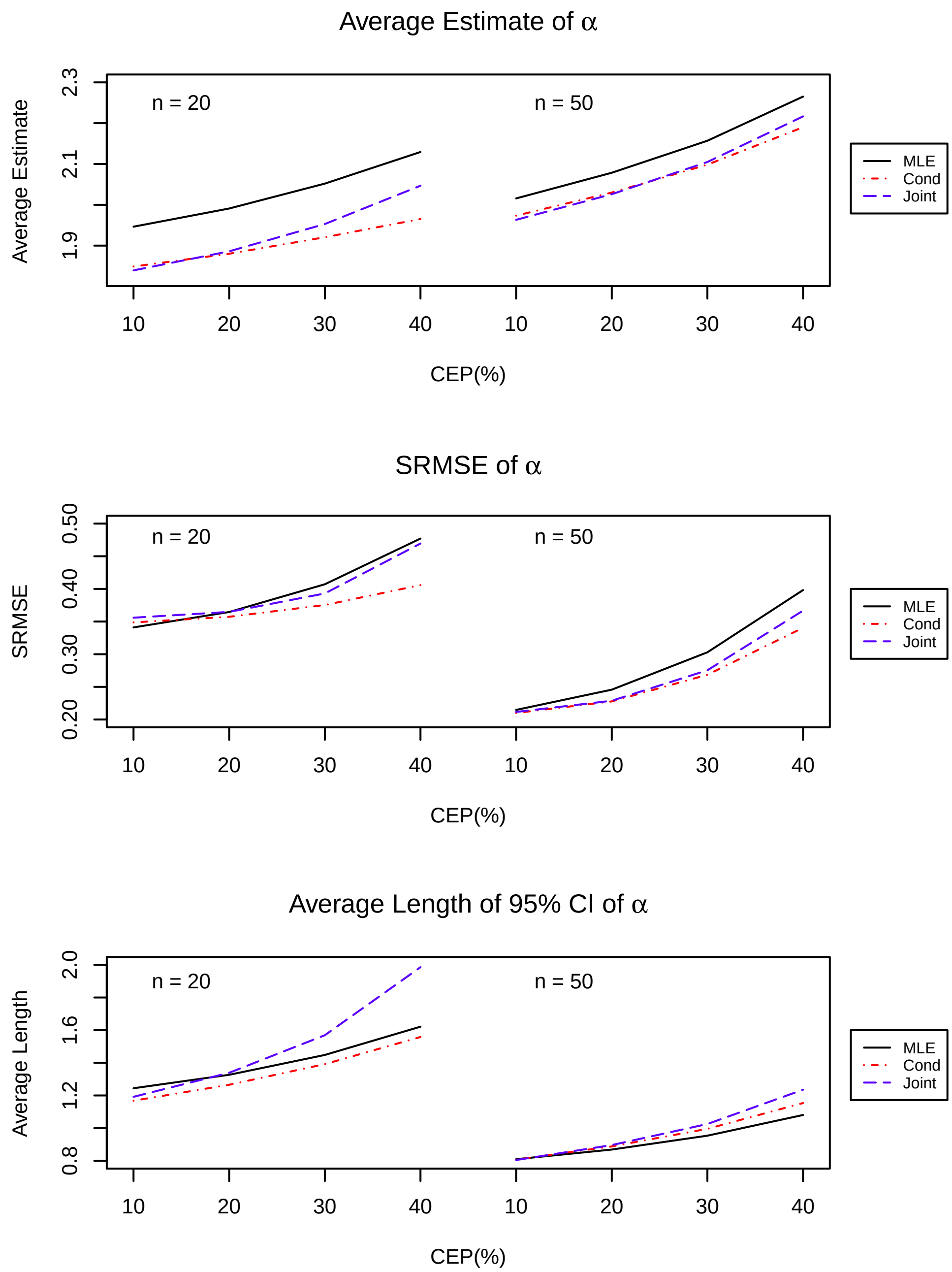

Figure 4.5: Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 3 ($\alpha = 2$, $\beta = 1$)

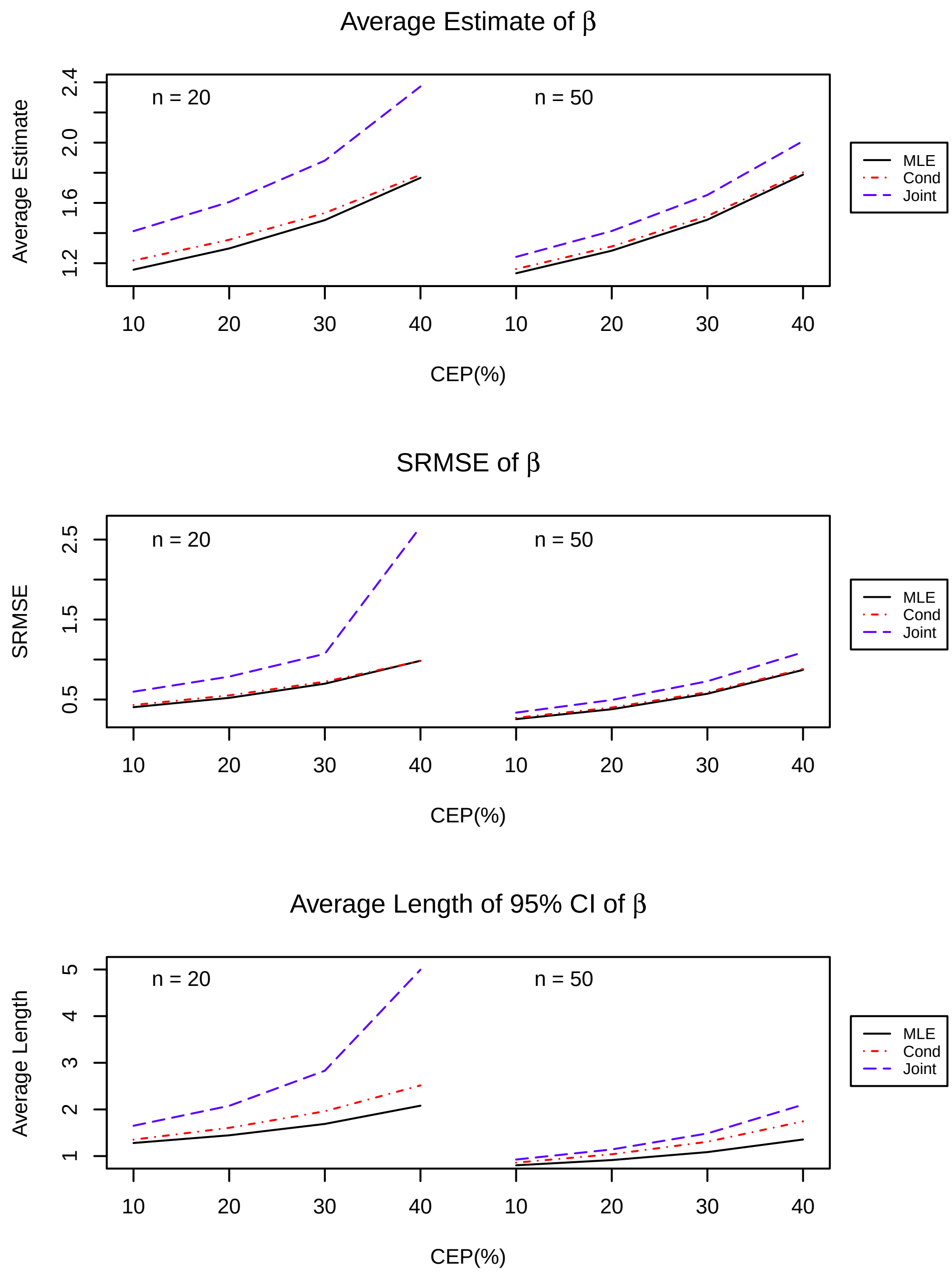

Figure 4.6: Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 3 ($\alpha = 2$, $\beta = 1$)

4.2 The Generalized BS Distribution

We perform several simulation examples to assess performance of the ML and Bayesian estimators. Recall that for the GBS distribution, β is no longer a scale parameter. In these simulation examples, we use the following three parameter settings:

$$(\alpha, \beta, \kappa) = \{(0.5, 1, 0.5), (1, 5, 0.8), (2, 5, 0.2)\},$$

under 4 random right censoring percentages (CEP) at 10%, 20%, 30% and 40% for two sample sizes, $n = 20$ and 50.

For each of these 24 parameter settings, we generate 10,000 data sets, and apply the MLE and conditional sampling algorithms, as explained in Chapter 3, to obtain the point estimates and interval estimates of the parameters α, β and κ for these simulated data sets. Then, we compute the average point estimates, average length (AL) of the 95% credible intervals (for conditional sampling) and average length (AL) of the 95% confidence intervals (for MLE), coverage probability (CP), and square root of mean squared error (SRMSE) of the parameters.

Here, we only consider conditional sampling as our Bayesian method. We adopt the same hyperparameters $a_0 = a_1 = b_0 = b_1 = 5$, as well as $d_0 = d_1 = 1$, which corresponds to a standard uniform prior for κ, under the assumption that we have little prior knowledge about the parameters. The Gibbs sampler has 8,000 iterations with a burn-in period of 2,000. The tuning parameters γ_β and γ_κ for the proposal distributions are calibrated to ensure acceptance rates of the Metropolis-Hastings steps to be around 40% to 47%, as suggested by Gelman et al. (2004).

We separate the simulation results into three tables, namely Table 4.1 for parameter setting ($\alpha = 0.5, \beta = 1, \kappa = 0.5$), Table 4.2 for parameter setting ($\alpha = 1, \beta = 5, \kappa = 0.8$), and Table 4.3 for parameter setting ($\alpha = 2, \beta = 5, \kappa = 0.2$). Each table contains results for different CEP and sample sizes. To aid visualization, we also plot the average estimates, AL of 95% CI and SRMSE for α, β and κ under different CEP and sample sizes for each of the 3 parameter settings.

From these tables and plots, some features can be summarized as follows:

1. Bias, SRMSE and average length (AL) of 95% CI decrease with sample size n but increase with censoring percentage (CEP).

2. Note that 95% asymptotic confidence interval is obtained from the formula: point estimate $\pm 1.96 \times \sqrt{\text{asymptotic variance}}$. This may result in confidence interval covering values outside parameter space. For example, some 95% confidence intervals for κ may have covered values outside the range $(0, 1)$, thus explaining the fact that some reported average length of 95% confidence intervals of κ are longer than 1. The 95% credible intervals obtained from the Bayesian method will be inside parameter space due to the choice of priors and proposals.

3. Under the parameter setting $(\alpha = 1, \beta = 5, \kappa = 0.8)$, MLE tends to overestimate both α and β while underestimating κ. More simulations could be done to ascertain if MLE consistently produces such biases when κ is relatively larger. Our Bayesian approach obtains less bias for all three parameters.

4. Overall, even with relatively "less" informative priors, our Bayesian approach outperforms MLE for all parameter settings in terms of bias, SRMSE and 95% CI width, especially when sample size is small.

Table 4.4: Simulation Results for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$)

		Method	MLE				Bayesian			
n	*Pm	CEP(%)	10	20	30	40	10	20	30	40
20	α	Mean	0.4741	0.4836	0.4953	0.5104	0.5056	0.5182	0.5333	0.5521
		AL	0.3205	0.3498	0.3898	0.4474	0.3227	0.3502	0.3851	0.4323
		CP(%)	88.64	90.49	91.86	93.69	97.67	97.92	97.37	96.85
		SRMSE	0.0882	0.0919	0.0993	0.1124	0.0699	0.0756	0.0849	0.0988
	β	Mean	1.0483	1.0971	1.1552	1.2284	1.0362	1.0855	1.1465	1.2232
		AL	0.4856	0.5322	0.5914	0.6739	0.4850	0.5360	0.6053	0.7064
		CP(%)	92.94	91.40	86.92	79.54	95.24	93.40	87.79	77.04
		SRMSE	0.1361	0.1659	0.2123	0.2769	0.1212	0.1491	0.1967	0.2649
	κ	Mean	0.4911	0.4861	0.4856	0.4880	0.4916	0.4893	0.4820	0.4823
		AL	0.9563	1.0308	1.1256	1.2469	0.7163	0.7375	0.7515	0.7733
		CP(%)	96.13	97.16	97.34	97.59	99.04	99.30	99.44	99.64
		SRMSE	0.2155	0.2251	0.2422	0.2570	0.1331	0.1299	0.1260	0.1204
50	α	Mean	0.4978	0.5115	0.5275	0.5482	0.5093	0.5244	0.5421	0.5648
		AL	0.2087	0.2293	0.2563	0.2952	0.2097	0.2302	0.2572	0.2949
		CP(%)	93.68	95.17	95.73	95.90	95.77	94.85	92.31	87.71
		SRMSE	0.0536	0.0594	0.0701	0.0877	0.0507	0.0591	0.0724	0.0924
	β	Mean	1.0487	1.1025	1.1682	1.2515	1.0428	1.0964	1.1624	1.2468
		AL	0.3217	0.3562	0.4009	0.4631	0.3196	0.3548	0.4018	0.4696
		CP(%)	92.82	83.76	66.53	42.73	93.31	83.47	63.49	34.19
		SRMSE	0.0947	0.1339	0.1922	0.2711	0.0887	0.1267	0.1848	0.2648
	κ	Mean	0.4939	0.4909	0.4904	0.4901	0.5002	0.4990	0.5004	0.5011
		AL	0.5595	0.5995	0.6487	0.7096	0.5286	0.5564	0.5868	0.6189
		CP(%)	95.69	96.19	96.29	96.10	96.58	97.27	97.48	97.76
		SRMSE	0.1330	0.1412	0.1510	0.1650	0.1206	0.1245	0.1277	0.1316

*Pm = Parameter

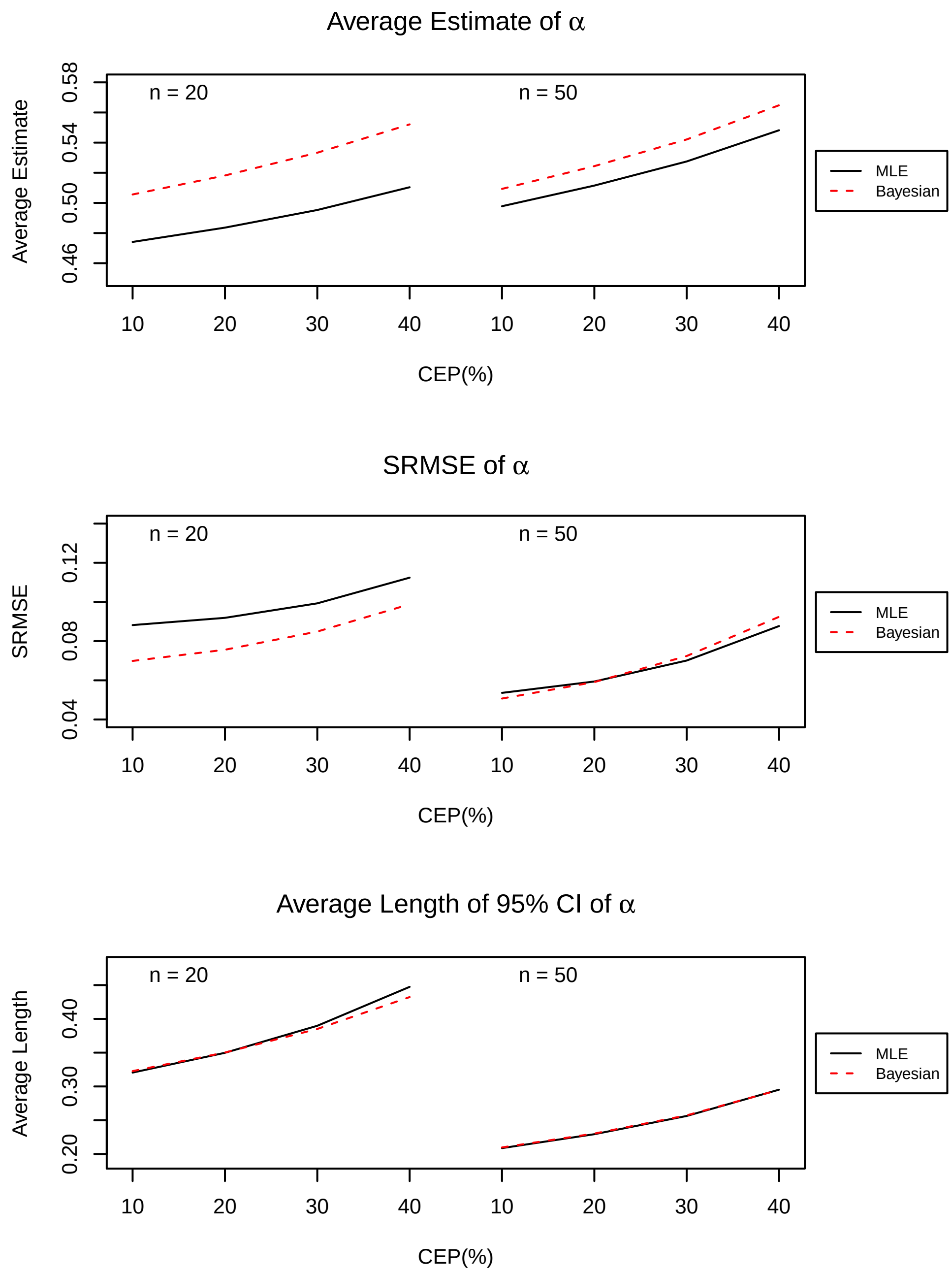

Figure 4.7: Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$)

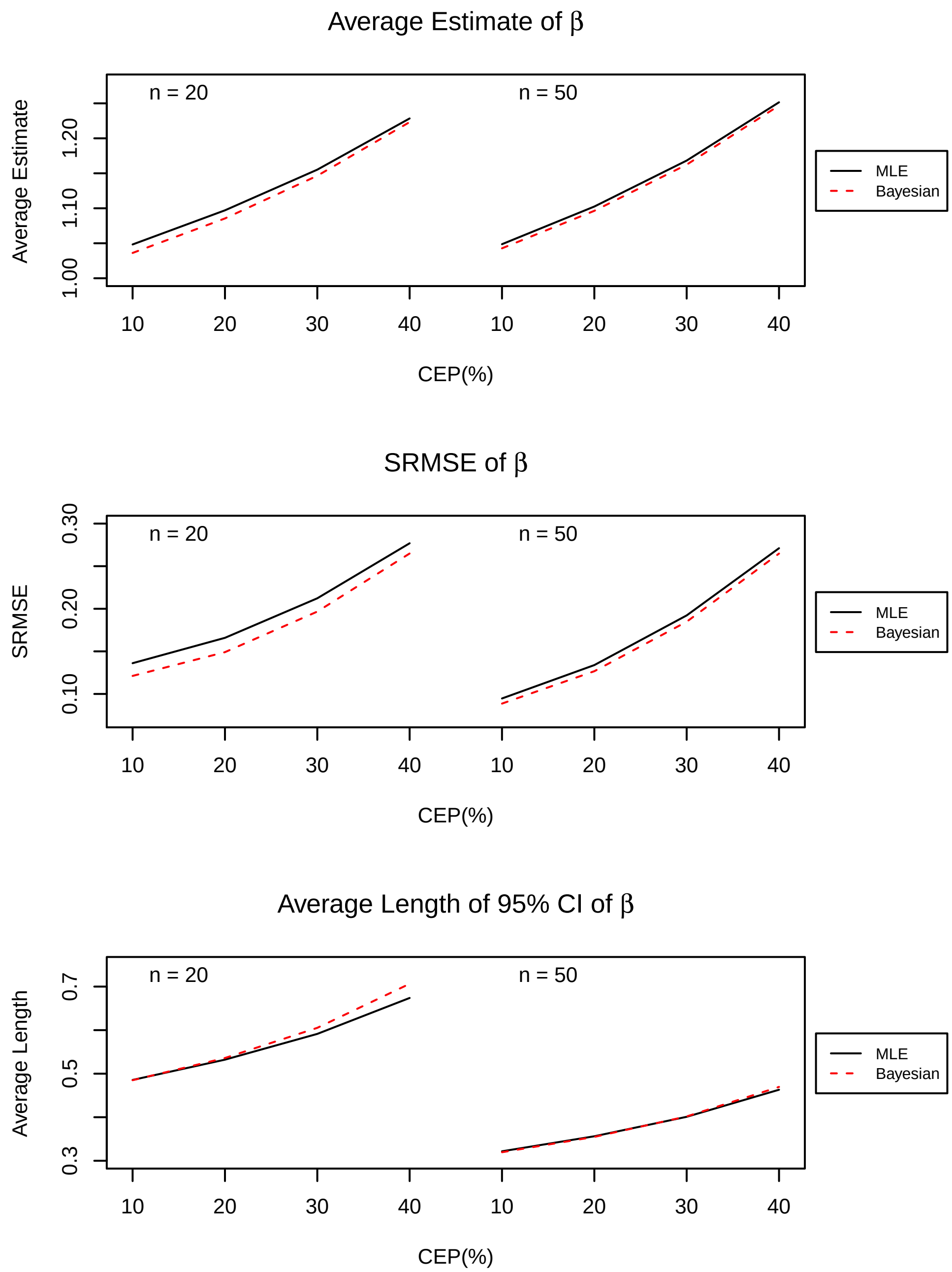

Figure 4.8: Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$)

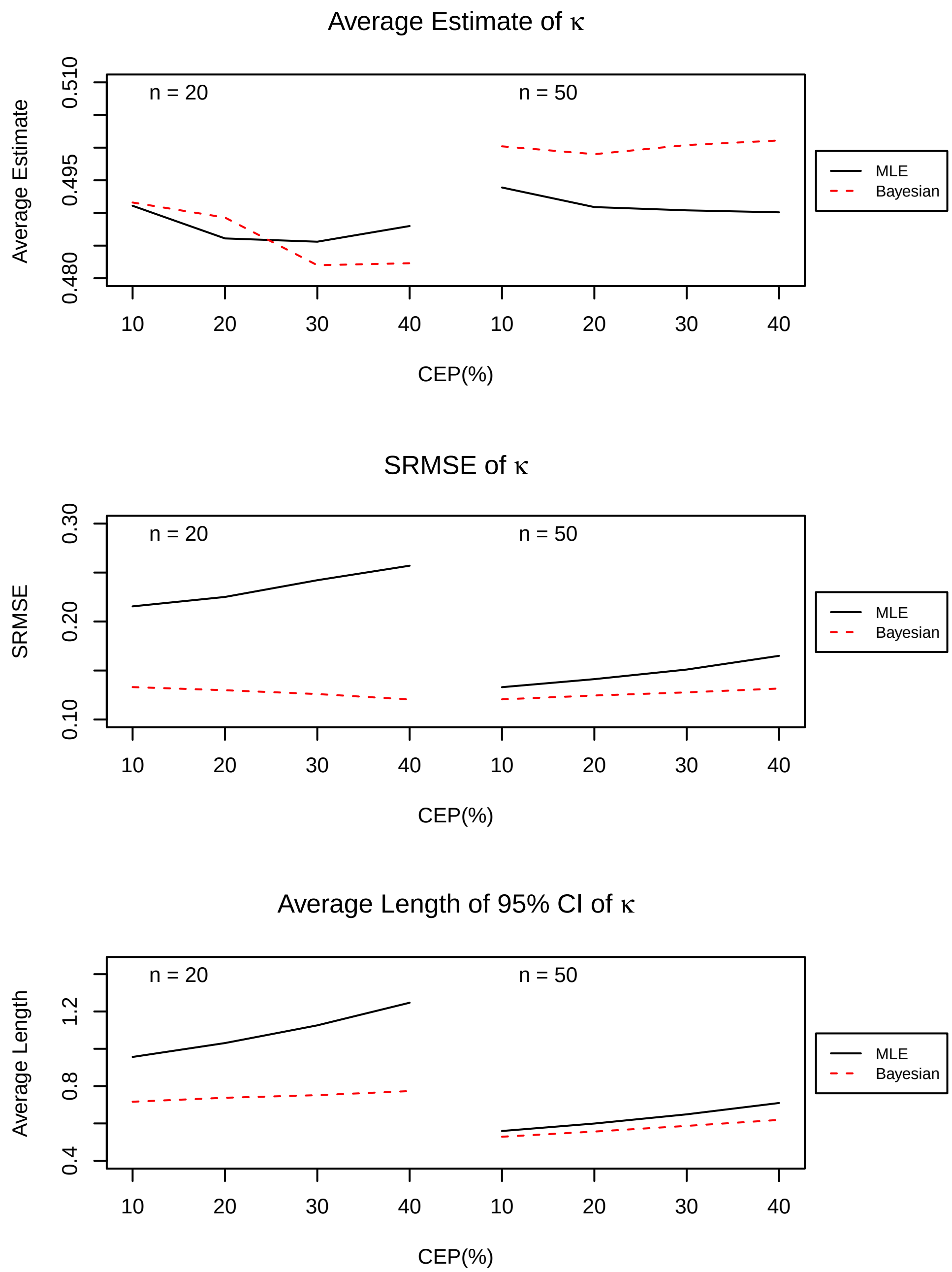

Figure 4.9: Average Estimates, SRMSE and Average Length of 95% CI of κ for Parameter Setting 1 ($\alpha = 0.5$, $\beta = 1$, $\kappa = 0.5$)

Table 4.5: Simulation Results for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$)

		Method	MLE				Bayesian			
n	*Pm	CEP(%)	10	20	30	40	10	20	30	40
20	α	Mean	1.1646	1.3004	1.4830	1.7343	1.1032	1.1794	1.2661	1.3703
		AL	1.6427	2.0356	2.6115	3.4710	1.2077	1.3738	1.5763	1.8308
		CP(%)	95.37	97.16	98.18	98.92	98.93	98.47	97.58	95.90
		SRMSE	0.4575	0.6013	0.8153	1.1246	0.2540	0.3097	0.3827	0.4778
	β	Mean	7.1976	8.9611	11.7815	16.7003	5.6410	6.4548	7.5050	8.9949
		AL	14.5686	21.1321	32.7303	54.9316	9.5989	12.1196	15.6025	20.8502
		CP(%)	95.75	98.18	99.27	99.65	97.24	98.28	97.50	94.91
		SRMSE	4.8881	7.4574	12.0429	20.1768	2.1679	2.8873	3.9751	5.6452
	κ	Mean	0.7544	0.7349	0.7131	0.6871	0.7837	0.7753	0.7682	0.7607
		AL	0.3405	0.3856	0.4418	0.5103	0.2624	0.2853	0.3103	0.3381
		CP(%)	95.33	94.72	93.59	91.36	97.73	97.99	97.97	97.99
		SRMSE	0.0972	0.1149	0.1369	0.1644	0.0632	0.0692	0.0753	0.0838
50	α	Mean	1.1207	1.2515	1.4388	1.7126	1.0988	1.1998	1.3308	1.4971
		AL	0.9492	1.1887	1.5623	2.1702	0.8386	1.0018	1.2303	1.5431
		CP(%)	97.68	98.62	98.68	98.18	96.50	92.83	84.99	71.68
		SRMSE	0.2668	0.3839	0.5743	0.8700	0.2095	0.2915	0.4143	0.5795
	β	Mean	6.3284	7.7865	10.1616	14.4751	5.8219	6.9071	8.4887	10.9294
		AL	7.4256	10.7689	16.9587	29.7127	6.4988	8.8249	12.6252	19.0087
		CP(%)	98.22	99.64	99.91	99.85	95.79	90.70	76.90	53.55
		SRMSE	2.3597	3.9109	6.6676	11.8703	1.7187	2.7319	4.3779	7.0172
	κ	Mean	0.7757	0.7586	0.7364	0.7081	0.7890	0.7780	0.7653	0.7515
		AL	0.1938	0.2207	0.2560	0.3014	0.1726	0.1914	0.2150	0.2422
		CP(%)	95.36	93.49	89.71	84.45	96.62	95.86	94.21	92.22
		SRMSE	0.0541	0.0682	0.0885	0.1162	0.0419	0.0493	0.0596	0.0729

*Pm = Parameter

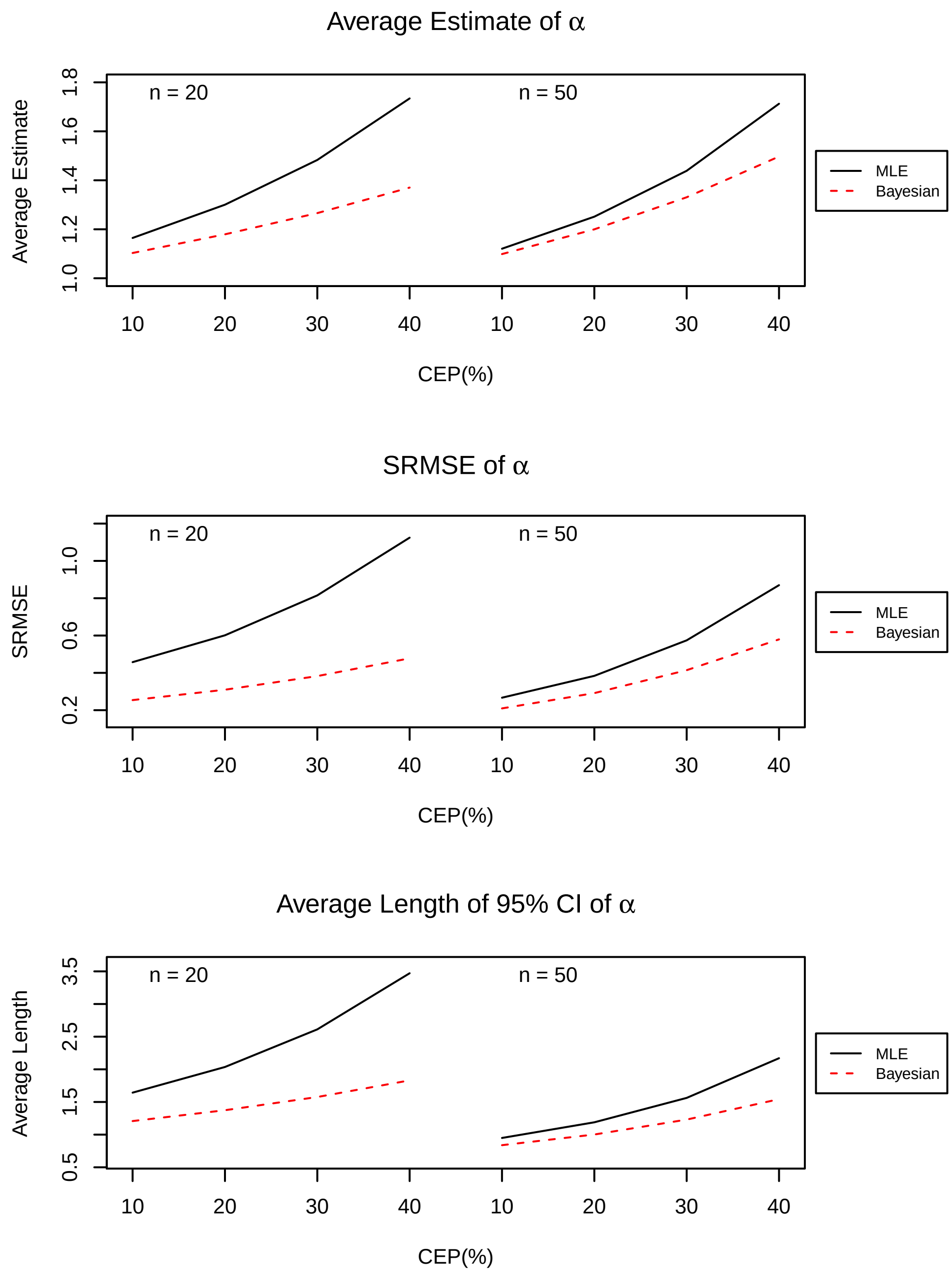

Figure 4.10: Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$)

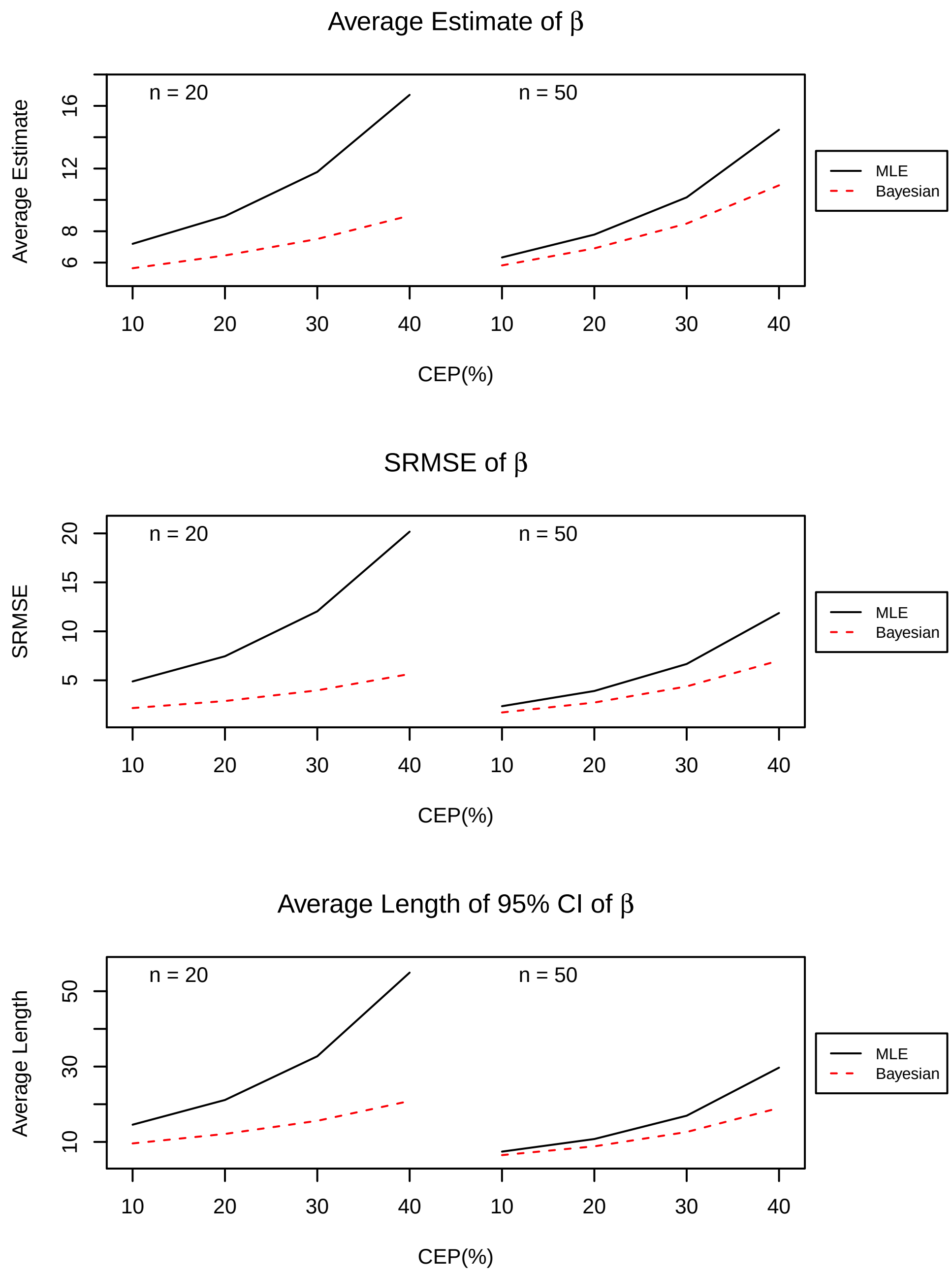

Figure 4.11: Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$)

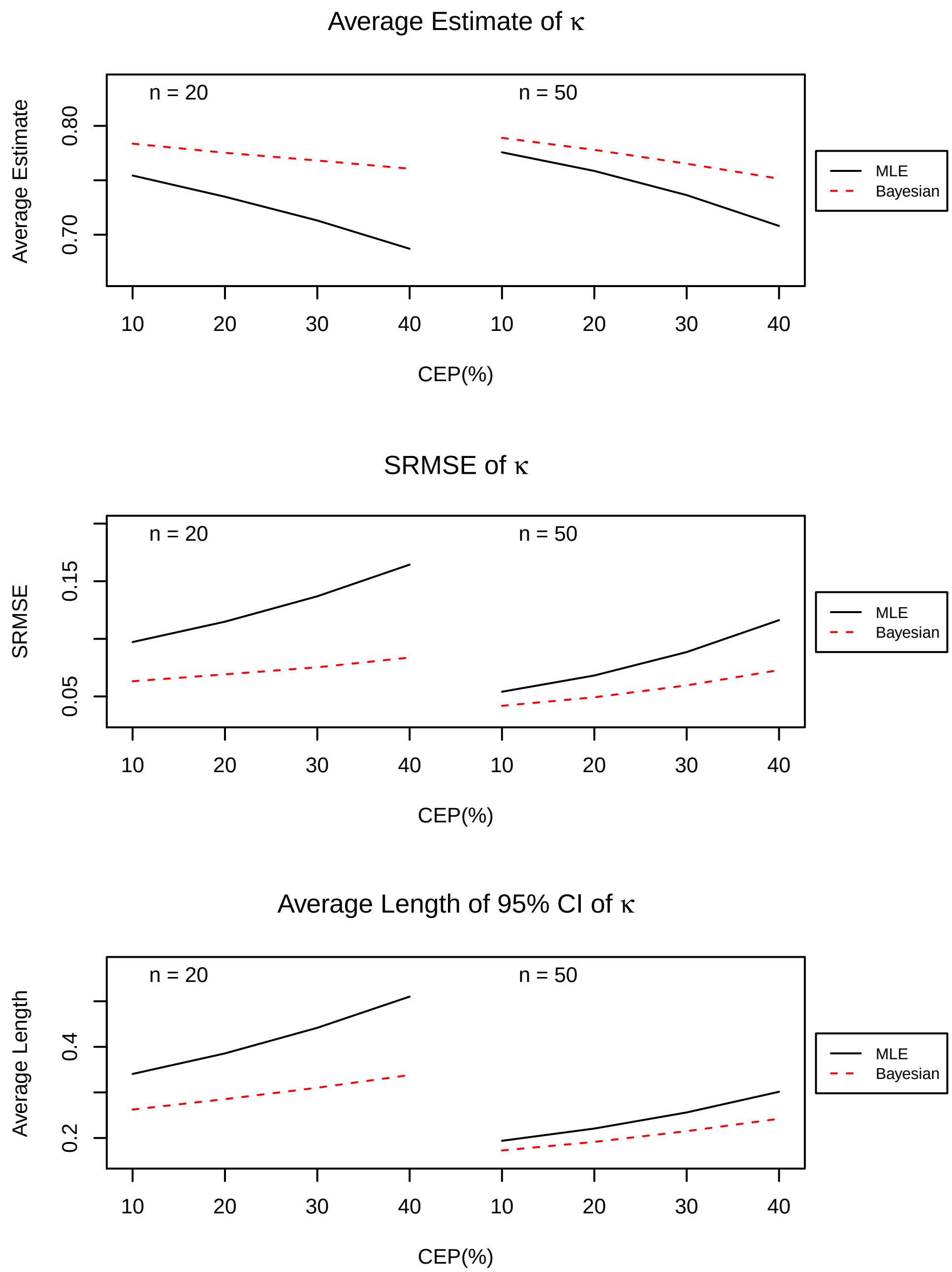

Figure 4.12: Average Estimates, SRMSE and Average Length of 95% CI of κ for Parameter Setting 2 ($\alpha = 1$, $\beta = 5$, $\kappa = 0.8$)

Table 4.6: Simulation Results for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$)

n	*Pm	Method	MLE				Bayesian			
		CEP(%)	10	20	30	40	10	20	30	40
20	α	Mean	1.8482	1.8521	1.8553	1.8653	1.8101	1.8142	1.8184	1.8303
		AL	1.3058	1.3766	1.4674	1.5957	1.2160	1.2748	1.3495	1.4516
		CP(%)	86.48	86.80	87.12	87.84	86.77	86.51	87.13	88.07
		SRMSE	0.4061	0.4259	0.4509	0.4836	0.4070	0.4199	0.4343	0.4495
	β	Mean	5.4049	5.9894	6.6844	7.5644	4.9604	5.4408	6.0163	6.7359
		AL	5.8884	6.4589	7.1767	8.1791	5.4311	6.0066	6.7627	7.8288
		CP(%)	92.62	91.39	87.29	80.84	95.11	95.83	93.86	89.97
		SRMSE	1.5724	1.8921	2.4005	3.1405	1.3320	1.4870	1.8287	2.3796
	κ	Mean	0.2293	0.2244	0.2215	0.2195	0.2681	0.2679	0.2700	0.2736
		AL	0.3519	0.3698	0.3933	0.4256	0.3452	0.3633	0.3855	0.4137
		CP(%)	95.65	95.47	94.73	94.42	90.65	91.45	91.39	92.37
		SRMSE	0.0957	0.0991	0.1060	0.1156	0.1204	0.1250	0.1333	0.1423
50	α	Mean	1.9544	1.9724	1.9935	2.0218	1.9425	1.9605	1.9817	2.0102
		AL	0.8382	0.8847	0.9442	1.0242	0.8193	0.8633	0.9198	0.9953
		CP(%)	91.84	92.78	93.20	94.03	92.16	93.09	92.85	93.46
		SRMSE	0.2297	0.2382	0.2543	0.2757	0.2307	0.2377	0.2514	0.2695
	β	Mean	5.4927	6.1477	6.9505	7.9653	5.3207	5.9322	6.6838	7.6331
		AL	3.8895	4.2911	4.7939	5.4800	3.7713	4.1841	4.7070	5.4342
		CP(%)	93.11	85.08	66.83	42.48	94.71	88.30	72.01	46.25
		SRMSE	1.0921	1.5411	2.2372	3.1877	0.9846	1.3559	1.9869	2.8669
	κ	Mean	0.2058	0.1987	0.1916	0.1843	0.2186	0.2130	0.2074	0.2020
		AL	0.1956	0.2019	0.2095	0.2191	0.1951	0.2020	0.2100	0.2202
		CP(%)	95.15	94.36	93.01	90.87	93.76	94.20	93.80	93.24
		SRMSE	0.0507	0.0513	0.0547	0.0588	0.0569	0.0569	0.0596	0.0629

*Pm = Parameter

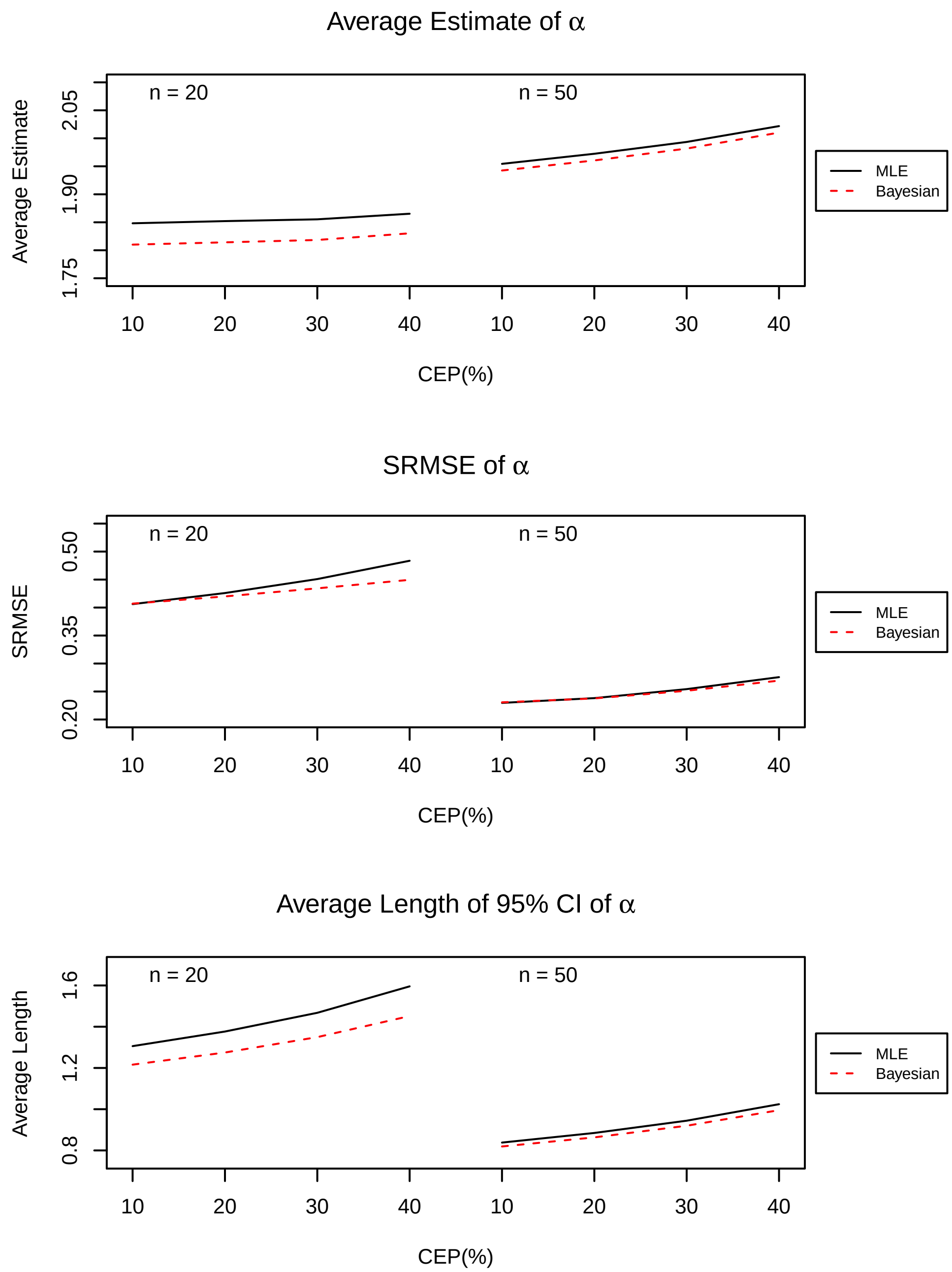

Figure 4.13: Average Estimates, SRMSE and Average Length of 95% CI of α for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$)

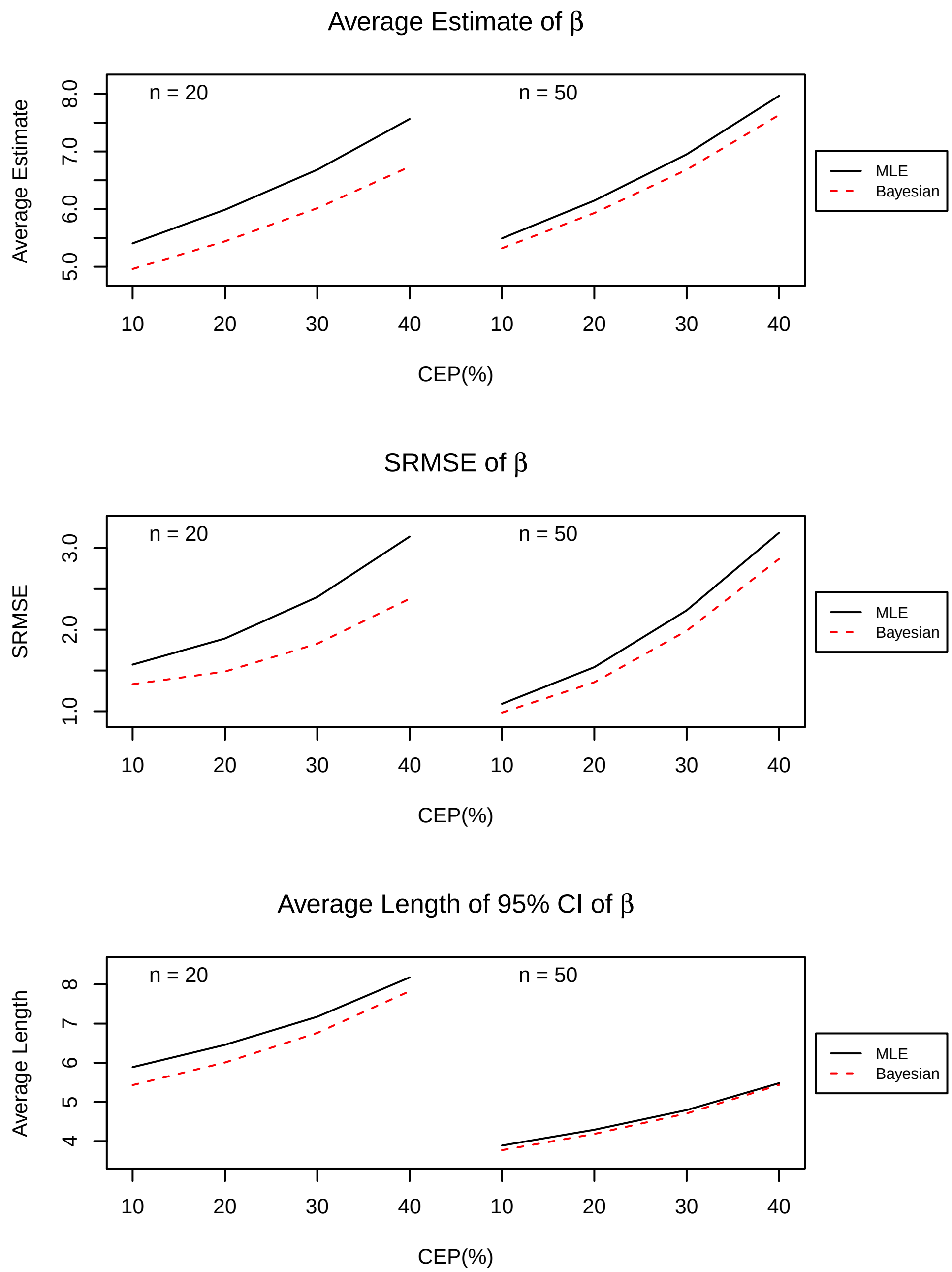

Figure 4.14: Average Estimates, SRMSE and Average Length of 95% CI of β for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$)

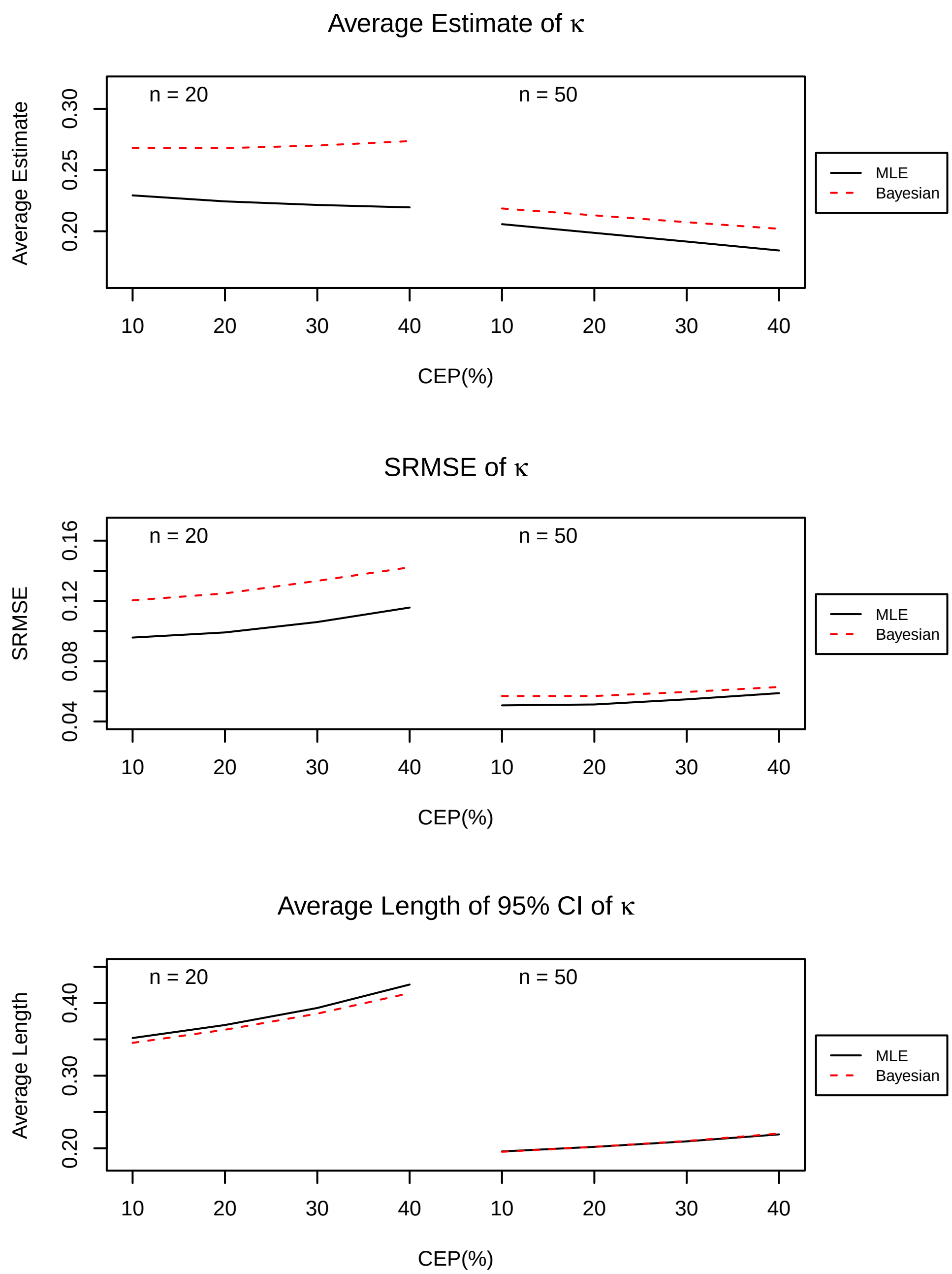

Figure 4.15: Average Estimates, SRMSE and Average Length of 95% CI of κ for Parameter Setting 3 ($\alpha = 2$, $\beta = 5$, $\kappa = 0.2$)

Chapter 5

Real Data Analyses

In this chapter, we will analyze two real data sets.

5.1 Data Set 1

Table 5.1: Lifetimes (in months) of 20 cancer patients receiving a new treatment

3	5	6	7	8
9	10	10+	12	15
15+	18	19	20	22
25	28	30	40	45

We first analyze the lifetimes (in months) of 20 cancer patients receiving a new treatment, where the symbol + denotes a right-censored observation. This data can be found in Achcar and Moala (2010), and is presented in Table 5.1. Assuming an underlying GBS distribution for the data, we adopt the same prior distributions as discussed in Chapter 3 with the following hyperparameters:

$$a_0 = 10$$

$$a_1 = 19$$

$$b_0 = 10$$

$$b_1 = 0.083$$

$$d_0 = d_1 = 1.$$

Here, we explain the reasons behind the choices for our hyperparameters. The standard uniform prior is adopted for κ as we assume we do not have much prior information on κ. Since β is the median of the GBS distribution, the hyperparameters b_0 and b_1 are chosen such that the prior mean of β is close to the sample median of the data. Assuming we also do not have much prior knowledge on α, we choose a_0 and a_1 such that the conditional prior mean of $(\alpha^2|\beta = 15)$ equals to 1, which is close to $\widehat{\alpha}^2_{\text{MLE}}$. Since the coefficient of variation (CV) for a standard uniform distribution is $\frac{1}{\sqrt{3}}$, we specify the hyperparameters such that the inverse gamma priors also have CVs close to $\frac{1}{\sqrt{3}}$.

Maximum likelihood and conditional sampling algorithms, as explained in Chapter 3, are employed to estimate the parameters of interest for this data. The Gibbs sampler has an iteration size of 20,000 with a burn-in period of 5,000. To minimize the possible correlation between consecutive random numbers generated by the (pseudo) random number generator in the software, every 5th sample of the remaining 15,000 samples are chosen to estimate the posterior mean and 95% credible interval of the parameters. Convergence of the algorithm is monitored by trace plots of the simulated samples. The tuning parameters are fixed at $\gamma_\beta = 2.4$ and $\gamma_\kappa = 15$, such that the acceptance rates for the Metropolis-Hastings steps hover around 40%, as recommended by Gelman et al. (2004).

The results are tabulated in Table 5.2. The point estimates obtained by both methods are very close to one another. The 95% credible intervals produced by the Bayesian method are slightly narrower than the corresponding 95% confidence intervals obtained from ML estimation, due to the relatively small sample size in this data set. Our results are also pretty close to the posterior estimates obtained by Achcar and Moala (2010), who fitted a BS distribution for this data. Recall that $BS(\alpha, \beta) = GBS(\alpha, \beta, \frac{1}{2})$, and our 95% credible interval for κ includes $\frac{1}{2}$. The posterior estimates $\widehat{\alpha}_{\text{Gibbs}} = 0.8854$ and $\widehat{\beta}_{\text{Gibbs}} = 16.03$ obtained by Achcar and Moala (2010) also lie within our corresponding 95% credible interval for α and β. Our 95% credible interval for α is slightly wider than the one obtained by

Achcar and Moala (2010), as we adopt a "flatter" prior for α^2. On the other hand, our 95% credible interval for β is narrower than the one obtained by Achcar and Moala (2010), since they specified a flat prior $U(0, 500)$ for β.

Using the ML estimates and posterior estimates, the fitted reliability curves are compared to the Kaplan-Meier estimates. These reliability curves, as well as the histograms for the simulated samples of the parameters from Gibbs sampling, are shown in Figure 5.1. It is worth noting that the two fitted reliability curves are very similar to the Kaplan-Meier plot. The posterior distributions for β and κ are relatively symmetric at their respective posterior mean, while the posterior distribution for α is slightly positively-skewed.

Table 5.2: Point and Interval Estimates for Data Set 1

Parameter	α		β		κ	
Method	MLE	Bayesian	MLE	Bayesian	MLE	Bayesian
Mean	0.9740	0.9619	15.6289	15.4105	0.4195	0.4558
95% CI lower bound	0.1273	0.6035	9.6137	10.4887	0.0833	0.2472
95% CI upper bound	1.8207	1.5103	21.6441	21.6960	0.7558	0.6736
Length of 95% CI	1.6934	0.9068	12.0304	11.2073	0.6726	0.4264

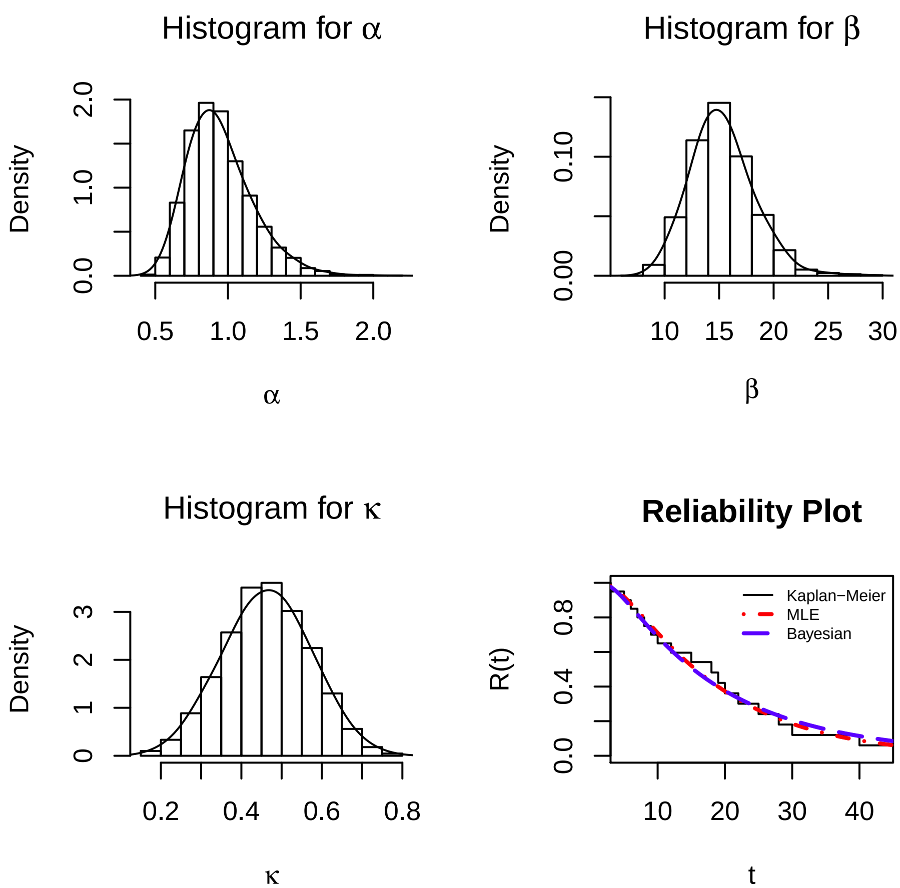

Figure 5.1: Histograms of the simulated samples for the parameters and the reliability plots

5.2 Data Set 2

Table 5.3: Fatigue life of 6061-T6 aluminium coupons exerted with maximum stress per cycle of 21,000 psi

370	706	716	746	785	797	844	855
858	886	886	930	960	988	990	1000
1010	1016	1018	1020	1055	1085	1102	1102
1108	1115	1120	1134	1140	1199	1200	1200
1203	1222	1235	1238	1252	1258	1262	1269
1270	1290	1293	1300	1310	1313	1315	1330
1355	1390	1416	1419	1420	1420	1450	1452
1475	1478	1481	1485	1502	1505	1513	1522
1522	1530	1540	1560	1567	1578	1594	1602
1604	1608	1630	1642	1674	1730	1750	1750
1763	1768	1781	1782	1792	1820	1868	1881
1890	1893	1895	1910	1923	1940	1945	2023
2100	2130	2215	2268	2440			

This data set is given by Birnbaum and Saunders (1969b) about the fatigue life of 6061-T6 aluminium coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set, as presented in Table 5.3, consists of 101 observations with maximum stress per cycle of 21,000 psi.

This is a complete data set. For the purpose of illustration, we artificially applied Type-II right censoring to the data with 4 different censoring thresholds $n = 90$, 80, 70 and 60. For instance, when $n = 90$, the first 90 samples are treated as fully observed data, and the remaining 11 observations are right censored at the value of the 90th sample.

Again, assuming an underlying GBS distribution for the data, we adopt the same prior distributions for the parameters, with the following hyperparameter values:

$$a_0 = 10$$
$$a_1 = 55$$
$$b_0 = 10$$
$$b_1 = 0.00088$$
$$d_0 = d_1 = 1.$$

Here, we explain the reasons behind the choices for our hyperparameters. Assuming little prior information about κ, the standard uniform prior is adopted for κ. Since β is the median of the GBS distribution, the hyperparameters b_0 and b_1 are chosen such that the prior mean of β is close to 1420, which is the sample median of the fully observed data (when $n = 101$). Assuming little prior knowledge about α, we choose a_0 and a_1 such that the conditional prior mean of $(\alpha^2|\beta = 1420)$ is close to $\widehat{\alpha}^2_{\mathrm{MLE}} \approx 5.7^2$ when $n = 101$. Since the coefficient of variation (CV) for a standard uniform distribution is $\frac{1}{\sqrt{3}}$, we specify the hyperparameters such that the inverse gamma priors also have CVs close to $\frac{1}{\sqrt{3}}$.

Maximum likelihood as well as conditional sampling algorithms, as explained in Chapter 3, are employed to estimate the parameters of interest for this data under different censoring thresholds. Again, the Gibbs sampler has an iteration size of 20,000 with a burn-in period of 5,000. To minimize the possible correlation between consecutive random numbers generated by the (pseudo) random number generator in the software, every 5th sample of the remaining 15,000 samples are chosen to estimate the posterior mean and 95% credible interval of the parameters. Convergence of the algorithm is monitored by trace plots of the simulated samples. The tuning parameters are fixed at $\gamma_\beta = 2.4$ and $\gamma_\kappa = 200$, such that the acceptance rates for the Metropolis-Hastings steps hover around 40%, as recommended by Gelman et al. (2004).

The results are tabulated in Table 5.4. The point estimates for β and κ obtained by both methods are very close to one another. Note that the asymptotic 95% confidence interval for MLE is calculated using the formula: point estimate $\pm 1.96 \times \sqrt{\text{asymptotic variance}}$. Using this formula, some of the calculated asymptotic confidence intervals may cover values beyond the parameter space. For instance, some confidence intervals for α may have covered negative values, and some of the confidence intervals for κ may have covered values outside the parameter space $(0, 1)$. The 95% credible intervals produced by the Bayesian method are within the corresponding parameter space, and are narrower than the corresponding 95% confidence intervals obtained from ML estimation.

We also compare our findings under the case of fully observed data (when $n = 101$) with the results obtained by Owen (2006), whose ML point estimates $\widehat{\alpha}_{\text{MLE}} = 6.605$, $\widehat{\beta}_{\text{MLE}} = 1393.42$ and $\widehat{\kappa}_{\text{MLE}} = 0.064$ are very close to ours. In fact, our numerical solutions for the normal equations (as shown in the Appendix section) are more accurate, since the values of the normal equations are much closer to zero by plugging in our ML estimates instead of his ML estimates. Owen (2006) did not proceed with interval estimation, nor did he consider the case where data is right-censored.

For each censoring threshold, the fitted reliability curves based on the ML estimates and posterior estimates are compared to the Kaplan-Meier plot. These reliability curves, as well as the histograms for the simulated samples of the parameters from Gibbs sampling, are shown. It is worth noting that, under each censoring threshold, the two fitted reliability curves are very similar to the Kaplan-Meier plot. The posterior distribution of α is slightly positively-skewed; the posterior distribution of β is rather symmetric at its posterior mean; and the posterior distribution of κ is slightly negatively-skewed.

Table 5.4: Point and Interval Estimates for Data Set 2

Parameter	Method	n	Mean	95% CI	CI Length
α	MLE	101	5.7112	(-4.0904 , 15.5127)	19.6031
		90	4.7668	(-4.3600 , 13.8937)	18.2537
		80	3.7136	(-3.9684 , 11.3956)	15.3640
		70	6.0901	(-7.6038 , 19.7841)	27.3879
		60	5.1077	(-7.3191 , 17.5344)	24.8535
	Bayesian	101	5.2754	(3.5531 , 8.0930)	4.5399
		90	5.3465	(3.4941 , 8.3382)	4.8441
		80	5.3613	(3.5583 , 8.5863)	5.0280
		70	5.2940	(3.5319 , 7.8951)	4.3632
		60	5.1435	(3.4069 , 8.0708)	4.6639
β	MLE	101	1391.1037	(1309.5219 , 1472.6856)	163.1637
		90	1391.0140	(1307.8489 , 1474.1791)	166.3302
		80	1392.3865	(1306.5400 , 1478.2330)	171.6930
		70	1384.8141	(1303.6873 , 1465.9408)	162.2535
		60	1389.0569	(1301.8535 , 1476.2604)	174.4069
	Bayesian	101	1387.7813	(1309.7578 , 1467.5915)	157.8337
		90	1390.9510	(1311.9698 , 1475.2519)	163.2821
		80	1393.6893	(1315.4993 , 1481.9475)	166.4482
		70	1383.2160	(1303.5612 , 1465.4094)	161.8481
		60	1391.5434	(1304.2480 , 1484.3510)	180.1030
κ	MLE	101	0.0844	(-0.1569 , 0.3257)	0.4826
		90	0.1119	(-0.1615 , 0.3853)	0.5468
		80	0.1504	(-0.1488 , 0.4495)	0.5983
		70	0.0727	(-0.2570 , 0.4023)	0.6593
		60	0.1007	(-0.2608 , 0.4622)	0.7230
	Bayesian	101	0.1005	(0.0347 , 0.1563)	0.1215
		90	0.1006	(0.0295 , 0.1619)	0.1324
		80	0.1032	(0.0295 , 0.1623)	0.1328
		70	0.0991	(0.0336 , 0.1610)	0.1275
		60	0.1069	(0.0333 , 0.1728)	0.1395

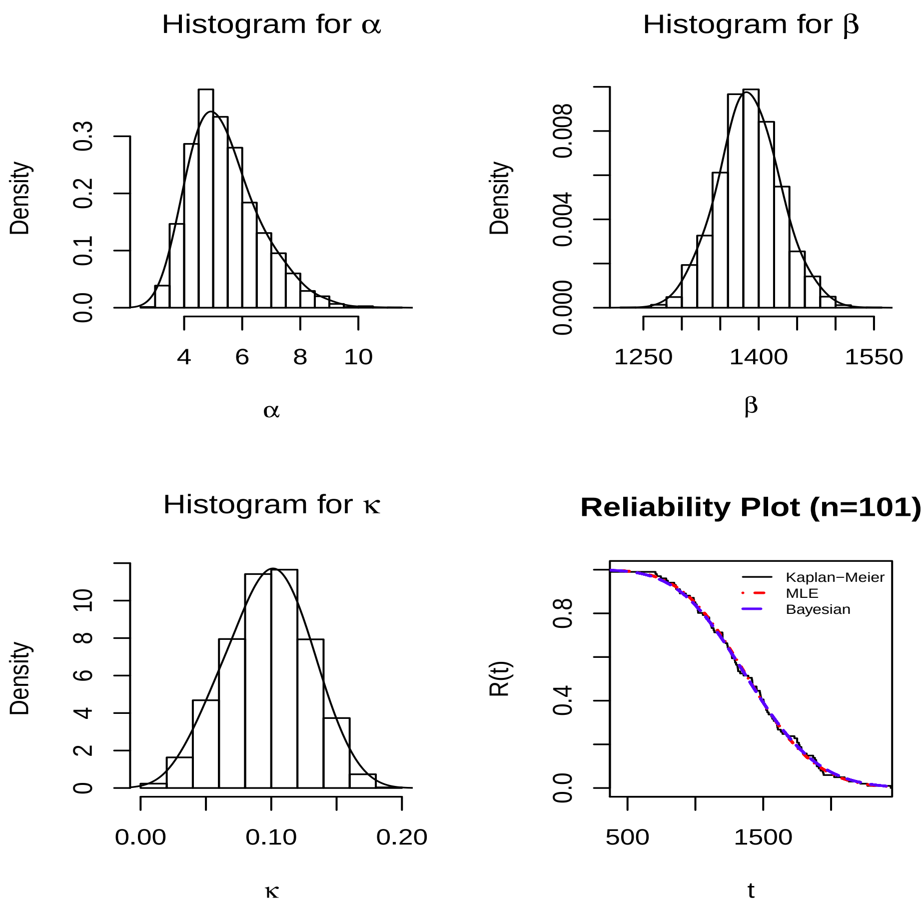

Figure 5.2: Histograms of the simulated samples for the parameters and the reliability plots for $n = 101$

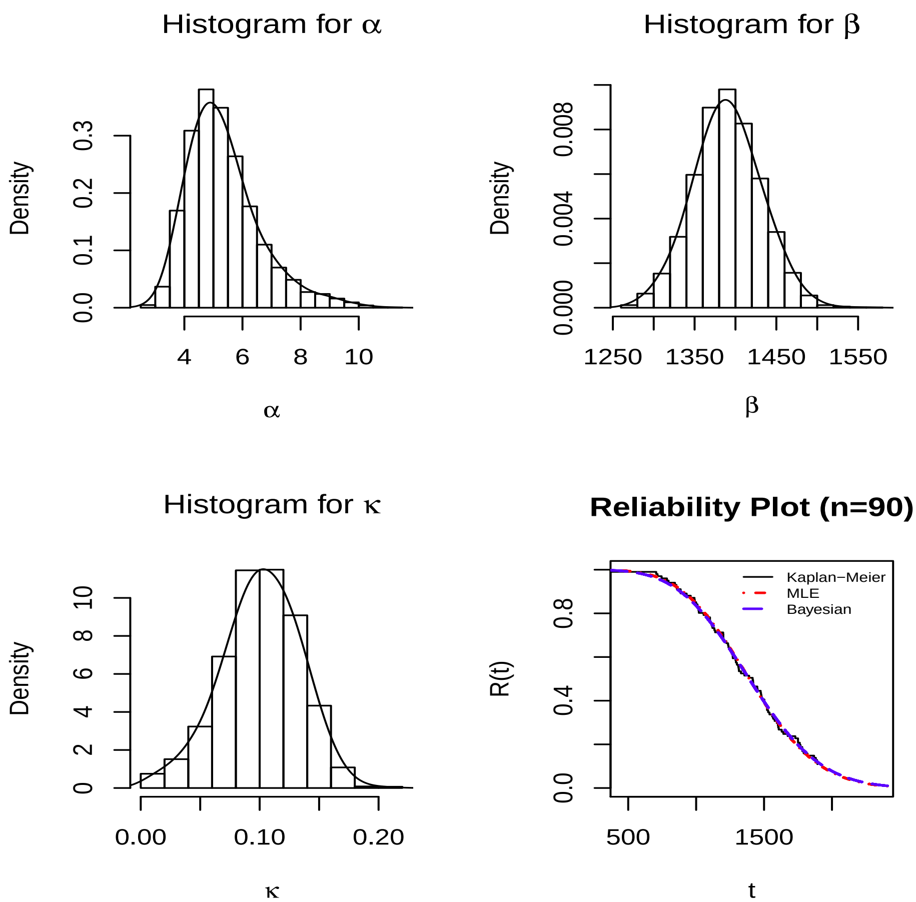

Figure 5.3: Histograms of the simulated samples for the parameters and the reliability plots for $n = 90$

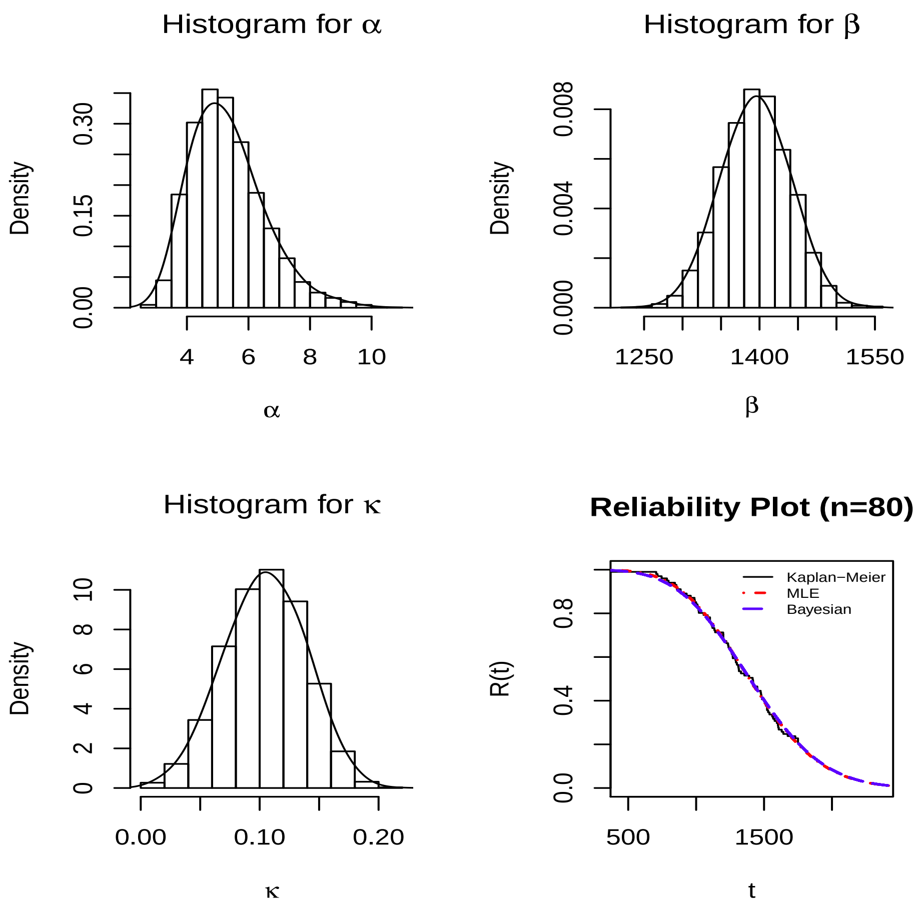

Figure 5.4: Histograms of the simulated samples for the parameters and the reliability plots for $n = 80$

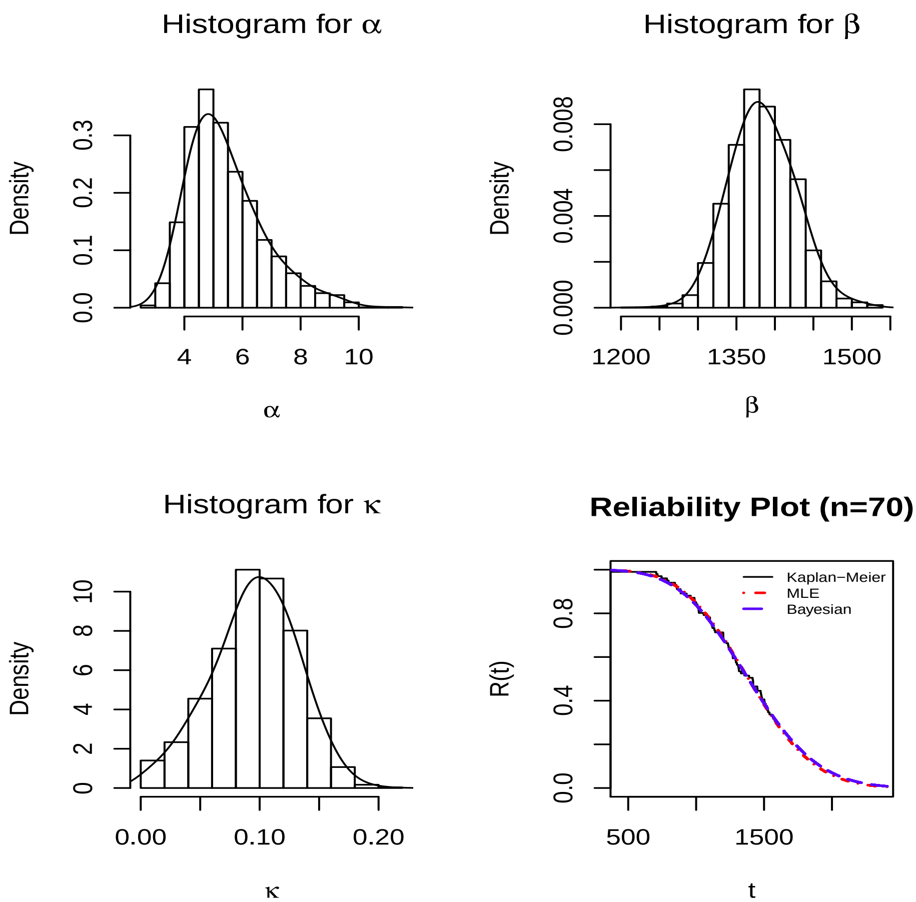

Figure 5.5: Histograms of the simulated samples for the parameters and the reliability plots for $n = 70$

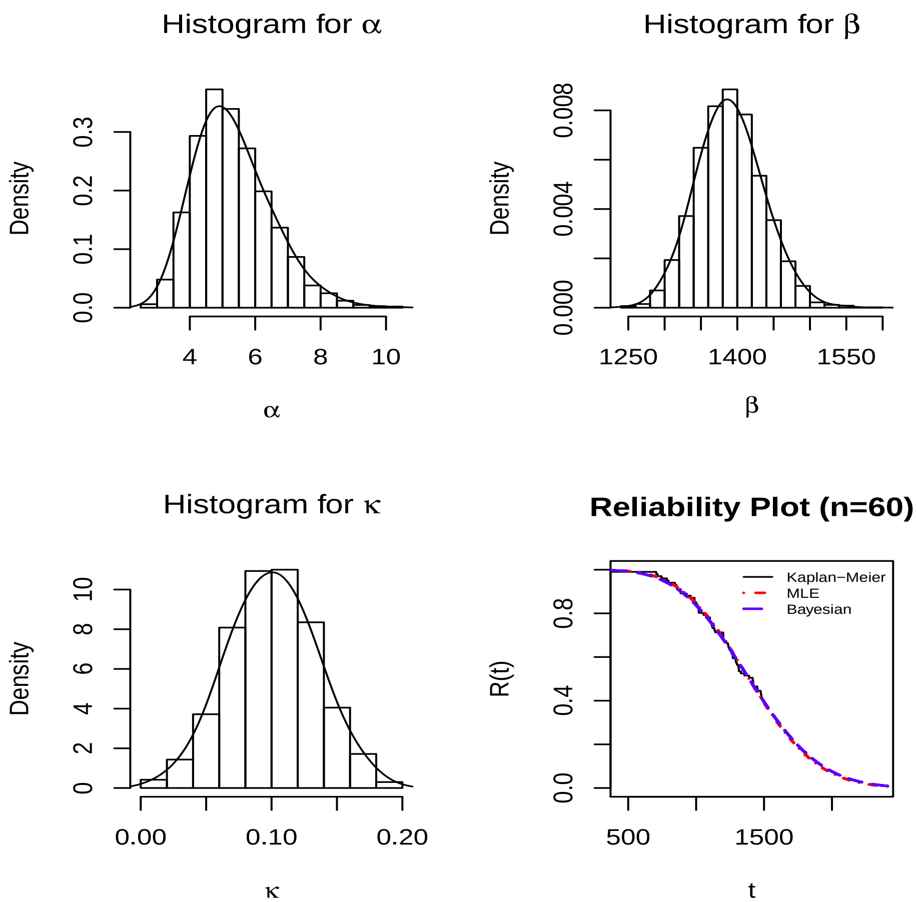

Figure 5.6: Histograms of the simulated samples for the parameters and the reliability plots for $n = 60$

Chapter 6

Concluding Remarks

Using the classical MLE and (mainly) the Bayesian approach, we obtain point and interval estimation for the parameters of the Birnbaum-Saunders (BS) distribution devised by Birnbaum and Saunders (1969a), as well as the Generalized Birnbaum-Saunders (GBS) distribution obtained by Owen (2006), in the presence of random right censored data.

For the classical MLE method, we derive the expressions for the observed Information of the GBS distribution. Where Bayesian approach is concerned, new sets of priors, which are mentioned in Subsections 3.1.3 and 3.2.3, are considered based on the model assumptions adopted by Birnbaum and Saunders (1969a) and Owen (2006). To handle the presence of random right censored observations, we utilize the data augmentation technique introduced by Tanner and Wong (1987), where the right-censored data is treated as latent/unobserved variables which are sampled together with the parameters. This method enables us to circumvent the arduous expressions involving the censored data in obtaining posterior inferences.

From our simulation study, we find that, accuracy of parameter estimation improves with larger sample size but deteriorates in the presence of more censored observations. Overall, conditional sampling outperforms joint sampling in terms of bias and credible interval width in the BS distribution parameter estimation. Conditional sampling also consistently outperforms MLE in terms of bias, square-root of mean squared error (SRMSE) and interval width in the GBS distribution parameter estimation, especially when sample size is small. We have also illustrated, with a real data set, that our posterior inference can be readily applied to the case of Type-II right censored data under the independent censoring assumption (Lawless, 2003).

References

Achcar, J. A. (1993), "Inferences for the BirnbaumSaunders fatigue life model using Bayesian methods," *Computational Statistics and Data Analysis*, 15, 367–380.

Achcar, J. A. and Moala, F. A. (2010), "Use of MCMC methods to obtain Bayesian inferences for the Birnbaum-Saunders distribution in the presence of censored data and covariates," *Advances and Applications in Statistics*, 17, 1–27.

Balakrishnan, N. and Zhu, X. (2014), "An improved method of estimation for the parameters of the BirnbaumSaunders distribution," *Journal of Statistical Computation and Simulation*, 84, 2285–2294.

Barndorff-Nielsen, O. E. (1991), "Modified signed log likelihood ratio," *Biometrika*, 78, 557–563.

Beran, J. (1994), *Statistics for Long Memory Processes*, London, U.K.: Chapman & Hall.

Bernardo, J. M. (1979), "Reference Posterior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society*, 41, 113–147.

Birnbaum, Z. W. and Saunders, S. C. (1969a), "A New Familty of Life Distributions," *Journal of Applied Probability*, 6, 319–327.

— (1969b), "Estimation for a Family of Life Distributions with Applications to Fatigue," *Journal of Applied Probability*, 6, 328.

Casella, G. and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167 – 174.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327 – 335.

Desmond, A. F. (1985), "Stochastic models of failure in random environments," *The Canadian Journal of Statistics*, 13, 171 – 183.

— (1986), "On the Relationship Between Two Fatigue-Life Models," *IEEE TRANSACTIONS ON RELIABILITY*, R-35, 167 – 169.

Diaz-Garcia, J. A. and Leiva-Sanchez, V. (2005), "A new family of life distributions based on the elliptically contoured distributions," *Journal of Statistical Planning and Inference*, 128, 445–457.

Dupuis, D. J. and Mills, J. E. (1998), "Robust Estimation of the Birnbaum-Saunders Distribution," *IEEE Transactions on Reliability*, 47, 88 – 95.

Engelhardt, M., Bain, L. J., and Wright, F. T. (1981), "Inferences on the Parameters of the Birnbaum-Saunders Fatigue Life Distribution Based on Maximum Likelihood Estimation," *Technometrics*, 23, 251.

Fierro, R., Leiva, V., Ruggeri, F., and Sanhueza, A. (2012), "On a BirnbaumSaunders distribution arising from a non-homogeneous Poisson process," *Statistics and Probability Letters*, 83, 1233 – 1239.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Boca Raton, Florida: Chapman & Hall, 2nd ed.

Gomez, H. W., Olivares-Pacheco, J. F., and Bolfarine, H. (2009), "An extension of the generalized BirnbaumSaunders distribution," *Statistics & Probability Letters*, 79, 331 – 338.

Hirose, H. (1999), "Bias Correction for the Maximum Likelihood Estimates in the Two-parameter Weibull Distribution," *IEEE Transactions on Dielectrics and Electrical Insulation*, 6, 66–69.

Kundu, D., Kannan, N., and Balakrishnan, N. (2008), "On the hazard function of BirnbaumSaunders distribution and associated inference," *Computational Statistics & Data Analysis*, 52, 2692 – 2702.

Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, New Jersey: John Wiley & Sons, 2nd ed.

Leiva, V., Riquelme, M., Balakrishnan, N., and Sanhueza, A. (2008), "Lifetime analysis based on the generalized BirnbaumSaunders distribution," *Computational Statistics & Data Analysis*, 52, 2079–2097.

Lemonte, A. J., Cribari-Neto, F., and Vasconcellos, K. L. P. (2007), "Improved statistical inference for the two-parameter BirnbaumSaunders distribution," *Computational Statistics and Data Analysis*, 51, 4656–4681.

Lindley, D. V. (1980), "Approximate Bayesian Methods," *Trabajos de estadstica y de investigacin operativa*, 31.

Ng, H. K. T., Kundu, D., and Balakrishnan, N. (2003), "Modified moment estimation for the two-parameter BirnbaumSaunders distribution," *Computational Statistics and Data Analysis*, 43, 283–298.

— (2006), "Point and interval estimation for the two-parameter BirnbaumSaunders distribution based on Type-II censored samples," *Computational Statistics and Data Analysis*, 50, 3222–3242.

Owen, W. J. (2006), "A New Three-Parameter Extension to the Birnbaum-Saunders Distribution," *IEEE Transactions on Reliability*, 55, 475–479.

Owen, W. J. and Padgett, W. J. (2000), "A Birnbaum-Saunders accelerated life model," *IEEE Transactions on Reliability*, 49, 224–229.

Rieck, J. R. (1999), "A moment generating function with applications to the birnbaum-saunders distributions," *Communications in Statistics-Theory and Methods*, 28, 2213–2222.

Rieck, J. R. and Nedelman, J. R. (1991), "A Log-Linear Model for the Birnbaum-Saunders Distribution," *Technometrics*, 33, 51–60.

Sanhueza, A., Leiva, V., and Balakrishnan, N. (2008), "The Generalized BirnbaumSaunders Distribution and Its Theory, Methodology, and Application," *Communications in Statistics - Theory and Methods*, 37, 645 – 670.

Steven, G. and Li, L. (2006), "Estimation of the Parameters of the BirnbaumSaunders Distribution," *Communications in Statistics - Theory and Methods*, 35, 2157–2169.

Sun, D. and Berger, J. O. (1998), "Reference Priors with Partial Information," *Biometrika*, 85, 55–71.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528 – 540.

Wang, M., Sun, X., and Park, C. (2015), "Bayesian analysis of BirnbaumSaunders distribution via the generalized ratio-of-uniforms method," *Computational Statistics*.

Wang, Z., Desmond, A. F., and Lu, X. (2006), "Modified censored moment estimation for the two-parameter BirnbaumSaunders distribution," *Computational Statistics and Data Analysis*, 50, 1033–1051.

Wu, J. and Wong, A. C. M. (2004), "Improved interval estimation for the two-parameter BirnbaumSaunders distribution," *Computational Statistics and Data Analysis*, 47, 809–821.

Xu, A. and Tang, Y. (2010), "Reference analysis for Birnbaum-Saunders distribution," *Computational Statistics and Data Analysis*, 54, 185–192.

— (2011), "Bayesian analysis of BirnbaumSaunders distribution with partial information," *Computational Statistics and Data Analysis*, 55, 2324–2333.

Appendix

First, we will explain the classical MLE inferential procedures for the case of the GBS distribution. To obtain the maximum likelihood estimates of α, β and κ (denoted by $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\kappa}$ respectively), we solve the following normal equations numerically:

$$\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial \alpha} = \frac{1}{\alpha}\sum_{i=1}^{n}\left[z^2(t_i)-1\right] + \sum_{j=1}^{m}\frac{\partial \log S(c_j)}{\partial \alpha},$$

$$\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial \beta} = -\frac{n}{2\beta} + \sum_{i=1}^{n}\left[\frac{\kappa}{\eta(t_i)} + \frac{t_i^{2-2\kappa}-\beta^2 t_i^{-2\kappa}}{2\alpha^2\beta^2}\right] + \sum_{j=1}^{m}\frac{\partial \log S(c_j)}{\partial \beta},$$

$$\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial \kappa} = \sum_{i=1}^{n}\left\{\frac{\beta - t_i}{\eta(t_i)} + \left[z^2(t_i)-1\right]\log t_i\right\} + \sum_{j=1}^{m}\frac{\partial \log S(c_j)}{\partial \kappa}.$$

Let $\widehat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\kappa})^T$ be the MLE vector. From the asymptotic normality of the MLE, it follows that

$$\widehat{\boldsymbol{\theta}} \sim \mathcal{N}_3\left(\boldsymbol{\theta}, \hat{I}^{-1}\right),$$

where $\hat{I}^{-1}$ denotes the inverse of the observed information matrix

$$\hat{I} = \begin{bmatrix} \hat{I}_{11} & \hat{I}_{12} & \hat{I}_{13} \\ \hat{I}_{12} & \hat{I}_{22} & \hat{I}_{23} \\ \hat{I}_{13} & \hat{I}_{23} & \hat{I}_{33} \end{bmatrix}$$

given by the following expressions:

$$\hat{I}_{11} = \frac{2n}{\alpha^2} - \sum_{j=1}^{m}\frac{\partial^2 \log S(c_j)}{\partial \alpha^2},$$

$$\hat{I}_{12} = -\frac{1}{\alpha^3}\sum_{i=1}^{n} t_i^{-2\kappa} + \frac{1}{\alpha^3\beta^2}\sum_{i=1}^{n} t_i^{2-2\kappa} - \sum_{j=1}^{m}\frac{\partial^2 \log S(c_j)}{\partial \alpha \partial \beta},$$

$$\hat{I}_{13} = \frac{2}{\alpha}\sum_{i=1}^{n}\left[z^2(t_i)\log t_i\right] - \sum_{j=1}^{m}\frac{\partial^2 \log S(c_j)}{\partial \alpha \partial \kappa},$$

$$
\begin{aligned}
\hat{I}_{22} &= -\frac{n}{2\beta^2} + \sum_{i=1}^{n}\left[\left(\frac{\kappa}{\eta(t_i)}\right)^2 + \frac{t_i^{2-2\kappa}}{\alpha^2\beta^3}\right] - \sum_{j=1}^{m}\frac{\partial^2 \log S(c_j)}{\partial\beta^2}, \\
\hat{I}_{23} &= -\sum_{i=1}^{n}\left\{\frac{t_i}{[\eta(t_i)]^2} + \frac{(\beta^2 - t_i^2)\log t_i}{\alpha^2\beta^2 t_i^{2\kappa}}\right\} - \sum_{j=1}^{m}\frac{\partial^2 \log S(c_j)}{\partial\beta\partial\kappa}, \\
\hat{I}_{33} &= \sum_{i=1}^{n}\left\{\left[\frac{\beta - t_i}{\eta(t_i)}\right]^2 + 2[z(t_i)\log t_i]^2\right\} - \sum_{j=1}^{m}\frac{\partial^2 \log S(c_j)}{\partial\kappa^2}.
\end{aligned}
$$

No closed form expressions are available for the Fisher information, so it could only be estimated by using the observed Fisher information.

We will now derive the first and second derivatives of $\sum_{j=1}^{m}\log S(c_j)$ with respect to the parameters. Let $\Phi\{.\}$ and $\phi[.]$ be the cdf and pdf for the standard normal distibution. Recall that the functions $z(t)$ and $S(t)$ are given by the following:

$$
\begin{aligned}
z(t) &= \frac{1}{\alpha}\left(\frac{t^{1-\kappa}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{t^{\kappa}}\right), \\
S(t) &= 1 - \Phi\{z(t)\},
\end{aligned}
$$

and here, we also define:

$$
\zeta(t) = \frac{1}{\alpha}\left(\frac{t^{1-\kappa}}{\sqrt{\beta}} + \frac{\sqrt{\beta}}{t^{\kappa}}\right).
$$

To further simplify expressions for the first derivatives of $\sum_{j=1}^{m}\log S(c_j)$, we introduce the following functions:

$$
\begin{aligned}
S_1(t) &\stackrel{\text{def}}{=\!=} \frac{\partial}{\partial\alpha}S(t) = \frac{z(t)\times\phi[z(t)]}{\alpha}, \\
S_2(t) &\stackrel{\text{def}}{=\!=} \frac{\partial}{\partial\beta}S(t) = \frac{\zeta(t)\times\phi[z(t)]}{2\beta}, \\
S_3(t) &\stackrel{\text{def}}{=\!=} \frac{\partial}{\partial\kappa}S(t) = (\log t)\times z(t)\times\phi[z(t)].
\end{aligned}
$$

Then, we obtain the first derivatives of $\sum_{j=1}^{m}\log S(c_j)$ as follows:

$$\frac{\partial}{\partial\alpha}\sum_{j=1}^{m}\log S(c_j) = \sum_{j=1}^{m}\frac{S_1(c_j)}{S(c_j)},$$
$$\frac{\partial}{\partial\beta}\sum_{j=1}^{m}\log S(c_j) = \sum_{j=1}^{m}\frac{S_2(c_j)}{S(c_j)},$$
$$\frac{\partial}{\partial\kappa}\sum_{j=1}^{m}\log S(c_j) = \sum_{j=1}^{m}\frac{S_3(c_j)}{S(c_j)}.$$

Again, to simplify the expressions for the second derivatives of $\sum_{j=1}^{m}\log S(c_j)$, we introduce the following functions:

$$S_{11}(t) \overset{\text{def}}{=\!=} \frac{\partial^2 S(t)}{\partial\alpha^2} = \frac{z(t)\times\phi[z(t)]\times[z^2(t)-2]}{\alpha^2},$$
$$S_{12}(t) \overset{\text{def}}{=\!=} \frac{\partial^2 S(t)}{\partial\alpha\partial\beta} = \frac{\zeta(t)\times\phi[z(t)]\times[z^2(t)-1]}{2\alpha\beta},$$
$$S_{13}(t) \overset{\text{def}}{=\!=} \frac{\partial^2 S(t)}{\partial\alpha\partial\kappa} = \frac{z(t)\times(\log t)\times\phi[z(t)]\times[z^2(t)-1]}{\alpha},$$
$$S_{22}(t) \overset{\text{def}}{=\!=} \frac{\partial^2 S(t)}{\partial\beta^2} = \frac{\phi[z(t)]\times\{z(t)\times\zeta^2(t)-2\zeta(t)-z(t)\}}{4\beta^2},$$
$$S_{23}(t) \overset{\text{def}}{=\!=} \frac{\partial^2 S(t)}{\partial\beta\partial\kappa} = \frac{\zeta(t)\times(\log t)\times\phi[z(t)]\times[z^2(t)-1]}{2\beta},$$
$$S_{33}(t) \overset{\text{def}}{=\!=} \frac{\partial^2 S(t)}{\partial\kappa^2} = z(t)\times(\log t)^2\times\phi[z(t)]\times\left[z^2(t)-1\right].$$

Therefore, we obtain the second derivatives of $\sum_{j=1}^{m}\log S(c_j)$ as follows:

$$\frac{\partial^2}{\partial\alpha^2}\sum_{j=1}^{m}\log S(c_j) = \sum_{j=1}^{m}\frac{S(c_j)\times S_{11}(c_j)-S_1^2(c_j)}{S^2(c_j)},$$
$$\frac{\partial^2}{\partial\alpha\partial\beta}\sum_{j=1}^{m}\log S(c_j) = \sum_{j=1}^{m}\frac{S(c_j)\times S_{12}(c_j)-S_1(c_j)\times S_2(c_j)}{S^2(c_j)},$$
$$\frac{\partial^2}{\partial\alpha\partial\kappa}\sum_{j=1}^{m}\log S(c_j) = \sum_{j=1}^{m}\frac{S(c_j)\times S_{13}(c_j)-S_1(c_j)\times S_3(c_j)}{S^2(c_j)},$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\beta^2}\sum_{j=1}^{m}\log S(c_j) &= \sum_{j=1}^{m}\frac{S(c_j)\times S_{22}(c_j)-S_2^2(c_j)}{S^2(c_j)},\\
\frac{\partial^2}{\partial\beta\partial\kappa}\sum_{j=1}^{m}\log S(c_j) &= \sum_{j=1}^{m}\frac{S(c_j)\times S_{23}(c_j)-S_2(c_j)\times S_3(c_j)}{S^2(c_j)},\\
\frac{\partial^2}{\partial\kappa^2}\sum_{j=1}^{m}\log S(c_j) &= \sum_{j=1}^{m}\frac{S(c_j)\times S_{33}(c_j)-S_3^2(c_j)}{S^2(c_j)}.
\end{aligned}
$$

For the BS distribution, to obtain the maximum likelihood estimates of α and β (denoted by $\hat{\alpha}$ and $\hat{\beta}$ respectively), we solve the following normal equations numerically:

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial\alpha} &= \frac{1}{\alpha}\sum_{i=1}^{n}\left[z^2(t_i)-1\right]+\sum_{j=1}^{m}\frac{\partial\log S(c_j)}{\partial\alpha},\\
\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{D})}{\partial\beta} &= -\frac{n}{2\beta}+\sum_{i=1}^{n}\left[\frac{1}{t_i+\beta}+\frac{t_i}{2\alpha^2\beta^2}-\frac{1}{2\alpha^2 t_i}\right]+\sum_{j=1}^{m}\frac{\partial\log S(c_j)}{\partial\beta}.
\end{aligned}
$$

Let $\widehat{\boldsymbol{\theta}}=(\hat{\alpha},\hat{\beta})^T$ be the MLE vector. From the asymptotic normality of the MLE, it follows that

$$
\widehat{\boldsymbol{\theta}}\sim\mathcal{N}_2\left(\boldsymbol{\theta},\hat{I}^{-1}\right),
$$

where $\hat{I}^{-1}$ denotes the inverse of the observed information matrix

$$
\hat{I}=\begin{bmatrix}\hat{I}_{11} & \hat{I}_{12}\\ \hat{I}_{12} & \hat{I}_{22}\end{bmatrix}
$$

given by the following expressions:

$$
\begin{aligned}
\hat{I}_{11} &= \frac{2n}{\alpha^2}-\sum_{j=1}^{m}\frac{\partial^2\log S(c_j)}{\partial\alpha^2},\\
\hat{I}_{12} &= -\sum_{j=1}^{m}\frac{\partial^2\log S(c_j)}{\partial\alpha\partial\beta},\\
\hat{I}_{22} &= \frac{n[\alpha(2\pi)^{-1/2}h(\alpha)+1]}{\alpha^2\beta^2}-\sum_{j=1}^{m}\frac{\partial^2\log S(c_j)}{\partial\beta^2},
\end{aligned}
$$

where $h(\alpha) = \alpha\sqrt{(\pi/2)} - \pi e^{2/\alpha^2}[1 - \Phi(2/\alpha)]$. Recall that $GBS(\alpha, \beta, \frac{1}{2}) = BS(\alpha, \beta)$. Hence, for the case of the BS distribution, substitute $\kappa = \frac{1}{2}$ into the expressions for $\frac{\partial}{\partial \alpha} \sum_{j=1}^{m} \log S(c_j)$ and $\frac{\partial}{\partial \beta} \sum_{j=1}^{m} \log S(c_j)$ from above in order to obtain the first derivatives of the censored loglikelihood $\sum_{j=1}^{m} \log S(c_j)$.

Similarly, to obtain the second derivatives of the censored loglikelihood $\sum_{j=1}^{m} \log S(c_j)$, substitute $\kappa = \frac{1}{2}$ into the expressions for $\frac{\partial^2}{\partial \alpha^2} \sum_{j=1}^{m} \log S(c_j)$, $\frac{\partial^2}{\partial \alpha \beta} \sum_{j=1}^{m} \log S(c_j)$ and $\frac{\partial^2}{\partial \beta^2} \sum_{j=1}^{m} \log S(c_j)$ from above, and we are done.

Notice that the Fisher information for the observed data has closed form expressions which were derived by Lemonte et al. (2007), whereas the Fisher information for the censored data does not have closed form expression, and could only be estimated by using the observed Fisher information (Ng et al., 2003).

Curriculum Vitae

Tun Lee Ng was born and raised in Penang, Malaysia. Since a young age, he found his passion in mathematics as he gained inspiration from his father, a mathematics high school teacher. After graduating as valedictorian from Chung Ling High School in Malaysia, he received a competitive scholarship to pursue a baccalaureate in Actuarial Science at the University of Melbourne in Australia. He made the Deans List twice, and received a Statistics subject prize.

After completing his undergraduate studies with First Class Honors, he worked as an actuarial analyst in Malaysia for eighteen months. He helped to code programming templates in statistical package and spreadsheet to automate claims analyses in the company.

In the fall of 2014, he entered the Graduate School of The University of Texas at El Paso. While pursuing a master's degree in Statistics, he worked as a Teaching Assistant, and as a summer intern at the Center of Evaluation and Planning, University of Texas at El Paso during Summer 2015. He was also a member of the Phi Kappa Phi Honors Society.

Permanent address: 17-05-12, Lintang Paya Terubong 1
Georgetown, Penang, Malaysia 11060