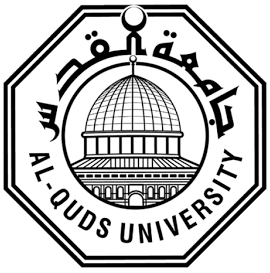
**Deanship of Graduate Studies**

**Al-Quds University**




# Bayesian and Non-Bayesian Inference for Survival Data Using Generalized Exponential Distribution: A Comparison Study

**Firyal Mohammad Khalaf Saraheen**

**M.Sc. Thesis**

**Jerusalem- Palestine**

**1437/2016**

# Bayesian and Non-Bayesian Inference for Survival Data Using Generalized Exponential Distribution:

# A Comparison Study

**Prepared By**

**Firyal Mohammad Khalaf Saraheen**

**B. Sc. Mathematics, Bethlehem University**

**Palestine**

**Supervisor: Dr. Khalid Salah**

**A thesis Submitted in Partial Fulfillment of the requirements for the Degree of Master of Mathematics at Al-Quds University**

**1437/2016**

Al-Quds University

Deanship of Graduate Studies

Graduate Studies / Mathematics


## Thesis Approval

**Bayesian and Non-Bayesian Inference for Survival Data Using Generalized Exponential Distribution: A Comparison Study**

**Prepared By : Firyal Mohammad Khalaf Saraheen**

**Registration No : 21310121**

**Supervisor: Dr. Khalid Ali Salah**

Master thesis submitted and accepted, Date : 9/01/2016.

The names and signatures of the examining committee members are as follows

1. Dr. Khalid Ali Salah Head of Committee, Signatures.....
2. Dr. Sa'di Alkronz Internal Examiner, Signatures.....
3. Dr. Bader Aljawadi External Examiner, Signatures.....

Jerusalem-Palestine

1437/2016

## Dedication

I dedicate this thesis to my parents and brothers who led me through this darkness with their light of hope and support.

To my friends who touched my life with their love and passion, especially my dearest friend Bushra, who stands with me when things look bleak.

## Declaration

I certify that this thesis submitted for the degree of Master, is the result of my own research, except where otherwise acknowledged, and that this study has not been submitted for a higher degree to any other university or institution.

**Signature :** 

**Student's name :** Firyal Mohammad Khalaf Saraheen

**Date : 9/1/2016**

## Acknowledgement

I am thankful to all teachers in mathematics department who helped me to know that as lifelong learners of science, I can live more productive, responsible and fulgent. While special thanks go to Dr. Khalid Salah.

Lastly, I offer my regards to all of those who supported me during the completion of this thesis, especially Miss Ibtisam.

# Table of Contents

## Contents Page

## List of Tables

## List of Figures

## List of Abbreviations

| Abbreviations | Meaning |
|---|---|
| K-M | Kaplan-Meier |
| MLE | Maximum Likelihood Estimation |
| NPMLE | Nonparametric Maximum Likelihood Estimator |
| EM | Expectation-Maximization |
| GE | Generalized Exponential |
| MPLE | Maximum Partial Likelihood Estimate |
| LR | likelihood-ratio |
| CI | Confidence Interval |
| MCMC | Markov Chain Monte Carlo |
| M-H | Metropolis-Hastings |
| Log | Logarithm |
| BSE | Bayes estimate using Lindley's approximation |
| BGE | Bayes estimates under the general entropy loss functions |
| MSE | Mean Squared Errors |
| DPC | Diagnosis Procedure Combination |
| QIP | Quality Indicator/Improvement Project |
| NPML | Non-Parametric Maximum Likelihood |
| AIC | Akaike’s Information Criterion |

## Abstract

In this thesis we consider the Bayesian and non-Bayesian estimation of the unknown parameters of the Generalized Exponential (GE) distribution. Our aim is to compare the estimates of parameters and to observe the performance of the methods used for estimation.

By the developed methodology for MLE and Bayesian estimation has been demonstrated on a real data set when both the shape $(p)$ and scale $(\theta)$ parameters of the GE distribution are unknown under informative set of independent priors. It is observed that the parameter estimates under the classical maximum likelihood method could not be obtained in close form; we therefore employed Newton- Raphson iterative approach via the Hessian matrix.

In this study following *C. Guure and S. Bosomprah* (2013), we consider the Bayesian estimation of the unknown parameters of the GE distribution. We have also assumed a gamma prior on both parameters, and we provide the Bayesian estimators under the assumptions of squared error and general entropy loss functions. We see that the Bayesian estimators cannot be obtained in explicit forms, due to the complex nature of the posterior distribution of which Bayesian inference is drawn. Therefore, Lindley's numerical approximations procedure is used.

Results show that the Bayesian estimator under general entropy loss function performed quiet better than Bayesian under squared error loss function and that of maximum likelihood estimator for estimating the scale parameter with both MSE and absolute bias.

# Chapter One
# Introduction

## 1.1 Survival Data

Survival analysis is a branch of statistics which includes a variety of "statistical methods designed to describe, explain or predict the occurrence of events". It is widely applied in many fields such as biology, medicine, public health, and epidemiology. In survival analysis, our objective is to model the survival time, i.e. the time to the occurrence of a given event. The event could be just about anything. Within the medical field, common examples are the time to development of a disease, response to a treatment, and of course death. The available data often include the survival time, patient characteristics (such as gender, age, and blood pressure), disease information, treatment information, examination data and much more. Often we attempt to predict the probability of survival, response, or mean lifetime given a set of observed variables and compare survival distributions.

## 1.2 Survival Function

For matters of simplicity we assume time $T$ ( where $T$ is the random variable representing survival time) to be continuous. The distribution of survival times is described by three mathematically equivalent functions: survival , hazard and cumulative hazard functions . A very simple way to specify the probability distribution of continuous durations $T$ is the distribution function

$$F(t) = P(T \leq t) \tag{1.1}$$

The distribution function of $t$ represents the probability that a realization of the random variable $T$ is less than a value $t$. Furthermore $f(t)$ is the density function corresponding to (1.1) and thus can be written as

$$f(t) = dF(t)/dt \tag{1.2}$$

An alternative specification of the probability distribution of duration and an important concept in survival analysis is the survivor function, $S(t)$, defined as

$$S(t) = P(T > t) = 1 - F(t) = 1 - \int_{-\infty}^{t} f(x)dx = \int_{t}^{\infty} f(x)dx \tag{1.3}$$

which is the probability that a realization of the random variable $T$ is greater than or equals to $t$. Or in other words: the probability that the event has not yet occurred by time $t$. Theoretically, the survival curve $S(t)$ can be plotted graphically to represent the probability of an individual's survival at varying time points. As $t$ ranges from 0 to ∞ all survival curves have the following properties:

i. $S(t)$ is monotone

ii. $S(t)$ is non-increasing

iii. At time $t = 0$, $S(t) = 1$ (i.e. the probability of surviving past time 0 is 1)

iv. At time $t = \infty$, $S(t) = 0$ (i.e. as time goes to infinity, the survival curve goes to 0)

(See Figure 1.1).

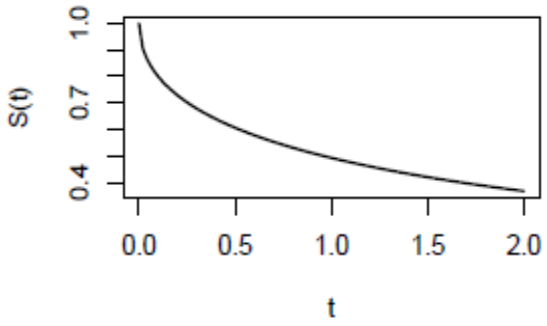



**Figure 1.1:** The survival function

### 1.3 Hazard Function

The hazard function $h(t)$ is the instantaneous rate at which events occur, given no previous events, defined as:

$$\begin{aligned} h(t) &= \lim_{dt \to 0} \frac{Pr\{t \leq T < t + dt / T \geq t\}}{dt} \\ &= \lim_{dt \to 0} \frac{Pr\{t \leq T < t + dt\}}{dt \; Pr(T \geq t)} \\ &= \frac{1}{S(t)} \lim_{dt \to 0} \frac{F(t + dt) - F(t)}{dt} \\ &= \frac{f(t)}{S(t)} \\ &= -\frac{d}{dt} \log(S(t)) \end{aligned} \tag{1.4}$$

from the definition; the hazard function is the 'chance' of failure (though it is a normalized probability, not a probability) at time *t* , given that the individual has survived until time *t*. We see that the hazard function is similar to the density in the sense that it is a positive function. However it does not integrate to one. Indeed, it is not integrable.

*The cumulative hazard function ,* $H(t)$, define as:

$$H(t) = \int_0^t h(u)du = -log\, S(t) \tag{1.5}$$

#### 1.3.1 Relationship between survival function and hazard function

From (1.3) and (1.4), we get the relationship

$$h(t) = \frac{f(t)}{S(t)} \tag{1.6}$$

Furthermore, since the density function is defined as the derivative of the cumulative distribution function, we get

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t) \tag{1.7}$$

Inserting (1.7) in (1.6), we have

$$h(t) = \frac{-S'(t)}{S(t)} = \frac{-d}{dt} log\ S(t) \tag{1.8}$$

Using (1.5) we get

$$S(t) = \exp[-H(t)] = \exp[-\int_0^t h(u)du] \tag{1.9}$$

Inserting (1.9) in (1.6) yields

$$f(t) = h(t)\ exp[-H(t)] \tag{1.10}$$

Hence, we have shown that it is possible to derive any of the three functions given the two others are known.

### 1.4 Censoring in Survival Data (left-right-interval)

A key characteristic that distinguishes survival analysis from other areas in statistics is that survival data are usually censored. Censoring is probably most well known because of survival analysis, which studies time until an event. There are usually some individuals who do not experience the event during the study, so the time to event is incomplete for these cases**.** Subjects are said to be censored if they are lost to follow up or drop out of the study, or if the study ends before they die or have an outcome of interest. The most common censoring models are:

1- Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. The only information we have is this right bound. This is very important in study of survival time, because data are often right-censored. (An example of right censoring data are shown in Figure 1.2).

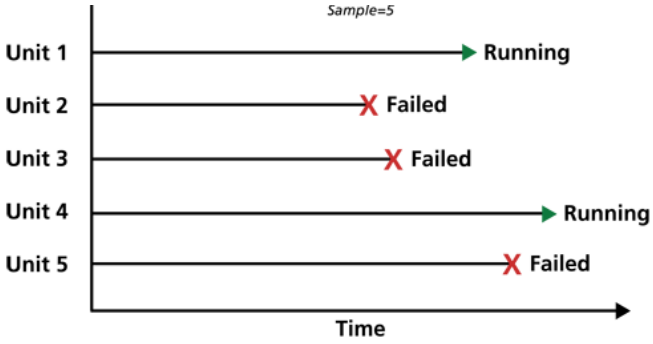



**Figure 1.2:** Right-censoring example

2- Left censoring occurs when the event of interest has already occurred before recording. It happens, for example, when we know the date of a medical exam that revealed a disease, but we don't know when the patient has been infected.( An example of left censoring data are shown in Figure 1.3).

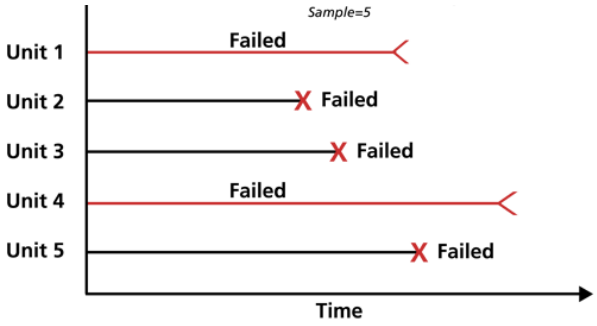



**Figure 1.3:** Left-censoring example

3- Interval censoring: when the event occurs between two times , but the exact time of failure is not known.(i.e. the event occurred between date A and date B). It could occurs, for example, when a patient is regularly checked, and one time we discover a medical deterioration. The only information we have is that the deterioration appears between two checks. Usually Turnbull gives an algorithm using to find a nonparametric estimator for interval censored data.(An example of interval censoring data are shown in Figure 1.4).

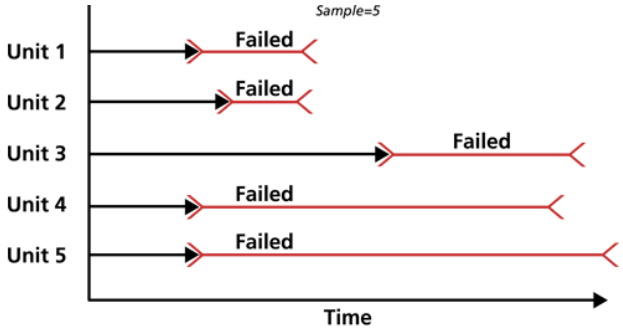



**Figure 1.4:** Interval-censoring example

Left and right censoring are special cases of interval censoring, with the beginning of the interval at zero or the end at infinity, respectively.

## 1.5 Estimation of survival function

Survival analysis in a brief is to estimate the three survival (survivorship, density, and hazard) functions as defined before. There exist parametric as well as non-parametric methods for this purpose. In case we do not know the exact survival times, estimation of the survival functions becomes much more difficult.

### 1.5.1 Parametric approach

In this case we consider the Bayes and non-Bayes estimation of the unknown parameters.

1. *Non- Bayesian Inference* (Maximum Likelihood Approach): It provides a consistent approach to parameter estimation problems. This means that maximum likelihood estimates can be developed for a large variety of estimation situations. Also it has desirable mathematical and optimality properties. The disadvantages of this method are: The likelihood equations need to be specifically worked out for a given distribution and estimation problem , the numerical estimation is usually non-trivial , it can be heavily biased for small samples. The optimality properties may not apply for small samples, and it sensitive to the choice of starting values.
2. *Bayesian Inference* :In Bayesian Inference, the parameter of interest is always considered to be a random variable with a prior distribution. The prior distribution is

the distribution of the parameter before any data is observed. Distributions that are commonly used in survival analysis are the Exponential, Weibull, Gamma and Log-normal. Because of its historical significance and mathematical simplicity.

- Exponential distribution, with density function $f(t) = \alpha e^{-\alpha t}$ and survival function $S(t) = e^{-\alpha t}$.
- Weibull distribution, with density function $f(t;\alpha,\beta) = \begin{cases} \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^{\alpha}} & t \geq 0 \\ 0 & t < 0 \end{cases}$
- Gamma distribution, with density function $f(t;\alpha,\beta) = \frac{\beta^{\alpha} t^{\alpha-1} e^{-\beta t}}{\gamma(\alpha)} \quad t, \beta, \alpha > 0$

  where $\gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad \alpha > 0$
- Log-normal distribution, with density function $f(t) = \frac{1}{t.\sigma\sqrt{2\pi}} e^{\left(-\frac{1}{2\sigma^2}(\log(t)-\mu)^2\right)}$

The exponential distribution is one of the most popular parametric models and play a central role in analyses of lifetime or survival data, in part because of their convenient statistical theory, their important 'lack of memory' property and their constant hazard rates. As shown in the following example.

**Example 1.1**

Consider a random variable $T$ with an exponential probability distribution with parameter $\theta$: $f(t) = \frac{1}{\theta} e^{-t/\theta}$ .The formula for the cumulative distribution function of the exponential distribution is $F(t) = \int_t^0 \frac{1}{\theta} e^{-y/\theta} dy = 1 - e^{-t/\theta}$ .

The formula for the survival function is $S(t) = 1 - (1 - e^{-\frac{t}{\theta}}) = e^{-t/\theta}$.

The formula for the hazard function is $h(t) = \frac{\frac{1}{\theta} e^{-t/\theta}}{e^{-t/\theta}} = \frac{1}{\theta}$.

The formula for the cumulative hazard function is $H(t) = \int_0^t \frac{1}{\theta} dt = \frac{t}{\theta}$.

The following is the plot of the exponential survival function (Figure 1.5).

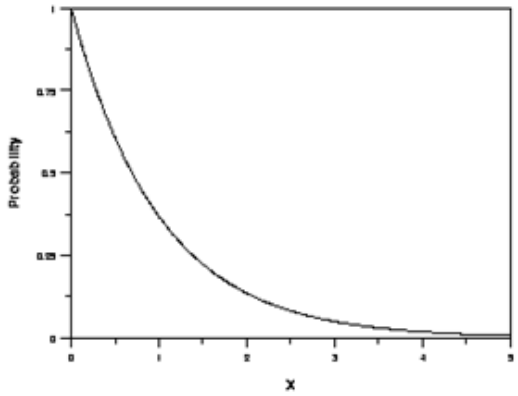



**Figure 1.5:** Exponential Survival Function

### 1.5.2 Non-parametric approach

Nonparametric estimator does not assume that the data come from a specified distribution, so we use it when we cannot know the distribution of the data. Now we will discuss the Kaplan-Meier estimator and the Turnbull estimator as the nonparametric estimators of the survival function.

1. **Kaplan-Meier (K-M) estimator**

   The standard nonparametric estimator of the survival function is the Kaplan-Meier (K-M) estimator, also known as the product-limit estimator. This estimator is defined as:

$$\hat{S}(t) = \begin{cases} 1 & if\ t < t_1, \\ \prod_{t_i \leq t} \left[1 - \frac{d_i}{Y_i}\right] & if\ t_1 < t\ , \end{cases} \qquad (1.11)$$

   where $t_1$ denotes the first observed failure time, $d_i$ represents the number of failures at time $t$, and $Y_i$ indicates the number of individuals who have not experienced the event of interest, and have also not been censored, by time $t$.

   From the function given in Equation (1.3), we notice that before the first failure happens, the survival probability is always 1. As failures occur, the K-M estimator of the survival function decreases. A step function with jumps at the observed event

times will be obtained by using K-M method to estimate the survival function. The jumps on the survival curve depend not only on the number of events observed at each event time, but also on the pattern of the censored observations before the event time.

**Example 1.2**

Consider the 10-year follow-up study, where we are interested in knowing how long people will survive after a kidney transplant. Suppose there are a total of 50 patients in the 10-year study. Also suppose that six of them died at 0.5 years, and two are lost to follow up during the half year after transplant. Therefore, at 0.5 years after the transplant, there are 42 patients still in this study. Similarly, we have some deaths at 1 year after transplant and so on, until the end of the study period, at which time there are 22 patients still alive and enrolled in the study. Data from this hypothetical study are given in Table 1.1, along with K-M estimates of the survival function at the various death times.

**Table 1.1**: Construction of the Kaplan-Meier estimator.

| **Time** $t_i$ | **Number of events** $d_i$ | **Number at risk** $Y_i$ | **K-M Estimator** $\widehat{S}(t) = \prod_{t_i \le t} \left[1 - \frac{d_i}{Y_i}\right]$ |
|---|---|---|---|
| 0.5 | 6 | 42 | $[1-\frac{6}{42}] = 0.857$ |
| 1 | 5 | 35 | $[0.857](1-\frac{5}{35}) = 0.735$ |
| 2 | 3 | 32 | $[0.735](1-\frac{3}{32}) = 0.666$ |
| 3.5 | 2 | 30 | $[0.666]( 1-\frac{2}{30}) = 0.622$ |
| 5 | 1 | 28 | $[0.622]( 1-\frac{1}{28}) = 0.600$ |
| 6.5 | 1 | 27 | $[0.600]( 1-\frac{1}{27}) = 0.578$ |
| 8.5 | 2 | 25 | $[0.578]( 1-\frac{2}{25}) = 0.532$ |
| 9.5 | 2 | 22 | $[0.532]( 1-\frac{2}{22}) = 0.484$ |

Table 1.2 shows the K-M estimates for all times, and the corresponding graph of the K-M function is given in Figure 1.6.

**Table 1.2:** Kaplan-Meier survival estimates

| **Time on study** $(t)$ | **K-M Estimator** $\widehat{S}(t)$ |
|---|---|
| $0 \leq t < 0.5$ | 1.000 |
| $0.5 \leq t < 1$ | 0.857 |
| $1 \leq t < 2$ | 0.735 |
| $2 \leq t < 3.5$ | 0.666 |
| $3.5 \leq t < 5$ | 0.622 |
| $5 \leq t < 6.5$ | 0.600 |
| $6.5 \leq t < 8.5$ | 0.578 |
| $8.5 \leq t < 9.5$ | 0.532 |
| $9.5 \leq t < 10$ | 0.484 |

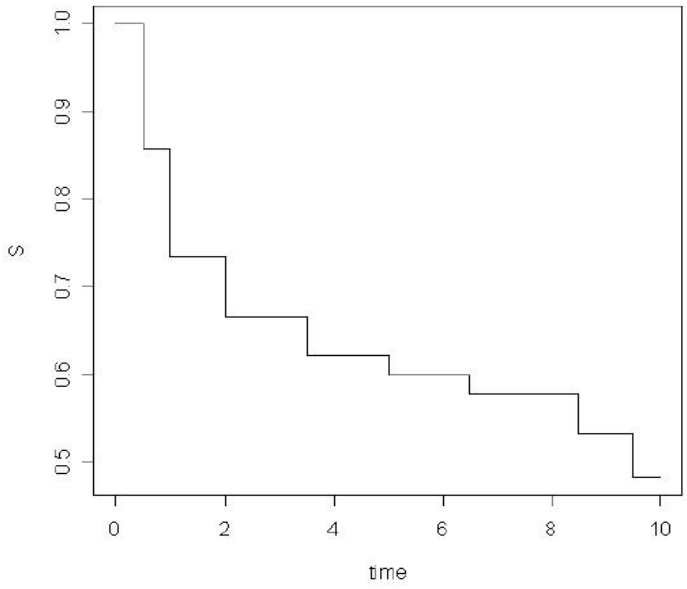



**Figure 1.6**: Kaplan-Meier survival function for right-censored data

The K-M estimator is a common nonparametric estimator. It is efficient and easy to use, and it is available in many statistical software programs such as SAS and S-Plus.

## 2. Turnbull estimator

An estimator of the survival function is available for interval-censored data. Richard Peto developed a Newton-Raphson method to estimate the nonparametric maximum likelihood estimator (NPMLE) for interval-censored data in (1973). Then in 1976 Richard Turnbull formulated an Expectation-Maximization (EM) algorithm which also estimated the NPMLE for interval-censored data. The NPMLE for interval-censored data is based on n independent, arbitrarily interval-censored observations. The NPMLE can be estimated using Turnbull's algorithm in R software. (An example of Turnbull survival function for interval-censored data are shown in Figure 1.7)

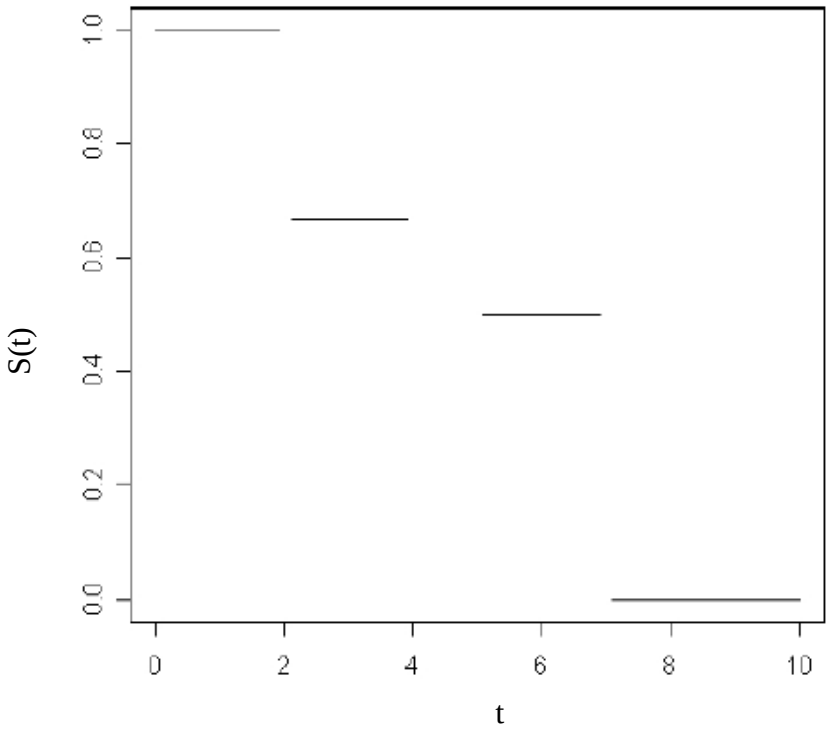



**Figure 1.7**: Turnbull survival function for interval-censored data

## 1.6 Problem statement

Generating survival data means observing a sample of research subjects (individuals) over a predefined time period and recording whether and when the individuals experience the event. Basic survival data consist of a variable measuring the time that has passed (the duration) before an individual experiences the event (or until the study ends) and a variable indicating if the individual experiences that event during the observation period or not. Survival analysis estimates by estimating a survival time.

In this thesis both Bayesian and Non- Bayesian approaches will be used to estimate the unknown parameters and compared it to determine the best method (with less Standard Errors, Absolute Bias, and Mean Squared Errors) that can be used to estimate the parameters of the generalized exponential distribution and survival function.

## 1.7 Literature review

In the past five decades, survival analysis has become one of the most frequently used methods for analyzing data in various disciplines. Introductory treatments of survival analysis for social scientists can be found in Allison (1984, 1995), Tuma and Hannan (1984), Kiefer (1988), Blossfeld and Rohwer (2001). For a biostatistical point of view, see Collett (2003), Hosmer and Lemeshow (2003). Soliman et al. (2006) estimated the Weibull distribution by using the maximum likelihood estimator and Bayesian estimator under squared error loss function and Linex loss function for a given shape parameter and several unknown parameters. Gupta and Kundu (1999) recently proposed the two parameter generalized exponential distribution ($GE$) as an alternative to the lognormal, gamma, and Weibull distributions and did some studies on its properties. Some references on $GE$ distribution are Raqab (2002), Zheng (2002), and Kundu and Gupta (2008). According to Gupta and Kundu (2001), the two-parameter $GE(\theta, p)$ can have increasing and decreasing

failure rates depending on the shape parameter. Some research has been done to compare MLE to that of the Bayesian approach in estimating the survival function and the parameters of the Weibull distribution which are similar to the GE distribution. Amongst others, Sinha (1986) determined the Bayesian estimates of the reliability function and the hazard rate of the Weibull failure time distribution by employing only squared error loss function. Singh et al. (2008) estimated generalized-exponential by maximum likelihood and obtained Bayes estimator using Lindley's expansion. Preda et al. (2010) used maximum likelihood and Bayesian methods to estimate the modified Weibull by Lindley's expansion under various loss functions.

## 1.8 Objectives

The main objectives of this thesis can be summarized as follows:

1- Estimation the parameters of generalized exponential distribution for survival data using Bayesian estimation method.
2- Estimation the parameters of generalized exponential distribution for survival data using Non-Bayesian estimation method.
3- A comparison study is made between Bayesian and Non-Bayesian estimators to determine the best method that can be used to estimate the parameters of the generalized exponential distribution.

## 1.9 Thesis Structure

This thesis consists of four chapters: Basic concepts and an introduction to survival analysis are described In Chapter 1, Chapter 2 deals with analysis of non-Bayesian inference for survival data. Bayes' theorem and Bayesian survival analysis are discussed in Chapter 3. A comparison study through a simulation study and real data analysis followed by conclusion of this study in the final chapter.

# Chapter Two

# Non-Bayesian Estimation

## 2.1 Maximum Likelihood Estimation

In this section we discuss the maximum likelihood estimators of the unknown parameters of the GE model. The GE can be used in situation where a skewed distribution for a non-negative random variable is needed. The two parameters of a GE distribution represent the scale and the shape parameters and because of the scale and shape parameters, it has quite a bit of flexibility to analyze any positive real data. Due to the simple structure of its distribution function, the GE can be used quite effectively in analyzing any lifetime data, especially in the presence of censoring or if the data is grouped. It has increasing as well as decreasing failure rate depending on the shape parameter. The distribution function of GE is:

$$F(t;\theta,p) = \left\{1 - e^{-\theta t}\right\}^{p} \qquad ;\theta\ ,p\ ,t > 0 \tag{2.1}$$

The probability density function of GE is written as:

$$f(t;\theta,p) = p\theta\left\{1 - e^{-\theta t}\right\}^{p-1} e^{-\theta t} \qquad ;\theta\ ,p\ ,t > 0 \tag{2.2}$$

where $\theta$ is the scale parameter and $p$ is the shape parameter.

Let $T = (t_1, \dots, t_n)$ be the set of $n$ random lifetimes with respect to the generalized exponential distribution, with $p$ and $\theta$ as the parameters, the survival function is:

$$S(t;\theta,p) = 1 - \left\{1 - e^{-\theta t}\right\}^{p} \tag{2.3}$$

Let the $GE$ distribution with the shape parameter $p$ and the scale parameter $\theta$ be denoted by $GE(\theta,p)$.

Since $T = (t_1, \dots, t_n)$ is the set of $n$ random lifetimes from the generalized exponential distribution with parameters $\theta$ and $p$. The likelihood function is

$$L(t;\theta,p)=\prod_{i=1}^{n} f(t_i)=\prod_{i=1}^{n}\left[p\theta\{1-e^{-\theta t_i}\}^{p-1}e^{-\theta t_i}\right] \tag{2.4}$$

Taking log of equation (2.4) we get

$$l(t;\theta,p)=n\ln p+n\ln\theta+(p-1)\sum_{i=1}^{n}\ln\left(1-e^{-\theta t_i}\right)-\theta\sum_{i=1}^{n}t_i \tag{2.5}$$

Differentiate $l(t;\theta,p)$ partially with respect to the unknown parameters we obtain ;

$$\frac{\partial l}{\partial p}=\frac{n}{p}+\sum_{i=1}^{n}\ln\left(1-e^{-\theta t_i}\right)\ =0 \tag{2.6}$$

and

$$\frac{\partial l}{\partial \theta}=\frac{n}{\theta}+(p-1)\sum_{i=1}^{n}\frac{t_i e^{-\theta t_i}}{1-e^{-\theta t_i}}-\sum_{i=1}^{n}t_i=0 \tag{2.7}$$

Solve (2.6) for $p$ we get

$$p=\frac{-n}{\sum_{i=1}^{n}\ln(1-e^{-\theta t_i})} \tag{2.8}$$

Substituting (2.8) into (2.7) , we obtain

$$\frac{n}{\theta}+\left(\frac{-n}{\sum_{i=1}^{n}\ln(1-e^{-\theta t_i})}-1\right)\sum_{i=1}^{n}\frac{t_i e^{-\theta t_i}}{1-e^{-\theta t_i}}-\sum_{i=1}^{n}t_i=0 \tag{2.9}$$

Since it is difficult to solve equation (2.9) for $\theta$. Newton-Raphson method is employed in order to estimate the unknown parameters.

By Newton-Raphson method, $\theta$ can be estimated by iteration as follows:

$$\theta_{m+1}=\theta_m-\frac{g(\theta_m)}{g'(\theta_m)}$$

where,

$$g(\theta)=\frac{n}{\theta}+\left(\frac{-n}{\sum_{i=1}^{n}\ln(1-e^{-\theta t_i})}-1\right)\sum_{i=1}^{n}\frac{t_i e^{-\theta t_i}}{1-e^{-\theta t_i}}-\sum_{i=1}^{n}t_i$$

and

$$g'(\theta) = -\frac{n}{\theta^2} + \left(\frac{-n}{\sum_{i=1}^{n} \ln(1 - e^{-\theta t_i})} - 1\right) \sum_{i=1}^{n} \frac{-t_i^2 e^{-\theta t_i}}{(1 - e^{-\theta t_i})^2}$$

$$+ n \left[ \frac{\sum_{i=1}^{n} \frac{t_i e^{-\theta t_i}}{1 - e^{-\theta t_i}}}{\sum_{i=1}^{n} \ln(1 - e^{-\theta t_i})} \right]^2$$

This method need an initial value for $\theta$ , say $\theta_0$ ,

To prove that $\hat{\theta}$ make likelihood function is maximum, substituting (2.8) into (2.5) we obtain

$$l(t; \theta, p) = C + n \ln \theta - n \ln \left( - \sum_{i=1}^{n} \ln\left(1 - e^{-\theta t_i}\right) \right) - \sum_{i=1}^{n} \ln\left(1 - e^{-\theta t_i}\right) - \theta \sum_{i=1}^{n} t_i \tag{2.10}$$

where $C$ is a constant independent of $\theta$.

$l(t; \theta, p)$ is need to prove it as unimodal function of $\theta$. Prove it is out of theorem in an article entitled " Generalized Exponential Distribution: Statistical Inferences" published by Gupta and Kundu (2003).

**Theorem**: If $n = 1$,

$$g(\theta) = l(t; \theta, p) = C + \ln \theta - \ln\left(-\ln\left(1 - e^{-\theta t}\right)\right) - \ln\left(1 - e^{-\theta t}\right) - \theta t$$

is unimodal function of $\theta$.

**Proof**: Note that it is equivalent to prove that

$$g(\theta) = \ln \theta - \ln\left(-\ln\left(1 - e^{-\theta}\right)\right) - \ln\left(1 - e^{-\theta}\right) - \theta$$

is unimodal function of $\theta$. Consider the second derivative of $g(\theta)$,

$$g''(\theta) = \left[ \frac{e^{-2\theta}}{(1 - e^{-\theta})^2 [\ln(1 - e^{-\theta})]^2} + \frac{e^{-\theta}}{(1 - e^{-\theta})^2 \ln(1 - e^{-\theta})} \right]$$

$$+ \left[ \frac{e^{-\theta}}{(1 - e^{-\theta})^2} - \frac{1}{\theta^2} \right]$$

say, $g''(\theta) = g_1(\theta) + g_2(\theta)$. We show that $g_1(\theta) \leq 0$ and $g_2(\theta) \leq 0$. It would imply that $g(\theta)$ is a concave function. Now the result follows from the fact that $g(0) \to -\infty$ and $g(\infty) \to -\infty$, therefore $g(.)$ has to be unimodal. Therefore, the proof will be complete if we can show that $g_1(\theta) \leq 0$ and $g_2(\theta) \leq 0$.

Now to prove $g_1(\theta) \leq 0$, it is enough to prove that for $\theta \geq 0$

$$\frac{e^{-\theta}}{\ln(1-e^{-\theta})} + 1 \geq 0 \Leftrightarrow u(x) = x + \ln(1-x) \leq 0\,; \quad 0 \leq x \leq 1$$

Since $u(x)$ is a decreasing function and $u(0) = 0$, implies $u(x) \leq 0$.

Now to prove $g_2(\theta) \leq 0$, it is enough to prove that for $\theta \geq 0$

$$\frac{e^{-\frac{\theta}{2}}}{(1-e^{-\theta})} \leq \frac{1}{\theta} \Leftrightarrow 1 - e^{-\theta} - \theta e^{-\frac{\theta}{2}} \geq 0 \Leftrightarrow u(\theta) = e^{\frac{\theta}{2}} - e^{-\frac{\theta}{2}} - \theta \geq 0$$

Since $u'(\theta) = \frac{1}{2} e^{-\frac{\theta}{2}} \left(1 - e^{-\frac{\theta}{2}}\right)^2 \geq 0$ and $u(0) = 0$, therefore $u(\theta) \geq 0$.

So $g_2(\theta) \leq 0$. ■

Maximizing (2.10) using Newton-Raphson method.

## 2.2 Fisher Information matrix

Since the MLEs of the unknown parameters $p, \theta$ cannot be in closed forms, it is not easy to derive the exact distributions of the MLEs. We can derive the asymptotic confidence intervals of these parameters when $p > 0$, and $\theta > 0$. The large sample approach is to assume that the MLE $(\hat{p}, \hat{\theta})$ are approximately bivariate normal with mean $(p, \theta)$ and covariance matrix $\mathrm{I}(p, \theta)$.

where $\mathrm{I}(p, \theta)$ is the Fisher Information matrix , defined as:

$$\mathrm{I}(p,\theta) = -\frac{1}{n}\begin{bmatrix} E\left(\frac{\partial^2 l}{\partial p^2}\right) & E\left(\frac{\partial^2 l}{\partial p \partial \theta}\right) \\ E\left(\frac{\partial^2 l}{\partial \theta \partial p}\right) & E\left(\frac{\partial^2 l}{\partial \theta^2}\right) \end{bmatrix}$$

It is the variance of the score, or the expected value of the observed information, it is used to calculate the covariance matrices associated with maximum-likelihood estimates.

The elements of the Fisher Information matrix are as follows,

For $p > 2$;

$$E\left(\frac{\partial^2 l}{\partial p^2}\right) = -\frac{n}{p^2}$$

$$\begin{aligned} E\left(\frac{\partial^2 l}{\partial p \partial \theta}\right) &= nE\left(\frac{te^{-t\theta}}{(1-e^{-t\theta})}\right) \\ &= nE\left(t\left[\frac{1}{1-e^{-t\theta}} - 1\right]\right) = \frac{n}{\theta}\left[\frac{p}{p-1}\big(\psi(p) - \psi(1)\big) - \big(\psi(p+1) - \psi(1)\big)\right] \end{aligned}$$

$$\begin{aligned} E\left(\frac{\partial^2 l}{\partial \theta^2}\right) &= -n\left[\frac{1}{\theta^2} + (p-1)E\left(\frac{t^2 e^{-t\theta}}{(1-e^{-t\theta})^2}\right)\right] \\ &= -n\left[\frac{1}{\theta^2} + (p-1)E\left(\frac{t^2}{(1-e^{-t\theta})^2} - \frac{t^2}{(1-e^{-t\theta})}\right)\right] \\ &= -\frac{n}{\theta^2}\left[1 + \frac{p(p-1)}{p-2}\big(\psi'(1) - \psi'(p-1)\big) + \Big(\big(\psi(p-1) - \psi(1)\big)^2\Big)\right] \\ &\quad -\frac{np}{\theta^2}\left[\big(\psi'(1) - \psi'(p)\big) + \Big(\big(\psi(p) - \psi(1)\big)^2\Big)\right] \end{aligned}$$

For $0 < p \leq 2$,

$$E\left(\frac{\partial^2 l}{\partial p^2}\right) = -\frac{n}{p^2}$$

$$E\left(\frac{\partial^2 l}{\partial p \partial \theta}\right) = \frac{np}{\theta}\int_0^\infty te^{-2t}(1-e^{-t})^{p-2}dx \ < \infty$$

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = -\frac{n}{\theta^2} - \frac{np(p-1)}{\theta^2}\int_0^\infty t^2 e^{-2t}(1-e^{-t})^{p-3}dx \ < \infty$$

For $p > 0$, the GE family satisfies all the regularity conditions and therefore, we have the following result ; which is published in article entitled " Generalized Exponential Distribution: Statistical Inferences" by Gupta and Kundu (2003);

**Theorem**: For $p > 0$ the maximum-likelihood estimators, $(\hat{p}, \hat{\theta})$, of $(p, \theta)$ are consistent and $\sqrt{n}(\hat{p} - p, \hat{\theta} - \theta)$ is asymptotically normal with mean vector zero and dispersion matrix $I^{-1}$ .

We provide the elements of the Fisher Information matrix, when the data are type I censored. Note that it is possible to obtain the Fisher Information matrix in terms of

$$E(T/T \leq L) = \psi(p, L) \ \ (say) \ \ and$$

$$E(T^2/T \leq L) = \tilde{\psi}(p, L) \ \ (say)$$

where $T$ is a $GE(p, 1)$ random variable. The explicit expressions of $\psi(p, L)$ and $\tilde{\psi}(p, L)$ are as follows;

$$\psi(p, L) = \frac{p}{(1-e^{-L})^p}\sum_{j=0}^{\infty}(-1)^j c(p-1, j)\left[\frac{1}{(j+1)^2}\left(1-e^{-(j+1)L}\right) - \frac{Le^{-(j+1)L}}{j+1}\right]$$

and

$$\tilde{\psi}(p, L) = \frac{p}{(1-e^{-L})^p}\sum_{j=0}^{\infty}(-1)^j c(p-1, j)\left[\frac{2}{(j+1)^2}\left(1-e^{-(j+1)L}\right) - \frac{2Le^{-(j+1)L}}{(j+1)^2} - \frac{L^2 e^{-(j+1)L}}{j+1}\right]$$

where $c(p,i) = \frac{p(p-1)\ldots(p-i+1)}{i!}$.

Then

$$E\left(\frac{\partial^2 l}{\partial p^2}\right) = -\frac{r}{p^2} + g_1^{'}(p,\theta)$$

$$E\left(\frac{\partial^2 l}{\partial p \partial \theta}\right) = \frac{1}{\theta}\sum_{i\in D}\left[\frac{p}{(p-1)(1-e^{-L_i\theta})}\psi(p-1,L_i\theta) - \psi(p,L_i\theta)\right] + g_1^{'}(p,\theta)$$

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = -\frac{r}{\theta^2} + g_2^{'}(p,\theta) + \frac{(p-1)}{\theta^2} \times \sum_{i\in D}\left[\frac{p}{(p-2)(1-e^{-L_i\theta})^2}\tilde{\psi}(p-2,L_i\theta) - \frac{p}{(p-1)(1-e^{-L_i\theta})} \times \tilde{\psi}(p-1,L_i\theta)\right]$$

Therefore, the survival function can be obtained as

$$\hat{S}(t) = 1 - \left\{1 - e^{-\hat{\theta}t}\right\}^{\hat{p}} \tag{2.11}$$

where $\hat{\theta}$ and $\hat{p}$ are the maximum likelihood estimates of the parameters.

# Chapter Three
# Bayesian Estimation

## 3.1 Bayes' Theorem

The foundation of Bayesian statistics is Bayes' theorem. Suppose we observe a random variable $y$ and wish to make inferences about another random variable $\theta$, where $\theta$ is drawn from some distribution $p(\theta)$. From the definition of conditional probability,

$$Pr(\theta/y) = \frac{Pr(y,\theta)}{Pr(y)} \tag{3.1}$$

Again from the definition of conditional probability, we can express the joint probability by conditioning on $\theta$ to give

$$Pr(y,\theta) = Pr(y/\theta)\, Pr(\theta) \tag{3.2}$$

Substituting (3.2) into (3.1) together gives Bayes' theorem:

$$Pr(\theta/y) = \frac{Pr(y/\theta)\, Pr(\theta)}{Pr(y)} \tag{3.3}$$

With $n$ possible outcomes $(\theta_1, \dots\ , \theta_n)$,

$$Pr(\theta_j/y) = \frac{Pr(y/\theta_j)\, Pr(\theta_j)}{Pr(y)} = \frac{Pr(y/\theta_j)}{\sum_{i=1}^{n} Pr(\theta_i)\ Pr(y/\theta_i)} \tag{3.4}$$

$Pr(\theta)$ is the *prior distribution* of the possible $\theta$ values, while $Pr(\theta/y)$ is the *posterior distribution* of $\theta$ given the observed data $y$.

The continuous multivariate version of Bayes' theorem is:

$$p(\Theta/y) = \frac{p(y/\Theta)\, p(\Theta)}{p(y)} = \frac{p(y/\Theta)\, p(\Theta)}{\int p(y/\Theta)\, d\Theta} \tag{3.5}$$

where $\Theta = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(k)})$ is a vector of $k$ (potentially) continuous variables. As with the univariate case, $p(\Theta)$ is the assumed prior distribution of the unknown parameters, while $p(\Theta/\mathrm{y})$ is the posterior distribution given the prior $p(\Theta)$ and the data $y$.

### 3.1.1 From Likelihood to Bayesian analysis

The method of maximum likelihood and Bayesian analysis are closely related. Suppose $\ell(\Theta/x)$ is the assumed likelihood function. Under ML estimation, we would compute the mode (the maximal value of $\ell$, as a function of $\Theta$ given the data $x$) of the likelihood function, and use the local curvature to construct confidence intervals. Hypothesis testing follows using likelihood-ratio (LR) statistics. The strengths of ML estimation rely on its large-sample properties, namely that when the sample size is sufficiently large, we can assume both normality of the test statistic about its mean and that LR tests follow $\chi^2$ distributions. These nice features don't necessarily hold for small samples.

An alternate way to proceed is to start with some initial knowledge/guess about the distribution of the unknown parameter(s), $p(\Theta)$. From Bayes' theorem, the data (likelihood) augment the prior distribution to produce a posterior distribution,

$$\begin{aligned} p(\Theta/x) &= \frac{1}{p(x)} \,.\, p(x/\Theta)\, p(\Theta) \\ &= \binom{normalizing}{constant} . p(x/\Theta)\, p(\Theta) \\ &= constant.likehood.prior \end{aligned} \tag{3.6}$$

as $p(x/\Theta) = \ell(\Theta/x)$ is just the likelihood function.$1/p(x)$ is a constant (with respect to $\Theta$), because our concern is the distribution over $\theta$. Because of this, the posterior distribution is often written as

$$\mathrm{p}(\Theta/\mathrm{x}) \propto \ell(\Theta/x)\mathrm{p}(\Theta) \tag{3.7}$$

where the symbol $\propto$ means "proportional to" (equal up to a constant). Note that the constant $p(x)$ normalizes $p(x/\Theta)\,.\,p(\Theta)$ to one, and hence can be obtained by integration,

$$p(x) = \int_{\Theta} p(x/\Theta) \, . \, p(\Theta) \, d\Theta \tag{3.8}$$

The dependence of the posterior on the prior (which can easily be assessed by trying different priors) provides an indication of how much information on the unknown parameter values is contained in the data. If the posterior is highly dependent on the prior, then the data likely has little signal, while if the posterior is largely unaffected under different priors, the data are likely highly informative. To see this, taking logs on Equation (3.6) (and ignoring the normalizing constant) gives

$$log(posterior) = log(likelihood) + log(prior) \tag{3.9}$$

### 3.1.2 Marginal Posterior Distributions

Often, only a subset of the unknown parameters is really of concern to us, the rest being nuisance parameters that are really of no concern to us. A very strong feature of Bayesian analysis is that we can remove the effects of the nuisance parameters by simply integrating them out of the posterior distribution to generate a marginal posterior distribution for the parameters of interest.

The marginal posterior may involve several parameters (generating joint marginal posteriors).Write the vector of unknown parameters as $\Theta = (\Theta_1, \Theta_n)$, where $\Theta_n$ is the vector of nuisance parameters. Integrating over $\Theta_n$ gives the desired marginal as

$$p(\Theta_1/y) = \int_{\Theta_n} p(\Theta_1, \Theta_n/y) \; d\Theta_n \tag{3.10}$$

### 3.1.3 Summarizing the posterior distribution

How do we extract a Bayes estimator for some unknown parameter ? If our mindset is to use some sort of point estimator (as is usually done in classical statistics), there are a

number of candidates. We could follow maximum likelihood and use the mode of the distribution (its maximal value), with

$$\hat{\theta} = \max_{\theta}[p(\theta/x)] \tag{3.11}$$

We could take the expected value of $\theta$ given the posterior,

$$\hat{\theta} = E[\theta/x] = \int \theta p(\theta/x) d\theta \tag{3.12}$$

Another candidate is the median of the posterior distribution, where the estimator satisfies $Pr(\theta > \hat{\theta}/x) = Pr(\theta < \hat{\theta}/x) = 0.5$ , hence

$$\int_{\hat{\theta}}^{+\infty} p(\theta/x) d\theta = \int_{-\infty}^{\hat{\theta}} p(\theta/x) d\theta = \frac{1}{2} \tag{3.13}$$

However, using any of the above estimators, or even all three simultaneously, loses the full power of a Bayesian analysis, as the full estimator is the entire posterior density itself . If we cannot obtain the full form of the posterior distribution, it may still be possible to obtain one of the three above estimators. However, as we will see later, we can generally obtain the posterior by simulation using Gibbs sampling, and hence the Bayes estimate of a parameter is frequently presented as a frequency histogram from (Gibbs) samples of the posterior distribution.

### 3.1.4 The choice of a prior

Obviously, a critical feature of any Bayesian analysis is the choice of a prior. The key here is that when the data have sufficient signal, even a bad prior will still not greatly influence the posterior. In a sense, this is an asymptotic property of Bayesian analysis in that all but pathological priors will be overcome by sufficient amounts of data. If the posterior is highly dependent on the prior, then the data (the likelihood function) may not contain

sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contain significant information.

The location of a parameter (mean or mode) and its precision (the reciprocal of the variance) of the prior is usually more critical than its actual shape in terms of conveying prior information. The shape of the prior distribution is often chosen to facilitate calculation of the prior, especially through the use of conjugate priors that, for a given likelihood function, return a posterior in the same distribution family as the prior.

### 3.2 Markov Processes

Arguably the simplest type of dependency that can be exhibited by the variables of a random process is the one found in first-order Markov processes: each variable $S_i$ depends only the preceding one, $S_{i-1}$; moreover, conditionally on $S_{i-1}$, it is independent of all other preceding variables. Formally, the process is called a first-order Markov process when

$$PS_n(s_n/s_{n-1}, s_{n-2}, \dots, s_1) = PS_n(s_n/s_{n-1}) \tag{3.14}$$

The joint probability function of any process (of any set of random variables) can be factored as

$$PS_1, \dots, S_n(s_1, \dots, s_n) = PS_n(s_n/s_{n-1}, \dots, s_1)PS_{n-1}(s_{n-1}/s_{n-2}, \dots, s_1) \dots PS_2(s_2/s_1)PS_1(s_1)$$

which is a trivial chain application of $p(A/B)\,p(B) = p(A, B)$. One of the most important consequence of the Markovianity of a process is that its factorization becomes simply

$$PS_1, \dots, S_n(s_1, \dots, s_n) = PS_n(s_n/s_{n-1})PS_{n-1}(s_{n-1}/s_{n-2}) \dots PS_2(s_2/s_1)PS_1(s_1) \tag{3.15}$$

Accordingly, a Markov process is completely characterized (i.e., it is possible to compute any joint probability function) once the initial probability function $S_1(s_1)$ , and the sequence of transition probability functions $PS_i(s_i/s_{i-1})$ are given.

Consider a Markov process such that each $S_i$ can take values on a finite set (the $i$th state space) $s_i = \{1,2,\dots,M_i\}$ (without loss of generality here identified with sets of integers; notice that these are merely labels). In this case, the process is called a finite Markov process, $PS_1(s_1)$ is a set of $M_1$ probability values, and the transition probability functions $PS_i(s_i/s_{i-1})$ define $M_{i-1} \times M_i$ transition matrices $\mathrm{P}(i) = [P_{kl}(i)]$ according to

$$P_{kl}(i) = PS_i(s_i = l/s_{i-1} = k) \geq 0 \tag{3.16}$$

Given their meaning, these matrices must verify

$$\sum_{l=1}^{M_i} P_{kl}(i) = \sum_{l=1}^{M_i} PS_i(s_i = l/s_{i-1} = k) = 1 \tag{3.17}$$

and are called stochastic matrices. If everything in the previous definitions is index-invariant, i.e.,$S_i = S$ (the state space, of $M_i = M$ course with ) and $P(i) = P$, we have a so-called time-invariant or homogeneous Markov chain. If the probability function of variable $S_n$ is $\mathrm{P}S_n(s_n)$ , then that of the "next" variable, $S_{n+1}$, can easily be obtained by noting that

$$\begin{aligned} \mathrm{P}S_{n+1}(s_{n+1}) &= \sum_{s_n \in S_n} \mathrm{P}S_n, S_{n+1}(s_n, s_{n+1}) \\ &= \sum_{s_n \in S_n} \mathrm{P}S_{n+1}(s_{n+1}/s_n) PS_n(s_n) \end{aligned} \tag{3.18}$$

(with integrals taking place of the summations in the case of continuous state spaces). If we are the presence of a time invariant chain (or process), then a probability function that remains unchanged from index $n$ to the next index $n + 1$, i.e., such that

$$\mathrm{P}S_{n+1}(b) = \sum_{s_n \in S_n} \mathrm{P}S_{n+1}(b/s_n) PS_n(s_n) = PS_n(b) \tag{3.19}$$

(again, with integrals instead of summations in the case of continuous state spaces), is called a stationary distribution.

### 3.2.1 Bayesian-MCMC and Gibbs Sampling

**Bayesian Methods**

Traditional maximum likelihood approach delivers only point estimate and associated asymptotic standard error estimates for the model parameters. This motivates the use of Bayesian analysis, as the development of computing power and improved scope for estimation via iterative sampling methods. Bayesian analysis of data in health, social and physical sciences has been greatly facilitated in the last decade. The new estimation methods Markov Chain Monte Carlo (MCMC) may be used to augment the data and this provides and analogue to the classical Expectation Maximization (EM) method.

*Priors for Parameters*: In classical inference the sample are considered as random while the parameter $\theta$ considered as fixed. In Bayesian analysis, parameters themselves follow a probability distribution, and is summarized in a prior distribution $[\theta]$ before considering the data at hand. In many situation, existing knowledge may be difficult to summarize in the form of an informative prior; resort is made to non-informative priors such as flat priors (Uniform Distribution). However some priors which are improper (don't integrate to 1 over their ranges) may add to identifiability problems. Minimally informative priors (just proper priors) are preferred such as normal distribution with mean zero and large variance.

*Posterior and likelihood*: In maximum likelihood approaches, inferences are based on the likelihood of the data alone. In Bayesian models, the likelihood of the observed data $y$

given parameter $\theta$, i.e. $f(y/\theta)$ or $L(\theta/y)$ is used to modify the prior $[\theta]$ , with the updated posterior density $[\theta/y]$. The relation can be written as

$$[\theta/y] = \frac{f(y/\theta)[\theta]}{m(\theta)} = \frac{f(y/\theta)[\theta]}{\int f(y/\theta)[\theta]\, d\theta} \tag{3.20}$$

and we can simplify it as

$$[\theta/y] \propto f(y/\theta)[\theta] \tag{3.21}$$

*Sampling Parameters*: In most of the situations, with many parameter $\theta$ and with possibly non-conjugate priors, the goal is to summarize the marginal posterior of a particular parameter $\theta_k$ given the data. This involves integrating out all parameters but this one

$$P(\theta_k/y) \propto \int \dots \int f(\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_p/y) d\theta_1 \dots d\theta_{k-1} d\theta_{k+1} \dots d\theta_p \tag{3.22}$$

Such integrations using classic approaches involved demanding methods such as numerical quadrature.

**Markov Chain Monte Carlo (MCMC)**

A major limitation towards more widespread implementation of Bayesian approaches is that obtaining the posterior distribution often requires the integration of high-dimensional functions. MCMC methods aim to simulate direct draws from some complex distribution of interest MCMC approaches are so-called because one uses the previous sample values to randomly generate the next sample value, generating a Markov chain as the transition probabilities between sample values are only a function of the most recent sample value. The Gibbs sampler is very widely applicable to a broad class of Bayesian problems it has sparked a major increase in application of Bayesian analysis. MCMC methods have their roots in Metropolis algorithm Metropolis et al. (1953), which computes complex integrals

by expressing them as expectations for some distribution and then estimate this expectation by drawing samples from that distribution. The Gibbs sampler, Geman and Geman (1984) is a special case of Metropolis-Hastings sampling.

**Monte Carlo Integration**

The original Monte Carlo approaches was a method developed by physicists to use random number generation to compute integrals. Suppose we wish to compute a complex integral

$$\int_a^b h(x)dx\,,$$

if we can decompose $h(x)$ into the production of a function $f(x)$ and a probability density function $P(x)$ defined over the interval $(a, b)$, then

$$\int_a^b h(x)dx = \int_a^b f(x)P(x)dx = E_{p(x)}[f(x)] \tag{3.23}$$

that is, the integral may be expressed as an expectation of $f(x)$ over the density $P(x)$. Thus, if we draw a large number $x_1, \ldots, x_n$ of random variables from the density $P(x)$, then

$$\int_a^b h(x)dx = E_{p(x)}[f(x)] \simeq \frac{1}{n}\sum_{i=1}^{n} f(x_i) \tag{3.24}$$

this is referred to as Monte Carlo integration, this integration can be used to approximate posterior or marginal posterior distributions required for Bayesian analysis. Consider the integral $I(y) = \int f(y/x)\,P(x)dx$ which we approximate by

$$\hat{I}(y) = \frac{1}{n}\sum_{i=1}^{n} f(y/x_i) \tag{3.25}$$

where $x_i$ are draws from the density $P(x)$. The estimated Monte Carlo standard error is given by

$$SE^2\left[\hat{I}(y)\right] = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(f(y/x_i) - \hat{I}(y))^2\right) \quad (3.26)$$

**Markov Chains**

Before introducing the Metropolis-Hastings algorithm and Gibbs sampler, a few introductory comments on Markov Chains are in order. Let $X_t$ denote the value of a random variable at time $t$, and let the state space refer to the range of possible $X$ values. The random variable is a Markov process if the transition probabilities between different values in the state space depend only on the random variable's current state, that is

$$P(X_{t+1} = s_i/X_0 = s_k, \dots, X_t = s_i) = P(X_{t+1} = s_j/X_t = s_i) \quad (3.27)$$

Thus for a Markov random variable the only information about the past needed to the future is the current state of the random variable. Knowledge of the values of earlier states do not change the transition probability. A Markov chain refers to a sequence of random variables $(X_0, \dots, X_n)$ generated by a Markov process, a particular chain is defined most critically by its transition probabilities, $P(i,j) = P(i \to j)$ , which is the probability that a process at state space $s_i$ moves to state $s_j$ in a single step,

$$P(i,j) = P(i \to j) = P(X_{t+1} = s_j/X_t = s_i) \quad (3.28)$$

Let $\pi_j(t) = P(X_t = s_i)$ denote the probability that the chain is in state $j$ at time $t$, and let $\pi(t)$ denote the row vector of the state space probabilities at step $t$. We start the chain by specifying a starting vector $\pi(0)$. Often all the elements of $\pi(0)$ are zero except for a single element of 1, corresponding to the process starting in the particular state. As the chain progresses, the probability values gets spread out over the possible state space. The probability that the chain has state value $s_i$ at time $t+1$ is given by the Chapman-

Kolomogrov equation, which sums over the probability of being in a particular state at the current step and the transition probability from that state into state $s_i$

$$\begin{aligned}\pi_i(t+1) &= P\left(X_{t+1} = s_j\right) \\ &= \sum_k P(X_{t+1} = s_i / X_t = s_k) P(X_t = s_k) \\ &= \sum_k P(k \to i)\pi_k(t) = \sum_k P(k,i)\pi_k(t)\end{aligned} \tag{3.29}$$

Finally, a Markov chain is said to be irreducible if there exists a positive integer such that $\left(P_{ij}^{n_{ij}} = P\left(X_{t+n} = s_j / X_t = s_i\right)\right) > 0$ for all $i$ and $j$.

That is, all states communicates with each other, as one can always go from any state to any other state. Likewise, a chain is said to be a periodic when the number of steps required to move between two states (say $x$ and $y$) is not required to be multiple of some integer. But another way, the chain is not forced into some cycle of fixed length between certain states.

**Metropolis-Hastings Algorithm**

One problem with applying Monte Carlo integration is in obtaining samples from some complex probability distribution $p(x)$. Attempts to solve this problem are the roots of MCMC methods. In particular, they trace to attempt by mathematical physicists to integrate very complex functions by random sampling Metropolis et al. (1953), Hastings (1970), and the resulting Metropolis-Hastings (M-H) algorithm. Suppose our goal is to draw samples from some distribution $p(\theta)$ where $p(\theta) = \frac{f(\theta)}{K}$ where the normalizing constant $K$ may not be known, and very difficult to compute. The Metropolis algorithm generates a sequence of draws from this distribution as follows:

**Algorithm 3.1 M-H Algorithm**

**Step 1**, Start with any initial value $\theta_0$ satisfying $f(\theta_0) > 0$.

**Step 2**, Using current $\theta$ value, sample a candidate point $\theta^*$ from some jumping distribution $q(\theta_1, \theta_2)$, which is the probability of returning a value of $\theta_2$ given a previous value of $\theta_1$. This distribution is also referred to as the proposal or candidate-generating distribution. The only restriction on the jump density in the Metropolis algorithm is that it is symmetric, $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$.

**Step 3**, Given the candidate point $\theta^*$, calculate the ratio of the density at the candidate $(\theta^*)$ and current $(\theta_{t-1})$ points

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})} \tag{3.30}$$

**Step 4**, If the jump increases the density $(\alpha > 1)$, then accept the candidate point ( $set\ \theta_t = \theta^*$) and return to step 2.

**Step 5**, If the jump decreases the density $(\alpha < 1)$, then with probability $\alpha$ accept the candidate point, else reject it and return to step 2.

This algorithm generates a Markov chain $(\theta_0, \theta_1, \dots, \theta_k, \dots)$, as the transition probabilities from $\theta_t$ to $\theta_{t+1}$ depends only on $\theta_t$ and not $(\theta_0, \dots, \theta_{t-1})$. Following a sufficient burn-in period (of, say, $k$ steps), the chain approaches its stationary distribution and samples from the vector $(\theta_{k+1}, \dots, \theta_{k+n})$ are samples from $p(x)$. Hasting (1970) generalized the Metropolis algorithm by using an arbitrary transition probability function $q(\theta_1, \theta_2) = P(\theta_1 \to \theta_2)$ and setting the acceptance probability for a candidate point as

$$\alpha = \min\left\{\frac{f(\theta^*)q(\theta^*, \theta_{t-1})}{f(\theta_{t-1})q(\theta_{t-1}, \theta^*)}, 1\right\} \tag{3.31}$$

Assuming that the proposal distribution is symmetric in M-H, recovers the original Metropolis algorithm.

**Convergence Diagnostics**

A key issue in the successful implementation of M-H or any other MCMC sampler is the number of runs (steps) until the chain approaches stationary which is the length of the

burn-in period, typically the first 1000 to 5000 elements are thrown out, and then convergence tests can be used to assess whether stationary distribution has indeed been reached. Adjacent members from a M-H sequence are expected to be positively correlated, and we can quantify the nature of this correlation by using an autocorrelation function. Consider a sequence $(\theta_1, \dots, \theta_n)$ of length $n$. Correlations can occur between adjacent members $(\rho(\theta_t, \theta_{t+1}) \neq 0)$ and more generally between more distant members $(\rho(\theta_t, \theta_{t+k}) \neq 0)$ . The $k$th order autocorrelation $\rho_k$ can be estimated by

$$\hat{\rho}_k = \frac{Cov(\theta_t, \theta_{t+k})}{Var(\theta_t)} = \frac{\sum_{t=1}^{n-k}(\theta_t - \bar{\theta})(\theta_{t-k} - \bar{\theta})}{\sum_{t=1}^{n-k}(\theta_t - \bar{\theta})^2}, \ \ where\ \bar{\theta} = \frac{1}{n}\sum_{t=1}^{n}\theta_t \tag{3.32}$$

An important result from the theory of time series analysis is that if the $\theta_t$ are from a stationary and correlated process, correlated draws still provide an unbiased picture of the distribution provided the sample size is sufficiently large. Some indication of the required sample size comes from the theory of a first order autoregressive process (or $AR_1$), where

$$\theta_t = \mu + \alpha(\theta_{t-1} - \mu) + \epsilon \tag{3.33}$$

where $\epsilon$ is white noise, that is $\epsilon \sim N(0\,, \sigma^2)$. Here $\rho_1 = \alpha$ and the $k$th order autocorrelation is given by $\rho_k = \rho_1^k$. Under this process $(AR_1)$, $E(\bar{\theta}) = \mu$ with standard error

$$SE(\bar{\theta}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho_1}{1-\rho_1}} \tag{3.34}$$

the first ratio is the standard error for white noise, while the second ratio is the sample size inflation factor, or SSIF, which shows how the autocorrelation inflates the sampling variance. One strategy for reducing autocorrelation is thinning the output, storing only every $m$th point after the burn-in period.

### 3.2.2 The Gibbs Sampling

The Gibbs sampler introduced in the context of image processing by Geman and Geman (1984), is a special case of M-H sampling wherein the random value is always accepted($i.e.\ \alpha = 1$). The task remains to specify how to construct a Markov Chain whose values converge to the target distribution. The key to the Gibbs sampler is that one only considers univariate conditional distribution − the distribution when all of the random variables but one are assigned fixed values. Such conditional distributions are far easier to simulate than complex joint distributions and usually have simple forms, such as normal, inverse $\chi^2$ or other common prior distributions. Thus, one simulates $n$ random variables sequentially from the $n$ univariate conditionals rather than generating a single $n$-dimensional vector in a single pass using the full joint distribution.

To introduce the Gibbs sampler, consider a bivariate random variable $(x, y)$, and suppose we wish to compute one or both marginals, $p(x)$ and $p(y)$. The idea behind the sampler is that it is far easier to consider a sequence of conditional distributions, $p(x/y)$ and $p(y/x)$, than it is to obtain the marginal by integration of the joint density $p(x, y)$, e.g. $p(x) = \int p(x, y)dy$. The sampler start with some initial value $y_0$ for $y$ and obtain $x_0$ by generating a random variable from the conditional distribution $p(x/y = y_0)$. The sampler then uses $x_0$ to generate a new value of $y_1$, drawing from the conditional distribution based on the value of $x_0$, $p(y/x = x_0)$. The sampler proceeds as follows:

$$x_i \sim p(x/y = y_{i-1}), y_i \sim p(y/x = x_i) \tag{3.35}$$

Repeating this process $k$ times, generates a Gibbs sequence of length $k$, where a subset of points $(x_j, y_j)$ for $1 \leq j \leq m < k$ are taken as the simulated draws from the full joint distribution. One iteration of all the univariate distributions is often called a scan of the sampler. To obtain desired total of $m$ sample points, one samples the chain (i) after a

sufficient burn-in to remove the effects of the initial sampling values and (ii) at set time points (say every $n$ samples) following the burn-in. The Gibbs sequence converges to a stationary (equilibrium) distribution that is independent of the starting values, and by construction this stationary distribution is the target distribution we were trying to simulate. When more than two variables are involved, the sampler is extended in the obvious fashion. In particular, the value of the $k$th variable is drawn from the distribution $p(\theta^{(k)}/\Theta^{(-k)}$ where $\Theta^{(-k)}$ denotes a vector containing all off the variables but $k$. Thus, during the $i$th iteration of the sample, to obtain the value of $\theta_i^{(k)}$ we draw from the distribution

$$\theta_i^{(k)} \sim p(\theta^{(k)}/\theta^{(1)} = \theta_i^{(1)}, \dots, \theta^{(k-1)} = \theta_i^{(k-1)}, \theta^{(k+1)} = \theta_{i-1}^{(k+1)}, \dots, \theta^{(n)} = \theta_{i-1}^{(n)} \quad (3.36)$$

### 3.3 Bayesian Estimation

In Bayesian analysis, the parameter of interest is always considered to be a random variable with a prior distribution. The prior distribution is the distribution of the parameter before any data is observed. The selection of prior distribution is most often than not based on the type of prior information that is available to us. When we have little or no information about the parameter, a non-informative prior should be used else an informative prior. In analyzing data from medical, engineering, or biological studies, it is possible to obtain information with respect to similar studies in the past, and if that is even unattainable, information from an expert could be modeled to fit an appropriate prior distribution. This can be referred to as prior elicitation.

We let the two unknown parameters take on the gamma prior distributions by assuming that the hyper parameters are all known and greater than zero, that is, $a\,, b\,, c\,, d\ > 0$. Where, the joint density function for Gamma distribution is $Ga(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta}$ , and the marginal distribution function is $\int_0^{\infty}\int_0^{\infty} Ga(\alpha,\beta)d\alpha\, d\beta$. The gamma prior is assumed for this distribution because both the scale and shape parameters are greater than zero:

$$v_1(\theta) = \frac{a^b}{\Gamma(b)}\theta^{b-1}\exp(-\theta a) \Longrightarrow v_1(\theta) \propto \theta^{b-1}\exp(-\theta a)\,, \theta, a, b > 0,$$

$$v_2(p) = \frac{c^d}{\Gamma(d)}p^{d-1}\exp(-pc) \Longrightarrow v_2(p) \propto p^{d-1}\exp(-pc)\,,\ \ p, c, d > 0, \qquad (3.37)$$

Bayesian inference is based on the posterior distribution which is obtained by dividing the joint density function to the marginal distribution function as given below:

$$\pi^*(\theta, pt_i)\ \propto \frac{v_1(\theta)v_2(p)L(t_i;\theta,p)}{\int_0^{\infty}\int_0^{\infty} v_1(\theta)v_2(p)L(t_i;\theta,p)d\theta dp} \qquad (3.38)$$

Due to the complex nature of the posterior distribution given in (3.38), Lindley approximation is employed in order to estimate the unknown parameters.

The Bayesian estimator is considered under two loss functions. Since in drawing conclusions about the survival or duration of a living organism, an overestimation could be more detrimental to underestimation or vice versa, we have considered both asymmetric (general entropy) loss function and symmetric (squared error) loss function.

**3.3.1 Lindley Approximation**. Lindley (1980) suggested an asymptotic approximation to compute the ratio of two integrals of the form

$$\frac{\int w(\alpha)exp\{\ell(\alpha)\}d\alpha}{\int v(\alpha)exp\{\ell(\alpha)\}d\alpha} \tag{3.39}$$

where $\ell(\alpha)$ is the log likelihood and $w(\alpha), v(\alpha)$ are arbitrary functions of $\alpha$ , $v(\alpha)$ is the prior distribution for $\alpha$, and $w(\alpha) = u(\alpha).v(\alpha)$ with $u(\alpha)$ being some function of interest as seen in (3.40). The posterior expectation according to Sinha is

$$E\{u(\alpha)/x\} = \frac{\int u(\alpha)exp\{\ell(\alpha) + \rho(\alpha)\}d\alpha}{\int exp\{\ell(\alpha)\rho(\alpha)\}d\alpha} \tag{3.40}$$

where $\rho(\alpha) = log\{v(\alpha)\}$ and $\ell(\alpha)$ represent the log-likelihood function. Considering the Bayesian estimator under the squared error loss function, which is the posterior mean, the posterior expectation can be approximated asymptotically with respect to the two parameters by (3.42):

$$\begin{aligned}\hat{u} = u\left(\hat{\theta}, \hat{p}\right) + \frac{1}{2}[(u_{11}\sigma_{11}) + (u_{22}\sigma_{22})] + u_1\rho_1\sigma_{11} + u_2\rho_2\sigma_{22} \\ + \frac{1}{2}[(\ell_{30}u_1\sigma_{11}^2) + (\ell_{03}u_2\sigma_{22}^2)]\end{aligned} \tag{3.41}$$

$$\begin{aligned} u = \theta\,, \qquad u_1 = \frac{\partial u}{\partial \theta} = 1\,, \qquad u_{11} = 0, \\ u = p\,, \qquad u_2 = \frac{\partial u}{\partial p} = 1\,, \qquad u_{22} = 0, \\ \rho = \ln v_1(\theta) + \ln v_2(p) \end{aligned} \tag{3.42}$$

$$\rho_1 = \frac{b-1}{\theta} - a\,, \qquad \rho_2 = \frac{d-1}{p} - c$$

$$\sigma_{11} = (-\ell_{20})^{-1}, \qquad \sigma_{22} = (-\ell_{02})^{-1}$$

The second and third derivatives with respect to the scale and shape parameters are

$$\begin{aligned} \ell_{20} \quad &= -\frac{n}{\theta^2} - \sum_{i=1}^{n} \frac{(p-1)[(t_i)^2 \exp(-\theta t_i)]}{(1-\exp(-\theta t_i))} \\ &\quad - \sum_{i=1}^{n} \frac{(p-1)[(t_i)^2 (\exp(-\theta t_i))^2]}{(1-\exp(-\theta t_i))^2} \end{aligned} \tag{3.43}$$

$$\ell_{30} = \frac{2n}{\theta^3} - \sum_{i=1}^{n} \frac{(p-1)[(t_i)^3 \exp(-\theta t_i)]}{(1-\exp(-\theta t_i))}$$

$$+ \sum_{i=1}^{n} \frac{3(p-1)[(t_i)^3 (\exp(-\theta t_i))^2]}{(1-\exp(-\theta t_i))^2}$$

$$+ \sum_{i=1}^{n} \frac{2(p-1)[(t_i)^3 (\exp(-\theta t_i))^3]}{(1-\exp(-\theta t_i))^3}$$

$$\ell_{02} = -\frac{n}{p^2}\,, \qquad \ell_{03} = -\frac{2n}{p^3}$$

To estimate the survival function, under the squared error loss function, we let

$$u(S) = 1 - \left\{1 - e^{-\theta t}\right\}^p \tag{3.44}$$

where

$$u_1 = \frac{\partial u(S)}{\partial \theta}, \qquad u_{11} = \frac{\partial^2 u(S)}{\partial \theta^2},$$

$$u_2 = \frac{\partial u(S)}{\partial p}, \qquad u_{22} = \frac{\partial^2 u(S)}{\partial p^2}, \tag{3.45}$$

### 3.3.2 General Entropy Loss Function.

This is another useful asymmetric loss function that is used to determine whether there is overestimation or underestimation. The general entropy loss function is a generalization of the entropy loss function. The Bayes estimator $\hat{\alpha}_{BG}$ of $\alpha$ under the general entropy loss is

$$\hat{\alpha}_{BG} = [E_{\alpha}(\alpha^{-k})]^{-1/k} \tag{3.46}$$

Provided $E_{\alpha}(.)$ exists and is finite. The Bayes estimator for this loss function with respect to the parameters and the survival function are

$$E\{[(\theta)^{-k},(p)^{-k}]\} = \frac{\int\int u\,[(\theta)^{-k},(p)^{-k}]v_1(\theta)v_2(p)L(t_i;\theta,p)d\theta dp}{\int\int v_1(\theta)v_2(p)L(t_i;\theta,p)d\theta dp}$$

$$E\{[(S)^{-k}]\} = \frac{\int\int u\,[1-\{1-\exp(-\theta t)\}^{p}]^{-k}v_1(\theta)v_2(p)L(t_i;\theta,p)d\theta dp}{\int\int v_1(\theta)v_2(p)L(t_i;\theta,p)d\theta dp} \tag{3.47}$$

A similar Lindley approach is used for the general entropy loss function as in the squared error loss function, with

$$\begin{aligned} u &= (\theta)^{-k}, \qquad u_1 = \frac{\partial u}{\partial \theta} = -k(\theta)^{-k-1}, \\ u_{11} &= \frac{\partial^2 u}{\partial(\theta)^2} = -(-k^2-k)(\theta)^{-k-2}, \\ u &= (p)^{-k}, \qquad u_2 = \frac{\partial u}{\partial p} = -k(p)^{-k-1}, \\ u_{22} &= \frac{\partial^2 u}{\partial(p)^2} = -(-k^2-k)(p)^{-k-2}, \end{aligned} \tag{3.48}$$

For the general entropy loss function, the posterior expectation according to Lindley can be approximated by using (3.47)

$$\begin{aligned} \hat{u} = \Big\{ u(\hat{\theta},\hat{p}) &+ \frac{1}{2}[(u_{11}\sigma_{11}) + (u_{22}\sigma_{22})] + u_1\rho_1\sigma_{11} + u_2\rho_2\sigma_{22} \\ &+ \frac{1}{2}[(\ell_{30}u_1\sigma_{11}^2) + (\ell_{03}u_2\sigma_{22}^2)] \Big\}^{-1/k} \end{aligned} \tag{3.49}$$

# Chapter Four

# Simulation and Real case study

## 4.1 Simulation Study

In this simulation study we propose MCMC (Gibbs) sampling procedure to generate samples from the posterior density functions described in the previous chapters under the assumptions that $\theta$ and $p$ follow Gamma($a$; $b$) and Gamma($c$; $d$) respectively and they are independent. Now using Lindley approach or entropy loss functions described in the previous chapter, we propose the following scheme to generate $(\theta, p)$ from their posterior density functions. Once we have the mechanism to generate samples given the data, we can use the samples to compute the approximate Bayesian estimates and also the corresponding descriptive statistics.

**Algorithm:**

Step 1: Generate $\theta_1$ from the Gamma(a; b)

Step 2: Generate $p_1$ from the Gamma(c; d)

Step 3: Obtain the posterior samples $(\theta_1,\ p_1); \ldots; (\theta_M,\ p_M)$ by repeating the Steps 1 and 2, $M$ times.

Step 4: The Bayes estimates of $\theta$ and $p$ then obtained by

$\hat{E}(\theta|data) = \frac{1}{M-N}\sum_{i=1}^{M-N}\theta_i$ and $\hat{E}(p|data) = \frac{1}{M-N}\sum_{i=1}^{M-N}p_i$

where $N$ is the burn-in period

Step 5: Obtain the posterior variance of $\theta$ and $p$ as

$$\hat{V}(\theta|data) = \frac{1}{M-N}\sum_{i=1}^{M-N}\left(\theta_i - \hat{E}(\theta|data)\right)^2 ; \text{and}$$

$$\hat{V}(p|data) = \frac{1}{M-N}\sum_{i=1}^{M-N}\left(p_i - \hat{E}(p|data)\right)^2$$

where $N$ is the burn-in period

Step 6: A lifetime $T$ is generated from $GE(\theta, p)$ as follows:

i. Generate $U$ from the Uniform(0, 1)

ii. Let $T = \left(-\ln(1 - U^{1/\hat{p}}\right)/\hat{\theta}$

iii. Repeating Steps i and ii, $n$ times (sample size).

The simulation study is carried out for different sample size and with different hyper parameter values. In particular we take sample sizes $n$ = 20, 40 and 100. Informative priors are used for the shape and scale parameters, we chose a = b = c = d = 0.0001 in order to obtain proper priors as suggested by Guure and Bosomprah (2013). In all these cases, we generate observation from a gamma distribution with assumed actual shape parameter ($p$) of the $GE$ distribution were taken to be 0.75, 1.5, and 2.5. Also, the scale parameter ($\theta$) was considered throughout this simulation to be 0.5, 1.0, and 10. The values of the loss parameter for the general entropy loss function are $k = \pm 0.5$, which can be extended for other values of the loss parameter.

For comparison purpose we compute maximum likelihood estimates (MLE), Bayes estimate using Lindley's approximation (BSE), and Bayes estimates under the general entropy loss functions (BGE). In all cases Bayes estimate using 5000 MCMC samples were obtained with 1500 iterations after 500 iterations were burn-in.

Note that the parameter estimates under the classical maximum likelihood method could not be obtained in close form, and we therefore employed Newton-Raphson iterative approach via the Hessian matrix. This can simply be implemented in the $R$ programming language with *vglm* function under the package *VGAM*. (See Appendix A).

1. For the scale parameter ($\theta = 1$), the average estimates obtained by all the methods along with mean squared error, the absolute bias values and the average 95% confidence intervals are determined and presented in Tables 4.1 – 4.6.

**Table 4.1**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 0.75$ and $\theta = 1$

| *n* | Method | $\hat{\theta}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 1.255219 | 0.110249 | 0.255219 | 0.872099 | 1.638339 |
| | BSE | 1.255095 | 0.110098 | 0.255095 | 0.872212 | 1.637978 |
| | BGE ($k$ = 0.5) | 1.251456 | 0.111166 | 0.251456 | 0.869158 | 1.633754 |
| | BGE ($k$ = –0.5) | 1.246358 | 0.101183 | 0.246358 | 0.878400 | 1.614316 |
| 40 | MLE | 1.170036 | 0.041618 | 0.170036 | 0.978673 | 1.361399 |
| | BSE | 1.170024 | 0.041613 | 0.170024 | 0.978673 | 1.361375 |
| | BGE ($k$ = 0.5) | 1.163792 | 0.040048 | 0.163792 | 0.977576 | 1.350008 |
| | BGE ($k$ = –0.5) | 1.163408 | 0.038896 | 0.163408 | 0.978896 | 1.347920 |
| 100 | MLE | 1.119557 | 0.016901 | 0.119557 | 1.017933 | 1.221181 |
| | BSE | 1.119556 | 0.016900 | 0.119556 | 1.017934 | 1.221178 |
| | BGE ($k$ = 0.5) | 1.115652 | 0.014504 | 0.115652 | 1.019062 | 1.212242 |
| | BGE ($k$ = –0.5) | 1.113236 | 0.013092 | 0.113236 | 1.019789 | 1.206683 |

**Table 4.2**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 1.5$ and $\theta = 1$

| *n* | Method | $\hat{\theta}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.853394 | 0.070492 | 0.146606 | 0.461282 | 1.245506 |
| | BSE | 0.854497 | 0.069543 | 0.145503 | 0.465089 | 1.243905 |
| | BGE ($k$ = 0.5) | 0.851304 | 0.068749 | 0.148696 | 0.462113 | 1.240495 |
| | BGE ($k$ = –0.5) | 0.856258 | 0.068157 | 0.143742 | 0.470902 | 1.241614 |
| 40 | MLE | 0.923829 | 0.024940 | 0.076171 | 0.803420 | 1.044238 |
| | BSE | 0.923974 | 0.024859 | 0.076026 | 0.803771 | 1.044177 |
| | BGE ($k$ = 0.5) | 0.920380 | 0.025602 | 0.079620 | 0.797161 | 1.043599 |
| | BGE ($k$ = –0.5) | 0.900474 | 0.026197 | 0.099526 | 0.766338 | 1.034610 |
| 100 | MLE | 0.962301 | 0.008282 | 0.037699 | 0.914159 | 1.010443 |
| | BSE | 0.962325 | 0.008273 | 0.037675 | 0.914211 | 1.010439 |
| | BGE ($k$ = 0.5) | 0.965101 | 0.007617 | 0.034899 | 0.919560 | 1.010642 |
| | BGE ($k$ = –0.5) | 0.967674 | 0.007042 | 0.032326 | 0.924500 | 1.010848 |

**Table 4.3**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 2.5$ and $\theta = 1$

| *n* | Method | $\hat{\theta}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 1.102155 | 0.055014 | 0.102155 | 0.769435 | 1.434875 |
| | BSE | 1.100471 | 0.053753 | 0.100471 | 0.771840 | 1.429102 |
| | BGE ($k$ = 0.5) | 1.097077 | 0.053014 | 0.097077 | 0.772063 | 1.422091 |
| | BGE ($k$ = –0.5) | 1.091142 | 0.052206 | 0.091142 | 0.771211 | 1.411073 |
| 40 | MLE | 1.042563 | 0.020509 | 0.042563 | 0.946628 | 1.138498 |
| | BSE | 1.042287 | 0.020373 | 0.042287 | 0.946738 | 1.137836 |
| | BGE ($k$ = 0.5) | 1.042121 | 0.020713 | 0.042121 | 0.946037 | 1.138205 |
| | BGE ($k$ = –0.5) | 1.039761 | 0.020135 | 0.039761 | 0.945911 | 1.133611 |
| 100 | MLE | 1.000584 | 0.005621 | 0.000584 | 0.976160 | 1.025008 |
| | BSE | 1.000533 | 0.005606 | 0.000533 | 0.976167 | 1.024899 |
| | BGE ($k$ = 0.5) | 1.000935 | 0.005640 | 0.000935 | 0.976295 | 1.025575 |
| | BGE ($k$ = –0.5) | 1.001552 | 0.006044 | 0.001552 | 0.975752 | 1.027352 |

**Table 4.4**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 0.75$ and $\theta = 1$

| *n* | Method | $\hat{p}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.826655 | 0.184489 | 0.076655 | 0.272569 | 1.380741 |
| | BSE | 0.826655 | 0.184489 | 0.076655 | 0.272569 | 1.380741 |
| | BGE ($k$ = 0.5) | 0.814999 | 0.181371 | 0.064999 | 0.271118 | 1.358880 |
| | BGE ($k$ = –0.5) | 0.811920 | 0.175658 | 0.061920 | 0.277697 | 1.346143 |
| 40 | MLE | 0.775291 | 0.111177 | 0.025291 | 0.588832 | 0.961750 |
| | BSE | 0.775291 | 0.111177 | 0.025291 | 0.588832 | 0.961750 |
| | BGE ($k$ = 0.5) | 0.776927 | 0.116500 | 0.026927 | 0.585537 | 0.968317 |
| | BGE ($k$ = –0.5) | 0.776832 | 0.119055 | 0.026832 | 0.583549 | 0.970115 |
| 100 | MLE | 0.761682 | 0.078559 | 0.011682 | 0.665625 | 0.857739 |
| | BSE | 0.761682 | 0.078559 | 0.011682 | 0.665625 | 0.857739 |
| | BGE ($k$ = 0.5) | 0.761904 | 0.077485 | 0.011904 | 0.666354 | 0.857454 |
| | BGE ($k$ = –0.5) | 0.761483 | 0.077037 | 0.011483 | 0.666403 | 0.856563 |

**Table 4.5**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 1.5$ and $\theta = 1$

| *n* | Method | $\hat{p}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 1.990047 | 0.469041 | 0.490047 | 0.922655 | 3.057439 |
| | BSE | 1.990047 | 0.469041 | 0.490047 | 0.922655 | 3.057439 |
| | BGE ($k$ = 0.5) | 2.053140 | 0.455767 | 0.553140 | 0.965922 | 3.140358 |
| | BGE ($k$ = –0.5) | 1.981065 | 0.444008 | 0.481065 | 0.940410 | 3.021720 |
| 40 | MLE | 1.647109 | 0.276404 | 0.147109 | 1.299493 | 1.994725 |
| | BSE | 1.647109 | 0.276404 | 0.147109 | 1.299493 | 1.994725 |
| | BGE ($k$ = 0.5) | 1.649840 | 0.281081 | 0.149840 | 1.298549 | 2.001131 |
| | BGE ($k$ = –0.5) | 1.659436 | 0.287383 | 0.159436 | 1.300266 | 2.018606 |
| 100 | MLE | 1.560677 | 0.179674 | 0.060677 | 1.393902 | 1.727452 |
| | BSE | 1.560677 | 0.179674 | 0.060677 | 1.393902 | 1.727452 |
| | BGE ($k$ = 0.5) | 1.557956 | 0.180913 | 0.057956 | 1.392072 | 1.723840 |
| | BGE ($k$ = –0.5) | 1.559022 | 0.181074 | 0.059022 | 1.392544 | 1.725500 |

**Table 4.6**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 2.5$ and $\theta = 1$

| *n* | Method | $\hat{p}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 3.243582 | 0.654769 | 0.741582 | 1.901152 | 4.586012 |
| | BSE | 3.243582 | 0.654769 | 0.741582 | 1.901152 | 4.586012 |
| | BGE ($k$ = 0.5) | 3.410514 | 0.676817 | 0.908514 | 1.968394 | 4.852634 |
| | BGE ($k$ = –0.5) | 3.398282 | 0.640145 | 0.896282 | 1.989414 | 4.807150 |
| 40 | MLE | 2.864344 | 0.438533 | 0.362344 | 2.337967 | 3.390721 |
| | BSE | 2.864344 | 0.438533 | 0.362344 | 2.337967 | 3.390721 |
| | BGE ($k$ = 0.5) | 2.868843 | 0.431086 | 0.366843 | 2.343160 | 3.394526 |
| | BGE ($k$ = –0.5) | 2.864682 | 0.428062 | 0.362682 | 2.342282 | 3.387082 |
| 100 | MLE | 2.630699 | 0.271799 | 0.128699 | 2.398542 | 2.862856 |
| | BSE | 2.630699 | 0.271799 | 0.128699 | 2.398542 | 2.862856 |
| | BGE ($k$ = 0.5) | 2.641918 | 0.279051 | 0.139918 | 2.401927 | 2.881909 |
| | BGE ($k$ = –0.5) | 2.626457 | 0.270577 | 0.124457 | 2.396798 | 2.856116 |

Some of the points are quite clear from the numerical results. As expected it is observed that the performances of all estimators become better when the sample size increases. It is

also observed that both in terms of biases and mean squared errors (MSE), for large sample sizes the Bayes estimates and the MLEs become closer. When $\theta = 1$, the Bayes estimates of $\theta$ perform marginally better than the MLEs in terms of biases and MSE for all cases considered. In general the Bayes estimates of $p$ perform better than the MLEs for $p \leq 1$ and for $p > 1$ it is the other way. Particularly, from Tables 1 – 3 , it is very clear that the most dominant estimator that had the smallest MSE with regard to the absolute biases for the scale parameter $(\theta)$ is Bayesian under general entropy loss function. This is followed closely by Bayes under squared error loss function. What has been observed again is that, as the sample size increases, the MSE of all the estimators decrease unswervingly. This is simply an indication of how good and reliable the estimators are.

Results in Tables 4 – 6 contain the MSE and the absolute biases of the estimated shape parameter $(\hat{p})$ when $\theta = 1$, it is noticed that the MSE and the absolute biases of the two estimators, that is, maximum likelihood and Bayes under squared error loss function, have the same values for the estimated shape parameter. This is expected in that the priors used for the Bayesian analysis are noninformative. With regards to the survival function, Bayes estimator under the general entropy loss function gives a minimum bias with relatively small samples. Maximum likelihood estimator is slightly ahead of the other estimators with respect to the MSE.

2. For the scale parameter $(\theta = 0.5)$, the average estimates obtained by all the methods along with mean squared error, the absolute bias values and the average 95% confidence intervals are determined and presented in Tables 4.7 – 4.12.

**Table 4.7**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 0.75$ and $\theta = 0.5$

| *n* | Method | $\hat{\theta}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.73419 | 0.11180 | 0.23419 | 0.55225 | 0.91613 |
| | BSE | 0.73409 | 0.11165 | 0.23409 | 0.55227 | 0.91590 |
| | BGE ($k$ = 0.5) | 0.73034 | 0.11268 | 0.23034 | 0.54768 | 0.91299 |
| | BGE ($k$ = –0.5) | 0.72624 | 0.10265 | 0.22624 | 0.55191 | 0.90057 |
| 40 | MLE | 0.65587 | 0.04232 | 0.15587 | 0.54864 | 0.76311 |
| | BSE | 0.65586 | 0.04231 | 0.15586 | 0.54863 | 0.76309 |
| | BGE ($k$ = 0.5) | 0.64979 | 0.04069 | 0.14979 | 0.54464 | 0.75493 |
| | BGE ($k$ = –0.5) | 0.64952 | 0.03953 | 0.14952 | 0.54588 | 0.75316 |
| 100 | MLE | 0.60787 | 0.01710 | 0.10787 | 0.54132 | 0.67441 |
| | BSE | 0.60787 | 0.01710 | 0.10787 | 0.54132 | 0.67441 |
| | BGE ($k$ = 0.5) | 0.60420 | 0.01466 | 0.10420 | 0.54258 | 0.66582 |
| | BGE ($k$ = –0.5) | 0.60193 | 0.01322 | 0.10193 | 0.54340 | 0.66045 |

**Table 4.8**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 1.5$ and $\theta = 0.5$

| *n* | Method | $\hat{\theta}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.33634 | 0.07096 | 0.16366 | 0.19140 | 0.48129 |
| | BSE | 0.33754 | 0.07000 | 0.16246 | 0.19358 | 0.48150 |
| | BGE ($k$ = 0.5) | 0.33443 | 0.06924 | 0.16557 | 0.19125 | 0.47761 |
| | BGE ($k$ = –0.5) | 0.33944 | 0.06859 | 0.16056 | 0.19693 | 0.48195 |
| 40 | MLE | 0.41134 | 0.02470 | 0.08867 | 0.32941 | 0.49326 |
| | BSE | 0.41149 | 0.02462 | 0.08851 | 0.32970 | 0.49328 |
| | BGE ($k$ = 0.5) | 0.40782 | 0.02540 | 0.09218 | 0.32474 | 0.49090 |
| | BGE ($k$ = –0.5) | 0.38785 | 0.02619 | 0.11215 | 0.30349 | 0.47222 |
| 100 | MLE | 0.45147 | 0.00766 | 0.04853 | 0.40693 | 0.49601 |
| | BSE | 0.45150 | 0.00765 | 0.04850 | 0.40699 | 0.49601 |
| | BGE ($k$ = 0.5) | 0.45434 | 0.00697 | 0.04566 | 0.41186 | 0.49682 |
| | BGE ($k$ = –0.5) | 0.45697 | 0.00637 | 0.04303 | 0.41637 | 0.49757 |

**Table 4.9**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 2.5$ and $\theta = 0.5$

| *n* | Method | $\hat{\theta}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.58665 | 0.05504 | 0.08665 | 0.45900 | 0.71431 |
| | BSE | 0.58510 | 0.05376 | 0.08510 | 0.45893 | 0.71126 |
| | BGE ($k$ = 0.5) | 0.58178 | 0.05298 | 0.08178 | 0.45652 | 0.70703 |
| | BGE ($k$ = –0.5) | 0.57592 | 0.05212 | 0.07592 | 0.45170 | 0.70014 |
| 40 | MLE | 0.53051 | 0.01993 | 0.03051 | 0.45691 | 0.60411 |
| | BSE | 0.53025 | 0.01980 | 0.03025 | 0.45691 | 0.60359 |
| | BGE ($k$ = 0.5) | 0.53005 | 0.02013 | 0.03005 | 0.45608 | 0.60402 |
| | BGE ($k$ = –0.5) | 0.52775 | 0.01953 | 0.02775 | 0.45489 | 0.60060 |
| 100 | MLE | 0.49002 | 0.00463 | 0.00998 | 0.45540 | 0.52464 |
| | BSE | 0.48997 | 0.00461 | 0.01003 | 0.45541 | 0.52453 |
| | BGE ($k$ = 0.5) | 0.49037 | 0.00465 | 0.00963 | 0.45567 | 0.52507 |
| | BGE ($k$ = –0.5) | 0.49095 | 0.00506 | 0.00905 | 0.45475 | 0.52715 |

**Table 4.10**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 0.75$ and $\theta = 0.5$

| *n* | Method | $\hat{p}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.80821 | 0.18272 | 0.05821 | 0.57561 | 1.04080 |
| | BSE | 0.80821 | 0.18272 | 0.05821 | 0.57561 | 1.04080 |
| | BGE ($k$ = 0.5) | 0.79686 | 0.17972 | 0.04686 | 0.56618 | 1.02754 |
| | BGE ($k$ = –0.5) | 0.79435 | 0.17404 | 0.04435 | 0.56735 | 1.02136 |
| 40 | MLE | 0.76417 | 0.10992 | 0.01417 | 0.59134 | 0.93700 |
| | BSE | 0.76417 | 0.10992 | 0.01417 | 0.59134 | 0.93700 |
| | BGE ($k$ = 0.5) | 0.76528 | 0.11523 | 0.01528 | 0.58832 | 0.94223 |
| | BGE ($k$ = –0.5) | 0.76493 | 0.11779 | 0.01493 | 0.58602 | 0.94383 |
| 100 | MLE | 0.75383 | 0.07744 | 0.00383 | 0.61220 | 0.89545 |
| | BSE | 0.75383 | 0.07744 | 0.00383 | 0.61220 | 0.89545 |
| | BGE ($k$ = 0.5) | 0.75416 | 0.07637 | 0.00416 | 0.61352 | 0.89480 |
| | BGE ($k$ = –0.5) | 0.75378 | 0.07592 | 0.00378 | 0.61355 | 0.89401 |

**Table 4.11**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 1.5$ and $\theta = 0.5$

| *n* | Method | $\hat{p}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 1.90314 | 0.45414 | 0.40314 | 1.53645 | 2.26983 |
| | BSE | 1.90314 | 0.45414 | 0.40314 | 1.53645 | 2.26983 |
| | BGE ($k$ = 0.5) | 1.96756 | 0.44024 | 0.46756 | 1.60653 | 2.32860 |
| | BGE ($k$ = –0.5) | 1.89666 | 0.42920 | 0.39666 | 1.54018 | 2.25314 |
| 40 | MLE | 1.57947 | 0.26493 | 0.07947 | 1.31115 | 1.84778 |
| | BSE | 1.57947 | 0.26493 | 0.07947 | 1.31115 | 1.84778 |
| | BGE ($k$ = 0.5) | 1.58173 | 0.26958 | 0.08173 | 1.31107 | 1.85239 |
| | BGE ($k$ = –0.5) | 1.59070 | 0.27579 | 0.09070 | 1.31694 | 1.86445 |
| 100 | MLE | 1.50271 | 0.16907 | 0.00271 | 1.29345 | 1.71197 |
| | BSE | 1.50271 | 0.16907 | 0.00271 | 1.29345 | 1.71197 |
| | BGE ($k$ = 0.5) | 1.49986 | 0.17033 | 0.00014 | 1.28982 | 1.70991 |
| | BGE ($k$ = –0.5) | 1.50091 | 0.17048 | 0.00091 | 1.29078 | 1.71105 |

**Table 4.12**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 2.5$ and $\theta = 0.5$

| *n* | Method | $\hat{p}$ | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 3.17811 | 0.67218 | 0.67811 | 2.73199 | 3.62422 |
| | BSE | 3.17811 | 0.67218 | 0.67811 | 2.73199 | 3.62422 |
| | BGE ($k$ = 0.5) | 3.34283 | 0.69590 | 0.84283 | 2.88891 | 3.79675 |
| | BGE ($k$ = –0.5) | 3.33427 | 0.65911 | 0.83427 | 2.89251 | 3.77602 |
| 40 | MLE | 2.82049 | 0.45216 | 0.32049 | 2.46996 | 3.17102 |
| | BSE | 2.82049 | 0.45216 | 0.32049 | 2.46996 | 3.17102 |
| | BGE ($k$ = 0.5) | 2.82573 | 0.44475 | 0.32573 | 2.47809 | 3.17338 |
| | BGE ($k$ = –0.5) | 2.82188 | 0.44169 | 0.32188 | 2.47543 | 3.16832 |
| 100 | MLE | 2.60352 | 0.28309 | 0.10352 | 2.33274 | 2.87430 |
| | BSE | 2.60352 | 0.28309 | 0.10352 | 2.33274 | 2.87430 |
| | BGE ($k$ = 0.5) | 2.61401 | 0.29045 | 0.11401 | 2.33973 | 2.88829 |
| | BGE ($k$ = –0.5) | 2.59940 | 0.28182 | 0.09940 | 2.32922 | 2.86957 |

When $(\theta = 0.5) < 1$, results in Tables 7, 8, and 9 show that the average biases and the average MSE's decrease as sample size increases. It is observed that the average biases and the average MSE's of $\hat{\theta}$ depend on $p$. For all the methods as $p$ increases the average relative MSE's of $\hat{\theta}$ decrease and the same thing is true for the average biases also for most of the methods. Moreover, with respect to the MSE's it is clear that when $\theta < 1$ the performances of all approaches in estimating $\theta$ are quite close to that when $\theta = 1$.

On the other hand there is no pattern observed for the average biases of $\hat{p}$ and the corresponding average MSE's. It observed that for most of the methods the biases are quite severe for small sample sizes and large $p$. Considering only MSE's it can be said that the estimation of $p$'s are more accurate for $p < 2$ and $\theta = 0.5$.

3. For the scale parameter $(\theta = 10)$, the average estimates obtained by all the methods along with mean squared error, the absolute bias values and the average 95% confidence intervals are determined and presented in Tables 4.13 – 4.18.

**Table 4.13**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 0.75$ and $\theta = 10$

| ***N*** | **Method** | | **MSE** | **Bias** | **95% C. I.** | |
|---|---|---|---|---|---|---|
| | | | | | **Lower** | **Upper** |
| 20 | MLE | 11.04537 | 0.13622 | 1.04537 | 10.84454 | 11.04537 |
| | BSE | 11.04526 | 0.13606 | 1.04526 | 10.84454 | 11.04526 |
| | BGE ($k$ = 0.5) | 11.04161 | 0.13671 | 1.04161 | 10.84042 | 11.04161 |
| | BGE ($k$ = –0.5) | 11.03651 | 0.12627 | 1.03651 | 10.84315 | 11.03651 |
| 40 | MLE | 10.96010 | 0.05891 | 0.96010 | 10.83358 | 10.96010 |
| | BSE | 10.96009 | 0.05890 | 0.96009 | 10.83358 | 10.96009 |
| | BGE ($k$ = 0.5) | 10.95386 | 0.05667 | 0.95386 | 10.82977 | 10.95386 |
| | BGE ($k$ = –0.5) | 10.95347 | 0.05548 | 0.95347 | 10.83069 | 10.95347 |
| 100 | MLE | 10.90958 | 0.02889 | 0.90958 | 10.82308 | 10.90958 |
| | BSE | 10.90958 | 0.02889 | 0.90958 | 10.82308 | 10.90958 |
| | BGE ($k$ = 0.5) | 10.90567 | 0.02608 | 0.90567 | 10.82348 | 10.90567 |
| | BGE ($k$ = –0.5) | 10.90325 | 0.02441 | 0.90325 | 10.82373 | 10.90325 |

**Table 4.14**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 1.5$ and $\theta = 10$

| *N* | Method | | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 10.64344 | 0.09733 | 0.64344 | 10.47368 | 10.81319 |
| | BSE | 10.64454 | 0.09625 | 0.64454 | 10.47573 | 10.81335 |
| | BGE ($k$ = 0.5) | 10.64135 | 0.09580 | 0.64135 | 10.47294 | 10.80977 |
| | BGE ($k$ = –0.5) | 10.64630 | 0.09465 | 0.64630 | 10.47890 | 10.81370 |
| 40 | MLE | 10.71381 | 0.04357 | 0.71381 | 10.60500 | 10.82262 |
| | BSE | 10.71395 | 0.04347 | 0.71395 | 10.60527 | 10.82264 |
| | BGE ($k$ = 0.5) | 10.71036 | 0.04462 | 0.71036 | 10.60025 | 10.82047 |
| | BGE ($k$ = –0.5) | 10.69047 | 0.04741 | 0.69047 | 10.57697 | 10.80397 |
| 100 | MLE | 10.55224 | 0.02251 | 0.55224 | 10.47587 | 10.62860 |
| | BSE | 10.55227 | 0.02250 | 0.55226 | 10.47593 | 10.62860 |
| | BGE ($k$ = 0.5) | 10.55504 | 0.02154 | 0.55504 | 10.48035 | 10.62972 |
| | BGE ($k$ = –0.5) | 10.55761 | 0.02067 | 0.55761 | 10.48443 | 10.63078 |

**Table 4.15**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{\theta})$ with $p = 2.5$ and $\theta = 10$

| *N* | Method | | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 10.44107 | 0.09187 | 0.44107 | 10.27614 | 10.60600 |
| | BSE | 10.44221 | 0.09083 | 0.44221 | 10.27821 | 10.60620 |
| | BGE ($k$ = 0.5) | 10.43905 | 0.09028 | 0.43905 | 10.27555 | 10.60254 |
| | BGE ($k$ = –0.5) | 10.44401 | 0.08929 | 0.44401 | 10.28141 | 10.60661 |
| 40 | MLE | 10.41299 | 0.04061 | 0.41299 | 10.30794 | 10.51804 |
| | BSE | 10.41313 | 0.04052 | 0.41313 | 10.30820 | 10.51806 |
| | BGE ($k$ = 0.5) | 10.40951 | 0.04155 | 0.40951 | 10.30326 | 10.51577 |
| | BGE ($k$ = –0.5) | 10.38960 | 0.04367 | 0.38960 | 10.28066 | 10.49853 |
| 100 | MLE | 10.04381 | 0.02090 | 0.04381 | 9.97024 | 10.11738 |
| | BSE | 10.04385 | 0.02088 | 0.04385 | 9.97030 | 10.11740 |
| | BGE ($k$ = 0.5) | 10.04737 | 0.02001 | 0.04737 | 9.97537 | 10.11937 |
| | BGE ($k$ = –0.5) | 10.05060 | 0.01924 | 0.05060 | 9.98001 | 10.12119 |

**Table 4.16**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 0.75$ and $\theta = 10$

| *N* | Method | | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 0.79907 | 0.17233 | 0.04907 | 0.57319 | 1.02496 |
| | BSE | 0.79907 | 0.17233 | 0.04907 | 0.57319 | 1.02496 |
| | BGE ($k$ = 0.5) | 0.78787 | 0.16941 | 0.03787 | 0.56391 | 1.01183 |
| | BGE ($k$ = –0.5) | 0.78565 | 0.16374 | 0.03565 | 0.56546 | 1.00583 |
| 40 | MLE | 0.75867 | 0.09983 | 0.00867 | 0.59397 | 0.92338 |
| | BSE | 0.75867 | 0.09983 | 0.00867 | 0.59397 | 0.92338 |
| | BGE ($k$ = 0.5) | 0.75952 | 0.10513 | 0.00952 | 0.59050 | 0.92854 |
| | BGE ($k$ = –0.5) | 0.75904 | 0.10769 | 0.00904 | 0.58797 | 0.93011 |
| 100 | MLE | 0.74996 | 0.06741 | 0.00004 | 0.61782 | 0.88210 |
| | BSE | 0.74996 | 0.06741 | 0.00004 | 0.61782 | 0.88210 |
| | BGE ($k$ = 0.5) | 0.75034 | 0.06634 | 0.00034 | 0.61926 | 0.88143 |
| | BGE ($k$ = –0.5) | 0.74998 | 0.06589 | 0.00002 | 0.61934 | 0.88063 |

**Table 4.17**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 1.5$ and $\theta = 10$

| *N* | Method | | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 1.85231 | 0.46011 | 0.35231 | 1.48322 | 2.22141 |
| | BSE | 1.85231 | 0.46011 | 0.35231 | 1.48322 | 2.22141 |
| | BGE ($k$ = 0.5) | 1.91951 | 0.44556 | 0.41951 | 1.55630 | 2.28272 |
| | BGE ($k$ = –0.5) | 1.85082 | 0.43523 | 0.35082 | 1.49184 | 2.20980 |
| 40 | MLE | 1.56648 | 0.27414 | 0.06648 | 1.29355 | 1.83942 |
| | BSE | 1.56648 | 0.27414 | 0.06648 | 1.29355 | 1.83942 |
| | BGE ($k$ = 0.5) | 1.56781 | 0.27876 | 0.06781 | 1.29259 | 1.84304 |
| | BGE ($k$ = –0.5) | 1.57554 | 0.28488 | 0.07554 | 1.29731 | 1.85378 |
| 100 | MLE | 1.50890 | 0.17904 | 0.00890 | 1.29355 | 1.72424 |
| | BSE | 1.50890 | 0.17904 | 0.00890 | 1.29355 | 1.72424 |
| | BGE ($k$ = 0.5) | 1.50579 | 0.18033 | 0.00579 | 1.28968 | 1.72191 |
| | BGE ($k$ = –0.5) | 1.50681 | 0.18047 | 0.00681 | 1.29061 | 1.72302 |

**Table 4.18**: The average values of MLE, BSE, and BGE along with average mean squared errors, absolute biases and 95% C. I. of $(\hat{p})$ with $p = 2.5$ and $\theta = 10$

| *N* | Method | | MSE | Bias | 95% C. I. | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| 20 | MLE | 3.24533 | 0.78999 | 0.74533 | 2.76169 | 3.72896 |
| | BSE | 3.24533 | 0.78999 | 0.74533 | 2.76169 | 3.72896 |
| | BGE ($k$ = 0.5) | 3.41242 | 0.83018 | 0.91242 | 2.91664 | 3.90820 |
| | BGE ($k$ = –0.5) | 3.40018 | 0.79254 | 0.90018 | 2.91577 | 3.88459 |
| 40 | MLE | 2.86571 | 0.53421 | 0.36571 | 2.48470 | 3.24671 |
| | BSE | 2.86571 | 0.53421 | 0.36571 | 2.48470 | 3.24671 |
| | BGE ($k$ = 0.5) | 2.87021 | 0.52732 | 0.37021 | 2.49166 | 3.24875 |
| | BGE ($k$ = –0.5) | 2.86605 | 0.52388 | 0.36605 | 2.48874 | 3.24335 |
| 100 | MLE | 2.63183 | 0.34344 | 0.13183 | 2.33358 | 2.93008 |
| | BSE | 2.63183 | 0.34344 | 0.13183 | 2.33358 | 2.93008 |
| | BGE ($k$ = 0.5) | 2.64306 | 0.35185 | 0.14306 | 2.34117 | 2.94494 |
| | BGE ($k$ = –0.5) | 2.62758 | 0.34176 | 0.12758 | 2.33006 | 2.92510 |

When $(\theta = 10) \gg 1$, results in Tables 13, 14, and 15 show that the average biases and the average MSE's decrease as sample size increases. It is observed that the average biases and the average MSE's of $\hat{\theta}$ depend on $p$. For all the methods as $p$ increases the average relative MSE's of $\hat{\theta}$ decrease and the same thing is true for the average biases also for most of the methods. It observed that for most of the methods the biases are quite severe for small sample sizes and small $p$. Moreover, with respect to the MSE's it is clear that when $\theta \gg 1$ the performances of all approaches in estimation of $\theta$'s are more accurate for large $p$.

On the other hand when $\theta = 10$ an increasing pattern observed for the average biases of $\hat{p}$ and the corresponding average MSE's. It observed that for most of the methods the biases are quite severe for small sample sizes and large $p$. Considering only MSE's it can be said that the estimation of $p$'s are more accurate for $p < 2$ and $\theta = 10$.

## 4.2 Real Data Analysis

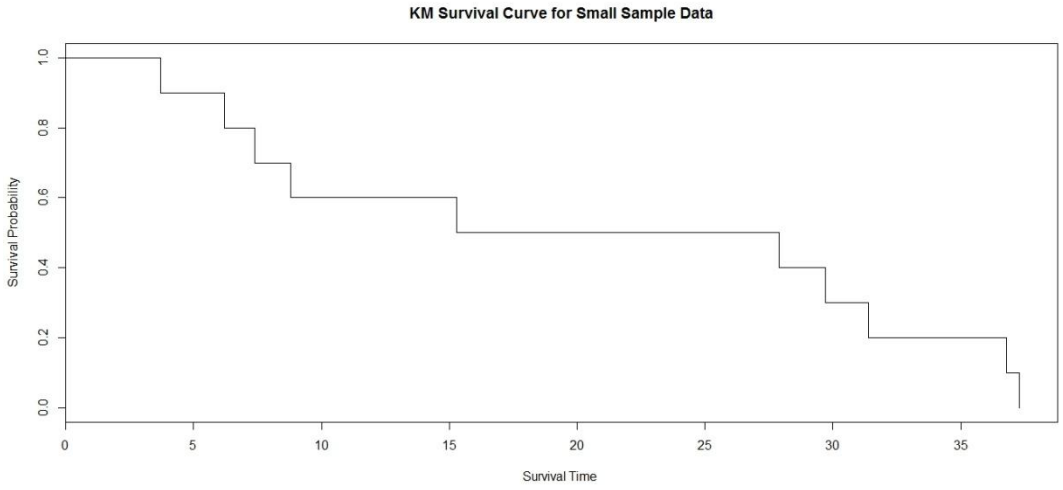
In this section we analyze three data sets in which we have considered to be relatively small, moderate, and large for illustration and comparison purposes. We analyzed the postoperative survival of non-small cell lung cancer patients using administrative data. The data were obtained from the Diagnosis Procedure Combination (DPC). DPC-formatted database administered by the Quality Indicator/Improvement Project (QIP). This database is very different from those of clinical registries: registries usually collect specific data for predetermined purposes and these data are submitted intentionally for analysis. In contrast, the DPC database uses medical claims data, which are routinely produced for all medical services with the primary intended purpose of reimbursement.

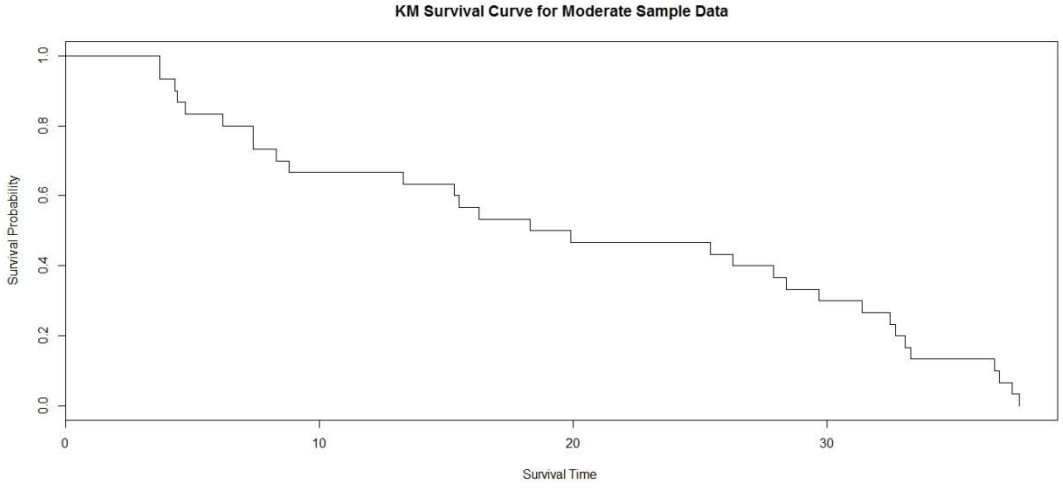
All the patients died by the end of the experiment, so there is no censoring. For the purpose of our study, we have considered the first set of data to be relatively small with $n$ = 10 and the second which seem relatively moderate with $n$ = 30 and the third which seem relatively large with $n$ = 150.

Since we do not have any prior information on the hyper parameters, we assume a = b = c = d = 0.0001. This makes the priors proper on $\hat{\theta}$ and $\hat{p}$ and the corresponding posteriors also proper. Also the values of the loss parameter for the general entropy loss function are $k$ = ±0.5, which can be extended for other values of the loss parameter.

Since there is no censoring, the Kaplan-Meier (KM) estimate coincides with the empirical survival function. Figure 4.1 shows Kaplan-Meier estimates for all groups in this study. Note that the KM estimator has a nice interpretation as a non-parametric maximum likelihood estimator (NPML) which gives a good and first impression of the behavior of the survival function.

Using the iterative procedure suggested in the previous sections the maximum likelihood estimates (MLE), Bayesian estimate using Lindley's approximation (BSE), and Bayesian estimates under the general entropy loss functions (BGE) of $\hat{\theta}$ and $\hat{p}$ were calculated. The results are presented in Tables 4.19 and 4.20.



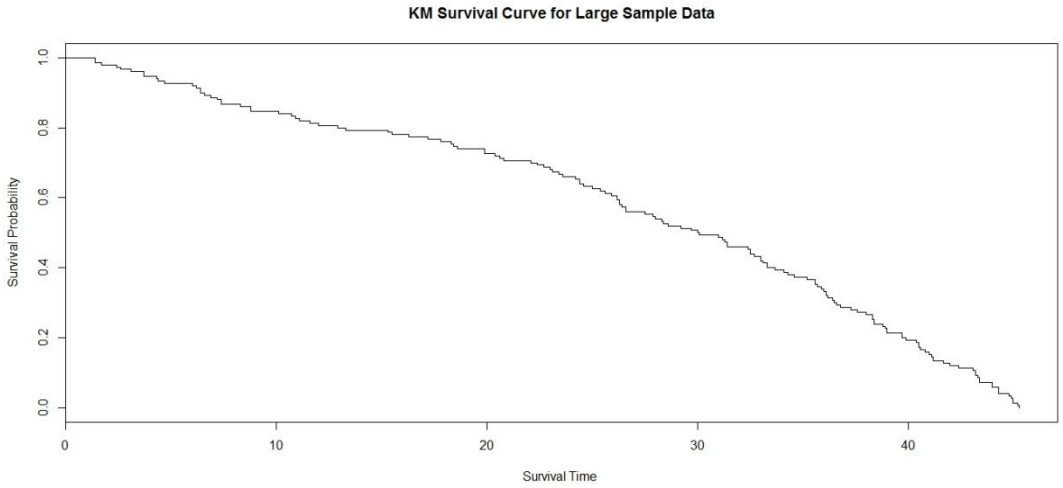



**Figure 4.1:** KM Survival Curves for the Three Data Sets

**Table 4.19:** Average parameters estimates and their corresponding standard error

| ***n*** | **Par** | **MLE** | **BSE** | **BGE (*k* = 0.5)** | **BGE (*k* = –0.5)** |
|---|---|---|---|---|---|
| 10 | $\hat{p}$ (shape) | 2.013757 | 2.013741 | 2.103082 | 2.012830 |
| | | 0.332790 | 0.332784 | 0.288041 | 0.296107 |
| | $\hat{\theta}$ (scale) | 0.073131 | 0.070231 | 0.068382 | 0.0690713 |
| | | 0.103280 | 0.101827 | 0.090814 | 0.090637 |
| 30 | $\hat{p}$ (shape) | 2.123059 | 2.124003 | 2.183105 | 2.190426 |
| | | 0.191745 | 0.191666 | 0.174825 | 0.181635 |
| | $\hat{\theta}$ (scale) | 0.075919 | 0.074246 | 0.068352 | 0.067747 |
| | | 0.056615 | 0.055721 | 0.047206 | 0.046261 |
| 150 | $\hat{p}$ (shape) | 2.497158 | 2.497542 | 2.492109 | 2.494437 |
| | | 0.100275 | 0.100021 | 0.099038 | 0.099875 |
| | $\hat{\theta}$ (scale) | 0.059869 | 0.059868 | 0.058958 | 0.059671 |
| | | 0.003171 | 0.003150 | 0.003013 | 0.003103 |

**Table 4.20:** Mean Square Errors (MSE) and (AIC) of the Survival Function

| | $n = 10$ | | $n = 30$ | | $n = 150$ | |
|---|---|---|---|---|---|---|
| | MSE | AIC | MSE | AIC | MSE | AIC |
| MLE | 0.023728 | 72.34 | 0.022819 | 71.77 | 0.024351 | 74.93 |
| BSE | 0.024102 | 73.01 | 0.023594 | 70.85 | 0.023364 | 74.71 |
| BGE ($k = 0.5$) | 0.013725 | 54.13 | 0.011552 | 49.17 | 0.004471 | 42.67 |
| BGE ($k = -0.5$) | 0.014089 | 57.64 | 0.012339 | 51.73 | 0.004028 | 41.03 |

Results in Table 4.19 show that the Bayes estimator under squared error loss for the shape parameter ($p$) approximately has the same estimate and standard error as compared to that of the classical maximum likelihood estimator but with the scale parameter ($\theta$), and Bayes under squared error has a smaller standard error in comparison or contrasting effect of MLE specially when *n* get larger. Moreover, Observing from the same Table, it is evident that the estimator with the smallest parameter estimate and having a corresponding smaller standard error is Bayesian with the generalized entropy loss function. This occurred for both parameters with a positive and negative loss parameter, that is, $\pm 0.5$

The importance of the survival function cannot be ignored; therefore, the correctness of its estimate is very crucial to both biological and medical studies. As Shown in Figures (4.2) - (4.4) and Table 4.20, the estimator with the smallest standard error and Akaike's Information Criterion (AIC) under small and moderate samples are the classical MLE and BSE; while the BGE is better than the others with large samples, therefore, the three estimator can be preferable upon the sample size, moreover, the BGE depend also on proper priors. Comparing all the estimators, it is clear from the results that Bayes estimator

under general entropy loss function with the loss parameter of $\pm 0.5$ has the smallest standard error and estimate for both the shape parameter ($p$) and the scale parameter ($\theta$).

### 4.3 Conclusions

In this thesis we consider the Bayes and non-Bayes estimation of the unknown parameters of the Generalized Exponential (GE) distribution. Our aim was to obtain the estimates of the parameters and to observe the performance of the methods used for estimation.

The developed methodology for MLE and Bayesian estimation has been demonstrated on a real data set when both the shape ($p$) and scale ($\theta$) parameters of the GE distribution are unknown under informative set of independent priors. It is observed that the parameter estimates under the classical maximum likelihood method could not be obtained in close form; we therefore employed Newton- Raphson iterative approach via the Fisher matrix.

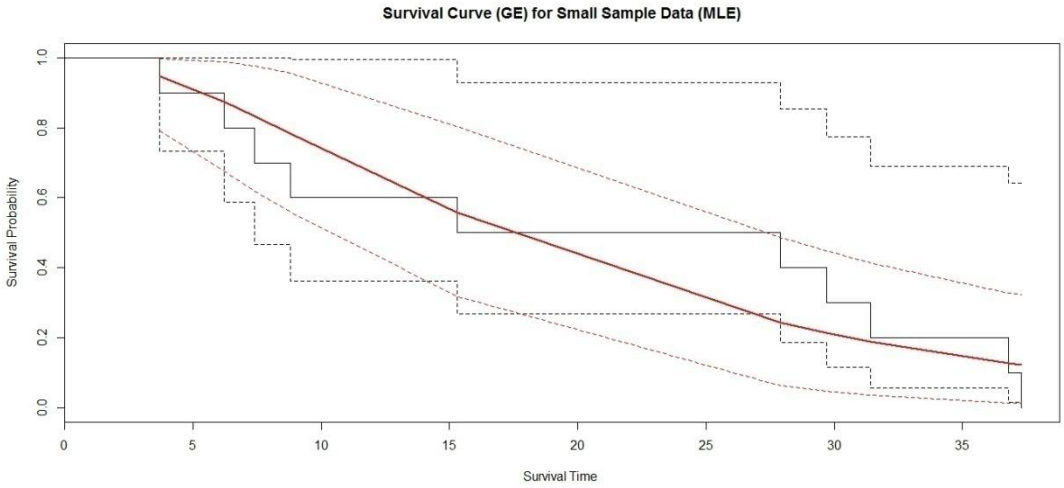
In Bayesian analysis, the parameter of interest is always considered to be a random variable with a prior distribution. The prior distribution is the distribution of the parameter before any data is observed. The selection of prior distribution is most often than not based on the type of prior information that is available to us. When we have little or no information about the parameter, a noninformative prior should be used else an informative prior.

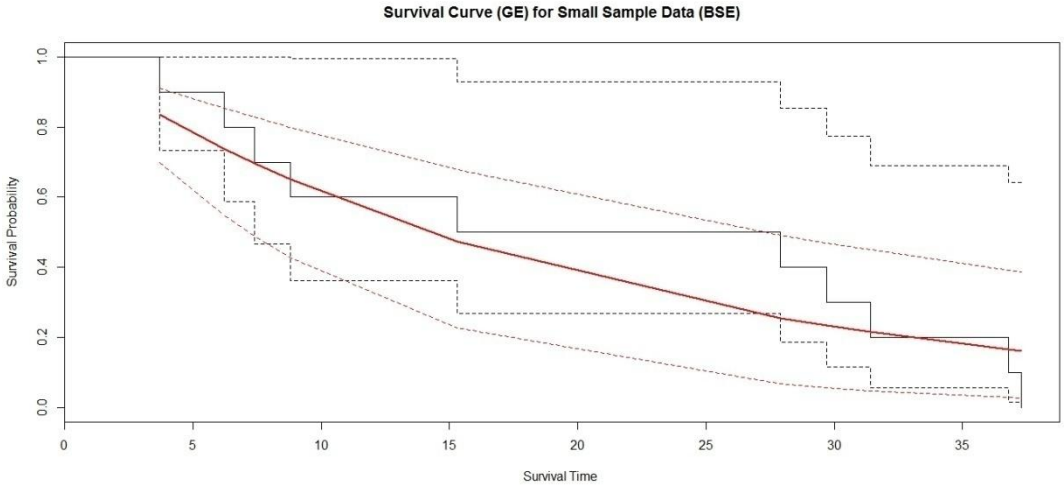
In this study we consider the Bayes estimation of the unknown parameters of the GE distribution. We have also assumed a gamma prior on both parameters, and we provide the Bayes estimators under the assumptions of squared error and general entropy loss functions. We see that the Bayes estimators cannot be obtained in explicit forms, due to the complex nature of the posterior distribution of which Bayes inference is drawn. Therefore, Lindley's numerical approximations procedure is used.

We have also used MCMC technique to generate posterior sample and it is observed that the estimated posteriors match quite well with the theoretical posterior density functions. Since we have an effective MCMC technique we can use any other loss function also. Moreover, this method can be easily extended for exponentiated Weibull distribution also.

From the results and discussions above, it is evident that the Bayesian estimator under general entropy loss function performed quiet better than Bayes under squared error loss function and that of maximum likelihood estimator for estimating the scale parameter with both MSE and absolute bias. In the case of the shape parameter, the Bayesian estimator under the squared error loss function and the maximum likelihood estimator are both almost tantamount in estimating it. For the survival function, maximum likelihood performed better than the other estimators for moderate and small samples.

Finally we should mention that, although we have used gamma priors on the shape parameter, but this method can be used for a more general class of priors also, for example priors with log-concave density functions. Choosing the proper priors is a challenging problem. More work is needed in that direction.



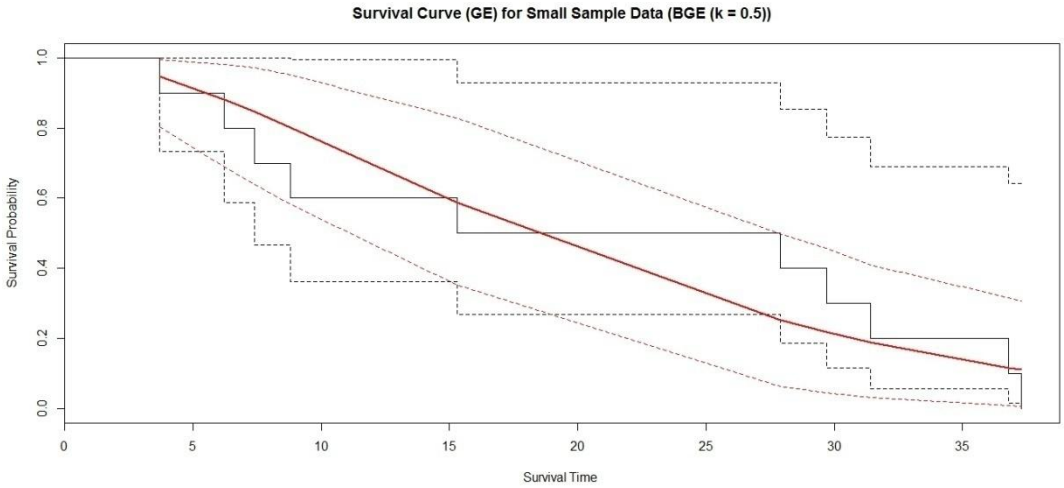
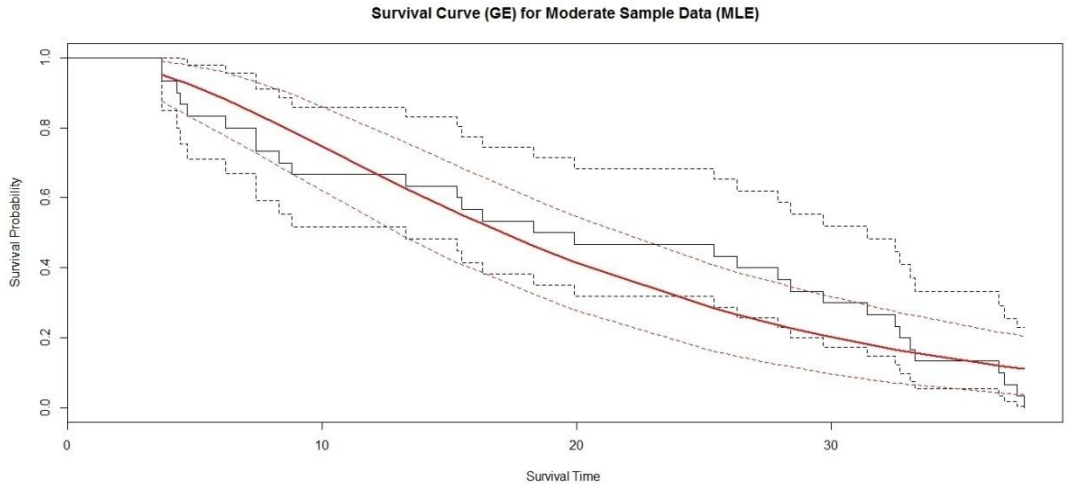
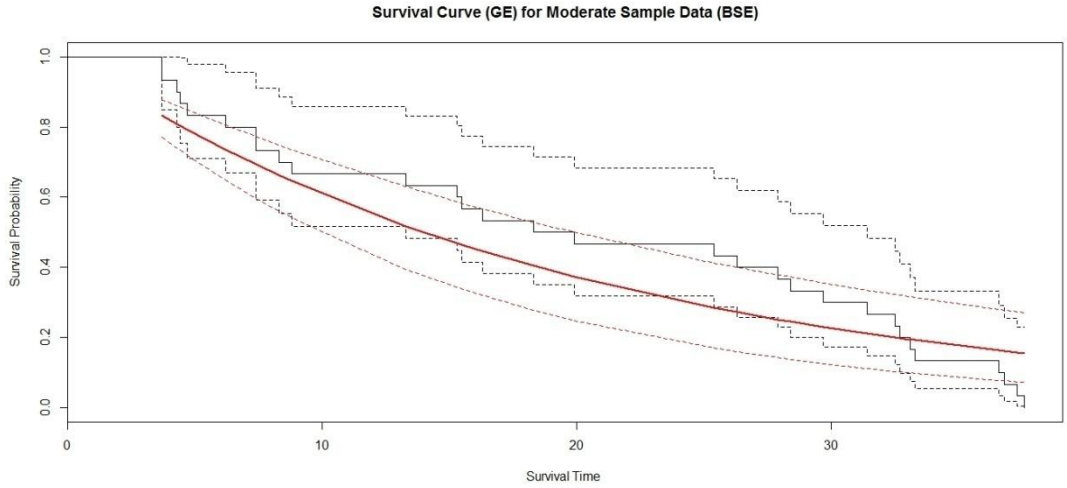




**Figure 4.2:** KM and Survival Curves for the Small Data Under (MLE), (BSE), and (BGE)



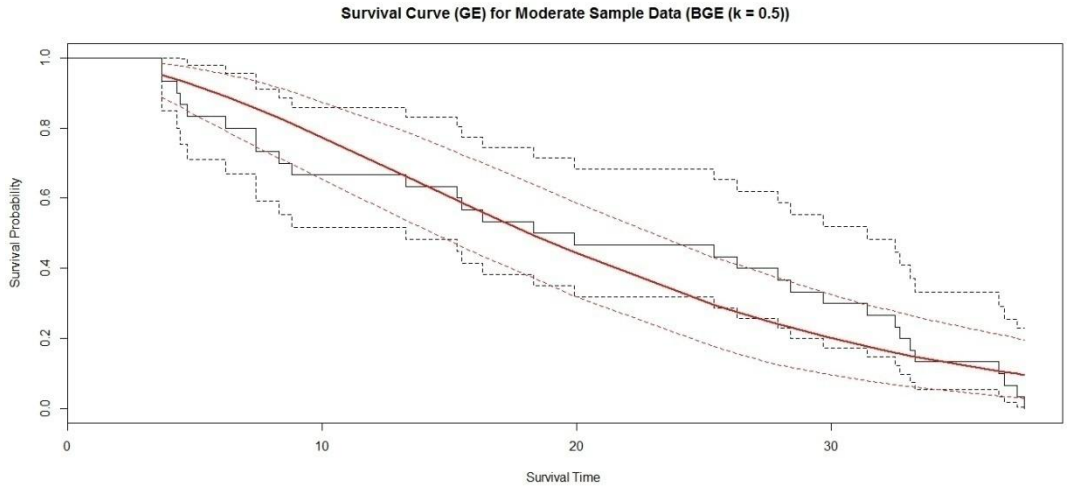
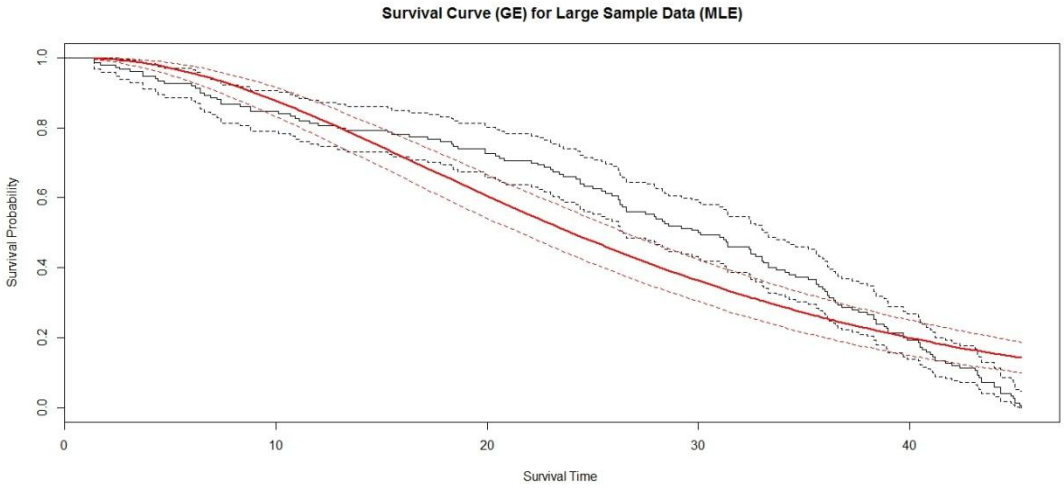
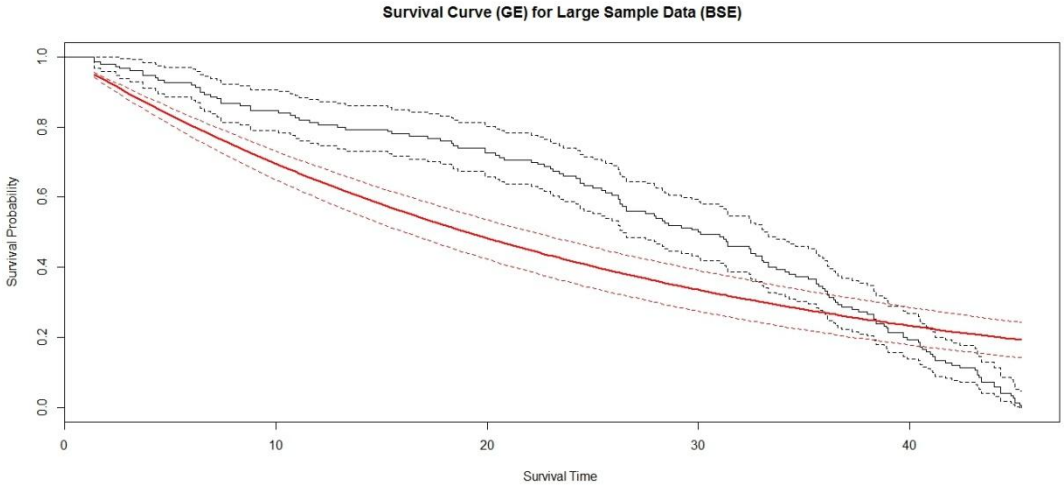




**Figure 4.3:** KM and Survival Curves for the Moderate Data Under (MLE), (BSE), and (BGE)



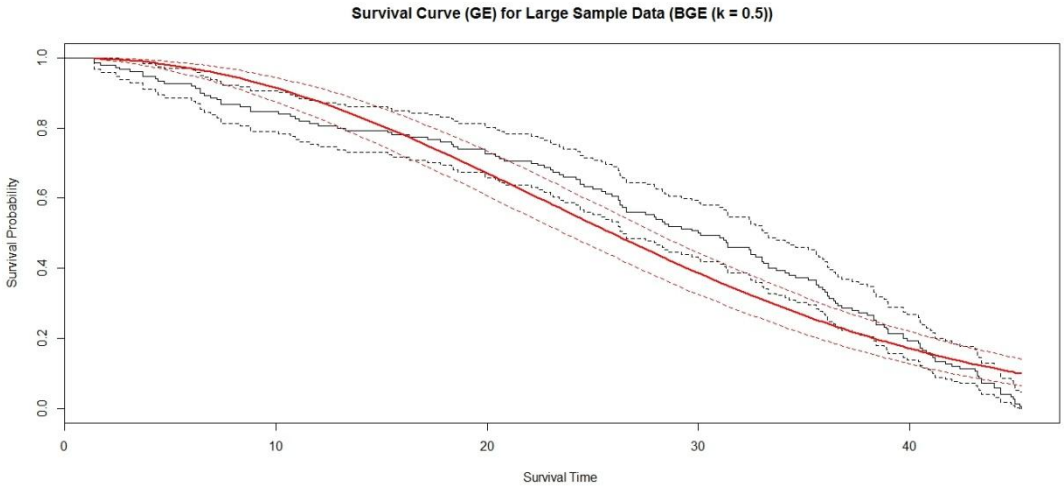




**Figure 4.4:** KM and Survival Curves for the Large Data Under (MLE), (BSE), and (BGE)

# References


A. A. Abdel-Wahid and A. Winterbottom, "*Approximate Bayesian estimates for the Weibull reliability function and hazard rate from censored data*," Journal of Statistical Planning and Inference, vol. 16, no. 3, pp. 277–283, 1987.

A. Zellner, "*Bayesian estimation and prediction using asymmetric loss functions*," Journal of the American Statistical Association, vol. 81, no. 394, pp. 446–451, 1986.

B. N. Pandey, N. Dwividi, and B. Pulastya, "*Comparison between Bayesian and maximum likelihood estimation of the scale parameter in weibull distribution with known shape under linex loss function*," Journal of Scientometric Research, vol. 55, pp. 163–172, 2011.

B. N. Pandey, N. Dwividi, and B. Pulastya, "*Comparison between Bayesian and maximum likelihood estimation of the scale parameter inweibull distribution with known shape under linex loss function*," Journal of Scientometric Research, vol. 55, pp. 163–172, 2011.

C. B. Guure and N. A. Ibrahim," *Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data*, " Mathematical Problems in Engineering, vol. 2012, Article ID 329489, 18 pages, 2012.

C. B. Guure, and S. Bosomprah, "*Bayesian and Non-Bayesian Inference for Survival Data Using Generalized Exponential Distribution*, " Journal of Probability and Statistics, Article ID 364705, 2013.

C. B. Guure, N. A. Ibrahim, and M. Bakri Adam, "*Bayesian inference of the Weibull model based on interval-censored survival data*," Computational and Mathematical Methods in Medicine, vol. 2013,Article ID 849520, 10 pages, 2013.

C. B. Guure, N. A. Ibrahim, M. B. Adam, S. Bosomprah, and A. O. M. Ahmed, "*Bayesian parameter and reliability estimate of Weibull failure time distribution*," Bulletin of the Malaysian Mathematical Sciences Society. In press.

D. Kundu and R. D. Gupta, "*Generalized exponential distribution: Bayesian estimations*," Computational Statistics & Data Analysis, vol. 52, no. 4, pp. 1873–1883, 2008.

D. V. Lindley, "*Approximate Bayesian methods*," Trabajos Estadist, vol. 31, pp. 223–245, 1980.

E. Limpert, W. A. Stahel, and M. Abbt, "*Log-normal distributions across the sciences: keys and clues*, " BioScience, vol. 51, no. 5, pp. 341–352, 2001.

F. M. Al-Aboud, "*Bayesian estimations for the extreme value distribution using progressive censored data and asymmetric loss*," International Mathematical Forum, vol. 4, no. 33–36, pp. 1603–1622, 2009.

F. M. Al-Athari, "*Parameter estimation for the double-pareto distribution*," Journal of Mathematics and Statistics, vol. 7, no. 4, pp. 289–294, 2011.

G. Zheng, "*On the Fisher information matrix in type II censored data from the exponentiated exponential family*," Biometrical Journal, vol. 44, no. 3, pp. 353–357, 2002.

J. F. Lawless, "*Statistical Models and Methods for Lifetime Data*, " Wiley, New York, NY, USA, 2003.

M. Z. Raqab and M. Ahsanullah, "*Estimation of the location and scale parameters of generalized exponential distribution based on order statistics*," Journal of Statistical Computation and Simulation, vol. 69, no. 2, pp. 109–123, 2001.

M. Z. Raqab, "*Inferences for generalized exponential distribution based on record statistics*, "Journal of Statistical Planning and Inference, vol. 104, no. 2, pp. 339–350, 2002.

R. Calabria andG. Pulcini, "*Point estimation under asymmetric loss functions for left-truncated exponential samples*," Communications in Statistics, vol. 25, no. 3, pp. 585–600, 1996.

R. D. Gupta and D. Kundu, "*Generalized exponential distribution: different method of estimations*," Journal of Statistical Computation and Simulation, vol. 69, no. 4, pp. 315–337, 2001.

R. D. Gupta and D. Kundu, "*Generalized exponential distributions*, " Australian & New Zealand Journal of Statistics, vol. 41, no. 2, pp. 173–188, 1999.

S. J. Press and J. M. Tanur, "*The Subjectivity of Scientists and the Bayesian Approach*, " Wiley, New York, NY,USA, 2001.

S. K. Sinha, "*Bayes estimation of the reliability function and hazard rate of a weibull failure time distribution*," Trabajos de Estadistica, vol. 1, no. 2, pp. 47–56, 1986.

S. R. Huang and S. J. Wu, "*Bayesian estimation and prediction for Weibull model with progressive censoring*," Journal of Statistical Computation and Simulation, vol. 82, no. 11, pp. 1607–1620, 2012.

T. Fagerstrom, P. Jagers, P. Schuster, and E. Szathmary, "*Biologists put on mathematical glasses*," Science, vol. 274, no. 5295, pp. 2039–2040, 1996.

دراسة مقارنة بين طريقتي Bayesian و Non-Bayesian لتحليل بيانات Survival باستخدام التوزيع الأسي المعمم

**إعداد:** فريال محمد خلف سراحين

**إشراف:** د. خالد صلاح

## الملخص

تناولت هذه الدراسة تقريب معالم المجتمع (Parameters) التي تتبع التوزيع الأسي المعمم (Generalized Exponential (GE) distribution) وذلك باستخدام الطريقة التقليدية (Maximum Likelihood Estimation ( وطريقة بيز (Bayes) لأغراض المقارنة.

الهدف من الدراسة هو تقريب معالم المجتمع (Parameters) بعدة طرق لمعرفة أفضل طريقة مناسبة للتقريب.

وبوساطة تطور الطرق السابقة قمنا باستخدام مجموعة بيانات حقيقية يكون توزيعها (GE) ومعالم هذا التوزيع ( shape parameters $(p)$ and scale $(\theta)$ ) غير معروفة، وليس لدينا أي معلومة من قبل عن توزيعها. لوحظ أن تقريب (Parameters) بطريقة MLE الكلاسيكية يعطي صور معقدة للمعلمات ونلجأ لاستخدام طرق التقريب مثل طريقة نيوتن رافسون.

في هذه الدراسة وتبعا لِـ C. Guure and S. Bosomprah (2013) اعتبرنا the Bayes estimation of the unknown parameters of the GE distribution ، كما فرضنا الـ gamma prior لكلا parameters. وبإيجاد the Bayes estimators بطريقتي squared error and general entropy loss functions وجدنا أنه من الصعب التوصل إلى صيغة سهلة لاقتران المعالم بسبب الطبيعة المعقدة لتوزيع هذه المعالم ؛ لذلك استخدمنا طريقة Lindley's numerical approximations.

من خلال النتائج توصلنا إلى أن طريقة the Bayesian estimator under general entropy loss function تعطي نتائج أفضل من طريقة MLE ، خاصة في تقدير المعالم للعينات الكبيرة نسبيا.