RISK AND ADMISSIBILITY FOR A WEIBULL CLASS OF DISTRIBUTIONS

EFREM OCUBAMICAEL NEGASH

Assignment presented in partial fulfilment of the requirements for the degree of Master of Science at the University of Stellenbosch

Supervisor: Professor Paul J. Mostert

December 2004

DECLARATION

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and that I have not previously in its entirely or in part submitted it at any university for a degree.

Signature:

Date:

SUMMARY

The Bayesian approach to decision-making is considered in this thesis for reliability/survival models pertaining to a Weibull class of distributions. A generalised right censored sampling scheme has been assumed and implemented. The Jeffreys' prior for the inverse mean lifetime and the survival function of the exponential model were derived. The consequent posterior distributions of these two parameters were obtained using this non-informative prior. In addition to the Jeffreys' prior, the natural conjugate prior was considered as a prior for the parameter of the exponential model and the consequent posterior distribution was derived. In many reliability problems, overestimating a certain parameter of interest is more detrimental than underestimating it and hence, the LINEX loss function was used to estimate the parameters and their consequent risk measures. Moreover, the same analogous derivations have been carried out relative to the commonly-used symmetrical squared error loss function. The risk function, the posterior risk and the integrated risk of the estimators were obtained and are regarded in this thesis as the risk measures. The performance of the estimators have been compared relative to these risk measures. For the Jeffreys' prior under the squared error loss function, the comparison resulted in crossing-over risk functions and hence, none of these estimators are completely admissible. However, relative to the LINEX loss function, it was found that a correct Bayesian estimator outperforms an incorrectly chosen alternative. On the other hand for the conjugate prior, crossing-over of the risk functions of the estimators were evident as a result. In comparing the performance of the Bayesian estimators, whenever closed-form expressions of the risk measures do not exist, numerical techniques such as Monte Carlo procedures were used. In similar fashion were the posterior risks and integrated risks used in the performance comparisons.

The Weibull pdf, with its scale and shape parameter, was also considered as a reliability model. The Jeffreys' prior and the consequent posterior distribution of the scale parameter of the Weibull model have also been derived when the shape parameter is known. In this case, the

estimation process of the scale parameter is analogous to the exponential model. For the case when both parameters of the Weibull model are unknown, the Jeffreys' and the reference priors have been derived and the computational difficulty of the posterior analysis has been outlined. The Jeffreys' prior for the survival function of the Weibull model has also been derived, when the shape parameter is known. In all cases, two forms of the scalar estimation error have been used to compare as much risk measures as possible. The performance of the estimators were compared for acceptability in a decision-making framework. This can be seen as a type of procedure that addresses robustness of an estimator relative to a chosen loss function.

OPSOMMING

Die Bayes-benadering tot besluitneming is in hierdie tesis beskou vir betroubaarheids- / oorlewingsmodelle wat behoort tot 'n Weibull klas van verdelings. 'n Veralgemene regs gesensoreerde steekproefnemingsplan is aanvaar en geïmplementeer. Die Jeffreyse prior vir die inverse van die gemiddelde leeftyd en die oorlewingsfunksie is afgelei vir die eksponensiële model. Die gevolglike aposteriori-verdeling van hierdie twee parameters is afgelei, indien hierdie nie-inligtingge-wende apriori gebruik word. Addisioneel tot die Jeffreyse prior, is die natuurlike toegevoegde prior beskou vir die parameter van die eksponensiële model en ooreenstemmende aposteriori-verdeling is afgelei. In baie betroubaarheidsprobleme het die oorberaming van 'n parameter meer ernstige nagevolge as die onderberaming daarvan en omgekeerd en gevolglik is die LINEX verliesfunksie gebruik om die parameters te beraam tesame met ooreenstemmende risiko maatstawwe. Soortgelyke afleidings is gedoen vir hierdie algemene simmetriese kwadratiese verliesfunksie. Die risiko funksie, die aposteriori-risiko en die integreerde risiko van die beramers is verkry en word in hierdie tesis beskou as die risiko maatstawwe. Die gedrag van die beramers is vergelyk relatief tot hierdie risiko maatstawwe. Die vergelyking vir die Jeffreyse prior onder kwadratiese verliesfunksie het op oorkruisbare risiko funksies uitgevloei en gevolglik is geeneen van hierdie beramers volkome toelaatbaar nie. Relatief tot die LINEX verliesfunksie is egter gevind dat die korrekte Bayes-beramer beter vaar as die alternatiewe beramer. Aan die ander kant is gevind dat oorkruisbare risiko funksies van die beramers verkry word vir die toegevoegde apriori-verdeling. Met hierdie gedragsvergelykings van die beramers word numeriese tegnieke toegepas, soos die Monte Carlo prosedures, indien die maatstawwe nie in geslote vorm gevind kan word nie. Op soortgelyke wyse is die aposteriori-risiko en die integreerde risiko's gebruik in die gedragsvergelykings.

Die Weibull waarskynlikheidsverdeling, met skaal- en vormingsparameter, is ook beskou as 'n betroubaarheidsmodel. Die Jeffreyse prior en die gevolglike aposteriori-verdeling van die skaalparameter van die Weibull model is afgelei, indien die vormingsparameter bekend is. In hierdie

geval is die beramingsproses van die skaalparameter analoog aan die afleidings van die eksponensiële model. Indien beide parameters van die Weibull model onbekend is, is die Jeffreyse prior en die verwysingsprior afgelei en is daarop gewys wat die berekeningskomplikasies is van 'n aposteriori-analise. Die Jeffreyse prior vir die oorlewingsfunksie van die Weibull model is ook afgelei, indien die vormingsparameter bekend is. In al die gevalle is twee vorms van die skalaar beramingsfoute gebruik in die vergelykings, sodat soveel as moontlik risiko maatstawwe vergelyk kan word. Die gedrag van die beramers is vergelyk vir aanvaarbaarheid binne die besluitnemingsraamwerk. Hierdie kan gesien word as 'n prosedure om die robuustheid van 'n beramer relatief tot 'n gekose verliesfunksie aan te spreek.

ACKNOWLEDGEMENTS

The input for this thesis stems from many dimensions, although my gratitude goes beyond the confines I can express. I would like to thank the following persons:

- First and foremost, I owe special debts of Praise to the Almighty God, the author and perfecter of my faith for bestowing His undeserved and abundant grace in my life.
- Special thanks and appreciation is due to my supervisor, professor Paul J. Mostert, for his invaluable and continual suggestions and for his open door and availing himself when I needed assistance.
- I would also like to thank the University of Asmara HRD for giving me the opportunity and sponsoring my postgraduate study. Also, Ms. Lula Gebreyesus for the efforts to enroll me at such a reputed university as the University of Stellenbosch.
- I am also indebted to professor Tertius de Wet, chairperson of the Department of Statistics and Actuarial Science, for awarding me an assistantship, which motivated me tremendously. I would also love to record thanks and due respect to all my instructors for being amenable. My sincere appreciation is due to professor Sarel J. Steel, for his continued paternal concerns and encouragements throughout my study period in Stellenbosch.
- Due respect, love and thanks go to my parents as well as the rest of my family for the initiative they took to send me to school despite the hardship they had experienced and for everything they invested in my life along the way. Multitudes of love and respect is extended to my brothers and sisters in the Lord and also all the friends for upholding me with their priceless prayers and encouragements. Next to God's grace, your company be it in person, via e-mails or phone calls, had played a vital role in my well-being during my stay away from home. Albeit, the space is too small to list your names here, be reminded that you will always remain imprinted in my heart. May the Almighty God bless you all the more!

DEDICATION

This work is dedicated to the glory of the Creator and Sustainer of the whole universe, the Lover of my soul and Saviour of my life, God Almighty!

TERMINOLOGY AND NOTATION

The various symbols, notations and abbreviations used throughout the thesis are defined and explained below. Note that some of the symbols might have been used to represent different notions in different parts of the thesis, for instance, θ has been used throughout to represent the parameter(s) of interest, be it a scalar or a vector in either the exponential or Weibull models.

pdf	probability density function
θ	parameter(s)
Θ	parameter space for θ
$S(t)$	the survival function or reliability function at time t
$h(t)$	the hazard rate or failure rate at time t
$\mathcal{F}$	family of distributions
$\mathcal{X}$	sample space for the data (observations)
$\mathcal{D}$	decision or estimation space
$f(x\|.)$	pdf of the variable X
$\ell(data\|.)$	likelihood function of the data
$\pi(\theta), \pi(\theta\|.)$	prior distribution of the parameter θ
$p(\theta\|data)$	posterior distribution of the parameter(s) of interest θ, given the data
$E[.]$	expected value
$E_{post}[.]$	expectation with respect to the posterior distribution
$var_{post}(.)$	variance with respect to the posterior distribution
$L(.)$	loss function
$L_S(.)$	squared error loss function
LINEX	linear exponential
$L_L(.)$	LINEX loss function
$\Delta_1 = \widehat{\theta} - \theta$	type I scalar estimation error
$\Delta_2 = \frac{\theta}{\widehat{\theta}} - 1$	type II scalar estimation error
$\widehat{\theta}$	estimator of θ
$\widehat{\theta}_{\mathrm{S}i}$	Bayesian estimator of θ under the squared error loss function when using Δ_i
$\widehat{\theta}_{Li}$	Bayesian estimator of θ under the LINEX loss function when using Δ_i
$R_{L,S}(\widehat{\theta}, \theta)$	risk function relative to (L=LINEX loss function, S= squared error loss function)
$R(\widehat{\theta}, \theta)$	risk function of some estimator $\widehat{\theta}$ relative to some loss function

$R(\widehat{\theta}, \theta)_{\min}$	the minimum value of the risk function of the estimator $\widehat{\theta}$ for some θ
$R_i^{L,S}(\widehat{\theta})$	posterior risk of the estimator $\widehat{\theta}$ relative to S or L using Δ_i
$r(\widehat{\theta})$	integrated risk of the estimator $\widehat{\theta}$ relative to any loss function
$r_{S,L}(\widehat{\theta})$	integrated risk of the estimator $\widehat{\theta}$ relative to S or L
$\Gamma(.)$	gamma function
$\mathcal{G}(\alpha, \beta)$	gamma pdf with parameters α and β
$\mathcal{U}_{[a,b]}$	uniform distribution between a and b
$\mathcal{I}_w$	indicator function for argument w

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In various fields where decision-making is involved, such as reliability studies in industrial settings i.e., reliability engineering, biomedical studies, etc., risk analysis is a fundamental stage. Most often, not all observations regarding the above areas of study are fully measured due to time, limited budget or any natural phenomenon for that matter. Hence, censored (type I or type II) or truncated lifetimes are considered despite the fact that they create analytical difficulties in making inferences. Furthermore, these inconveniences also dictate some analysts to use nonparametric procedures (Kvam *et al.*, 1999).

However, it is not unusual to assume parametric models which are believed to represent many settings in the reliability study. The focus of this thesis lies in the Bayesian approach to decision-making process, mainly risk analysis using the parametric models pertaining to the Weibull class of distributions.

In general, there are two schools of thoughts in statistical inference namely, the Bayesian approach and the traditional or frequentist approach which is solely based on sampling. In the sequel, the terms Bayesian and frequentist will be used for convenience purposes.

A frequentist seeks estimates or decisions of certain parameter(s) based only on the model obtained from the observable data, i.e. sampling. A Bayesian, on the other hand, seeks a decision that is based on the information prior to collecting the data and then combining this with the information obtained after data collection. Bayes' theorem is the link between the two sources of information for the Bayesian analyst, updating the total information.

Using only current products or observed data to perform statistically sufficient tests and the desired inferences are often impractical. Hence, a need arises to use past experiences or knowledge to incorporate in the conclusions from currently available data. By applying classical statistical theory, past experiences are set as assumptions and do not involve any risk analysis if not in the inferences. Therefore, one has to carry out the experiment a number of times, which makes testing or inferences a very lengthy task, as no prior knowledge has been used as part of the decision-making process. In the Bayesian paradigm, however, the prior knowledge is summarised in such a way that an appropriate prior distribution is chosen to incorporate the past experience in the risk analysis.

1.1 OVERVIEW OF THE DECISION-MAKING PROCESS

In the Bayesian approach to the decision-making process, the analyst summarises the prior information, which may be based on personal beliefs or the client's belief, past experience, etc. and usually represented it in the form of a so-called *prior distribution*. The information obtained after data collection is summarised in the form of the *likelihood*. The Bayes' theorem plays as a liaison between these two sources of information updating it into the *posterior distribution*. A certain loss function is chosen based on some criteria and then used along with the posterior distribution in Bayesian inference. More often, the choice of an appropriate loss function pertaining the problem at hand is a process that needs a decision. The loss functions, once chosen, are used to obtain the Bayesian estimators and their corresponding risk measures. These risk measures could be either the risk function also referred to as frequentist risk, the posterior risk or integrated risk. At the risk analyses stage, the risk measures are used to compare the performance of the various Bayesian estimators, thereby choosing an alternative that results in minimum risk. It should be noted that in the sequel, if an estimator that is obtained relative to a certain loss function is used to evaluate the risk relative to another loss function, such an estimator is identified as a "wrongly chosen alternative" or a "wrong estimator". However, an estimator corresponding to the same loss function under consideration is regarded as the "correct estimator" or the "correct Bayesian estimator". The final stage in this case is then to make inferences and draw conclusions from the estimators or any other predictive values or hypothesis tests. The flow chart given in figure 1 briefly demonstrates the decision-making process in the two schools of thoughts.

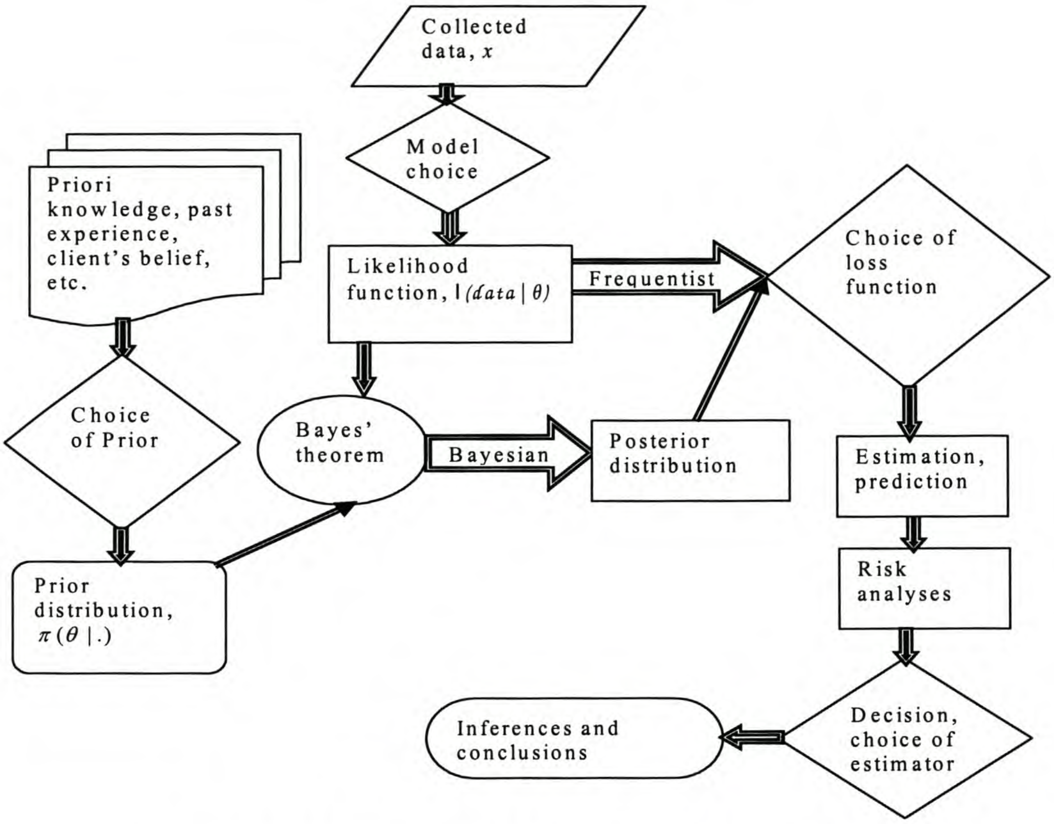

Figure 1: An overview of the decision-making process.

1.2 OUTLINE OF THE THESIS

The remainder of the thesis is schemed as follows. In chapter 2, the general concepts of a decision-making process is presented with the main focus being the Bayesian perspective. The reliability models of the Weibull class of distributions, their survival functions, likelihood and the failure rates are also discussed. In chapter 3, some Bayesian computational techniques, which are often used in Bayesian analyses or other analyses are discussed. The Bayesian analysis of the one-parameter exponential model regarding the parameter θ and the survival function is performed in chapter 4. Both the squared error loss and the LINEX loss functions (Varian, 1975) have been used. The two loss functions are used relative to two types of the estimation error in order for a decision-maker to have a bit of flexibility in choosing from as many alternatives as possible. In these analyses, both the Jeffreys' prior and the conjugate prior have been assumed as the prior distributions of the unknown parameter θ, while only the Jeffreys' prior has been used for the survival function. Chapter 5 presents the two-parameter Weibull model, considering both the scale and shape parameters, as well as the survival functions. Here, two situations have been considered: treating only the scale parameter to be unknown and secondly the more general case, where both scale and shape parameters are treated to be unknown. Chapter 6 concludes the thesis.

CHAPTER 2

DECISION-MAKING PROCESS

2.1 INTRODUCTION

In many dimensions of everyday life, human beings need to make crucial decisions, which quite often involve making choices among alternatives. Many examples of these choices exist and all these lie in the sphere of decision-making. In many cases, however, decisions are made without a thorough study or an analysis of consequences, which really form the base of any decision-making process. In this chapter, the basic concepts of the decision-making process will be discussed, the main focus being the Bayesian approach.

The concept of statistical decision theory was first introduced by Wald (1950) and is now commonly applied by various professionals of different fields in solving real problems. Some of these professionals include operational analysts, statisticians, computer scientists, chemical engineers, economists, epidemiologists and many more.

Decision-making or statistical inference is aimed at seeking a procedure or decision rule that is cost effective, either for the analyst or for the client. The decision reached is normally expected to meet, as much as possible, some degree of satisfaction for the decision-maker and/or the client. Cost effectiveness here means that the decision should have the least risk (or minimum risk) possible and thus bestows the utmost or optimal satisfaction. In obtaining a cost effective and optimal alternative, the analyst requires a summary of information, thoughts or ideas, which can be represented in mathematical functional forms or *models*. These models are normally indexed by parameters, which can be known or unknown, depending on the situation. The parameter space will be indicated by Θ. The space of all decisions or in terms of a model, the space of all possible estimators for the parameters, is indicated by $\mathcal{D}$. Collecting data from random samples will aid one to obtain effective estimators/decisions for the unknown parameters and the space of all possible samples relative to the model is indicated by $\mathcal{X}$. Consequently, a decision-making process involves these three spaces: Θ, $\mathcal{D}$ and $\mathcal{X}$.

The information summarised in a model can normally be obtained from two sources when using the Bayesian approach, firstly from experts in the field, prior to collecting the data and secondly, from the likelihood after collecting the data. Bayesian analysis is a statistical approach which combines the information from these two sources and as a result a *posterior* model is obtained. The posterior model is therefore an updated probabilistic model of all the information relevant to the parameters of interest (Jeffrey, 1997). Estimators or decisions obtained when using the posterior distribution should lead to a minimum expected loss over a certain distribution of the parameter of interest (Rice, 1995). The concept of the probability models, the priors, the loss functions, the risk measures and the relative comparisons of the estimators, will be explained in subsequent sections.

In section 2.2, the commonly used reliability models of the Weibull (named after the Swedish engineer Waloddi Weibull (1887-1979)) class of distributions are discussed. Two broad categories of prior distributions, which focus on the non-informative and the conjugate priors, are discussed in section 2.3. The concept of posterior analysis is presented in section 2.4. Section 2.5 discusses the concept of loss functions, which is incorporated in the estimation procedure of the decision-making process. In section 2.6, the different risk measures are explained and the criteria for estimation in applied Bayesian decision theory is discussed. The concepts of admissibility and inadmissibility and some background studies on the admissibility of estimators are given in section 2.7 and section 2.8 concludes the chapter.

2.2 RELIABILITY MODELS OF THE WEIBULL CLASS OF DISTRIBUTIONS

The Weibull class of distributions is part of a broader class, namely the exponential family of models. The Weibull class of distributions are the only type of models that will be considered in this thesis.

Suppose a non-negative monotonically increasing function of x, $g(x)$, is defined, such that $g(x)$ is zero when x is zero and diverges to infinity as x tends to infinity. Given such a function, the cumulative probability function can be defined as

$$F(x) = 1 - \exp[-g(x)]. \tag{2.1}$$

Clearly, (2.1) is zero, if x is zero and tends to unity, as x tends to infinity. Hence, a proper pdf can now be defined. The first derivative of (2.1) yields the corresponding pdf, given by

$$f(x) = \frac{dF(x)}{dx} = g'(x)\exp[-g(x)]. \tag{2.2}$$

The failure rate or hazard rate (also known as the occurrence rate or intensity) of (2.2), denoted by $h(x)$, is given as

$$h(x) = \frac{f(x)}{1 - F(x)} = g'(x). \tag{2.3}$$

The relationships (2.1) and (2.2) allow one to define any pdf on the non-negative real line with specified hazard rate of the form (2.3).

The hazard rate dictates in many cases the form of the pdf. Often, a flexible hazard rate is needed to successfully describe the information from a sample. Monotonically increasing and decreasing hazard rates are very popular choices and the hazard rate of the Weibull class of distributions satisfies this criteria. The hazard rate for the Weibull class of distributions is

$$h(x) = \lambda\theta^{\lambda}x^{\lambda-1}, \qquad \lambda, \theta > 0, \tag{2.4}$$

where λ and θ are the shape and scale parameters, respectively. The corresponding $g(x)$ of (2.4) is

$$g(x) = (\theta x)^{\lambda}.$$

Using (2.2) and (2.4), the corresponding pdf for the Weibull class is given by

$$f(x|\lambda, \theta) = \lambda\theta^{\lambda}x^{\lambda-1}\exp[-(\theta x)^{\lambda}]. \tag{2.5}$$

Other representations of (2.5) exist, which will be defined in a later section. Information from a random sample is summarised in the likelihood function, assuming these underlying models of the Weibull class of distributions. The sampling scheme considered for this thesis is where n components or objects are observed and only r lifetimes, $x_1 < x_2 < \ldots < x_r$, are fully measured, while the remainder, $(n - r)$, are censored. These $(n - r)$ censored lifetimes will be ordered separately and denoted by $x_{r+1} < x_{r+2} < \ldots < x_n$. This scheme is known as generalised right censoring (Dellaportas and Wright, 1991) and when a fully observed sample is considered, r is chosen to be equal to n. The likelihood function in the case of a generalised right censored sample is

$$\ell(data|\lambda,\theta) = \prod_{i=1}^{r} f(x_i|\lambda,\theta) \prod_{i=r+1}^{n} (1 - F(x_i|\lambda,\theta)). \tag{2.6}$$

The Weibull class of distributions is widely used in many types of analyses due to its mathematical tractability and also its applicability in reflecting many real world problems of repairable components (Ananda and Ravindra, 1998). The hazard rate in (2.4) portrays the growth of the components under study when $\lambda > 1$ or a decay in a situation when $\lambda < 1$. For $\lambda = 1$ (exponential pdf), a constant hazard rate is described and when $\lambda = 2$ (Rayleigh pdf) a linear increasing hazard is described. These choices of λ makes the Weibull class a very popular model to use and will be the focus of this thesis. The next sections explain in detail the pdfs that arise from the Weibull class, as well as some of the important parameters to be estimated from these models.

2.2.1 Exponential model

A constant hazard rate models a component's failure time to the lack of memory principle. Whenever a failure occurs and is repaired, the subsequent observed lifetime is independent of what was observed before failure. The one parameter exponential model has a constant hazard rate and is a special type of the Weibull class (2.5), with the shape parameter $\lambda = 1$. The exponential model to be considered in this thesis, is given by

$$f(x|\theta) = \theta \exp[-\theta x], \qquad x,\ \theta > 0. \tag{2.7}$$

The survival function, the other important parameter to be estimated, i.e. the probability that a certain component stays under operation at least up to a specified time x, is

$$\begin{aligned} S(x) &= \Pr(X > x) = 1 - F(x) \\ &= \exp[-\theta x], \qquad x,\ \theta > 0. \end{aligned} \tag{2.8}$$

Function (2.8) converges to 0 as x increases to infinity, which means that no component stays in operation for eternity.

The hazard rate of (2.7) is constant and equal to the inverse of the mean of the distribution, i.e.

$$h(x) = \theta. \tag{2.9}$$

Using (2.6) and (2.7), the likelihood function for the generalised right censored sample is

$$\ell(data|\theta) \propto \theta^r \exp[-T\theta], \tag{2.10}$$

where

$$T = \sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n} x_j = \sum_{i=1}^{n} x_i.$$

However, if there is no right censoring, $r = n$.

2.2.2 Weibull model

The Weibull model represents many reliability problems and consequently, it is widely used by many Bayesian analysts (Dellaportas and Wright, 1991; Martz and Waller, 1982). Mainly two ways of representation for the Weibull model exist by changing the functional form of the scale parameter. The two representations are

$$f(x|\lambda, \theta) = \lambda\theta x^{\lambda-1} \exp[-x^{\lambda}\theta] \tag{2.11}$$

and

$$f(x|\lambda, \theta) = \lambda\theta^{\lambda} x^{\lambda-1} \exp[-(\theta x)^{\lambda}], \qquad \lambda,\ \theta > 0. \tag{2.12}$$

The corresponding survival function of (2.11) is

$$S(x) = \exp[-x^{\lambda}\theta]$$

and that of (2.12) is

$$S(x) = \exp[-(\theta x)^{\lambda}].$$

The hazard rate of the Weibull model was defined in (2.4) and the hazard rate corresponding to the model (2.12) is given by

$$h(x) = \lambda\theta^{\lambda} x^{\lambda-1}.$$

Under the generalised right censored sampling scheme (2.6), the likelihood for model (2.11) is defined as

$$\ell(data|\lambda, \theta) \propto \lambda^r \theta^r \left(\prod_{i=1}^{r} x_i^{\lambda-1} \right) \exp[-T\theta], \tag{2.13}$$

with

$$T = \sum_{i=1}^{n} x_i^{\lambda},$$

and similarly for (2.12).

Remark 2.1 *The two models (2.11) and (2.12) can also be represented by substituting θ with $\frac{1}{\theta}$, respectively. The consequent survival functions, hazard rates and likelihood functions can easily be transformed for this choice of the parameter. These two representations of the Weibull model are needed in later work to obtain results in closed-form.*

Remark 2.2 *As stated earlier, when $\lambda = 2$ in the Weibull model, it reduces to the Rayleigh distribution. Lord Rayleigh (1919) derived and used this distribution in a problem of acoustics. If X_1 and X_2 are both normally distributed variables with mean zero and variance equal to σ^2, then $Y = \sqrt{X_1^2 + X_2^2}$ (distance from the origin to (X_1, X_2) in 2-dimensional Euclidian space) will be a variate from the Rayleigh distribution with parameter $\theta = \frac{1}{2\sigma^2}$. The Rayleigh distribution has been widely used in life testing problems and clinical studies dealing with cancer patients (Gross and Clark, 1975; Lee, 1980; Mostert et al., 1998).*

2.3 PRIOR INFORMATION

In the previous sections, the concept of the decision-making process and the models of the Weibull class of distributions have been discussed. The models are mathematical representations used to depict the summary of the information about the components under consideration. It has been pointed out that the information involved in a decision-making process has in essence two sources, the prior and the likelihood. In this section, the concept of prior information will be discussed, with special attention to non-informative and conjugate priors. The prior encapsulates the knowledge about the parameter of interest (a scalar or vector depending on the particular problem), which the analyst or client has summarised before collecting the data (Jeffrey, 1997). The information might be from past experience, the analyst's and/or the client's belief or presumptions or in the worst case, due to lack of information.

In most practical problems, the analyst knows beforehand certain information about the parameter of interest (Zellner, 1980) and that he/she would probably have an idea what values to choose for the parameter(s) through an appropriate pdf. It might also be possible to have knowledge, obtained prior to collecting the data about the parameter and the likelihood to attain certain value(s) in the subset of the parameter space (Bhattacharya, 1967). The prior knowledge is often represented in the form of a pdf, or a prior distribution. The choice of the prior distribution is often made subjectively, yet in a way that satisfies the specified prior summary and on the grounds of convenience both for the analyst and/or the client. On the other hand, prior distributions must meet some basic tests of credibility, for instance, they should not contradict and disallow the information from observed data (Box and Tiao, 1973). Uncertainty of the existence of any prior information should evolve in approximations of the choice of priors. A prior distribution can be either proper or improper, depending on whether it integrates to unity or not. Furthermore, it might be informative, conveying the information or non-informative, conveying ignorance (Jeffrey, 1997). The important point is, to choose the closest approximation that best represents reality. In the Bayesian paradigm, incorporating the prior distribution is advantageous, since it uncovers the assumptions of the underlying problem by recapping the form of a distribution, which might be hidden in the classical approach. Whenever possible, prior distributions should be specific and should reflect the situation under consideration, otherwise they fail to comply with reality and do not meet the satisfaction of the client and/or the analyst him/herself (Jeffrey, 1997). These two widely used types of priors will be discussed in the subsequent sections.

2.3.1 Non-informative priors

Even though in many statistical analyses enough information might be obtained from elicitation, yet in many others, analysts might also fall short of obtaining information about the parameter in their study (Zellner, 1980). An important point to rise here is, *what should Bayesian analysts do about this lack of information*. If no prior knowledge is available for the experiment, the Bayesian is obliged to perform the analysis under total or partial ignorance (Ghosal, 1997). There were times when Bayesian analysts had to subjectively choose a prior under the name of *subjective priors*. This method of choosing priors subjectively on the belief of the analyst, however, exposed this method to severe critics. Subsequently, tremendous work has been done in the past decades to find ways of obtaining non-subjective priors (Zellner, 1980). The quest of finding non-informative priors dates back to the time of Laplace and Bayes when they based their research under name the of *inverse probability,* instead of non-informative priors. The use of non-informative priors in Bayesian analysis often leads to procedures with approximate frequentist validity. Thus, while retaining the Bayesian flavour on the one hand, non-informative priors allow for some reconciliation between the two conflicting paradigms of statistics and therefore provide mutual justification. Many Bayesians also use different names for non-informative priors, such as conventional, default, flat, formal, neutral, non-subjective, objective and reference priors (Bernardo, 1997). Even though non-informative priors are most often improper, there exist priors that are proper, e.g., the Beta $(\frac{1}{2}, \frac{1}{2})$ is a Jeffreys' prior for data from a binomial distribution. Very little is known about non-informative priors in infinite dimensional problems (Ghosal, 1997). Gradually though, the concept of non-informative priors found a respected place in the statistical literature. The tremendous work of Jeffreys (1946, 1961) in obtaining non-informative priors, particularly the *Jeffreys' prior,* has played a distinguishable role. The introduction of the Jeffreys' prior has boosted the confidence of the Bayesian community and supplemented the credibility when choosing non-informative priors. This is due to the derivation of priors directly from the distribution of the data (Robert, 2001). The Jeffreys' non-informative prior is based on the *Fisher information matrix*. Under some regularity conditions and for any pdf, say $f(x|\theta)$, with parameter(s) $\theta \in \mathcal{R}^m$ $(m \geqslant 1)$, the Fisher information matrix, with ij^{th} element, is obtained from

$$I_{ij}(\theta) = -E_\theta \left[\frac{\partial^2 \{\log[f(x|\theta)]\}}{\partial\theta_i \partial\theta_j} \right]. \tag{2.14}$$

The expectation in (2.14) is taken with respect to the data. From (2.14), the Jeffreys' non-informative prior is given by

$$\pi(\theta) \propto (\det[I(\theta)])^{\frac{1}{2}}. \tag{2.15}$$

The choice of the prior based on $I(\theta)$ is justified by the fact that it is an indication of amount of information obtained directly from the data (Robert, 2001).

Although the Jeffreys' prior is useful in many Bayesian analyses, it is not very effective in multi-parameter models (Robert, 2001). Due to this shortcoming, the *reference prior* was introduced by Bernardo (1979). The reference prior is basically a modification of the Jeffreys' prior, where in the latter, the prior distribution merely depends on the data and in the former the prior depends both on the data and the inferences to be done (Robert, 2001). In a one parameter model, the reference prior and Jeffreys' prior are identical. The idea, in any experiment, is to maximise the information obtained from prior knowledge, as well as the information expected to be revealed when collecting the data. The more prior knowledge about the parameter(s) of interest is available, the less information is expected from the data (Berger and Bernardo, 1989). It is important to note that the information to be obtained from the observations overweighs the prior knowledge, especially in the presence of 'vague' information. Therefore, the need of the reference prior stems to alleviate the ambiguity that may arise from vague prior knowledge and is aimed at maximising the missing information about the parameter of interest (Bernardo and Ramon, 1998).

In order to lay the ground for procedures to obtain the reference prior, it is important that the summary of information be amassed in a mathematical model. Suppose that in an experiment, the underlying distribution of the data is given by $f(x|\theta)$ and the prior knowledge is summarised by $\pi(\theta)$. The summary of the information about the parameter(s) θ to be expected from the experiment, is given by

$$I^{\theta}\{f,\pi\} = \int\limits_{\mathcal{X}}\int\limits_{\Theta} f(x|\theta)\pi(\theta)\log\left[\frac{p(\theta|data)}{\pi(\theta)}\right] d\theta dx. \tag{2.16}$$

The prior distribution, $\pi(\theta)$, that maximises (2.16) is the desired reference prior (Berger and Bernard, 1989). The expected information of (2.16) is non-negative as well as invariant under one-to-one transformation of θ, even when any sufficient statistic of the data is used (Bernardo and Ramon, 1998). Repeating a certain experiment infinitely many times would disclose the missing information. Hence, the definition in (2.16) is often used for a sequence of k independent replications of the experiment under consideration. The corresponding information that the sequence of experiments would provide, is given by

$$I^{\theta}\{f(k),\pi\} = \int\limits_{\mathcal{X}}\int\limits_{\Theta} f(z_k|\theta)\pi(\theta)\log\left[\frac{p(\theta|z_k)}{\pi(\theta)}\right] d\theta dz_k, \tag{2.17}$$

where $z_k = (X_1, X_2, ..., X_k)$ is the sequence of independent replications of the experiment. In this case, the limit of the sequence of priors, $\pi_k(\theta)$ as $k \to \infty$, which maximises the missing information (2.17) in the domain of strictly positive priors, is the reference prior. The following simple procedure can be used to obtain the reference prior.

Suppose two parameters θ and λ in a model, $f(x|\lambda,\theta)$, where θ is the parameter of interest and

λ is a nuisance parameter (known or assumed a certain value). The corresponding expected Fisher information matrix can be partitioned and represented as

$$I(\lambda, \theta) = \begin{bmatrix} I_{11}(\lambda, \theta) & I_{12}(\lambda, \theta) \\ I_{21}(\lambda, \theta) & I_{22}(\lambda, \theta) \end{bmatrix}.$$

The reference prior is obtained with the following steps (Berger and Bernardo, 1989).

Step 1: Let $\pi(\theta|\lambda)$ be the usual reference prior for θ given λ, which is given by

$$\pi(\theta|\lambda) \propto (\det[I_{11}(\lambda, \theta)])^{\frac{1}{2}},$$

where

$$I_{11}(\lambda, \theta) = -E_\theta \left[\frac{\partial^2 \{\log[f(x|\lambda, \theta)]\}}{\partial \theta^2} \right].$$

Step 2: A sequence of subsets of the parameter space are chosen, such that $\Delta_1 \subset \Delta_2 \subset \Delta_3 ...$ and $\cup_i \Delta_i = \Theta$. Suppose the reference prior in step 1 has a finite mass on $\Omega_{i,\lambda} = \{\theta : (\lambda, \theta) \in \Delta_i\}$ for all λ. Normalise $\pi(\theta|\lambda)$ on each $\Omega_{i,\lambda}$ and obtain

$$\pi_i(\theta|\lambda) = K_i(\lambda)\pi(\theta|\lambda)1_{\Omega_{i,\lambda}}(\theta),$$

where 1_Ω is an indicator function on Ω and

$$K_i(\lambda) = \frac{1}{\int\limits_{\Omega_{i,\lambda}} \pi(\theta|\lambda) d\theta}.$$

Step 3: Find the marginal reference prior for λ with respect to $\pi_i(\theta|\lambda)$. This is given by

$$\pi_i(\lambda) = \exp \left[\frac{1}{2} \int\limits_{\Omega_{i,\lambda}} \pi_i(\theta|\lambda) \times \log \left[\frac{\det[I(\lambda, \theta)]}{\det[I_{11}(\lambda, \theta)]} d\theta \right] \right],$$

assuming that the integral exists.

Step 4: Assuming that the limit exists, the reference prior for the joint parameters, (λ, θ), is now defined as

$$\pi(\lambda, \theta) = \lim_{i \to \infty} \left[\frac{K_i(\lambda)\pi_i(\lambda)}{K_i(\theta_0)\pi_i(\theta_0)} \pi(\theta|\lambda) \right], \tag{2.18}$$

where θ_0 is any fixed point.

2.3.2 Conjugate priors

The conjugate priors are sometimes referred to as the priors *closed under sampling* due to Wetherill (1961). A given family of distributions, say $\mathcal{F}$, are said to be a conjugate family if for a given likelihood function and for any prior distribution in the family $\mathcal{F}$, the posterior distribution also belongs to that same family. Hence, a conjugate prior is one, when combined with the likelihood function and its parametric form leads to a posterior distribution that has a form that allows analytical inspection i.e., it is represented in a recognisable form of some known distribution (Jeffrey, 1997). The concept of a conjugate prior was introduced by Bernard (1954) and thoroughly studied by Raiffa and Schlaifer (1961). Before the recent advances of developing techniques for the derivation of the prior distributions, researchers had difficulty in obtaining a conjugate prior which suits the assumed likelihood. Nowadays, many analysts rather choose to obtain suitable priors befitting their prior beliefs using numerical techniques (Jeffrey, 1997). The main interest of conjugacy lies when $\mathcal{F}$ is as small as possible and parameterised, in which case the posterior distribution is merely an updated prior distribution of the parameter (Robert, 2001).

Suppose there is a family of pdfs of the parameter θ, then two conditions, which are sufficient for the existence of a conjugate family, are:

i. For any sample size n, the joint distribution of the sample $X_1, X_2, ..., X_n$ conditioned on θ, is proportional to one of the pdfs in the family.
ii. The family is closed under multiplication, i.e. a family of distributions is said to be closed under multiplication provided that for any two distributions of the parameter θ, say $f(\theta|\gamma_1)$ and $f(\theta|\gamma_2)$, then

$$f(\theta|\gamma_3) \propto f(\theta|\gamma_1) \times f(\theta|\gamma_2),$$

where γ_1, γ_2 and γ_3 are called hyperparameters.

Prior elicitation plays the most crucial role in Bayesian inference. Historical data from past similar studies can be very helpful in interpreting the result for a current study. The power prior is defined to be the likelihood function based on historical data, raised to a power, γ, where $0 \leq \gamma \leq 1$. The parameter γ, therefore, controls the influence of the historical data. In this case, the power prior has very often the form of the conjugate prior, due to the nature and likelihood of the problem and data.

2.4 POSTERIOR ANALYSIS

The Bayesian paradigm, as discussed in the above sections, is based on specifying a probability model for the data, given the unknown parameter(s), θ. This leads to defining a likelihood function $\ell(data|\theta)$ and assuming θ is a random variable with prior distribution denoted by $\pi(\theta)$. Inference concerning θ is then based on the posterior distribution, which is obtained via Bayes' theorem (Bayes, 1958)

$$p(\theta|data) \propto \ell(data|\theta)\pi(\theta). \tag{2.19}$$

Using (2.19) with the appropriate normalising constant, the posterior distribution is given by

$$p(\theta|data) = \frac{\ell(data|\theta)\pi(\theta)}{\int\limits_{\Theta} \ell(data|\theta)\pi(\theta)d\theta}. \tag{2.20}$$

The denominator in (2.20) is called the marginal of data or prior predictive distribution.

Apart from using it in inference, the posterior distribution has two other main focuses:

i. Obtaining the Bayesian point estimator that minimises the posterior expected loss. Provided that the conditions of differentiation inside the integration are met, the Bayesian estimator is obtained by solving for $\widehat{\theta}$ in

$$\frac{d}{d\widehat{\theta}}\left(\int\limits_{\Theta} L(\widehat{\theta},\theta)p(\theta|data)d\theta\right) = 0. \tag{2.21}$$

ii. Another major aspect within the Bayesian paradigm is prediction of a new observation. The posterior predictive distribution of a future lifetime of a component z given the data, is

$$p(z|data) = \int\limits_{\Theta} f(z|\theta)p(\theta|data)d\theta, \tag{2.22}$$

where $f(z|\theta)$ denotes the sampling density of z and $p(\theta|data)$ is the posterior distribution of θ.

The posterior predictive distribution (2.22) is the posterior expectation of $f(z|\theta)$ and sampling from (2.22) is easily accomplished via numerical techniques, such as Gibbs sampling.

2.5 LOSS FUNCTIONS

The notion of the decision-making process, the reliability models and the prior information were discussed in the previous sections. In this section, the concept of loss functions, which handles the measurement or performance of estimators will be discussed. Once the analyst selects an appropriate prior distribution, the need arises to select a suitable loss function to ensure that an optimal estimator is made. A loss function in this context is a function of the form

$$L(\widehat{\theta}, \theta) = \rho(\theta)\gamma(\widehat{\theta}, \theta), \tag{2.23}$$

where $\rho(\theta)$ is any non-negative function that reflects the relative seriousness of a given error for different values of the parameter and γ is a non-negative function of the error satisfying $\gamma(0) = 0$. The loss function of the form (2.23) measures or quantifies the loss or penalty incurred when the estimate of θ, $\widehat{\theta}$, is used, i.e. it evaluates the severity of the consequence of making a certain decision about the parameter. A systematic development of point estimation of a parameter can be achieved by introducing an appropriate loss function, which best suits to the decision-making problem. Various types of loss functions are commonly used in the estimation and prediction procedures of statistical inference. In practice, the choice of an appropriate loss function that best represents a certain scenario under consideration is not an easy task, especially when the parameter space is infinitely large. Quantifying a qualitative consequence of any decision further aggravates the situation.

Statisticians are often confined to use classical or canonical loss functions which are acceptable for various reasons, amongst others, for their simplicity and mathematical tractability. Some of the classical loss functions most often used are the squared error loss (also referred to as quadratic loss), the weighted quadratic loss, the absolute error loss (Laplace, 1773) or multilinear function and the 0-1 loss function (used in hypothesis testing). The linear-exponential loss function (Varian, 1975) is widely used in many Bayesian analyses, since its introduction. The choice of a loss function is sometimes criticised and it is as difficult as choosing the prior distribution of a parameter. The squared error loss and LINEX loss functions will be discussed in the following sections.

2.5.1 Squared error loss function

The squared error loss (Gauss, 1810) is one of the most predominantly used loss functions in statistical decision problems. This loss function is one of the many symmetrical loss functions, which equally penalise overestimation and underestimation of a certain parameter with the same magnitude (Varian, 1975). Mathematically, this implies that $L(\Delta) = L(-\Delta)$, where Δ is the scalar estimation error incurred. The squared error loss function is given by

$$L_S(\widehat{\theta}, \theta) = k(\widehat{\theta} - \theta)^2, \tag{2.24}$$

where k is any positive constant.

A more general form of (2.24) is

$$L_S(\widehat{\theta}, \theta) = w(\theta)(\widehat{\theta} - \theta)^2,$$

where $w(\theta)$ is any positive function of θ. The estimation error in (2.24) is of the form (type I)

$$\Delta_1 = \widehat{\theta} - \theta.$$

Remark 2.3 Another form of the estimation error which will be considered in this thesis is $\Delta_2 = \frac{\theta}{\widehat{\theta}} - 1$ *(called type II estimation error). The type I estimation error can also be generally represented by* $\Delta_1 = \widehat{\varphi}(\theta) - \varphi(\theta)$ *and the type II estimation error by* $\Delta_2 = \frac{\varphi(\theta)}{\widehat{\varphi}(\theta)} - 1$, *where* $\widehat{\varphi}(\theta)$ *is the estimator of the one-to-one function of* θ, $\varphi(\theta)$, *(Mostert et al., 1999).*

The Bayesian estimator relative to squared error loss function defined in terms of the type I estimation error is obtained by using (2.21) and consequently solving for $\widehat{\theta}_{S1}$ in

$$\frac{d}{d\widehat{\theta}_{S1}} \left(E_{post} \left[k(\widehat{\theta}_{S1} - \theta)^2 \right] \right) = 0$$

to obtain

$$\widehat{\theta}_{S1} = E_{post}[\theta]. \tag{2.25}$$

An analogous result of (2.25) can be obtained when using

$$\begin{aligned} E_{post} \left[L_S(\widehat{\theta}_{S1}, \theta) \right] &= k E_{post}[\widehat{\theta}_{S1} - \theta]^2 \\ &= k \left(\widehat{\theta}_{S1} - E_{post}[\theta] \right)^2 + k var_{post}(\theta). \end{aligned} \tag{2.26}$$

The minimum of (2.26) is obtained when $\widehat{\theta}_{S1} = E_{post}[\theta]$ and the minimum expected posterior loss is a scalar multiple of the posterior variance (when $k = 1$ in (2.26) the minimum expected posterior loss is equal to the variance of the posterior distribution).

Considering the type II estimation error, the Bayesian estimator under the squared error loss is obtained when solving $\widehat{\theta}_S$ in

$$-\frac{2}{\widehat{\theta}_{S2}^3} E_{post}\left[\theta^2\right] + \frac{2}{\widehat{\theta}_{S2}^2} E_{post}[\theta] = 0$$

to obtain

$$\widehat{\theta}_{S2} = \frac{E_{post}\left[\theta^2\right]}{E_{post}[\theta]}. \tag{2.27}$$

The loss function given in (2.24) has been widely used in statistical analyses due to its mathematical tractability (Singh *et al.*, 2004). The squared error loss has also been used in reliability problems over the years. The literature has been surging and all of the references are far too enormous to list here, amongst them are: El-Sayyad (1967) to obtain an unbiased estimator of the parameter of the exponential model; Basu and Ebrahimi (1991) in their study of the reliability exponential model and the comparison of the performance of Bayes point estimators; Samaniego and Reneau (1994) in the context of exponential families, under the conjugate priors, to compare Bayesian and classical estimators; Mostert *et al.* (1999) in the estimation of Bayesian estimators for the Rayleigh model; Chaturvedi *et al.* (2000) in analysis of disturbances variance in the linear regression models; Mostert *et al.* (1998) in the analysis of cancer survival times using the Weibull model; Farsipour and Asgharzadeh (2002) in an admissibility study of an estimator of the linear combination of the common mean of two populations from a normal distribution; Singh *et al.* (2004) in obtaining the Bayesian estimators of the exponentiated-Weibull distribution.

The squared error loss function has many important properties and some of them are:

- The penalty incurred due to estimation in (2.24) does not depend on the sign, but on the magnitude of the error. This incurring same penalty for underestimation and overestimation of equal magnitude is sometimes considered as a drawback.
- The estimate that minimises the expected loss is obtained by calculating the posterior mean in the Bayesian context.

- Loss functions that are strictly convex result in Bayesian estimators that are unique. Clearly, the squared error loss is strictly convex, since the second derivative of (2.24) yields $2k > 0$, hence, any Bayesian estimator with respect to the squared error loss is unique.
- The squared error loss avoids the concept of randomised estimators, i.e. estimators with an infinite Bayes risk. This is due to the property that the Bayesian estimator under the squared error loss, is the posterior mean.
- The squared error loss function can also be used as an approximation of more complex symmetric loss functions, via Taylor series expansions (Robert, 2001).

2.5.2 LINEX loss function

Many real world problems exist where symmetric loss functions would not be plausible to use. In fact, it has been noted by Bhattacharya *et al.* (2002) that in the study of hydrology, reliability engineering and pharmaceuticals, the use of asymmetric loss functions are believed to be more appropriate than symmetric loss functions. Nonetheless, in many problems of point estimation of location parameters, the symmetric loss functions are the most widely used. The assumption of symmetrical loss functions, however, in many instances compromised the reality and hence, might have led to erroneous conclusions. Therefore, the use of asymmetric loss functions is necessitated, since in reality when one predicts a given parameter of interest, overestimation may be more severe than underestimation or vice versa (Parsian and Farsipour, 1993). As a consequence, overestimating the survival function of an item might be much more severe than underestimating it. This in turn implies that underestimating the hazard rate results in a more serious consequence than an overestimating of the hazard rate. Therefore, if the reliability is overestimated, a system may fail prior to its predicted failure time causing serious consequences (Basu and Ebrahimi, 1991). Another example that can be mentioned was the disaster of a space shuttle, where it was believed that its failure rate was underestimated and thus failed prior to the time predicted (Feynman, 1987). Underestimating the peak of the water level in dam construction might also be disastrous than overestimation (Zellner, 1986). More references of such examples are in Pandy *et al.* (1996) and Singh *et al.* (2004). These examples justify that there are situations where the symmetric loss function will not be appropriate. To alleviate the problem of using unsuitable symmetric loss functions which do not represent the problem, many authors used *asymmetric linear* loss functions, namely, Ferguson (1967); Zellner and Geisel (1968); Aitchison and Dunsmore (1975); Berger (1985) and many others. Unlike many of these authors, Varian (1975) introduced the asymmetric LINEX loss function, whose modified functional form is given as

$$L_L(\Delta) = \exp[a\Delta] - a\Delta - 1\,, \qquad\qquad a \neq 0. \tag{2.28}$$

Remark 2.4 *A general form of (2.28) is defined as*

$$L_L(\Delta) = b\{\exp[a\Delta] - c\Delta - 1\},$$

where a and c are nonzero constants and b is positive.

The sign of the shape parameter in (2.28) reflects the direction of the asymmetry and its magnitude reflects the degree of asymmetry. Varian introduced the loss function (2.28) in the study of an appraisal of single-family homes in a real estate assessment. The problem stems from an assessor in need of estimating the current market value of a house taking into consideration certain characteristics like, total living area, number of bedrooms, etc. The motivation for the use of (2.28) was on the basis of a natural imbalance in the consequences of overestimation and underestimation of the same magnitude (Bhattacharya *et al.*, 2002). Overestimation was more severe than underestimation in this particular problem, since an underestimation of the house, means that the loss incurred is equal to the amount of the underestimate. If, however, the house is overestimated the owner might appeal and reverse the decision. In this case, the loss will be two-fold, both the amount of the over assessment and the cost involved in running the case and the appeal procedures to reverse the decision made. The LINEX loss function is well used in the field of Bayesian analyses, amongst them are: Zellner (1986); Parsian (1990b); Kuo and Dey (1990); Basu and Ebrahimi (1991); Parsian and Farsipour (1993); Pandy *et al.* (1996); Zou (1997); Mostert *et al.* (1998); Mostert *et al.* (1999); Chaturvedi, *et al.* (2000); Bhattacharya *et al.* (2002); Farsipour and Asgharzadeh (2002); Jaheen (2003); Singh *et al.* (2004) and many others.

The Bayesian estimator relative to the LINEX loss function defined in terms of the type I estimation error is obtained by minimising the expected loss given by

$$\begin{aligned} & E_{post}\left[\exp[a(\widehat{\theta}_{L1}-\theta)]-a(\widehat{\theta}_{L1}-\theta)-1\right] \\ = & \exp[a\widehat{\theta}_{L1}]E_{post}[\exp[-a\theta]]-a\widehat{\theta}_{L1}+aE_{post}[\theta]-1. \end{aligned} \tag{2.29}$$

Differentiating the expected loss (2.29) and solving for $\widehat{\theta}_{L1}$ as

$$\frac{d}{d\widehat{\theta}_{L1}}\left(\exp[a\widehat{\theta}_{L1}]E_{post}[\exp[-a\theta]]-a\widehat{\theta}_{L1}+aE_{post}[\theta]-1\right)=0$$

to obtain

$$\widehat{\theta}_{L1}=-\frac{1}{a}\log[E_{post}[\exp[-a\theta]]] \tag{2.30}$$

provided that $E_{post}[\exp[-a\theta]]$ exists and is finite.

Considering the type II estimation error of the expected posterior loss is given by

$$\begin{aligned} E_{post}\left[L_L\left(\frac{\theta}{\widehat{\theta}_{L2}}-1\right)\right] &= E_{post}\left[\exp\left[a\left(\frac{\theta}{\widehat{\theta}_{L2}}-1\right)\right]-a\left(\frac{\theta}{\widehat{\theta}_{L2}}-1\right)-1\right] \\ &= \exp[-a]E_{post}\left[\exp\left[\frac{a\theta}{\widehat{\theta}_{L2}}\right]\right]-\frac{a}{\widehat{\theta}_{L2}}E_{post}[\theta]+a-1. \end{aligned} \tag{2.31}$$

Differentiating (2.31) and equating to zero yields the Bayesian estimator. Hence, the Bayesian estimator under the LINEX loss function is obtained from

$$E_{post}\left[\theta\exp\left[\frac{a\theta}{\widehat{\theta}_{L2}}\right]\right]=\exp[a]E_{post}[\theta]. \tag{2.32}$$

The various forms of Bayesian estimators will be used in the subsequent chapters. Some important properties of the LINEX loss function are:

- The minimum of (2.28) is zero and is obtained when $\Delta = 0$.
- The second derivative of (2.28) is $a^2\exp[a\Delta] > 0$, hence, it is a strictly convex function, therefore, the Bayesian estimator obtained under (2.28) is unique.
- Using the Taylor series expansion about zero (also called Maclaurin's expansion) and for small $|a|$, the LINEX loss function reduces to the squared error loss function of the form

$$L_L(\Delta)\approx\frac{1}{2}a^2\Delta^2.$$

- The LINEX loss function is flexible, since the parameter can be chosen in such a way to provide a variety of asymmetric effects, which best explain a given scenario under consideration. The choice of the parameter a is important and if there is no idea what value of a

to choose, an alternative approach can be followed to construct actually a weighted LINEX loss function, with different values of a (Mostert *et al.*, 1999).

- If the type I estimation error is large positive i.e. highly overestimated, the exponential term of (2.28) dominates the linear term and hence (2.28) increases rapidly. The reverse effect is obtained for a negative estimation error.

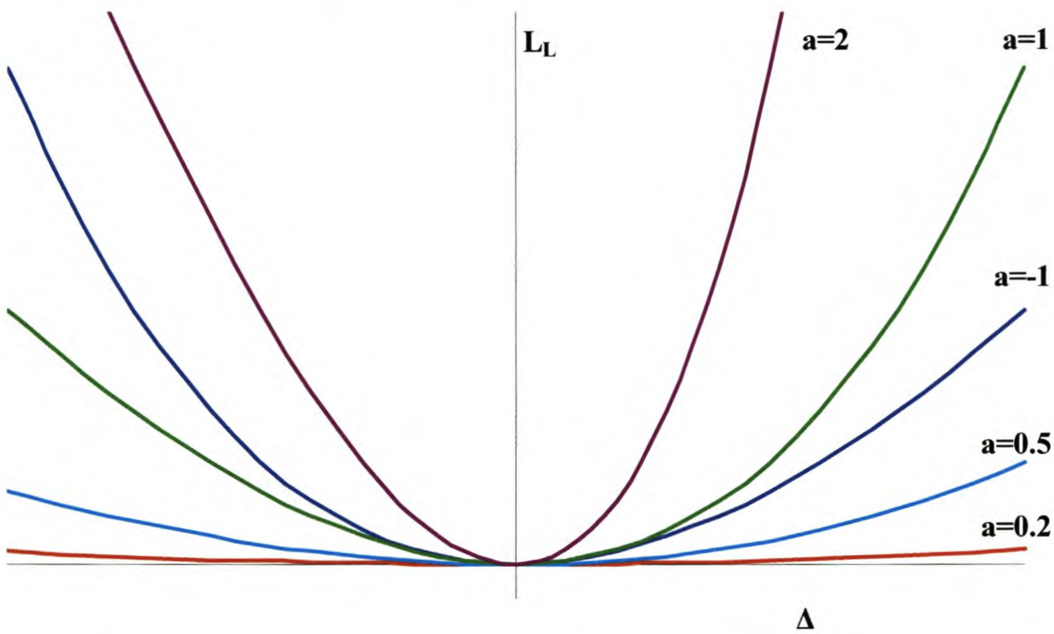

Figure 2: The LINEX loss function for different values of a

Figure 2 depicts the behaviour of the LINEX loss function for different values of a. For instance, when $a = 2$, the value of the LINEX loss function increases faster when $\Delta > 0$ than when $\Delta < 0$, and the opposite is true when a is negative.

Remark 2.5 *The LINEX loss function preserves the same property of resembling the squared error loss when $|a|$ is small even when using the type II estimation error. Parsian and Farsipour (1993) recommended that under the LINEX loss function, the type II estimation error is better in the estimation of a scale parameter, while for a location parameter, the type I estimation error is more suitable.*

2.6 RISK MEASURES

The quantification of uncertainty associated with risk estimates is an important part in the area of risk assessment (Nayak and Kundu, 2001). Risk analysis has been used as a tool of measuring the efficiency and productivity in many problems and is considered as an important safety management strategy (Aven and Pörn, 1998). Risk measurement has been applied as one of the most important issues in financial and insurance industries (Siu and Yang, 1999). Since risk is applied in various fields referring to many different things, be it with human beings or assets and financial interests or in decision-making, the interpretation also varies accordingly. Risk can be defined as an exposure to uncertainty (Holton, 1997), but since many different people might have different views of explaining uncertainty, the concept of risk might also vary. In general, there are two schools of thoughts in risk analysis, namely, the classical and the Bayesian approach (Aven and Kvalóy, 2002). The concept of risk measures is required in order to assess the goodness or performance of the estimators which are used in decisions.

In many problems of risk assessment, analysts employ the concept of subjectively choosing priors to incorporate some prior information, yet without exclusively adopting the Bayesian paradigm (Aven and Kvalóy, 2002). Even though many engineers, physical scientists or analysts are trained in the classical approach to risk analysis, in practice, they often use a mixture of both classical and the Bayesian approaches (Aven and Pörn, 1998). It has already been mentioned that the idea of Bayesian inference is to obtain an optimal estimator and therefore it is important to measure the risk incurred under these optimal estimators.

In this section the three types of risk measures namely, the risk function (also called frequentist risk), the posterior risk (also called expected posterior loss) and the integrated risk are defined and explained. Furthermore, the criteria for estimation in applied decision theory will be highlighted.

2.6.1 Risk function

In most problems of inference, the parameter for which an inference is sought, is unknown, hence, the need for inference. It is often difficult and even impossible to uniformly minimise the loss function in terms of $\widehat{\theta}$ for θ (Robert, 2001). The term uniformly minimising refers to the exercise of obtaining the minimum risk over all allowable ranges of the decision space, which may be the same as the parameter space. In such situations, an important aspect of the estimation is comparing the losses of the estimators made. The frequentist approach to risk analysis is therefore a crucial comparison criterion. The risk function (or sometimes refer to only as risk) is obtained as the expectation of the loss function with respect to the distribution of the statistic, i.e. the expectation is over the data space, $\mathcal{X}$.

Suppose that the distribution of the data conditioned on the parameter θ, is given by the pdf $f(x|\theta)$ and $\widehat{\theta}$ is any estimator, then the risk function is

$$R(\widehat{\theta},\theta) = E_{\mathcal{X}}\left[L(\widehat{\theta},\theta)\right] = \int_{\mathcal{X}} L(\widehat{\theta},\theta)f(x|\theta)dx. \tag{2.33}$$

Once the expectation is obtained, the analyst seeks a decision that minimises (2.33).

2.6.2 Posterior risk

In contrast to the risk function where the expectation is taken with respect to the likelihood, the posterior risk is obtained as the average of the loss function with respect to the posterior distribution of the parameter(s). For a given estimator $\widehat{\theta}$, a given posterior distribution $p(\theta|data)$ and any loss function $L(\widehat{\theta},\theta)$, the posterior expected loss is

$$R_{1,2}^{S,L}(\widehat{\theta}) = E_{post}\left[L(\widehat{\theta},\theta)\right] = \int_{\Theta} L(\widehat{\theta},\theta)p(\theta|data)d\theta. \tag{2.34}$$

Clearly, (2.34) is a function of the observed data and measures the average loss incurred over the whole parameter space, when an estimator is obtained. This means that (2.34) gives the average loss accrued over the parameter space Θ for every estimator $\widehat{\theta}$. Therefore, if the decision space is an infinitely large set, obtaining the posterior risk for all possible decisions is not simple, but the Bayesian always seeks the estimator for which (2.34) is minimised.

2.6.3 Integrated risk

The integrated risk, which is obtained by averaging the risk function over the domain of θ, with respect to the prior distribution, is another risk measure to be considered. Using (2.33) and the prior distribution of the parameter, $\pi(\theta)$, the integrated risk is given by

$$r(\widehat{\theta}) = \int_{\Theta} R(\widehat{\theta}, \theta)\pi(\theta)d\theta. \tag{2.35}$$

This means that the integrated risk is obtained by first obtaining the risk function in terms of the Bayesian estimator and then averaging with respect to the prior distribution.

An alternative way of expressing (2.35) is

$$r(\widehat{\theta}) = \int_{\Theta}\int_{\mathcal{X}} L(\widehat{\theta}, \theta)f(x|\theta)dx\ \pi(\theta)d\theta. \tag{2.36}$$

The value of the risk evaluated in (2.35), when the Bayesian estimator is used for $\widehat{\theta}$, is called the *Bayes risk* (Robert, 2001). This Bayes risk evaluated at any Bayesian estimator $\widehat{\theta}$ will henceforth be denoted as $r_B(\widehat{\theta})$.

2.6.4 Criteria for estimation in applied Bayesian decision theory

It has been already mentioned that in the Bayesian paradigm of decision theory, an estimator of the parameter is obtained based on the prior knowledge and loss function under consideration. One way of obtaining an estimator of the parameter is by minimising the expected loss with respect to the assumed distribution (prior to data collection), i.e. the estimator that minimises the risk function. This minimum value of the risk function is also called the *Expected Value of Perfect Information* (*EVPI*). On the other hand, the optimal estimator, called the Bayesian estimator minimises the posterior risk, i.e. after data collection. For this reason, it is more convenient to take the average with respect to the marginal distribution of the data to obtain the Bayes risk. The Bayes risk depends on the sample size and therefore helps to obtain the average value of a certain sample size, n. This means that one can make a pre-posterior analysis to obtain the EVPI and if the Bayes risk is obtained then the average value of a sample size n can be deduced by

$$EVPI - r_B(\widehat{\theta}). \tag{2.37}$$

The expression in (2.37) is called the *Expected Value of Sample Information* (*EVSI*), since it portrays the value of sampling. The EVSI also depicts the maximum value one should pay, when choosing a sample of size n in order to acquire as much information as possible, thereby reducing uncertainty.

In sampling, there is a cost attributed to sampling, since data collection in any procedure demands time or financial expenses. If the cost of sampling of the random variables, $X_1, X_2, ..., X_n$ is given by $C(X_1, X_2, ..., X_n)$, then the difference of the expected cost of sampling and EVSI is

$$ENGS(n) = E[C(X_1, X_2, ..., X_n)] - EVSI$$

where $ENGS(n)$ is the *Expected Net Gain in Sampling* a sample of size n. If $ENGS(n)$ is positive, then it is worth taking a sample. $ENGS(n)$ is usually a concave function whose maximum is attained at n_0, which is the optimal sample size, at which point one would gain the most out of sampling (Raiffa and Schlaifer, 2000).

In posterior analysis, an alternative way of measurement criterion, is the *highest posterior density* interval with credibility $(1 - \alpha)$, $0 \leq \alpha \leq 1$, denoted by $I^{hpd}_{(1-\alpha)}$. This gives the interval estimation of the parameter, where $Pr[\theta \in I^{hpd}_{(1-\alpha)}] = (1 - \alpha)$ and the probability of any point inside $I^{hpd}_{(1-\alpha)}$ is larger than any point outside it. Most often, however, point estimators are used, instead of highest posterior densities.

The aforementioned terms play a crucial role in measuring the cost of estimation, thereby rendering an idea whether or not it is appropriate to sample. Moreover, the optimal sample size yielding a minimum cost, can subsequently be obtained.

2.7 ADMISSIBILITY AND INADMISSIBILITY

In the preceding section the different risk measures were discussed. It was pointed out that these risk measures are helpful tools for quantifying the performance of the estimators and comparison for acceptability. This notion of acceptability of estimators lies in the area of admissibility and inadmissibility, which will be explained in this section.

In many decision problems involving risk measures, the analyst might be encountered with a situation of choosing decisions/estimators from crossing-over alternatives. Suppose for instance, there are two estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ and two values of the parameter θ_1 and θ_2. If the risk function is the tool used for comparison and $R(\widehat{\theta}_2, \theta_1) > R(\widehat{\theta}_1, \theta_1)$, while $R(\widehat{\theta}_2, \theta_2) < R(\widehat{\theta}_1, \theta_2)$, then the scenario renders crossing-over risk functions. In situations like this, the estimator $\widehat{\theta}_1$

is superior when the value of the parameter θ is equal to θ_1, while $\widehat{\theta}_2$ is superior when θ is equal to θ_2. This means that none of the two estimators is uniformly superior. In general, if $R(\widehat{\theta}_2, \theta) < R(\widehat{\theta}_1, \theta)$ for all θ provided strict inequality holds for at least one value of θ, it is said that $\widehat{\theta}_2$ is superior. An estimator, $\widehat{\theta}$, is said to be *admissible,* if there exists no estimator that is superior than $\widehat{\theta}$ (Kotz and Johnson, 1982). This means that a certain procedure (be it a point estimator, an interval of the estimator, a sample design, etc.) is said to be an admissible procedure if and only if there does not exist another procedure that uniformly performs at least as well as the procedure under consideration. The term *uniformly* is important and applies to the whole decision space $\mathcal{D}$, which in this case, is equal to Θ. Equivalently, this means that an estimator $\widehat{\theta}_1$ is said to be *inadmissible*, if there exists another estimator $\widehat{\theta}_2$, such that $R(\widehat{\theta}_1, \theta) \geq R(\widehat{\theta}_2, \theta)$ for all θ, where strict inequality holds for at least one θ, i.e. $\widehat{\theta}_1$ is inadmissible if $\widehat{\theta}_2$ is superior to it (Kotz and Johnson, 1983). An estimator which is not inadmissible, is admissible and the latter is acceptable in decision-making.

The concept of admissibility is applied not only to compare the performance of estimators, but also to compare preferences among possible priors as well as loss functions (Rice, 1995). It is also applied for various fields of statistics carrying correspondingly suitable interpretations for instance, the interpretation of admissibility in point estimation is not the same as that of admissibility in sample design. Nonetheless, in all procedures where the term admissibility is applied, it carries a similar criterion in such a way that the definition stated above is preserved within that procedure. Therefore, various classes of admissibility might be employed depending on the type of statistical inference. If the interest is to compare only point estimators, the admissibility of point estimators is applied. If the statistician is interested to collectively estimate and compare confidence intervals of estimators, then, there is a corresponding interpretation of the concept of admissibility. Similarly, there are admissibility of confidence sets, hypothesis tests, sampling designs, etc. (Kotz and Johnson, 1982).

2.7.1 Background on admissibility of estimators

In this section a brief background is given on the admissibility of estimators. In most of these references, the aforementioned risk measured were used for comparison purposes.

Farsipour and Asgharzadeh (2002) provided the inadmissibility criterion of the common mean of two normally distributed populations using the linear combination given by $\alpha_1\overline{X}+\alpha_2\overline{Y}+\alpha_3$, where α_1, α_2 and α_3 are known scalars. Both the squared error loss and LINEX loss functions were used in the study. It has been pointed out that the common mean of the two independent univariate normal populations is applied in the application of experimental design, clinical trials and many others. The Best Linear Risk Unbiased Equivariant Estimator (BLRUE) were proved to be admissible when both the aforementioned loss functions were used. The inadmissibility conditions of this linear combination depends on the magnitude and sign of the coefficients of the two sample means.

Zellner (1986) proved that the usual sample mean, $\overline{X}$, is inadmissible when estimating the normal mean with known variance under the LINEX loss function, since it is dominated by $\overline{X}-\frac{\sigma^2 a}{2n}$, where σ^2, a and n are the variance, LINEX shape parameter and sample size, respectively. Parsian (1990b) proved that $\overline{X} - \frac{\sigma^2 a}{2n}$ is the unique minimax and hence an admissible estimator. Rojo (1987) also studied the admissibility and inadmissibility of the linear estimator given by $c\overline{X} + d$, where c and d are known constants. Rojo generalised the result of Zellner using the natural conjugate prior, the normal distribution. Later, Bolfarine (1989) considered the estimation problems of the finite population total under the LINEX loss function and presented the Bayesian estimators of the parameter and discussed the admissibility of some of these estimators. The admissibility of the Pitman estimator under squared error loss for the variance of normal distribution with known mean was proved by Hodges and Lehmann (1951) and Girshick and Savage (1951). The same Pitman estimator with unknown mean was proved to be inadmissible by Stein (1964). Kuo and Dey (1990) studied an estimation of the mean of the Poisson distribution under the LINEX loss function. Necessary and sufficient conditions of inadmissibility were provided and the mean of the Poisson distribution was proved to be inadmissible under the LINEX loss function. Basu and Ebrahimi (1991) studied the admissibility of the Bayesian estimators obtained under the squared error ($\widehat{\theta}_S$) and the Bayesian estimator obtained under LINEX loss functions ($\widehat{\theta}_L$). Comparing the risk functions of the these two Bayesian estimators under the LINEX loss function, it was proved that risk function of $\widehat{\theta}_L$ uniformly dominates the risk function of $\widehat{\theta}_S$. In the case of the squared error loss, however, none of the two risk functions dominates the other uniformly.

2.7.2 Types of admissibility

Special types of admissibility exist and are discussed in this section. Depending on whether a condition is more or less stringent or binding, an admissible estimator can be called *strongly* or *weakly* admissible, whereby the former being the more stringent case. Another special type of admissibility is the so called $\varepsilon-$*admissibility*. This measures the extent of how much any decision rule or procedure (an estimator for example) is short of being admissible. Suppose decision about a point estimator is made under the risk function, an estimator $\widehat{\theta}_1$ is said to be $\varepsilon-$admissible if there is no other estimator $\widehat{\theta}_2$ such that $R(\theta, \widehat{\theta}_2) < R(\theta, \widehat{\theta}_1) - \varepsilon$, for all $\theta \in \Theta$, where $\varepsilon > 0$. At times it might be of importance to find an $\varepsilon-$admissible estimator in case admissibility conditions are not easily obtainable.

Uniform admissibility is another type of admissibility which is particularly applied to survey sampling theory. Depending on the constraints underlying sampling theory, i.e., cost and time, a sampling strategy is admissible in a given boundary of time and cost in such a way that the whole population be assessed with probability 1. This means that there should not be any other strategy that would outperform in satisfying the constraints of sample size, time or cost and should not exceed a certain threshold or limit. In some fields, there is a situation where joint admissibility is applied. Suppose for instance an estimator $\widehat{\theta}_1$ is obtained under a certain sampling design say, p_1, then the joint admissibility of $(\widehat{\theta}_1, p_1)$ is uniformly admissible if $\widehat{\theta}_1$ is an admissible estimator as well as $p_1$be admissible within the class of sampling designs (Kotz and Johnson,1982).

2.7.3 Sufficient conditions for admissibility

Certain conditions must be met in order for an estimator, a prior or a loss function to be admissible. The same are true for the conditions of inadmissibility. There are proved propositions that are stated to be sufficient for the admissibility of estimators. Some of these propositions are listed here without proof and for further details see Robert (2001).

- If a unique minimax estimator exists, then it will be admissible. It is important to explain the concept of minimaxity, a term predominantly used in game theory. Suppose for convenience purposes, the risk function is used as a performance measure. The idea of minimax rule is first to maximise the risk function with respect to the parameter and then for the maximised risk, find the decision which yields the minimum risk (Morris, 1994). This minimax rule has a weakness in the sense that it is a very conservative procedure that places all its emphasis on guarding against the worst possible cases.

- If the prior distribution of the parameter is strictly positive on the parameter space with a finite posterior risk and the risk function is continuous for every estimator of θ, then the Bayesian estimator is admissible.
- If the Bayesian estimator associated with the prior is unique, then it is admissible.
- If the Bayesian estimator associated with a (proper or improper) prior is such that the posterior risk is finite, then the estimator is admissible.

2.8 CONCLUSION

A decision-making process requires selecting a cost effective or an optimal action amongst alternatives. In the Bayesian paradigm in particular, the decision process incorporates the prior knowledge about the parameter(s) of interest, with the information obtained after collecting data. At the heart of this is Bayes' theorem, that plays an important role in consolidating the information from the two sources, updating it in the form of the posterior distribution. Some commonly used reliability models were discussed under the Weibull class of distributions. Two types of priors, the non-informative and conjugate priors were considered. The optimal estimators are derived by using loss functions, which are used to quantify the error incurred after obtaining an estimator of an unknown parameter. The performance of these estimators are measured with risk measures to determine the admissibility or inadmissibility thereof. An important point is the criterion of estimation which highlights whether sampling is appropriate or not. In the case where sampling is plausible, a far more appealing point is the fact that optimal sample size can also be obtained.

It was clear that when using risk measures in many practical problems, it is difficult to obtain the closed-form expressions of the of posterior moments. The main reason is due to censoring or indefinite multi-dimensional regions for the integrals. This problem is often encountered in finding the integrated risk, since it requires the integration over both regions of the parameter and the data. To this end, various numerical techniques of integration are employed by many Bayesian analysts and other researchers. Discussion of some of these techniques will be given in chapter 3.

CHAPTER 3

BAYESIAN COMPUTATION

3.1 INTRODUCTION

In the Bayesian paradigm, once the analyst chooses the loss function and prior distribution of the parameter(s) of interest, inferences about the parameter(s) or prediction of new observations are of importance. To this end, a user of the Bayesian paradigm in statistical inference should be able to compute posterior moments, marginal posterior densities, or other characteristics of the posterior distribution. Given the likelihood function and the prior distribution of the parameter(s), a Bayesian analyst is usually interested in computing a posterior moment of the form

$$E_{post}[g(\theta|data)] = \frac{1}{K}\int_{\Theta} g(\theta|data)\pi(\theta)\ell(data|\theta)d\theta, \tag{3.1}$$

where

$$K = \int_{\Theta} \pi(\theta)\ell(data|\theta)d\theta,$$

and $\pi(\theta)$ is the prior distribution of the parameter(s), $g(\theta|data)$ is any function of interest of the parameter(s) conditional on the data and $\ell(data|\theta)$ is the likelihood function. In the case of multi-parameter posterior analyses, another point of interest for the Bayesian could be to obtain the marginal posterior densities of the individual parameters. Suppose for instance, $\theta \in \mathcal{R}^m$ and can be partitioned as $\theta = (\theta_1, \theta_2)$, where $\theta_1 \in \mathcal{R}^j$ and $\theta_2 \in \mathcal{R}^{m-j}$, the marginal posterior density for θ_1 is given by

$$p(\theta_1|data) = \int_{\Theta} p(\theta_1, \theta_2|data)d\theta_2, \tag{3.2}$$

where $p(\theta_1, \theta_2|data)$ is the joint posterior density of θ (θ_1 and θ_2 may be scalars or vectors). The starting point for all subsequent inferences is often the joint posterior density.

As mentioned in chapter 2, quite often, performing the integration of K in (3.1), i.e. the normalising constant of the posterior distribution, has computational difficulties (Chib, 1995). Consequently, any analysis involving the posterior distribution, such as posterior moments, posterior quantiles, Bayesian point estimators, marginal posterior densities, or probability intervals, etc., may not be analytically tractable. This has contributed to the use of some classical loss functions, which lead to explicit solutions of Bayesian estimators. As a consequence, to avoid computational complexity or difficulty, inappropriate loss functions or prior distributions are chosen, since they lead to mathematically tractable solutions, yet without reflecting the reality (Robert, 2001). Another resort for the Bayesian practitioner, may be to turn back to the classical approach, leaving the beauty of the Bayesian approach behind. However, these constraints are no longer justified, as a result of the diverse computational tools, which have sprouted out over the years.

This chapter, therefore, discusses some of the approximation techniques needed for the apparent computational difficulties. In section 3.2, techniques to obtain posterior moments are discussed and in section 3.3 techniques to obtain marginal posterior densities are discussed. Section 3.4 concludes the chapter.

3.2 POSTERIOR MOMENTS

The implementation of the Bayesian paradigm requires the computation of the posterior moments. In general, a user of the Bayesian method in practice needs to evaluate various characteristics of the posterior and predictive distributions, especially in terms of means and variances. However, unless the conjugate priors are used, these tasks are most often not easily performed (Tierney and Kadane, 1986). The main contributing factor is that exact analytical solutions for the integral in (3.1) is impossible. Appropriate approximation techniques are therefore needed and some of them are discussed in the subsequent sections.

3.2.1 Tierney and Kadane approximation

Tierney and Kadane (1986) introduced a method of approximating the ratio of two integrals. Their approximation is based on the Laplace's method, which is generally used to solve integrals of the form

$$I = \int_{\Theta} f(\theta) \exp[-nh(\theta)] d\theta, \tag{3.3}$$

where $-nh(\theta)$ is a function having a maximum at some point $\widetilde{\theta}$ (in statistical applications, n is taken as the sample size). Therefore, applying Taylor series expansion for the functions h and f about $\widetilde{\theta}$, (3.3) yields

$$\begin{aligned} I \quad \approx \quad &= \exp[-nh(\widetilde{\theta})] \int_{\Theta} \left[f(\widetilde{\theta}) + f'(\widetilde{\theta}) \left(\theta - \widetilde{\theta} \right) + \frac{f''(\widetilde{\theta})}{2!} (\theta - \widetilde{\theta})^2 \right] \\ &\times \exp\left[-\frac{(\theta - \widetilde{\theta})^2}{2[nh''(\widetilde{\theta})]^{-1}} \right] d\theta. \end{aligned} \tag{3.4}$$

It can be noted that the last term of the integrand in (3.4) is proportional to a normal distribution with mean $\widetilde{\theta}$ and variance $[nh''(\widetilde{\theta})]^{-1}$. It can be easily seen that (3.4) is a *first order* Laplace's approximation and reduces to

$$I \approx \left(\frac{2\pi}{nh''(\widetilde{\theta})} \right)^{\frac{1}{2}} f(\widetilde{\theta}) \exp[-nh(\widetilde{\theta})] \left[1 + \frac{f''(\widetilde{\theta})}{2f(\widetilde{\theta})} var(\theta) \right], \tag{3.5}$$

where $var(\theta) = [nh''(\widetilde{\theta})]^{-1} = O(\frac{1}{n})$ (order of $\frac{1}{n}$ which entails the computational complexity or accuracy).

Further simplification of (3.5) yields

$$I = \exp[-nh(\widetilde{\theta}) f(\widetilde{\theta}) \left(\frac{2\pi}{nh''(\widetilde{\theta})} \right)^{\frac{1}{2}} \left[1 + O\left(\tfrac{1}{n}\right) \right].$$

An analogous result to (3.5) in an $m-$dimensional space is given by

$$I \approx (2\pi)^{\frac{m}{2}} \left(\det \left[nH_h^{-1} \right] \right)^{-\frac{1}{2}} f(\widetilde{\theta}) \exp[-nh(\widetilde{\theta})],$$

where H_h^{-1} is the inverse of the Hessian matrix of h at $\widetilde{\theta}$. The Hessian matrix of h is a matrix whose ij^{th} element is given by $\left(\frac{\partial^2 h}{\partial \theta_i \partial \theta_j} \right)$.

Tierney and Kadane (1986) devised their method of approximation for any positive function $g(\theta|data)$ of a real or vector-valued parameter based on this Laplace's method. It has been pointed out that the computational complexity of the method is minimal when compared to that of others, such as Lindley (1980), Mosteller and Wallace (1964). The method of Tierney and Kadane requires only the computation of the first and second derivatives of the posterior and is feasible for any problem where maximum likelihood estimators can be derived. In order to obtain the approximate posterior mean and variance of g, the logarithm is taken for both the numerator and denominator of (3.1). In order to apply the Laplace's approximation, (3.1) can be rewritten as

$$E_{post}[g(\theta|data)] = \frac{\int\limits_{\Theta} f_1(\theta)\exp[-nh(\theta)]d\theta}{\int\limits_{\Theta} f_2(\theta)\exp[-nh(\theta)]d\theta}, \tag{3.6}$$

where $f_1 = g$, $f_2 = 1$ and, $\exp[-nh(\theta)] = \pi(\theta)\ell(data|\theta)$. Hence, the modal approximation of (3.6) is given as

$$E_{post}[g(\theta|data)] = g(\overline{\theta})\left[1 + O\left(\tfrac{1}{n}\right)\right] \approx g(\overline{\theta}),$$

where $\overline{\theta}$ is the mode of the posterior distribution. Setting $f_1 = f_2 = 1$ in (3.6) yields an alternative *second order* approximation, but in this case the log-posterior is the ratio of

$$-nh_{\mathbf{n}}(\theta) = \log[g(\theta|data)] - nh_{\mathbf{d}}$$

and

$$-nh_{\mathbf{d}}(\theta) = \log[\pi(\theta)] + \log[\ell(data|\theta)].$$

Substituting these two expressions in (3.6) results

$$E_{post}[g(\theta|data)] = \frac{\int\limits_{\Theta} \exp[-nh_{\mathbf{n}}]d\theta}{\int\limits_{\Theta} \exp[-nh_{\mathbf{d}}]d\theta}. \tag{3.7}$$

Suppose $\overline{\theta}_{\mathbf{n}}$ and $\overline{\theta}_{\mathbf{d}}$ are the posterior modes of $h_{\mathbf{n}}$ and $h_{\mathbf{d}}$ respectively, $\sigma^2_{\mathbf{n}} = \left[h''_{\mathbf{n}}\left(\overline{\theta}_{\mathbf{n}}\right)\right]^{-1}$ and $\sigma^2_{\mathbf{d}} = \left[h''_{\mathbf{d}}\left(\overline{\theta}_{\mathbf{d}}\right)\right]^{-1}$. Applying Laplace's method of approximation to (3.7), yields

$$\begin{aligned} E_{post}[g(\theta|data)] &= \frac{\int\limits_{\Theta} \exp\left[nh_{\mathbf{n}}\left(\overline{\theta}_{\mathbf{n}}\right) - \frac{n(\theta-\overline{\theta}_{\mathbf{n}})^2}{2\sigma^2_{\mathbf{n}}}\right]d\theta}{\int\limits_{\Theta} \exp\left[nh_{\mathbf{d}}\left(\overline{\theta}_{\mathbf{d}}\right) - \frac{n(\theta-\overline{\theta}_{\mathbf{d}})^2}{2\sigma^2_{\mathbf{d}}}\right]d\theta} \\ &= \frac{\sigma_{\mathbf{n}}}{\sigma_{\mathbf{d}}}\exp\left[n\left[h_{\mathbf{n}}\left(\overline{\theta}_{\mathbf{n}}\right) - h_{\mathbf{d}}\left(\overline{\theta}_{\mathbf{d}}\right)\right]\right]. \end{aligned} \tag{3.8}$$

The result in (3.8) can be rewritten as

$$E_{post}[g(\theta|data)] = g(\overline{\theta})\left[1 + O\left(\tfrac{1}{n^2}\right)\right].$$

Analogous to the result in (3.8), the multiparameter equivalent is

$$\left(\frac{\det\left[H_{h_{\mathbf{n}}}^{-1}\right]}{\det\left[H_{h_{\mathbf{d}}}^{-1}\right]}\right)^{\frac{1}{2}} \exp\left[n\left[h_{\mathbf{n}}\left(\overline{\theta}_{\mathbf{n}}\right) - h_{\mathbf{d}}\left(\overline{\theta}_{\mathbf{d}}\right)\right]\right], \tag{3.9}$$

where $\overline{\theta}_{\mathbf{n}}$ and $\overline{\theta}_{\mathbf{d}}$ in this case are values that maximise $h_{\mathbf{n}}$ and $h_{\mathbf{d}}$ respectively and $H_{h_{\mathbf{n}}}^{-1}$ and $H_{h_{\mathbf{d}}}^{-1}$ are the negative of the inverse of the Hessian matrices of $h_{\mathbf{n}}\left(\overline{\theta}_{\mathbf{n}}\right)$ and $h_{\mathbf{d}}(\overline{\theta}_{\mathbf{d}})$ respectively. Tierney and Kadane also provided the computational complexity and accuracy of the Laplace's approximation. Both the first and second order approximations require only the first and second derivatives, where in the latter case, the differentiation is done for the log-posterior. In situations when n is considered as the sample size, the accuracy of Laplace's method increases with increasing n.

Laplace's method of approximation can also be used for a function, $g(\theta)$, that is not strictly positive. The trick in such situations is to add a large positive constant, c , thus apply Laplace's method and then subtract c from the final result. As an alternative, Laplace's method can be applied to the moment generating functions, whose integrand is always positive, where differentiation is applied to the result at the end (Tierney *et al.*, 1989).

In general, the Laplace's method provides a quick computational procedure since it does not require iterative algorithms. It needs only the first and second differentials. Another important aspect of the Laplace's method is the fact that it does not depend on random numbers, hence it gives deterministic results consistent for different analysts. It also reduces the computational complexity in robust studies.

3.2.2 Numerical methods

Many numerical methods of integration evolved over the past years. Some of these techniques have been used by Bayesian statisticians in posterior analyses. Due to the computational complexity, the use of numerical methods in approximation is restricted mainly to small number of parameters. Simpson's rule is one of the basic techniques of numerical integration. Simpson's method for one parameter is given by

$$\int_a^b f(\theta)d\theta = \frac{h}{3}(f(\theta_0)+4f(\theta_1)+2f(\theta_2)+...+2f(\theta_{2n-2})+4f(\theta_{2n-1})+2f(\theta_{2n}))+R_n, \quad (3.10)$$

where $h = \frac{b-a}{2n}$, the length of the sub-interval obtained by dividing the interval $[a, b]$ into $2n$ intervals of equal length and R_n is the approximation error, which is given by

$$R_n = -\frac{(b-a)^5 f^{(4)}(\xi)}{180(2n)^4}, \ a < \xi < b.$$

Simpson's rule depends on two conditions, firstly, the allowable range of the parameter must be finite and secondly, the function of interest must be differentiable. Most often, however, this might not be the case with problems involving Bayesian analysis. The *polynomial quadrature* is another important numerical approximation. This technique is mainly applied to approximate integrals of distributions resembling the normal distribution (Naylor and Smith, 1982). Given a pdf, $f(\theta)$, which is defined over the whole real line, the approximation is given by

$$\int_{-\infty}^{\infty} \exp\left[-\frac{\theta^2}{2}\right] f(\theta)d\theta \approx \sum_{i=1}^{n} u_i f(\theta_i), \quad (3.11)$$

where $u_i = \frac{2^{n-1}n!\sqrt{n}}{n^2[H_{n-1}(\theta_i)]^2}$ is the i^{th} coefficient for the Gauss-Hermite quadrature and θ_i is the i^{th} root of the polynomial equation of Hermite $H_n(\theta)$. If f is symmetrical around the origin, (3.11) is given as $\int_0^{\infty} \exp\left[-\frac{\theta^2}{2}\right] f(\theta)d\theta \approx \frac{1}{2}\sum_{i=1}^{n} u_i f(\theta_i)$. Hence, this latter case might be helpful to approximate the expectation of functions of the form $\exp\left[-\frac{\theta^2}{2}\right]$ defined only in the non-negative real line.

3.2.3 Monte Carlo procedures

The numerical methods, such as polynomial quadrature and others, are known to be inappropriate as the dimension of the parameter space increases. The rule of thumb is that numerical methods should not be applied for higher dimensions ($m \geq 5$), where m is the dimension of the parameter space. In such situations, alternative approximation techniques are needed.

The Monte Carlo procedure was introduced by Metropolis and Ulam (1949) and Von Neumann (1951). It has an advantage, especially when one is interested in obtaining the expectation of a function of interest with respect to a proper pdf. The key point in utilising the Monte Carlo procedure, is generating a sequence of independent random variates from a given pdf. The Monte Carlo sampling procedure involves the following steps for the expectation of a function $g(\theta|data)$ after guaranteeing the simulation of random variates.

Step 1: Generate $\theta^{(i)}$ for θ from the posterior distribution $p(\theta|data)$.

Step 2: Calculate $g(\theta^{(i)}|data)$.

Step 3: Repeat steps 1 and 2 for $i = 1, ..., k$ (k sufficiently large).

Step 4: The expectation of g is given as the limit of the arithmetic mean of step 2 as k tends to infinity, i.e.

$$E_{post}[g(\theta|data)] \approx \widehat{g}(\theta|data) = \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} g(\theta^{(i)}|data). \tag{3.12}$$

According to the law of large numbers, $\widehat{g}(\theta|data) \to E_{post}[g(\theta|data)]$ with probability 1, as $k \to \infty$ (Geweke, 1989). Moreover, once the random variates are simulated, any other characteristic of the function, such as the mode, the median, the variance, etc., can be easily obtained. Using (3.12), the standard error of the approximation is given by

$$\sqrt{\frac{\sum_{i=1}^{k}\left[(g\theta^{(i)}|data) - \widehat{g}(\theta|data)\right]^{\frac{1}{2}}}{k(k-1)}}. \tag{3.13}$$

It can be observed from (3.13) that the error is inversely proportional to the number of generated random variables, k. Hence, the fact that the accuracy of the approximation is entirely at the hand, it means the more the number of iterations, the better. Monte Carlo procedures are commonly used in the physical sciences and other fields for solving problems that are analytically impractical to tackle (Yang, 2002). The most important ingredient in applying Monte Carlo is the issue of a good randomness. The sequence of random variates generated must exhibit a uniform distribution in a certain interval. As a matter of fact, there are some algorithms, called *pseudo-random number generators,* which ensure this requirement of randomness. One example of these generators is the linear congruencial method (Lehmer, 1949).

3.3 MARGINAL POSTERIOR DENSITIES

In many situations, inference in the Bayesian paradigm involves the evaluation of the ratio of two integrals. Given the likelihood function and the prior distribution of the parameters, the principal for all subsequent inferences starts from the joint posterior density for the parameters of interest conditioned on the data (Bernardo, 1997). From section 2.4, it is clear that the denominator of the posterior distribution involves integration over the domain of θ. Moreover, there might arise a need to obtain the univariate or bivariate marginal posterior densities for components of θ, which require additional computations. Therefore, obtaining both the normalising constant of the posterior and the components of the marginal densities require successive integrations. If the joint distribution and the conditional marginal distributions are obtained, the unconditional marginal distributions can also obtained directly. However, these integrations might be onerous or in the worst case, the conditional marginal distributions might not be analytically obtainable, especially for a multi-dimensional parameter space. In such situations, efficient approximation techniques are indispensable. Some of these approximation techniques will thus be explained in the subsequent sections.

3.3.1 Tierney, Kass and Kadane approximation

In Bayesian analysis with multi-dimensional parameters, an important aspect of the posterior analysis is to obtain marginal posterior densities of the parameter(s) of interest. The method devised by Tierney and Kadane in section 3.2 can also be used to obtain these marginal posterior densities. The Laplace's approximation can also be used to obtain the marginal posterior density of partitioned parameter(s).

The joint posterior distribution in (3.2) can be represented as

$$p(\theta_1, \theta_2|data) = cf(\theta_1, \theta_2) \exp[-nh(\theta_1, \theta_2)],$$

where c is a constant. The application of Laplace's approximation involves the finding of posterior modes that maximise the product of the prior and the likelihood, as well as the negative of the inverse of the Hessian matrices evaluated at the mode. The usual Laplace's approximation can then be applied for (3.2) to give the respective approximate marginal posterior density of θ_1. A similar approach is applied for the marginal posterior density of θ_2. Likewise, if θ has more than two components, the marginal posterior densities of the individual parameters can be obtained by partitioning the parameter space into just two components (Tierney and Kadane, 1986; Tierney *et al.*, 1989).

3.3.2 Markov Chain Monte Carlo

In many problems within the Bayesian paradigm, the analyses of posterior moments, or marginal posterior densities or risks in many reliability studies, are mathematically intractable and complicated. Hence, there are situations where closed-forms of the marginal posterior densities are not obtainable. Therefore, techniques to alleviate the difficulty of obtaining posterior moments in stochastic processes are necessary. To this end, the Markov Chain Monte Carlo (MCMC) methods have been widely used for simulating these processes (Metropolis *et al.* 1953; Hastings, 1970). MCMC algorithms have broad applications in all areas of statistics with the main focus being in Bayesian applications (Geman and Geman, 1984; Gelfand and Smith, 1990). The advent of MCMC methods to simulate posterior moments or posterior distributions have, therefore, revolutionised the application of Bayesian statistics (Chib, 1995).

Unlike the simple Monte Carlo simulation, which directly generates independent data from the distribution, MCMC methods generate iteratively from the conditional distributions, where every subsequent draw depends on the previous generated data. Several Markov chain methods of sampling from the posterior distribution are available and some of the latest methods are discussed in Robert (2001). Some of these are discussed in the subsequent sections.

3.3.2.1 Gibbs sampler

The Gibbs sampler is one of the commonly used MCMC techniques of sampling. It is a computer intensive tool of iteratively generating random variates from a certain posterior distribution, without obtaining the density itself. Gibbs sampler is used to solve problems occurring in multi-dimensional numerical integration over awkwardly defined regions (Kuo and Smith, 1992). This method is often used in multi-parameter distributions, when their conditional counterparts are given and is a popular statistical technique both in applied and theoretical work

(Casella and George, 1992). The Gibbs sampler found its popularity with the pioneering paper of Geman and Geman (1984), who applied image-processing models in their study and later by Tanner and Wong (1987) and Gelfand and Smith (1990). The application of Gibbs sampler is common in Bayesian models when generating posterior distributions. On the other hand, non-Bayesians apply Gibbs sampling in the calculation of likelihoods and characteristics of the likelihood estimators.

To illustrate the Gibbs sampler, suppose the joint posterior density of the parameters $\theta_1, ..., \theta_m$ is given by $p(\theta_1, ..., \theta_m|data)$. Suppose one is interested in the marginal posterior density of any one of the parameters. More often than not, calculation of this integral is complicated and even sometimes analytically or numerically unobtainable. There are mainly two approaches of handling this one of which is a direct approximation, as discussed in section 3.3.1 and the other is to generate successive random variates from the marginal density of the parameter of interest. The latter alternative is what is called the Gibbs sampler. Considering the marginal posterior density of θ_i, $p(\theta_i|data)$, the Gibbs sampler effectively generates sequences of random variates (sample) $\theta_{i1}, \theta_{i2}, ..., \theta_{in}$ from $p(\theta_i|data)$. The efficiency of the Gibbs sampler depends on the rate of convergence. The usefulness of the Gibbs sampler increases greatly as the complexity of the problem increases, since in such situations analytical or numerical methods do not help. The Gibbs sampling goes through the following steps:

Step 1: Start with initial values (guesses) of the parameters of interest. Suppose a model has three unknown parameters, $(\theta_1, \theta_2, \theta_3)$, with their respective initial values denoted by $\theta_1^{(0)}$, $\theta_2^{(0)}$ and $\theta_3^{(0)}$.

Step 2: Given the posterior distribution, $p(\theta|data)$, generate random variates in sequence from the full conditional posterior distributions as follows:

$\theta_1^{(j+1)}$ from $p\left(\theta_1|data, \theta_2^{(j)}, \theta_3^{(j)}\right)$

$\theta_2^{(j+1)}$ from $p\left(\theta_2|data, \theta_1^{(j+1)}, \theta_3^{(j)}\right)$

$\theta_3^{(j+1)}$ from $p\left(\theta_3|data, \theta_1^{(j+1)}, \theta_2^{(j+1)}\right)$.

Step 3: Repeat step 2 for $j = 1, ..., k$, where k is sufficiently large. This means that the *Gibbs sequence* is obtained:

$$\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \theta_1^{(2)}, \theta_2^{(2)}, \theta_3^{(2)}, ..., \theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)} \tag{3.14}$$

Once this sequence is obtained, the first J triplets of $\theta_1^{(J)}$, $\theta_2^{(J)}$ and $\theta_3^{(J)}$ should be discarded to avoid dependence on the initial values.

Step 4: After obtaining the Gibbs sequence, a vital point is to check the chain for convergence and to ensure it is safe to terminate the process. This convergence depends on the number of iterations, i.e. as $k \to \infty$, the sequences generated in (3.14) will effectively reflect a sample from the joint distribution (Casella and George, 1992). It is therefore, very important to know what value of k gives a reasonable approximation of the sampling. An alternative way of employing the Gibbs sampler has been suggested by Gelfand and Smith (1990), which requires generating a large number of independent Gibbs sequences of length k and using the final realisation from each sequence.

3.3.2.2 Slice sampler

It has been illustrated that the Gibbs sampler is implemented in a hierarchical structure where one parameter is generated from the conditional distribution. This hierarchical algorithm of implementation might lead to the perception that Gibbs sampler would not be applied in unidimensional problems. However, the general implementation of the Gibbs sampler applies to any number of dimensions, which is actually the strength of the technique. The generality of the Gibbs sampler is exposed in the particular version called *slice sampler* (Wakefield *et al.*, 1991; Damien *et al.*, 1999). The foundation of the slice sampler stems from the following results. Any posterior distribution (any distribution for that matter) of the parameter(s) can be written as a product of k positive functions (not necessarily pdfs), i.e.,

$$p(\theta|data) = \prod_{i=1}^{k} \omega_i(\theta|data), \tag{3.15}$$

where $\omega_i^{'}s$ are positive functions. Equation (3.15) can be rewritten as

$$p(\theta|data) = \int_{\Theta} \prod_{i=1}^{k} \mathcal{I}_{0 \leqslant w_i \leqslant \omega_i(\theta|data)} dw_1 dw_2 ... dw_k \tag{3.16}$$

where $\mathcal{I}$ is the indicator function which attains 1 for $0 \leqslant w_i \leqslant \omega_i(\theta|data)$ and 0, otherwise. Therefore, to sample from $p(\theta|data)$, using the slice sampler needs the following steps.

Step 1: Write the posterior distribution in the form of (3.16).

Step 2: At iteration t, simulate

$$w_i^{(t+1)} \sim \mathcal{U}_{[0,\omega_i(\theta^{(t)}|data)]},\ i = 1, ..., k$$

and

$$\theta^{(t+1)} \sim \mathcal{U}_{A^{(t+1)}},\ A^{(t+1)} = \left\{\xi : \omega_i(\xi|data) \geqslant w_i^{(t+1)}, i = 1, 2, ..., k\right\}.$$

Step 3: Repeat step 2 for t infinitely large.

The $w_i^{'}s$ are auxiliary variables (also referred to as latent variables), which are chosen based on convenience. The main problem in utilising the slice sampler lies in choosing these auxiliary variables in order to construct the slice $A^{(t+1)}$. Once the construction for these slices are guaranteed, the slice sampler can then be applied for any distribution, be it unidimensional or multidimensional. Incorporating these auxiliary variables is mainly used in sampling from non-standard densities, which arise in the context of Bayesian analysis of non-conjugate and hierarchical models (Damien *et al.*, 1999). The Gibbs sampler has an extremely important application in many fields, since it can be used to generate random variates from a given distribution, possibly non-normalised ones. Even though there are other methods such as adaptive rejection sampling for log-concave densities, the method of introducing auxiliary variables is preferred for its simplicity and generality (Wakefield *et al.*, 1991).

3.4 CONCLUSION

Different approximation techniques have been used in many problems where analytical procedures are not appropriate. It should be noted that all these techniques have conditions under which they are suitable to apply. In general, the use of approximation techniques have simplified the work of many analysts and kindled the interest towards statistical inferences.

In general, the approximation techniques play an important role in the implementation of Bayesian methods for arbitrary forms of likelihood, prior specification and sample size. Due to the advent of these techniques, researchers got the liberty of not being confined in the analytical methods of solving problems concerning pdfs. Approximation techniques also avoid the use of simple and yet inappropriate loss functions or prior distributions for the parameter(s). The advantage of approximation techniques becomes vivid in situations when analytical computations are not helpful, hence, often with increasing dimensions of the parameter space. It should be noted, however, that some of the approximation techniques are restricted to fewer dimensions of the parameter space. Numerical techniques may lead to some approximation errors. Moreover, these techniques sometimes call for an enormous amount of computing time and can become very costly. In a lot of situations the error term depends on the number of variates generated, which is a big advantage. Therefore, reducing the approximation error remains entirely in the hand of the analyst.

CHAPTER 4

THE EXPONENTIAL MODEL

4.1 INTRODUCTION

The decision-making process within the Bayesian paradigm was discussed in chapter 2. It was clear that there are situations where closed-form expressions for elements in decision-making do not exist and alternative procedures have to be applied, which were discussed in chapter 3.

It was stated that the univariate exponential distribution is a well-known and a commonly used model in reliability theory. The literature is surging with papers regarding the exponential model, both in the classical and Bayesian perspectives. Some, amongst many of the papers relevant to the exponential model, are listed in this section. Enis and Geisser (1971) used the exponential model in a Bayesian approach to predict the probability, $Pr(Y < X)$ (given the available data), where X and Y are two future observations. Sinha and Guttman (1976) provided a Bayesian analysis of the reliability function for a non-truncated exponential distribution, using both an improper and the inverted gamma as prior distributions of the parameter. The Bayes risk of both the Bayesian and non-Bayesian estimators were calculated for comparisons. Basu and Ebrahimi (1991) carried out a Bayesian analysis for the exponential life testing model. Bayesian estimators were derived by using the non-informative priors of quasi-density and the inverted gamma pdf, which is a natural conjugate prior of the parameter of this model. Both the risk function and the Bayes risk were derived for comparison purposes, where the LINEX and the squared error loss functions were used. Basu and Ebrahimi (1992) also performed a Bayesian approach to the estimation of a parametric model under an asymmetric loss function as well as the reliability estimation of complex systems. Bhattacharya *et al.* (2002) assessed the comparative analysis of the performance of the Bayesian and classical point estimators of the parameter of the univariate exponential family relative to the LINEX loss function, using the Bayes risk and a fixed prior distribution. Trader (1985) provided a Bayesian inference for the left truncated exponential family of distributions, when its parameters have a truncated normal distribution. Shalaby (1993) performed a Bayesian analysis of the doubly truncated exponential model, whereby a comparison of the minimum posterior risk of the Bayesian estimator to the minimum risk of the Best Linear Unbiased Estimator (BLUE) was done. Many real world problems such as gunnery, bombing accuracy and life testing are said to exhibit a model of a doubly truncated exponential. The functional form for this doubly truncated distribution, the reliability function and the characteristic functions are given in Shalaby (1993).

The components related to the decision-making process within the Bayesian paradigm for the exponential model are discussed in this chapter. The remainder of the chapter is structured as follows. In section 4.2 the Jeffreys' prior is derived and the posterior distribution of the parameter of the exponential distribution is subsequently derived using the Jeffreys' prior. The Bayesian estimators relative to the squared error loss and the LINEX loss functions are also derived. The risk measures and the admissibility of the estimators are also discussed. Section 4.3 is devoted to the analogous derivation of the posterior distribution and the Bayesian estimators when using the natural conjugate distribution as the prior distribution for the parameter. Section 4.4 deals with the derivation of the Jeffreys' prior as well as the subsequent posterior distribution and the Bayesian estimators for the survival function of the exponential model. Section 4.5 concludes the chapter.

4.2 POSTERIOR ANALYSIS USING THE NON-INFORMATIVE PRIOR

In this section, the posterior analysis using the non-informative prior, mainly the Jeffreys' prior, is considered. Firstly, the Jeffreys' prior for the parameter of the exponential model and consequently the posterior distribution are derived.

Applying (2.14) to the exponential model (2.7) yields,

$$I_{11}(\theta) = -E_\theta\left[\frac{\partial^2\left\{\log\left[f\left(x|\theta\right)\right]\right\}}{\partial\theta^2}\right] = \frac{1}{\theta^2}.$$

Hence, using (2.15), the Jeffreys' prior for the parameter of the exponential model is given by

$$\pi(\theta) \propto \frac{1}{\theta}. \tag{4.1}$$

The Jeffreys' prior is of the same form as the mean of the exponential distribution, i.e., $E[x] = \frac{1}{\theta}$.

Given the Jeffreys' prior, the posterior distribution of the parameter θ is obtained using (2.10) and (4.1) in (2.20) to obtain

$$p(\theta|data) \propto \theta^{r-1}\exp[-T\theta], \tag{4.2}$$

with normalising constant $\frac{T^r}{\Gamma(r)}$, which is $\mathcal{G}(r, T)$.

4.2.1 Bayesian estimators using the Jeffreys' prior

The main objective is to obtain the various Bayesian estimators and compare their performances relative to the squared error loss and the LINEX loss functions. Hence, the various Bayesian estimators of the parameter θ are derived relative to both these loss functions.

Remark 4.6 *The two forms of the estimation error,* $\Delta_1 = \widehat{\theta} - \theta$ *(type I) and* $\Delta_2 = \frac{\widehat{\theta}}{\theta} - 1$ *(type II), will be considered since there are situations where* Δ_1 *is more appropriate than* Δ_2 *or vice versa.*

From (2.25), the Bayesian estimator relative to the squared error loss defined in terms of type I estimation error is the posterior mean. Therefore, the mean of the posterior (4.2) is given by

$$\widehat{\theta}_{S1} = E_{post}[\theta] = \frac{r}{T}. \tag{4.3}$$

Similarly, using (2.27) and (4.2), the Bayesian estimator is given by

$$\widehat{\theta}_{S2} = \frac{r+1}{T}. \tag{4.4}$$

In the same way, when the type I estimation error is considered, using (2.30) and (4.2), the Bayesian estimator relative to the LINEX loss function is obtained as

$$\begin{aligned}
\widehat{\theta}_{L1} &= -\frac{1}{a}\log[E_{post}[\exp[-a\theta]]] \\
&= -\frac{1}{a}\log\left[\int_0^\infty \frac{T^r\theta^{r-1}\exp[-[T+a]\theta]}{\Gamma(r)}d\theta\right] \\
&= \frac{r}{a}\log\left[\frac{T+a}{T}\right].
\end{aligned} \tag{4.5}$$

The Bayesian estimator relative to the LINEX loss function defined with the type II estimation error is obtained by using (2.32) and (4.2) from

$$E_{post}\left[\theta \exp\left[\frac{a\theta}{\widehat{\theta}_{L2}}\right]\right] = \frac{r}{T}\exp[a]. \tag{4.6}$$

The left hand side of (4.6) is given by

$$E_{post}\left[\theta \exp\left[\frac{a\theta}{\widehat{\theta}_{L2}}\right]\right] = \frac{T^r}{\Gamma(r)}\int_0^\infty \theta^r \exp\left[-\left(T-\frac{a}{\widehat{\theta}_{L2}}\right)\theta\right]d\theta = \frac{rT^r}{\left(T-\frac{a}{\widehat{\theta}_{L2}}\right)^{r+1}}. \tag{4.7}$$

Solving for $\widehat{\theta}_{L2}$ in

$$\frac{rT^r}{\left(T-\frac{a}{\widehat{\theta}_{L2}}\right)^{r+1}} = \frac{r}{T}\exp[a]$$

yields

$$\widehat{\theta}_{L2} = \frac{a}{zT}, \tag{4.8}$$

where

$$z = 1 - \exp\left[-\frac{a}{r+1}\right].$$

4.2.2 Risk measures and admissibility

Once the Bayesian estimators are derived, their respective risk measures have to be evaluated in order to be used for the performance comparisons. Since the non-informative prior considered here is improper, the integrated risk is not obtainable in this situation. Hence, in the subsequent sections, only the risk functions and the posterior risks for the Bayesian estimators will be derived and implemented in comparisons.

4.2.2.1 Risk functions

The risk function is one of the measures predominantly used to compare the performance of estimators for the parameter(s) of interest. In this section, the risk functions of the various Bayesian estimators are derived and consequently employed for performance comparisons.

Given the likelihood function (2.10), the pdf of the sufficient statistic, T, is given by

$$q(T|\theta) \propto \theta^r \exp[-\theta T], \tag{4.9}$$

which is a $\mathcal{G}(r, \theta)$.

Applying (2.33) relative to (2.24) and using (4.3) and (4.9), the risk function of $\widehat{\theta}_{S1}$ is

$$\begin{aligned} R_S(\widehat{\theta}_{S1}, \theta) &= \int_0^\infty \left(\frac{r}{T} - \theta\right)^2 q(T|\theta)dT \\ &= \frac{(r+2)\theta^2}{(r-1)(r-2)}, \qquad r > 2. \end{aligned} \tag{4.10}$$

An analogous computation of the risk function of the estimator $\widehat{\theta}_{L1}$, relative to the squared error loss function, is given by

$$R_S(\widehat{\theta}_{L1}, \theta) = \int_0^\infty \left(\frac{r}{a}\log\left[\frac{T+a}{T}\right] - \theta\right)^2 q(T|\theta)dT. \tag{4.11}$$

Both risk functions (4.10) and (4.11) are functions of the parameter θ. Furthermore, it can be noted from (4.10) that for any value of θ, $R_S(\widehat{\theta}_{S1}, \theta) \leqslant \frac{5}{2}\theta^2$, where equality holds if $r = 3$. Hence, if (4.11) is greater than $\frac{5}{2}\theta^2$, then $\widehat{\theta}_{S1}$ dominates $\widehat{\theta}_{L1}$, i.e., any choice of the estimator $\widehat{\theta}_{L1}$ in this scenario will result in a higher penalty.

The integral in (4.11) is not easily obtainable, one of the approximation techniques discussed in chapter 3 can be applied. However, since the ultimate goal here is to compare the two risk functions, a Monte Carlo simulation is carried out for the difference of the two risk functions, (4.10) and (4.11)

$$\begin{aligned} & R_S(\widehat{\theta}_{S1}, \theta) - R_S(\widehat{\theta}_{L1}, \theta) \\ = & \int_0^\infty \left\{\left(\frac{r}{T} - \theta\right)^2 - \left(\frac{r}{a}\log\left[\frac{T+a}{T}\right] - \theta\right)^2\right\} q(T|\theta)dT \\ = & \int_0^\infty \left\{\frac{r}{a}2\theta\log\left[\frac{T+a}{T}\right] - \frac{r^2}{a^2}\log^2\left[\frac{T+a}{T}\right]\right\} q(T|\theta)dT + \frac{(4r-r^2)\theta^2}{(r-1)(r-2)}, \end{aligned} \tag{4.12}$$

which can be written as

$$\int_0^\infty g(T|\theta)q(T|\theta)dT + g(\theta).$$

The Monte Carlo procedure is implemented for the integrand above. A sample of 10000 randomly generated variates from $\mathcal{G}(r, \theta)$ is used to simulate (4.12). It should be noted that the parameters a and θ are arbitrarily chosen, although from the Bayesian perspective, θ is supposed to be unknown or generated from an assumed prior distribution. Another point worth noting is, some of the 10000 randomly generated random variates were discarded when $a < 0$, due to the constraint that $T + a > 0$. Table 4.1 summarises the results.

Table 4.1: The difference of the risk functions relative to L_S using Δ_1.

	(r, θ)				
a	$(3, 1)$	$(3, 10)$	$(10, 1)$	$(10, 10)$	$(100, 3)$
1	2.03	152.83	0.04	−81.14	−7.00
8	2.26	152.34	0.08	−81.23	−7.02
10	2.22	152.23	0.07	−81.25	−7.03
−1	−1.11	153.05	−0.06	−81.11	−7.00
−8	2.32	153.66	−2.05	−80.99	−6.97

Clearly, the table shows that the risk functions of the two estimators are crossing-over. Hence, none of the estimators is admissible nor uniformly superior than the other, relative to the risk function measure. One might also use the estimator $\widehat{\theta}_{L2}$, while using the squared error loss function defined in terms of the type I estimation error. Therefore, the corresponding risk function of the estimator (4.8) is given by

$$\begin{aligned} R_S(\widehat{\theta}_{L2}, \theta) &= \int_0^\infty (\widehat{\theta}_{L2} - \theta)^2 q(T|\theta) dT \\ &= \frac{\left[a^2 - 2a(r-2)z + (r-1)(r-2)z^2\right]\theta^2}{(r-1)(r-2)z^2}, \qquad r > 2 \end{aligned} \tag{4.13}$$

which is once again a function of the parameter θ. It can be observed that even though (4.13) is differentiable with respect to θ, a stationary point cannot be obtained within the allowable range. Hence, the minimum of (4.13) is not obtainable. Comparison of (4.10) and (4.13) is done by using the ratio of the two risk functions. First, rewrite (4.10) and (4.13) as

$$\frac{R_S(\widehat{\theta}_{S1}, \theta)(r-1)(r-2)}{\theta^2} = r + 2$$

and

$$\frac{R_S(\widehat{\theta}_{L2}, \theta)(r-1)(r-2)}{\theta^2} = \frac{a^2 - 2a(r-2)z + (r-1)(r-2)z^2}{z^2}.$$

The ratio is given by

$$\frac{(r+2)z^2}{a^2 - 2a(r-2)z + (r-1)(r-2)z^2}.$$

For various values of the parameters, it was found that none of the two risk functions is uniformly larger than the other. Consider for instance, $a = 1$ and $r = 3$, which yields a ratio of 0.373, in which case the correct estimator is superior to $\widehat{\theta}_{L2}$. On the other hand, choosing the

LINEX parameter as $a = -3$, changes the ratio to 1.301, which implies that $\widehat{\theta}_{L2}$ is superior to the correct Bayesian estimator. Furthermore, from the fact that $R_S(\widehat{\theta}_{S1}, \theta) \leqslant \frac{5}{2}\theta^2$ for any θ, the correct Bayesian estimator is superior to $\widehat{\theta}_{L2}$, if the following condition is satisfied:

$$\frac{[a^2 - 2a(r-2)z + (r-1)(r-2)z^2]\,\theta^2}{(r-1)(r-2)z^2} > \frac{5}{2}\theta^2,$$

which (dividing both sides by by θ^2) reduces to

$$2a^2 - 4a(r-2)z - 3(r-1)(r-2)z^2 > 0.$$

In general, none of the two estimators dominates the other uniformly relative to the squared error loss function. This is analogous to the result obtained by Basu and Ebrahimi (1991) for another reparameterised exponential model.

Similarly, considering the type II estimation error, the risk function of the Bayesian estimator $\widehat{\theta}_{S2}$, is obtained using (2.24), (4.4) and (4.9), yielding

$$\begin{aligned} R_S(\widehat{\theta}_{S2}, \theta) &= \int_0^\infty \left(\frac{T\theta}{r+1} - 1\right)^2 q(T|\theta)dT \\ &= \frac{1}{r+1}. \end{aligned} \tag{4.14}$$

Suppose now, one uses the estimator $\widehat{\theta}_{L2}$ in place of $\widehat{\theta}_{S2}$, using (4.8), the risk function is

$$\begin{aligned} R_S(\widehat{\theta}_{L2}, \theta) &= \int_0^\infty \left(\frac{z\theta T}{a} - 1\right)^2 q(T|\theta)dT \\ &= \frac{r(r+1)z^2}{a^2} - \frac{2rz}{a} + 1. \end{aligned} \tag{4.15}$$

It is evident that both functions (4.14) and (4.15) are independent of θ. The risk function relative to the LINEX loss function is derived in similar fashion. Using (2.28), (4.5) and (4.9), the risk function is given by

$$\begin{aligned} R_L(\widehat{\theta}_{L1}, \theta) &= \int_0^\infty L_L\left(\frac{r}{a}\log\left[\frac{T+a}{T}\right] - \theta\right) q(T|\theta)dT \\ &= \exp[-a\theta]\int_0^\infty \left\{\left(\frac{T+a}{T}\right)^r - \log\left[\frac{T+a}{T}\right]^r\right\} q(T|\theta)dT + a\theta - 1. \end{aligned} \tag{4.16}$$

Applying the binomial theorem to $\left(\frac{T+a}{T}\right)^r$, (4.16) can be rewritten as

$$\begin{aligned}
R_L(\widehat{\theta}_{L1},\theta) &= \exp[-a\theta]\left\{\left[\sum_{k=0}^{r}\binom{r}{k}\int_0^\infty\left(\frac{a}{T}\right)^k q(T|\theta)dT\right]-\int_0^\infty \log\left[\frac{T+a}{T}\right]^r q(T|\theta)dT\right\} \\
&\quad +a\theta-1 \\
&= \exp[-a\theta]\left\{\sum_{k=0}^{r-2}\binom{r}{k}\frac{a^k\Gamma(r-k)\theta^k}{\Gamma(r)}+\frac{r(a\theta)^{r-1}}{\Gamma(r)}+\frac{a^r\theta^r}{\Gamma(r)}\int_0^\infty T^{-1}\exp[-\theta T]dT\right\} \\
&\quad -\exp[-a\theta]\int_0^\infty \log\left[\frac{T+a}{T}\right]^r q(T|\theta)dT+a\theta-1.
\end{aligned}$$

Similarly, the risk function corresponding to the estimator $\widehat{\theta}_{S1}$ is obtained by using (2.28), (4.3) and (4.9)

$$\begin{aligned}
R_L(\widehat{\theta}_{S1},\theta) &= \int_0^\infty L_L\left(\frac{r}{T}-\theta\right)q(T|\theta)dT \\
&= \int_0^\infty \left(\exp\left[a\left(\frac{r}{T}-\theta\right)\right]-a\left(\frac{r}{T}-\theta\right)-1\right)q(T|\theta)dT \\
&= \exp[-a\theta]\int_0^\infty \exp\left[\frac{ar}{T}\right]q(T|\theta)dT-\frac{ar\theta}{r-1}+a\theta-1. \qquad (4.17)
\end{aligned}$$

The Monte Carlo procedure is employed with 10000 randomly generated variates to determine whether the difference of the two risk functions (4.16) and (4.17) varies in sign or not. Hence, the difference is

$$\begin{aligned}
&R_L(\widehat{\theta}_{L1},\theta)-R_L(\widehat{\theta}_{S1},\theta) \\
&= \left\{\exp[-a\theta]\int_0^\infty \left[\left(\tfrac{T+a}{T}\right)^r-\exp\left[\tfrac{ar}{T}\right]\right]-r\log\left[\tfrac{T+a}{T}\right]\right\}q(T|\theta)dT+\frac{a\theta r}{r-1}. \qquad (4.18)
\end{aligned}$$

Since the values of the parameters are randomly chosen, some of the combinations resulted a negative risk function and hence are not valid for the comparison. Table 4.2 shows some of the results for which the simulation was done. Some of the values of the parameters chosen, whose risk function values are negative, have also been included.

Table 4.2: The difference of the risk functions relative to L_L using Δ_1.

	(r,θ)			
a	$(3,1)$	$(3,10)$	$(10,10)$	$(100,1)$
1	$-3.96E+47$	$R_L(\widehat{\theta}_{S1},\theta)<0$	*both negative*	$R_L(\widehat{\theta}_{L1},\theta)<0$
8	$-7.19E+106$	$R_L(\widehat{\theta}_{S1},\theta)<0$	$R_L(\widehat{\theta}_{S1},\theta)<0$	$R_L(\widehat{\theta}_{L1},\theta)<0$
10	$-1.51E+163$	$R_L(\widehat{\theta}_{S1},\theta)<0$	$R_L(\widehat{\theta}_{S1},\theta)<0$	$R_L(\widehat{\theta}_{L1},\theta)<0$
-1	$R_L(\widehat{\theta}_{L1},\theta)<0$	-128.32	-70.05	$R_L(\widehat{\theta}_{L1},\theta)<0$
-8	$-2.08E+02$	$-3.34E+33$	$-8.93E+32$	$R_L(\widehat{\theta}_{L1},\theta)<0$
-10	$-1.44E+03$	$-1.83E+42$	$-5.31E+41$	$R_L(\widehat{\theta}_{L1},\theta)<0$

It is evident that the correct Bayesian estimator dominates the other alternative, irrespective of the sign of the LINEX parameter.

Using (2.28) and (4.8) relative to the type II estimation error, the risk function is

$$\begin{aligned} R_L(\widehat{\theta}_{L2},\theta) &= \int_0^\infty L_L\left(\frac{\theta}{\widehat{\theta}_{L2}}-1\right) q(T|\theta)dT \\ & \exp[-a]\left(\frac{1}{1-z}\right)^r - rz + a - 1. \end{aligned} \tag{4.19}$$

Similarly, using (4.4), the risk function of the estimator $\widehat{\theta}_{S2}$, relative to the LINEX loss function is

$$\begin{aligned} R_L(\widehat{\theta}_{S2},\theta) &= \int_0^\infty L_L\left(\frac{\theta}{\widehat{\theta}_{S2}}-1\right) q(T|\theta)dT \\ &= \exp[-a]\left(\frac{r+1}{r+1-a}\right)^r - \frac{ar}{r+1} + a - 1, \qquad r+1>a. \end{aligned} \tag{4.20}$$

In order to compare the performance of the two estimators relative to the risk function measure, consider the difference between (4.19) and (4.20)

$$\begin{aligned} & R_L(\widehat{\theta}_{L2},\theta) - R_L(\widehat{\theta}_{S2},\theta) \\ =\; & \exp[-a]\left(\frac{1}{1-z}\right)^r - rz - \exp[-a]\left(\frac{r+1}{r+1-a}\right)^r + \frac{ra}{r+1} \\ =\; & \left\{(r+1)\exp\left[-\frac{a}{r+1}\right]\right\} + r\left(\frac{a-r-1}{r+1}\right) - \exp[-a]\left(\frac{r+1}{r+1-a}\right)^r. \end{aligned} \tag{4.21}$$

It can be easily observed that in (4.21), only the leading summand is positive. A numerical computation for various values of a and r shows that the expression (4.21) is negative irrespective of the sign of the LINEX parameter. It should be noted that the constraint $r+1>a$ is required in order to satisfy a positive expected loss. Table 4.3 summarises the differences in risk functions for different choices of r and a.

Table 4.3: The difference of the risk functions relative to L_L using Δ_2.

	r			
a	3	10	100	1000
1	$-6.81E-03$	$-9.86E-04$	$-1.22E-05$	$-1.25E-07$
1.5	$-3.98E-02$	$-5.19E-03$	$-6.20E-05$	$-6.31E-07$
2	$-1.57E-01$	$-1.73E-02$	$-1.97E-04$	$-2.00E-06$
3	$-2.05E+00$	$-1.01E-01$	$-1.01E-03$	$-1.01E-05$
-1	$-5.66E-03$	$-9.26E-04$	$-1.21E-05$	$-1.25E-07$
-1.5	$-2.90E-02$	$-4.71E-03$	$-6.13E-05$	$-6.31E-07$
-2	$-9.45E-02$	$-1.50E-02$	$-1.94E-04$	$-1.99E-06$

It is evident that $\widehat{\theta}_{L2}$ outperforms $\widehat{\theta}_{S2}$ relative to the LINEX loss function.

Consider now the estimator $\widehat{\theta}_{S1}$ to evaluate the risk function with respect to the LINEX loss function. The consequent risk function is obtained, using (4.3), as

$$\begin{aligned} R_L(\widehat{\theta}_{S1},\theta) &= \int_0^\infty L_L\left(\frac{\theta}{\widehat{\theta}_{S1}}-1\right) q(T|\theta)dT \\ &= \int_0^\infty \left(\exp\left[a\left(\frac{\theta T}{r}-1\right)\right] - a\left(\frac{\theta T}{r}-1\right)-1\right) q(T|\theta)dT \\ &= \exp[-a]\left(\frac{r}{r-a}\right)^r - 1, \qquad r>a. \end{aligned} \tag{4.22}$$

The performance of $\widehat{\theta}_{S1}$ and $\widehat{\theta}_{L2}$ relative to the LINEX loss function when considering the type II estimation error is evaluated as well. Hence, using (4.19) and (4.22), the difference is

$$\begin{aligned} & R_L(\widehat{\theta}_{L2},\theta) - R_L(\widehat{\theta}_{S1},\theta) \\ =\ & \exp[-a]\left[\left(\frac{1}{1-z}\right)^r - \left(\frac{r}{r-a}\right)^r\right] - rz + a \\ =\ & (r+1)\exp\left[-\frac{a}{r+1}\right] - \exp[-a]\left(\frac{r}{r-a}\right)^r + a - r. \end{aligned} \tag{4.23}$$

A numerical computation for various values of r and a shows that the expression (4.23) is always negative. Table 4.4 summarises these results for different values of r and a.

Table 4.4: Difference of the risk functions of $\widehat{\theta}_{S1}$ and $\widehat{\theta}_{L2}$ relative to L_L.

	r			
a	3	10	100	1000
1	$-1.26E-01$	$-1.10E-02$	$-1.12E-04$	$-1.12E-06$
1.5	$-5.36E-01$	$-3.56E-02$	$-3.45E-04$	$-3.45E-06$
2	$-2.23E+00$	$-8.91E-02$	$-8.06E-04$	$-8.01E-06$
2.5	$-1.61E+01$	$-1.94E-01$	$-1.60E-03$	$-1.58E-05$
-1	$-1.07E-02$	$-1.15E-03$	$-1.24E-05$	$-1.25E-07$
-1.5	$-7.94E-03$	$-7.21E-04$	$-7.06E-06$	$-7.03E-08$
-2	$-1.15E-03$	$-1.67E-05$	$-2.16E-09$	$-1.28E-13$
-3	$-4.27E-02$	$-8.01E-03$	$-1.09E-04$	$-1.12E-06$

4.2.2.2 Posterior risks

The posterior risk is obtained as the expectation of the loss function with respect to the posterior distribution of the parameter. For convenience purposes, the posterior risk of any estimator $\widehat{\theta}$ will first be derived. The relevant posterior risks of the different Bayesian estimators can then be obtained by replacing $\widehat{\theta}$ with the consequent estimator.

Using (2.24) (with $k = 1$), (2.34) and (4.2) the posterior risk of $\widehat{\theta}$ is given by

$$\begin{aligned} R_1^S(\widehat{\theta}) &= E_{post}\left[L_S(\widehat{\theta},\theta)\right] = \frac{T^r}{\Gamma(r)}\int_0^\infty (\widehat{\theta}-\theta)^2\theta^{r-1}\exp[-\theta T]d\theta \\ &= \widehat{\theta}^2 - \frac{2r\widehat{\theta}}{T} + \frac{r(r+1)}{T^2}. \end{aligned} \tag{4.24}$$

From (4.24), the posterior risk of the Bayesian estimator in (4.3) relative to the squared error loss function is given as

$$R_1^S(\widehat{\theta}_{S1}) = \frac{r}{T^2}, \tag{4.25}$$

which is equivalent to the variance of the posterior distribution.

Consider the situation when one uses another estimator say for instance, $\widehat{\theta}_{L1}$, to evaluate the posterior risk relative to the squared error loss function where in fact, the correct Bayesian estimator is $\widehat{\theta}_{S1}$. Consequently, the risk incurred due to the wrongly chosen estimator is obtained by substituting (4.5) in (4.24) to obtain

$$R_1^S(\widehat{\theta}_{L1}) = E_{post}[L_S(\widehat{\theta}_{L1},\theta)] = \left(\frac{r}{a}\log\left[\frac{T+a}{T}\right]\right)^2 - \frac{2r^2}{aT}\log\left[\frac{T+a}{T}\right] + \frac{r(r+1)}{T^2}. \tag{4.26}$$

It is also of importance to compare the two estimators relative to the LINEX loss function.

Using (2.28) and (4.2), with the type I estimation error, the general form of the posterior risk is given as

$$\begin{aligned} R_1^L(\widehat{\theta}) &= E_{post}\left[L_L(\widehat{\theta},\theta)\right] = \frac{T^r}{\Gamma(r)}\int_0^\infty L_L(\widehat{\theta}-\theta)\theta^{r-1}\exp[-T\theta]d\theta \\ &= \frac{T^r}{\Gamma(r)}\int_0^\infty [\exp[a(\widehat{\theta}-\theta)]-a(\widehat{\theta}-\theta)-1]\theta^{r-1}\exp[-T\theta]d\theta \\ &= \exp[a\widehat{\theta}]\left(\frac{T}{T+a}\right)^r - a\widehat{\theta}+\frac{ar}{T}-1. \end{aligned} \tag{4.27}$$

Substituting the results of (4.5) and (4.3) in (4.27) give

$$R_1^L(\widehat{\theta}_{L1}) = \frac{ar}{T} - r\log\left[\frac{T+a}{T}\right] \tag{4.28}$$

and

$$R_1^L(\widehat{\theta}_{S1}) = \exp\left[\frac{ar}{T}\right]\left(\frac{T}{T+a}\right)^r - 1, \tag{4.29}$$

respectively.

Remark 4.7 *Evaluating the posterior risk of $\widehat{\theta}_{S1}$ relative to the LINEX loss function is needed in the decision-making process, when there is a possibility that another estimator can be chosen incorrectly.*

An analogous derivation of the above results of the posterior risks is done when the loss functions are defined in terms of the type II estimation error. Moreover, there are situations when one of the two forms of error leads to a simpler posterior analysis than the other, or vice versa.

In general, relative to the squared error loss function when using the type II estimation error the posterior risk is given by

$$\begin{aligned} R_2^S(\widehat{\theta}) &= E_{post}\left[L_S(\widehat{\theta},\theta)\right] = \frac{T^r}{\Gamma(r)}\int_0^\infty L_S\left(\frac{\theta}{\widehat{\theta}}-1\right)\theta^{r-1}\exp[-T\theta]d\theta \\ &= \frac{T^r}{\Gamma(r)}\int_0^\infty \left(\frac{\theta}{\widehat{\theta}}-1\right)^2\theta^{r-1}\exp[-T\theta]d\theta \\ &= \frac{r(r+1)}{T^2\widehat{\theta}^2}-\frac{2r}{T\widehat{\theta}}+1. \end{aligned} \tag{4.30}$$

Substituting the estimator (4.4) in (4.30) yields the posterior risk corresponding to $\widehat{\theta}_{S2}$

$$R_2^S(\widehat{\theta}_{S2}) = \frac{1}{r+1}. \tag{4.31}$$

Suppose once again the case when one uses the estimator $\widehat{\theta}_{L2}$ instead of $\widehat{\theta}_{S2}$ to obtain the minimum posterior risk relative to the squared error loss function, the corresponding posterior risk is

$$R_2^S(\widehat{\theta}_{L2}) = \frac{r(r+1)z^2}{a^2}-\frac{2rz}{a}+1. \tag{4.32}$$

In a similar procedure, the posterior risk relative to the LINEX loss function using the type II estimation error, is obtained from (2.28) and (4.2) as

$$\begin{aligned} R_2^L(\widehat{\theta}) &= E_{post}\left[L_L(\widehat{\theta},\theta)\right] = \frac{T^r}{\Gamma(r)}\int_0^\infty L_L\left(\frac{\theta}{\widehat{\theta}}-1\right)\theta^{r-1}\exp[-T\theta]d\theta \\ &= \frac{T^r}{\Gamma(r)}\exp[-a]\int_0^\infty \theta^{r-1}\exp\left[-\left(T-\frac{a}{\widehat{\theta}}\right)\theta\right]d\theta - \frac{a}{\widehat{\theta}}\int_0^\infty \theta^{r+1-1}\exp[-T\theta]d\theta + a - 1 \\ &= \exp[-a]\left(\frac{T\widehat{\theta}}{T\widehat{\theta}-a}\right)^r - \frac{ar}{T\widehat{\theta}}+a-1. \end{aligned} \tag{4.33}$$

Replacing $\widehat{\theta}$ with (4.8) in (4.33), the posterior risk corresponding to $\widehat{\theta}_{L2}$ is

$$R_2^L(\widehat{\theta}_{L2}) = \exp[-a]\left(\frac{1}{1-z}\right)^r - rz + a - 1. \tag{4.34}$$

Analogously, replacing $\widehat{\theta}$ with (4.4) in (4.33) yields the posterior risk corresponding to $\widehat{\theta}_{S2}$

$$R_2^L(\widehat{\theta}_{S2}) = \exp[-a]\left(\frac{r+1}{r+1-a}\right)^r - \frac{ar}{r+1}+a-1, \qquad r+1>a. \tag{4.35}$$

Table 4.5 summarises the Bayesian estimators and their respective posterior risks.

Table 4.5: Bayesian estimators and their posterior risks

Estimator	R_1^S	R_1^L
$\widehat{\theta}$	$\widehat{\theta}^2 - \frac{2r\widehat{\theta}}{T} + \frac{r(r+1)}{T^2}$	$\exp[a\widehat{\theta}]\left(\frac{T}{T+a}\right)^r - a\widehat{\theta} + \frac{ar}{T} - 1$
$\widehat{\theta}_{S1} = \frac{r}{T}$	$\frac{r}{T^2}$	$\exp\left[\frac{ar}{T}\right]\left(\frac{T}{T+a}\right)^r - 1$
$\widehat{\theta}_{L1} = \frac{r}{a}\log\left[\frac{T+a}{T}\right]$	$\left(\frac{r}{a}\log\left[\frac{T+a}{T}\right]\right)^2 - \frac{2r^2}{aT}\log\left[\frac{T+a}{T}\right]$ $+\frac{r(r+1)}{T^2}$	$\frac{ar}{T} - r\log\left[\frac{T+a}{T}\right]$
Estimator	R_2^S	R_2^L
$\widehat{\theta}$	$\frac{r(r+1)}{T^2\widehat{\theta}^2} - \frac{2r}{T\widehat{\theta}} + 1$	$\exp[-a]\left(\frac{T\widehat{\theta}}{T\widehat{\theta}-a}\right)^r - \frac{ar}{T\widehat{\theta}} + a - 1$
$\widehat{\theta}_{S2} = \frac{r+1}{T}$	$\frac{1}{r+1}$	$\exp[-a]\left(\frac{r+1}{r+1-a}\right)^r - \frac{ar}{r+1} + a - 1$
$\widehat{\theta}_{L2} = \frac{a}{Tz}$	$\frac{r(r+1)z^2}{a^2} - \frac{2rz}{a} + 1$	$\exp[-a]\left(\frac{1}{1-z}\right)^r - rz + a - 1$
$\widehat{\theta}_{S1} = \frac{r}{T}$	$\frac{1}{r}$	$exp[-a]\left(\frac{r}{r-a}\right)^r - 1$

Performance comparison of the estimators is carried out by using the relevant posterior risks. Comparison could also be made relative to the estimation errors. For instance, if (4.25) and (4.31) are used, then $\widehat{\theta}_{S1}$ is superior to $\widehat{\theta}_{S2}$ if

$$R_2^S(\widehat{\theta}_{S2}) - R_1^S(\widehat{\theta}_{S1}) = \frac{1}{r+1} - \frac{r}{T^2} > 0, \tag{4.36}$$

i.e.

$$T^2 > r^2 + r.$$

In a situation where (4.36) is satisfied, the type I estimation error results in a smaller posterior risk relative to the squared error loss, the reverse is true otherwise. The difference between the posterior risks of (4.26) and (4.25), yields

$$\begin{aligned} R_1^S(\widehat{\theta}_{L1}) - R_1^S(\widehat{\theta}_{S1}) &= \left(\frac{r}{a}\log\left[\frac{T+a}{T}\right]\right)^2 - \frac{2r^2}{aT}\log\left[\frac{T+a}{T}\right] + \frac{r(r+1)}{T^2} - \frac{r}{T^2} \\ &= \left(\frac{r}{a}\log\left[\frac{T+a}{T}\right]\right)^2 - \frac{2r^2}{aT}\log\left[\frac{T+a}{T}\right] + \frac{r^2}{T^2} \\ &= \left(\frac{r}{a}\log\left[\frac{T+a}{T}\right] - \frac{r}{T}\right)^2 \geqslant 0, \end{aligned} \tag{4.37}$$

where equality holds when the two estimators are equal. From (4.37), it can be concluded that the Bayesian estimator $\widehat{\theta}_{S1}$ is superior to $\widehat{\theta}_{L1}$ irrespective of the value of the parameters. It means that the wrongly chosen estimator, $\widehat{\theta}_{L1}$, leads to a greater penalty or risk.

Similarly, the same estimators can also be compared relative to the LINEX loss function. It can be easily observed that

$$\exp\left[\frac{ar}{T} - r\log\left[\frac{T+a}{T}\right]\right] = \exp\left[\frac{ar}{T}\right]\left(\frac{T}{T+a}\right)^{r}. \tag{4.38}$$

Hence, using (4.28), (4.29) and (4.38), the following result is obtained:

$$R_1^L(\widehat{\theta}_{S1}) = \exp\left[R_1^L(\widehat{\theta}_{L1})\right] - 1. \tag{4.39}$$

This follows from

$$\exp[x] = 1 + x + \frac{x^2}{2!} + ... \tag{4.40}$$

that

$$\exp[x] > 1 + x, \qquad x \neq 0.$$

Replacing x with $R_1^L(\widehat{\theta}_{L1})$ in (4.40) it is obvious that

$$R_1^L(\widehat{\theta}_{S1}) > R_1^L(\widehat{\theta}_{L1}). \tag{4.41}$$

Therefore, (4.41) implies that $\widehat{\theta}_{L1}$ is superior to $\widehat{\theta}_{S1}$. This also implies that under the LINEX loss function, the wrongly chosen alternative, say $\widehat{\theta}_{S1}$, results in incurring a greater risk.

The same comparison of the performance of the estimators is done using (4.31) and (4.32), where the difference is given by

$$R_2^S(\widehat{\theta}_{L2}) - R_2^S(\widehat{\theta}_{S2}) = \frac{r(r+1)z^2}{a^2} - \frac{2rz}{a} + \frac{r}{r+1}. \tag{4.42}$$

Dividing (4.42) by r yields

$$\frac{(r+1)z^2}{a^2} - \frac{2z}{a} + \frac{1}{r+1},$$

which can be rewritten as

$$\left(\frac{z}{a} - \frac{1}{r+1}\right)\left((r+1)\frac{z}{a} - 1\right). \tag{4.43}$$

It can be easily observed that the two factors in (4.43) have the same sign, implying that (4.42) is non-negative. Therefore, when type II estimation error is considered, the correct Bayesian estimator, $\widehat{\theta}_{S2}$, is superior to the estimator obtained under the LINEX loss function.

Similarly, the performance of the same estimators can also be compared relative to the LINEX loss function. Using the posterior risks in (4.34) and (4.35), the difference is given by

$$R_2^L(\widehat{\theta}_{S2}) - R_2^L(\widehat{\theta}_{L2}) = \exp[-a]\left[\left(\frac{r+1}{r+1-a}\right)^r - \left(\frac{1}{1-z}\right)^r\right] + rz - \frac{ar}{r+1}. \tag{4.44}$$

Table 4.6 summarises (4.44), for different values of r and a.

Table 4.6: Difference of the posterior risks relative L_L using Δ_2.

	r			
a	1	2	7	20
0.001	$1.55E-14$	$9.14E-15$	$2.18E-15$	$3.19E-15$
0.1	$1.59E-06$	$9.37E-07$	$1.72E-07$	$2.70E-08$
1	$2.27E-02$	$1.15E-02$	$1.83E-03$	$2.77E-04$
1.5	$1.98E-01$	$7.29E-02$	$9.81E-03$	$1.43E-03$
-0.1	$1.54E-06$	$9.16E-07$	$1.70E-07$	$2.70E-08$
-1	$1.47E-02$	$8.86E-03$	$1.68E-03$	$2.68E-04$

It is evident that the correct Bayesian estimator dominates the wrongly chosen alternative. This is, however expected, since by definition the Bayesian estimator minimises the expected posterior loss. It is also evident that for a small absolute value of the LINEX parameter, the variation is negligible and hence both estimators perform almost equally, lending credibility to the property of the LINEX loss function discussed in chapter 2.

4.3 POSTERIOR ANALYSIS USING THE CONJUGATE PRIOR

In this section, the derivation of the posterior distribution, the Bayesian estimators and the risk measures are derived using the conjugate prior of the exponential model. The gamma distribution is a conjugate family of the sequence of random variables $X_1, X_2, ..., X_n$, whose pdf is of the form (2.7).

Remark 4.8 *Analogously, the inverted gamma distribution is a conjugate family for a sequence of random variables following the reparameterised form of the exponential distribution i.e. replacing θ by $\frac{1}{\theta}$ in (2.7).*

Suppose now the parameter of the exponential model (2.7) is assumed to be gamma distributed

$$\pi(\theta|\alpha,\beta) = \frac{\beta^{\alpha}\theta^{\alpha-1}\exp[-\beta\theta]}{\Gamma(\alpha)}, \qquad \theta \geqslant 0,\ \alpha,\ \beta > 0. \tag{4.45}$$

Using (2.10), (2.19) and (4.45), the corresponding posterior distribution is

$$p(\theta|data) = \frac{(T+\beta)^{r+\alpha}\,\theta^{r+\alpha-1}\exp[-(T+\beta)\theta]}{\Gamma(r+\alpha)} \tag{4.46}$$

which is $\mathcal{G}(r+\alpha, T+\beta)$.

4.3.1 Bayesian estimators using the conjugate prior

The general expressions of the Bayesian estimators relative to any arbitrary posterior distribution were derived in chapter 2. In this section, those expressions will be used in deriving the individual Bayesian estimators relative to the posterior distribution (4.46). The two forms of the estimation errors will interchangeably be used in defining the loss functions.

Using (2.25) and (4.46), the Bayesian estimator relative to the squared error loss function, which is the mean of the posterior, is given as

$$\widehat{\theta}_{S1} = E_{post}[\theta] = \frac{r+\alpha}{T+\beta}. \tag{4.47}$$

Relative to the type II estimation error, the corresponding Bayesian estimator is obtained using (2.27) and (4.46)

$$\widehat{\theta}_{S2} = \frac{E_{post}[\theta^2]}{E_{post}[\theta]} = \frac{r+\alpha+1}{T+\beta}. \tag{4.48}$$

The Bayesian estimator relative to the LINEX loss function when using the type I estimation error is

$$\begin{aligned}
\widehat{\theta}_{L1} &= -\frac{1}{a}\log[E_{post}[\exp[-a\theta]]] \\
&= -\frac{1}{a\Gamma(r+\alpha)}\log\left[\int_0^\infty (T+\beta)^{r+\alpha}\,\theta^{r+\alpha-1}\exp\left[-\left(T+\beta+a\right)\theta\right]d\theta\right] \\
&= \frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right].
\end{aligned} \tag{4.49}$$

Analogous to (4.49), when the type II estimation error is considered, the Bayesian estimator is

$$\widehat{\theta}_{L2} = \frac{a}{z_1\left(T+\beta\right)}, \tag{4.50}$$

where

$$z_1 = 1 - \exp\left[-\frac{a}{r+\alpha+1}\right].$$

4.3.2 Risk measures and admissibility

The risk functions, the posterior risks and the integrated risks will be derived in this section. In contrast to the non-informative scenario, the integrated risks can be obtained, since the prior is well defined here.

4.3.2.1 Risk functions

Since the likelihood is unchanged the distribution of the statistic T, is the same as in (4.9). Therefore, using (2.24), (4.9) and (4.47), the risk function of $\widehat{\theta}_{S1}$ is given as

$$\begin{aligned} R_S(\widehat{\theta}_{S1},\theta) &= \int_0^\infty \left(\frac{r+\alpha}{T+\beta}-\theta\right)^2 q(T|\theta)dT \\ &= \int_0^\infty \left[\left(\frac{r+\alpha}{T+\beta}\right)^2 - \frac{2(r+\alpha)\theta}{T+\beta}\right] q(T|\theta)dT + \theta^2. \end{aligned} \tag{4.51}$$

The risk function of $\widehat{\theta}_{L1}$ relative to the squared error loss function is also obtained using (4.49) and is given by

$$R_S(\widehat{\theta}_{L1},\theta) = \int_0^\infty \left(\frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right]-\theta\right)^2 q(T|\theta)dT. \tag{4.52}$$

Monte Carlo simulation can be applied to establish the trend of the of the risk functions of the two estimators relative to the squared error loss function. The difference is given by

$$\begin{aligned} & R_S(\widehat{\theta}_{L1},\theta) - R_S(\widehat{\theta}_{S1},\theta) \\ = & \int_0^\infty \left\{\left(\frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right]\right)^2 + \frac{2(r+\alpha)\theta}{T+\beta}\right\} q(T|\theta)dT \\ & -\int_0^\infty \left\{\frac{2(r+\alpha)\theta}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right] - \left(\frac{r+\alpha}{T+\beta}\right)^2\right\} q(T|\theta)dT. \end{aligned}$$

Similarly, the Monte Carlo procedures are implemented by generating 10000 variates from a $\mathcal{G}(r,\theta)$ for selected values of a, α and β. Table 4.7 summarises the results.

Table 4.7: Difference of the risk functions relative to L_S using Δ_1.

	(r,θ,α,β)			
a	$(1,0.5,0.01,0.5)$	$(2,1,0.5,1)$	$(3,2,2,3)$	$(4,3,4,10)$
0.01	-0.021	-0.0068	-0.00033	$-5.85E-05$
0.5	-1.139	-0.184	-0.032	$-3.35E-03$
-0.01	0.0336	0.0057	0.00068	$6.80E-05$
-0.5	8.98	0.55	4.12	0.0049
-1	3.00	0.889	0.056	0.0049

It is clear that crossing-over risk functions exist in some of the allowable ranges. Hence, none of the two estimators is uniformly superior. However, it is worth mentioning that for various values of the parameters, the difference attains positive when the LINEX parameter is negative. It might be the case that the correct Bayesian estimator outperforms the other alternative, when $a < 0$ and should be further investigated.

The corresponding risk function of $\widehat{\theta}_{S2}$ relative to the squared error loss function, which is defined in terms of the type II estimation error, is obtained using (2.24), (4.9) and (4.48)

$$\begin{aligned} R_S(\widehat{\theta}_{S2},\theta) &= \int_0^\infty \left(\frac{\theta\,(T+\beta)}{r+\alpha+1}-1\right)^2 q(T|\theta)dT \\ &= \frac{\theta^2}{(r+\alpha+1)^2}\left[\frac{r^2+r}{\theta^2}+2\beta\frac{r}{\theta}+\beta^2\right]+\frac{2\theta}{r+\alpha+1}\left(\frac{r}{\theta}+\beta\right)+1 \\ &= 3r^2\eta^2+3r\eta^2+4r\eta^2\beta\theta+\eta^2\beta^2\theta^2+2r\alpha\eta^2+2\alpha\eta^2\beta\theta+2\eta^2\beta\theta+1 \end{aligned} \quad (4.53)$$

where $\eta = \frac{1}{r+\alpha+1}$. It is evident that (4.53) increases when any one of the parameters increases, because of its polynomial property. Furthermore, the minimum of the (4.53) could not be obtained since there is no stationary point (a point where the first derivative with respect to θ vanishes). The risk function of $\widehat{\theta}_{L2}$ relative to the squared error loss function can also be obtained using (4.50) as follows

$$\begin{aligned} R_S(\widehat{\theta}_{L2},\theta) &= \int_0^\infty \left(\frac{z_1\theta\,(T+\beta)}{a}-1\right)^2 q(T|\theta)dT \\ &= \frac{z_1^2\theta^2}{a^2}\int_0^\infty (T+\beta)^2\,q\,(T|\theta)\,dT+\frac{2\theta z_1}{a}\int_0^\infty (T+\beta)\,q(T|\theta)dT+1 \\ &= \frac{z_1^2\theta^2}{a^2}\left[\frac{r^2+r}{\theta^2}+2\beta\frac{r}{\theta}+\beta^2\right]+\frac{2z_1\theta}{a}\left(\frac{r}{\theta}+\beta\right)+1. \end{aligned} \quad (4.54)$$

The expressions (4.53) and (4.54) are non-negative for any values of the parameters. To compare the performance of the two estimators $\widehat{\theta}_{L2}$ and $\widehat{\theta}_{S2}$, relative to the squared error loss function, the difference is

$$\begin{aligned} & R_S(\widehat{\theta}_{L2},\theta)-R_S(\widehat{\theta}_{S2},\theta) \\ = & \left\{\frac{z_1^2}{a^2}-\frac{1}{(r+\alpha+1)^2}\right\}\left(r^2+r+2r\beta\theta+\beta^2\theta^2\right)+\left[\frac{2z_1}{a}-\frac{2}{r+\alpha+1}\right](r+\beta\theta) \\ = & \left\{\tfrac{z_1}{a}-\tfrac{1}{r+\alpha+1}\right\}\left[\left(\tfrac{z_1}{a}+\tfrac{1}{r+\alpha+1}\right)\left(r^2+r+2r\beta\theta+\beta^2\theta^2\right)+2(r+\beta\theta)\right]. \end{aligned} \quad (4.55)$$

The expression in the second product is non-negative, hence, the sign of (4.55) depends on the sign of the expression

$$\frac{z_1}{a} - \frac{1}{r+\alpha+1}.$$

It can be observed that this expression is always negative when $a > 0$ and always positive when $a < 0$.

Therefore, relative to the squared error loss function, the correct estimator outperforms the other alternative if the LINEX parameter is negative and vice versa when the LINEX parameter is positive. Table 4.8 summarises the results of the difference for various chosen values of a, r, θ, α and β.

Table 4.8: Difference of the risk functions relative to L_S using Δ_2.

	$(r, \theta, \alpha, \beta)$			
a	$(1, 0.5, 0.01, 0.5)$	$(2, 1, 0.5, 1)$	$(3, 2, 2, 3)$	$(4, 3, 4, 10)$
-0.01	0.006	0.005	0.006	0.020
-0.1	0.064	0.051	0.064	0.202
-1	0.798	0.575	0.694	2.135
0.01	-0.006	-0.005	-0.006	-0.020
0.5	-0.280	-0.235	-0.307	-0.976
1	-0.504	-0.442	-0.590	-1.897

It is evident that none of the two risk functions is uniformly greater than the other. Hence, none of the two estimators outperforms the other uniformly relative to the squared error loss function. It also illustrates that the deviation of the two risk functions is negative whenever $a > 0$ and positive when $a < 0$.

Similar procedures are applied when using the LINEX loss function. Hence, using (4.49), the corresponding risk function is given by

$$\begin{aligned} R_L(\widehat{\theta}_{L1}, \theta) &= \int_0^\infty L_L\left(\frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right] - \theta\right) q(T|\theta)dT \\ &= \int_0^\infty \left[\exp[-a\theta]\left(\frac{T+\beta+a}{T+\beta}\right)^{r+\alpha} - (r+\alpha)\log\left[\frac{T+\beta+a}{T+\beta}\right]\right] q(T|\theta)dT \\ &\quad +a\theta - 1, \end{aligned} \tag{4.56}$$

which is not easily integrable and some of the Bayesian computation techniques discussed in chapter 3 can be implemented. The risk function corresponding to $\widehat{\theta}_{S1}$ relative to the LINEX loss function is obtained using (4.47)

$$\begin{aligned} R_L(\widehat{\theta}_{S1}, \theta) &= \int_0^\infty L_L\left(\frac{r+\alpha}{T+\beta} - \theta\right) q(T|\theta) dT \\ &= \left[\exp[-a\theta] \int_0^\infty \exp\left[\frac{a(r+\alpha)}{T+\beta}\right] - \frac{a(r+\alpha)}{T+\beta}\right] q(T|\theta) dT + a\theta - 1. \quad (4.57) \end{aligned}$$

At this stage, the performance of the two estimators might be compared using (4.56) and (4.57) by either approximating separately using the Monte Carlo procedure or by taking the difference and then using Monte Carlo procedures. In the latter case, care must be taken to verify that the two expressions must be non-negative, otherwise, the notion of a risk is not satisfied. In this case, the difference was taken in the range, where the two expressions attain a positive value, hence

$$\begin{aligned} & R_L(\widehat{\theta}_{L1}, \theta) - R_L(\widehat{\theta}_{S1}, \theta) \\ &= \int_0^\infty \left(L_L\left(\frac{r+\alpha}{a} \log\left[\frac{T+\beta+a}{T+\beta}\right] - \theta\right) - L_L\left(\frac{r+\alpha}{T+\beta} - \theta\right)\right) q(T|\theta) dT \\ &= \exp[-a\theta] \int_0^\infty \left\{ \left(\left(\frac{T+\beta+a}{T+\beta}\right)^{r+\alpha} - \exp\left[\frac{r+\alpha}{T+\beta}\right]\right)\right\} q(T|\theta) dT \\ &\quad + \exp[-a\theta] \int_0^\infty \left\{ \left(\frac{r+\alpha}{a} \log\left[\frac{T+\beta+a}{T+\beta}\right] - \frac{r+\alpha}{T+\beta}\right)\right\} q(T|\theta) dT. \quad (4.58) \end{aligned}$$

Using the Monte Carlo procedures, the expectation is obtained by generating random variates, T, from the $\mathcal{G}(r, \theta)$. Table 4.9 summarises the results.

Table 4.9: Difference of the risk functions relative L_L using Δ_1.

	(r,θ,α,β)		
a	$(1, 0.01, 1, 2)$	$(2, 0.5, 3, 1)$	$(3, 2, 10, 3)$
-0.001	$2.44E-10$	$2.05E-09$	$6.66E-11$
-0.1	$2.43E-04$	$1.58E-03$	$-2.65E-05$
-1	0.264	1.41	$3.73E-02$
0.01	$-2.44E-07$	$-2.85E-06$	$-7.95E-08$
1	-0.268	-13.48	-0.127
100	$-7.70E+42$	$-2.62E+161$	-9421.03

Table 4.9 shows that crossing-over risk functions are obtained relative to the LINEX loss func-

tion. Therefore, none of the two estimators dominates the other uniformly. It can also be noted that this is irrespective of the sign of the LINEX parameter.

The risk function of the Bayesian estimator $\widehat{\theta}_{L2}$ relative to the LINEX loss function is obtained using (4.50) and is given by

$$\begin{aligned} R_L(\widehat{\theta}_{L2}, \theta) &= \int_0^\infty \left(\exp\left[a\left(\frac{\theta}{\widehat{\theta}_{L2}} - 1 \right) \right] - a\left(\frac{\theta}{\widehat{\theta}_{L2}} - 1 \right) - 1 \right) q(T|\theta) dT \\ &= \exp[-a] \int_0^\infty \exp[z_1\theta\,(T+\beta)] q(T|\theta) dT - \int_0^\infty z_1\theta\,(T+\beta)\, q(T|\theta) dT + a - 1 \\ &= \exp[-a + z_1\beta\theta] \left(\frac{1}{1-z_1} \right)^r - z_1\beta\theta - rz_1 + a - 1. \end{aligned} \tag{4.59}$$

Sometimes, the performance of estimators might be compared using the minimum risk. The minimum risk incurred by choosing $\widehat{\theta}_{L2}$ relative to the LINEX loss function is obtained by partially differentiating (4.59) with respect to θ. Therefore,

$$\begin{aligned} \frac{\partial R_L(\widehat{\theta}_{L2}, \theta)}{\partial \theta} &= z_1\beta \exp\left[-a + z_1\beta\theta + \frac{ar}{r+\alpha+1} \right] - z_1\beta \\ &= z_1\beta \left(\exp\left[-a + z_1\beta\theta + \frac{ar}{r+\alpha+1} \right] - 1 \right), \end{aligned} \tag{4.60}$$

since $z_1\beta \neq 0$, (4.60) is zero only if

$$\exp\left[-a + z_1\beta\theta + \frac{ar}{r+\alpha+1} \right] - 1 = 0,$$

which implies that

$$-a + z_1\beta\theta + \frac{ar}{r+\alpha+1} = 0, \tag{4.61}$$

hence,

$$\theta = \frac{1}{z_1\beta}\left(a - \frac{ar}{r+\alpha+1} \right).$$

Substituting this in (4.59), gives the corresponding minimum risk function

$$\begin{aligned} R_L(\widehat{\theta}_{L2}, \theta)_{\min} &= \exp\left[-a + z_1\beta\left(\frac{1}{z_1\beta}\left(a - \frac{ar}{r+\alpha+1} \right) \right) + \frac{ar}{r+\alpha+1} \right] \\ &\quad -z_1\beta\left(\frac{1}{z_1\beta}\left(a - \frac{ar}{r+\alpha+1} \right) \right) - rz_1 + a - 1 \\ &= \frac{ar}{r+\alpha+1} - rz_1 = -r\log[1-z_1] - rz_1. \end{aligned} \tag{4.62}$$

The risk function of the estimator $\widehat{\theta}_{S2}$ relative to the LINEX loss function is obtained by using (4.48) and is given by

$$\begin{aligned} R_L(\widehat{\theta}_{S2}, \theta) &= \int_0^\infty \left(\exp\left[a \left(\frac{\theta}{\widehat{\theta}_{S2}} - 1 \right) \right] - a \left(\frac{\theta}{\widehat{\theta}_{S2}} - 1 \right) - 1 \right) q(T|\theta) dT \\ &= \exp\left[-a + \tfrac{a\beta\theta}{r+\alpha+1} \right] \left(\tfrac{r+\alpha+1}{r+\alpha+1-a} \right)^r - \tfrac{ar+a\beta\theta}{r+\alpha+1} + a - 1, \end{aligned} \tag{4.63}$$

where $(r + \alpha + 1 > a)$, since (4.63) must be non-negative. The minimum of (4.63) is obtained in a similar fashion as before. Hence, partially differentiating (4.63) with respect to θ

$$\frac{\partial R_L(\widehat{\theta}_{S2}, \theta)}{\partial \theta} = \exp\left[-a + \frac{a\beta\theta}{r+\alpha+1} \right] \left(\frac{r+\alpha+1}{r+\alpha+1-a} \right)^r \frac{a\beta}{r+\alpha+1} - \frac{a\beta}{r+\alpha+1}. \tag{4.64}$$

It can be observed easily that (4.64) reduces to zero only if

$$\exp\left[-a + \frac{a\beta\theta}{r+\alpha+1} \right] \left(\frac{r+\alpha+1}{r+\alpha+1-a} \right)^r - 1 = 0,$$

which implies that

$$\exp\left[\frac{a\beta\theta}{r+\alpha+1} \right] = \exp[a] \left(\frac{r+\alpha+1-a}{r+\alpha+1} \right)^r.$$

Solving θ yields

$$\theta = \frac{r+\alpha+1}{a\beta} \left(a + r \log\left[\frac{r+\alpha+1-a}{r+\alpha+1} \right] \right).$$

The corresponding minimum risk is given by

$$\begin{aligned} R_L(\widehat{\theta}_{S2}, \theta)_{\min} &= -\frac{ar}{r+\alpha+1} - r\log\left[\frac{r+\alpha+1-a}{r+\alpha+1} \right] \\ &= r\log[1 - z_1] - r\log[1 + \log[1 - z_1]], \end{aligned} \tag{4.65}$$

where it attains a positive value. The comparison of the performance of $\widehat{\theta}_{S2}$ and $\widehat{\theta}_{L2}$ can now be carried out in terms of the minimum risks (4.62) and (4.65)

$$\begin{aligned} & R_L(\widehat{\theta}_{S2}, \theta)_{\min} - R_L(\widehat{\theta}_{L2}, \theta)_{\min} \\ = \; & -r \left(\frac{2a}{r+\alpha+1} + \log\left[\frac{r+\alpha+1-a}{r+\alpha+1} \right] - 1 + \exp\left[-\frac{a}{r+\alpha+1} \right] \right). \end{aligned} \tag{4.66}$$

Table 4.10 summarises the results of the difference of the minimum of two risk functions for various r, α and a.

Table 4.10: Difference of the minimum risks relative to L_L using Δ_2.

	(r,α)				
a	$(3, 0.001)$	$(4, 0.5)$	$(5, 2)$	$(7, 8)$	$(30000, 300000)$
-0.001	$-2.34E-11$	$-1.20E-11$	$-4.88E-12$	$-8.55E-13$	$6.66E-12$
-0.1	$-2.32E-05$	$-1.19E-05$	$-4.86E-06$	$-8.52E-07$	0
-3	$-5.29E-01$	$-2.79E-01$	$-1.17E-01$	$-2.16E-02$	$-1.33E-11$
0.001	$2.34E-11$	$1.20E-11$	$4.88E-12$	$8.54E-13$	$-3.33E-12$
0.1	$2.37E-05$	$1.21E-05$	$4.91E-06$	$8.57E-07$	$-3.33E-12$
3	$2.33E-14$	$1.20E-14$	$5.00E-15$	$7.77E-16$	0

Table 4.10 shows that relative to the LINEX loss function, when defined in terms of type II estimation error, none of the two estimators dominates the other uniformly, irrespective of the sign of the LINEX parameter. It should also be noted that, the numerical computation was executed for more parameters, than given in the table. In this situation, an important idea is to investigate the sensitivity of the value of the LINEX parameter by changing the number of fully measured components, r. In general, from the pair-wise comparisons of the estimators, it could be observed that more often the correct estimator outperforms a wrongly chosen alternative. However, there are situations where none is superior than the other, i.e. crossing-over risks are obtained. For this reason, a further investigation on the sufficient or necessary conditions of admissibility or inadmissibility might always be an important idea. The risk function of the estimator $\widehat{\theta}_{S1}$ can also be obtained relative to the LINEX loss function, when defined in terms of the type II estimation error. Using (4.47), the corresponding risk function is given by

$$\begin{aligned} R_L(\widehat{\theta}_{S1},\theta) &= \int_0^\infty \left(\exp\left[a\left(\frac{\theta}{\widehat{\theta}_{S1}}-1\right)\right] - a\left(\frac{\theta}{\widehat{\theta}_{S1}}-1\right)-1\right) q(T|\theta)dT \\ &= \exp\left[-a+\frac{a\beta\theta}{r+\alpha}\right]\int_0^\infty \exp\left[\frac{a\theta T}{r+\alpha}\right] q(T|\theta)dT - \frac{ar}{r+\alpha} - \frac{a\beta}{r+\alpha} + a - 1 \\ &= \exp\left[-a+\frac{a\beta\theta}{r+\alpha}\right]\left(\frac{r+\alpha}{r+\alpha-a}\right)^r - \frac{ar}{r+\alpha} - \frac{a\beta\theta}{r+\alpha} + a - 1. \qquad (4.67) \end{aligned}$$

Hence, partially differentiating (4.67) yields

$$\frac{\partial R_L(\widehat{\theta}_{S1},\theta)}{\partial\theta} = \exp\left[-a+\frac{a\beta\theta}{r+\alpha}\right]\left(\frac{r+\alpha}{r+\alpha-a}\right)^r \frac{a\beta}{r+\alpha} - \frac{a\beta}{r+\alpha},$$

which by equating to zero and solving for θ gives the point at which the minimum is attained, namely

$$\theta = \frac{r+\alpha}{a\beta}\left(a + r\log\left[\frac{r+\alpha-a}{r+\alpha}\right]\right).$$

Substituting this in (4.67) yields the corresponding minimum risk

$$R_L(\widehat{\theta}_{S1},\theta)_{\min} = -\frac{ar}{r+\alpha} - r\log\left[\frac{r+\alpha-a}{r+\alpha}\right]. \tag{4.68}$$

The performance of $\widehat{\theta}_{S1}$ and $\widehat{\theta}_{L2}$ can now be compared in terms of the minimum risks numerically using (4.62) and (4.68)

$$\begin{aligned} & R_L(\widehat{\theta}_{S1},\theta)_{\min} - R_L(\widehat{\theta}_{L2},\theta)_{\min} \\ = & -r\left(\frac{a}{r+\alpha+1} + \frac{a}{r+\alpha} + \log\left[\frac{r+\alpha-a}{r+\alpha}\right] - 1 + \exp\left[-\frac{a}{r+\alpha+1}\right]\right). \end{aligned}$$

Choosing values for the unknowns r, α and a and calculating the differences in minimum risks yield the summary in table 4.11.

Table 4.11: Difference of the risk functions of $\widehat{\theta}_{S1}$ and $\widehat{\theta}_{L2}$ relative to L_L.

	(r,α)				
a	$(3, 0.001)$	$(4, 0.5)$	$(5, 2)$	$(7, 8)$	$(30000, 3000)$
-0.001	$7.28E-08$	$3.26E-08$	$1.20E-08$	$1.88E-09$	0
-0.1	$6.85E-04$	$3.08E-04$	$1.13E-04$	$1.79E-05$	$1.33E-11$
-3	$-1.80E-01$	$-9.64E-02$	$-4.05E-02$	$-7.36E-03$	$-3.76E-09$
0.001	$7.29E-08$	$3.27E-08$	$1.20E-08$	$1.88E-09$	$-3.33E-12$
0.1	$7.74E-04$	$3.45E-04$	$1.26E-04$	$1.98E-05$	$6.66E-12$
3	$2.04E+01$	$1.23E+00$	$3.44E-01$	$4.63E-02$	$1.88E-08$

Table 4.11 shows that over the allowable range of the parameters, none of the estimators dominates the other uniformly. If the LINEX parameter is positive, more often the correct estimator outperforms the other alternative except when r is extremely large.

4.3.2.2 Posterior risks

The same procedure will be employed as in section 4.2. First, the general expression of the posterior risk for any estimator $\widehat{\theta}$ will be derived and then the respective posterior risks. Using (2.24), (2.34) and (4.46), the posterior risk of any estimator $\widehat{\theta}$ is given by

$$\begin{aligned} R_1^S(\widehat{\theta}) &= E_{post}\left[L_S(\widehat{\theta},\theta)\right] = \frac{(T+\beta)^{r+\alpha}}{\Gamma(r+\alpha)}\int_0^\infty(\widehat{\theta}-\theta)^2\theta^{r+\alpha-1}\exp[-(T+\beta)\theta]d\theta \\ &= \widehat{\theta}^2 - \frac{2(r+\alpha)\widehat{\theta}}{T+\beta} + \frac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2}. \end{aligned} \tag{4.69}$$

Substituting (4.47) in (4.69), the posterior risk of the Bayesian estimator $\widehat{\theta}_{S1}$ is given by

$$R_1^S(\widehat{\theta}_{S1}) = \frac{r+\alpha}{(T+\beta)^2}, \tag{4.70}$$

which is equivalent to the variance of the posterior distribution. Considering the situation when one uses the estimator $\widehat{\theta}_{L1}$ for $\widehat{\theta}$ in (4.69), the consequent posterior risk is

$$R_1^S(\widehat{\theta}_{L1}) = \left(\tfrac{r+\alpha}{a}\log\left[\tfrac{T+\beta+a}{T+\beta}\right]\right)^2 - \tfrac{2(r+\alpha)}{a(T+\beta)}\log\left[\tfrac{T+\beta+a}{T+\beta}\right] + \tfrac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2}. \tag{4.71}$$

Similarly, the posterior risks relative to the LINEX loss function are obtained. Using (2.28) and (4.46), the posterior risk of any estimator $\widehat{\theta}$ is

$$\begin{aligned} R_1^L(\widehat{\theta}) &= E_{post}\left[L_L(\widehat{\theta},\theta)\right] = \frac{(T+\beta)^{r+\alpha}}{\Gamma(r+\alpha)}\int_0^\infty L_L(\widehat{\theta}-\theta)\theta^{r+\alpha-1}\exp[-(T+\beta)\,\theta]d\theta \\ &= \exp[-a\widehat{\theta}]\left(\frac{T+\beta}{T+\beta+a}\right)^{r+\alpha} - a\widehat{\theta} + \frac{a(r+\alpha)}{T+\beta} - 1. \end{aligned} \tag{4.72}$$

Using (4.49), the respective posterior risk of $\widehat{\theta}_{L1}$ is given by

$$R_1^L(\widehat{\theta}_{L1}) = \frac{a(r+\alpha)}{T+\beta} - (r+\alpha)\log\left[\frac{T+\beta+a}{T+\beta}\right]. \tag{4.73}$$

If the estimator (4.47) is used in (4.72), the respective posterior risk is given by

$$R_1^L(\widehat{\theta}_{S1}) = \exp\left[\frac{a(r+\alpha)}{T+\beta}\right]\left(\frac{T+\beta}{T+\beta+a}\right)^{r+\alpha} - 1. \tag{4.74}$$

An alternative approach is to derive the posterior risks using the type II estimation error. Hence, using (4.46) and (4.48), the respective posterior risk of any estimator $\widehat{\theta}$, is given by

$$\begin{aligned} R_2^S(\widehat{\theta}) &= E_{post}\left[L_S(\widehat{\theta},\theta)\right] = \frac{(T+\beta)^{r+\alpha}}{\Gamma(r+\alpha)}\int_0^\infty \left(\frac{\theta}{\widehat{\theta}}-1\right)^2 \theta^{r+\alpha-1}\exp[-(T+\beta)\,\theta]d\theta \\ &= \frac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2\widehat{\theta}^2} - \frac{2(r+\alpha)}{(T+\beta)\widehat{\theta}} + 1. \end{aligned} \tag{4.75}$$

Substituting (4.48) in (4.75), the posterior risk of the Bayesian estimator $\widehat{\theta}_{S2}$ is given by

$$R_2^S(\widehat{\theta}_{S2}) = \frac{1}{r+\alpha+1}. \tag{4.76}$$

If the estimator $\widehat{\theta}_{L2}$ is used, the posterior risk is obtained by substituting (4.50) in (4.75) to obtain

$$R_2^S(\widehat{\theta}_{L2}) = \frac{(r+\alpha)(r+\alpha+1)z_1^2}{a^2} - \frac{2rz_1}{a} + 1. \tag{4.77}$$

The same derivation of the posterior risk is performed relative to the LINEX loss function. Hence, using (2.28) and (4.46), the posterior risk of any given estimator, $\widehat{\theta}$, is given by

$$\begin{aligned} R_2^L(\widehat{\theta}) &= E_{post}\left[L_L(\widehat{\theta},\theta)\right] = \frac{(T+\beta)^{r+\alpha}}{\Gamma(r+\alpha)}\int_0^\infty L_L\left(\frac{\theta}{\widehat{\theta}}-1\right)\theta^{r+\alpha-1}\exp[-(T+\beta)\theta]d\theta \\ &= \frac{(T+\beta)^{r+\alpha}}{\Gamma(r+\alpha)}\exp[-a]\int_0^\infty \theta^{r+\alpha-1}\exp\left[-\left(T+\beta-\frac{a}{\widehat{\theta}}\right)\theta\right]d\theta \\ &\quad -\frac{a}{\widehat{\theta}}\int_0^\infty \theta^{r+\alpha-1}\exp[-(T+\beta)\theta]d\theta + a - 1 \\ &= \exp[-a]\left(\frac{(T+\beta)\,\widehat{\theta}}{(T+\beta)\,\widehat{\theta}-a}\right)^{r+\alpha} - \frac{a(r+\alpha)}{(T+\beta)\,\widehat{\theta}} + a - 1. \end{aligned} \tag{4.78}$$

Substituting $\widehat{\theta}_{L2}$ from (4.50) in (4.78), gives the consequent posterior risk, which is equivalent to

$$R_2^L(\widehat{\theta}_{L2}) = \exp[-a]\left(\frac{1}{1-z_1}\right)^{r+\alpha} - (r+\alpha)z_1 + a - 1. \tag{4.79}$$

The posterior risk corresponding to $\widehat{\theta}_{S2}$ relative to the LINEX loss function is obtained in similar fashion by substituting (4.48) in (4.78) to obtain

$$R_2^L(\widehat{\theta}_{S2}) = \exp[-a]\left(\frac{r+\alpha+1}{r+\alpha+1-a}\right)^{r+\alpha} - a\left(\frac{r+\alpha}{r+\alpha+1}\right) + a - 1. \tag{4.80}$$

Similarly, substituting the estimator given by (4.47) in (4.78), the posterior risk is

$$R_2^L(\widehat{\theta}_{S1}) = \exp[-a]\left(\frac{r+\alpha}{r+\alpha-a}\right)^{r+\alpha} - 1, \quad r+\alpha > a. \tag{4.81}$$

A summary of the estimators and their posterior risks haven been provided in table 4.12.

Table 4.12: Summary of estimators and their respective posterior risks

Estimator	R_1^S	R_1^L
$\widehat{\theta}$	$\widehat{\theta}^2 - \frac{2(r+\alpha)\widehat{\theta}}{T+\beta} + \frac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2}$	$\exp[-a\widehat{\theta}]\left(\frac{T+\beta}{T+\beta+a}\right)^{r+\alpha} - a\widehat{\theta}$ $+\frac{a(r+\alpha)}{T+\beta} - 1$
$\widehat{\theta}_{S1} = \frac{r+\alpha}{T+\beta}$	$\frac{r+\alpha}{(T+\beta)^2}$	$\exp\left[\frac{ar}{T}\right]\left(\frac{T}{T+a}\right)^r - 1$
$\widehat{\theta}_{L1} = \frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right]$	$\left(\frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right]\right)^2 - \frac{2(r+\alpha)^2}{a(T+\beta)}$ $\times\log\left[\frac{T+\beta+a}{T+\beta}\right] + \frac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2}$	$\frac{a(r+\alpha)}{T+\beta} - (r+\alpha)\log\left[\frac{T+\beta+a}{T+\beta}\right]$
Estimator	R_2^S	R_2^L
$\widehat{\theta}$	$\frac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2\widehat{\theta}^2} - \frac{2(r+\alpha)}{(T+\beta)\widehat{\theta}} + 1$	$\exp[-a]\left(\frac{(T+\beta)\widehat{\theta}}{(T+\beta)\widehat{\theta}-a}\right)^{r+\alpha}$ $-\frac{a(r+\alpha)}{(T+\beta)\widehat{\theta}} + a - 1$
$\widehat{\theta}_{S2} = \frac{r+1}{T}$	$\frac{1}{r+\alpha+1}$	$\exp[-a]\left(\frac{r+\alpha+1}{r+\alpha+1-a}\right)^{r+\alpha}$ $-a\left(\frac{r+\alpha}{r+\alpha+1}\right) + a - 1$
$\widehat{\theta}_{L2} = \frac{a}{Tz}$	$\frac{(r+\alpha)(r+\alpha+1)z_1^2}{a^2} - \frac{2rz_1}{a} + 1$	$\exp[-a]\left(\frac{1}{1-z_1}\right)^{r+\alpha}$ $-(r+\alpha)z_1 + a - 1$
$\widehat{\theta}_{S1} = \frac{r+\alpha}{T+\beta}$	$\frac{1}{r+\alpha}$	$\exp[-a]\left(\frac{r+\alpha}{r+\alpha-a}\right)^{r+\alpha} - 1$

Once the posterior risks of all the estimators are obtained, comparison of the performance of the estimators is carried out in order for a decision with minimum risk to be made. Hence, a pair-wise comparison of the estimators is performed obtaining the differences of the posterior risks. Using (4.70) and (4.71)

$$\begin{aligned}
& R_1^S(\widehat{\theta}_{L1}) - R_1^S(\widehat{\theta}_{S1}) \\
= & \left(\frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right]\right)^2 - \frac{2(r+\alpha)^2}{a(T+\beta)}\log\left[\frac{T+\beta+a}{T+\beta}\right] + \frac{(r+\alpha)(r+\alpha+1)}{(T+\beta)^2} \\
& -\frac{r+\alpha}{(T+\beta)^2},
\end{aligned}$$

which simplifies to

$$\left(\frac{r+\alpha}{a}\log\left[\frac{T+\beta+a}{T+\beta}\right] - \frac{r+\alpha}{T+\beta}\right)^2 \geqslant 0. \tag{4.82}$$

Therefore, from (4.82), it can be easily observed that with respect to the squared error loss function, $\widehat{\theta}_{S1}$ always performs better than the wrongly chosen estimator, $\widehat{\theta}_{L1}$. Hence, if a decision-maker chooses $\widehat{\theta}_{L1}$, instead, a greater risk will be incurred.

The performance comparison of the estimators is also done when the type II estimation error is used. Hence from (4.76) and (4.77), the difference is

$$\begin{aligned} R_2^S(\widehat{\theta}_{L2}) - R_2^S(\widehat{\theta}_{S2}) &= \frac{(r+\alpha)(r+\alpha+1)z_1^2}{a^2} - \frac{2rz_1}{a} - \frac{1}{r+\alpha+1} + 1 \\ &= \frac{(r+\alpha)(r+\alpha+1)z_1^2}{a^2} - \frac{2(r+\alpha)z_1}{a} + \frac{r+\alpha}{r+\alpha+1}, \quad (4.83) \end{aligned}$$

and dividing by $(r+\alpha)$ reduces to

$$\left(\frac{z_1}{a} - \frac{1}{r+\alpha+1}\right)\left(\frac{(r+\alpha+1)z_1}{a} - 1\right). \quad (4.84)$$

It can be easily shown that the two factors of the expression (4.84) have the same sign. Hence, (4.83) is non-negative which in turn implies that the Bayesian estimator $\widehat{\theta}_{S2}$ is always superior to $\widehat{\theta}_{L2}$ relative to the squared error loss function. It is, therefore, inevitable for one to incur a greater risk, if a wrong estimator is used in the decision-making process, under the squared error loss function.

Similar comparisons can be performed when using the LINEX loss function. Hence using (4.74) and (4.73), the difference is

$$\begin{aligned} & R_1^L(\widehat{\theta}_{S1}) - R_1^L(\widehat{\theta}_{L1}) \\ &= \exp\left[\frac{a(r+\alpha)}{T+\beta}\right]\left(\tfrac{T+\beta}{T+\beta+a}\right)^{r+\alpha} - \frac{a(r+\alpha)}{T+\beta} + (r+\alpha)\log\left[\tfrac{T+\beta+a}{T+\beta}\right] - 1. \quad (4.85) \end{aligned}$$

The same techniques used in (4.39)-(4.41), yield

$$R_1^L(\widehat{\theta}_{S1}) = \exp\left[R_1^L(\widehat{\theta}_{L1})\right] - 1,$$

implying that (4.85) is non-negative and entails that $R_1^L(\widehat{\theta}_{S1})$ is uniformly greater than $R_1^L(\widehat{\theta}_{L1})$.

Similarly, if the LINEX loss function is defined in terms of the type II estimation error, the performance of the two estimators can be compared, using (4.79) and (4.80)

$$\begin{aligned} & R_2^L(\widehat{\theta}_{S2}) - R_2^L(\widehat{\theta}_{L2}) \\ &= \exp[-a]\left\{\left(\tfrac{r+\alpha+1}{r+\alpha+1-a}\right)^{r+\alpha} - \left(\frac{1}{1-z_1}\right)^{r+\alpha}\right\} + (r+\alpha)z_1 - a\left(\frac{r+\alpha}{r+\alpha+1}\right). (4.86) \end{aligned}$$

Various values of r, a and α are used to evaluate (4.86) and the results are given in table 4.13.

Table 4.13: Difference of the posterior risks relative to L_L using Δ_2.

	(r,α)			
a	$(3,2)$	$(4,4)$	$(5,8)$	$(7,12)$
-0.001	$3.20E-15$	$3.06E-15$	$3.59E-15$	$3.55E-15$
-0.1	$2.88E-07$	$1.37E-07$	$5.91E-08$	$2.96E-08$
-1	$2.82E-03$	$1.35E-03$	$5.86E-04$	$2.94E-04$
-2	$4.64E-02$	$2.20E-02$	$9.48E-03$	$4.75E-03$
0.01	$2.90E-11$	$1.37E-11$	$5.92E-12$	$2.97E-12$
0.5	$1.88E-04$	$8.78E-05$	$3.76E-05$	$1.88E-0$
1	$3.18E-03$	$1.46E-03$	$6.15E-04$	$3.05E-04$
10	$1.75E-02$	$7.75E-03$	$3.21E-03$	$1.57E-03$

Table 4.13 shows that the two posterior risk evaluated at the correct Bayesian estimator dominates the other uniformly.

In most of the situations, the pair-wise comparisons of the performance of the various estimators entails that the correct estimators perform better than the other alternatives. However, some crossing-over risk functions are obtained. Therefore, a decision-maker who is encountered with these scenarios need to study meticulously the consequences of every choice.

4.3.2.3 Integrated risks

The integrated risk is the expectation of the risk function with respect to the prior distribution under consideration. In this section, given the conjugate prior in (4.9), the integrated risk for every risk function is defined, but not all have been obtained in closed-form. The two ways of defining the integrated risk, (2.35) and (2.36) will be interchangeably used depending on the situation.

Using (4.46) and (4.51), the integrated risk relative to the squared error loss function defined in terms of the type I estimation resulted in expressions that are not in closed-form solutions. The comparisons can, however, be made if the type II estimation error is considered.

Using (4.53) and (4.54), the integrated risk of the estimators $\widehat{\theta}_{S2}$ and $\widehat{\theta}_{L2}$ relative to the squared error loss function are

$$
\begin{aligned}
r_S(\widehat{\theta}_{S2}) &= \int_0^\infty \left[3r^2\eta^2 + 3r\eta^2 + 4r\eta^2\beta\theta + \eta^2\beta^2\theta^2 + 2r\alpha\eta^2 + 2\alpha\eta^2\beta\theta + 2\eta^2\beta\theta + 1\right] \pi(\theta|\alpha,\beta)d\theta \\
&= 3r^2\eta^2 + 3r\eta^2 + 4r\alpha\eta^2 + \eta^2(\alpha^2+\alpha) + 2r\eta^2\alpha + 2\alpha^2\eta^2 + 2\alpha^2\eta^2 + 1 \qquad (4.87)
\end{aligned}
$$

and

$$
\begin{aligned}
r_S(\widehat{\theta}_{L2}) &= \int_0^\infty \left[\frac{z_1^2}{a^2}\left[r^2 + r + 2\beta\theta r + \beta^2\theta^2\right] + \frac{2z_1}{a}(r+\beta\theta) + 1\right] \pi(\theta|\alpha,\beta)d\theta \\
&= \frac{z_1^2}{a^2}\left[r^2 + r + 2\alpha^2 r + (\alpha^2+\alpha)\right] + \frac{2z_1}{a}(r+\alpha) + 1, \qquad (4.88)
\end{aligned}
$$

respectively. The estimators $\widehat{\theta}_{S2}$ and $\widehat{\theta}_{L2}$ can now be compared. Hence, the difference of the two expression yields

$$
\begin{aligned}
r_S(\widehat{\theta}_{L2}) - r_S(\widehat{\theta}_{S2}) &= \frac{z_1^2}{a^2}\left[r^2 + r + 2r\alpha^2 + (\alpha^2+\alpha)\right] + \frac{2(r+\alpha)z_1}{a} - \\
&\quad \eta^2\left(3r^2 + 3r + 4r\alpha + (\alpha^2+\alpha) + 2r\alpha + 2\alpha^2 + 2\alpha^2\right). \qquad (4.89)
\end{aligned}
$$

The expression (4.89) can further be rewritten as

$$
\left\{\frac{z_1}{a} - \frac{1}{r+\alpha+1}\right\} \times
\left[\left(\frac{z_1}{a} + \frac{1}{r+\alpha+1}\right)\left(r^2 + r + 2r\alpha + \beta^2(\alpha^2+\alpha)\right) + 2(r+\alpha)\right]. \qquad (4.90)
$$

Whether the expression (4.90) is negative or positive depends again on the sign of the leading factor (as in the case of the risk functions mentioned before), since the second product is always non-negative. Hence, the correct Bayesian estimator, $\widehat{\theta}_{S2}$, outperforms $\widehat{\theta}_{L2}$ with respect to the Bayes risk, if $a < 0$, otherwise $\widehat{\theta}_{L2}$ is superior to $\widehat{\theta}_{S2}$. Therefore, within the allowable range of the hyperparameters of gamma distribution and for a negative LINEX parameter, $\widehat{\theta}_{S2}$ uniformly dominates the other relative to the Bayes risk when evaluated relative to the squared error loss function. This was confirmed using numerical computation for various parameters. The table provided below is a sample of the results obtained from EXCEL sheet computation. The table also illustrates for these particular values of the parameters, where $a > 0$, the wrong alternative is superior to the correct Bayesian estimator while the correct estimator outperforms the other when $a < 0$. Table 4.14 summarises the result for various values of r, a, α and β.

Table 4.14: Difference of the integrated risk relative to L_S using Δ_2.

	(r,α,β)				
a	$(2,0.5,1)$	$(3,2,3)$	$(4,4,10)$	$(7,13,25)$	$(100,0.001,0.01)$
-0.00001	$3.21E-06$	$4.07E-06$	$2.84E-05$	$1.23E-04$	$9.79E-08$
-0.01	$3.21E-03$	$4.08E-03$	$2.84E-02$	$1.23E-01$	$1.00E-04$
-200	$1.48E+46$	$1.97E+26$	$1.04E+18$	$5.34E+08$	$9.03E+00$
0.5	$-1.48E-01$	$-1.94E-01$	$-1.37E+00$	$-6.07E+00$	$-4.98E-03$
1	$-2.73E-01$	$-3.70E-01$	$-2.66E+00$	$-1.20E+01$	$-9.94E-03$
100	$-1.12E+00$	$-2.44E+00$	$-2.53E+01$	$-2.47E+02$	$-6.03E-01$
1000	$-1.12E+00$	$-2.44E+00$	$-2.55E+01$	$-2.58E+02$	$-9.99E-01$

The integrated risks of these estimators relative to the LINEX loss function is also another point of interest. Hence, from (4.59) and (4.63) these integrated risks of $\widehat{\theta}_{L2}$ and $\widehat{\theta}_{S2}$ are

$$\begin{aligned} r_L(\widehat{\theta}_{L2}) &= \int_0^\infty \left[\exp[-a+z_1\beta\theta]-z_1\beta\theta-rz_1+a-1\right]\pi(\theta|\alpha,\beta)d\theta \\ &= \exp[-a]\left(\frac{1}{1-z_1}\right)^{r+\alpha}-(r+\alpha)z_1+a-1 \end{aligned} \tag{4.91}$$

and

$$\begin{aligned} r_L(\widehat{\theta}_{S2}) &= \int_0^\infty \left[\exp\left[-a+\frac{a\beta\theta}{r+\alpha+1}\right]\left(\frac{r+\alpha+1}{r+\alpha+1-a}\right)^r \pi(\theta|\alpha,\beta)d\theta\right] \\ &\quad -\int_0^\infty \left[\frac{ar}{(r+\alpha+1)}-\frac{a\beta\theta}{r+\alpha+1}+a-1\right]\pi(\theta|\alpha,\beta)d\theta \\ &= \exp[-a]\left(\frac{r+\alpha+1}{r+\alpha+1-a}\right)^{r+\alpha}-\frac{a(r+\alpha)}{(r+\alpha+1)}+a-1 \end{aligned} \tag{4.92}$$

respectively, which are also the same as (4.79) and (4.80) respectively and it was already shown the two posterior risks cross-over each other.

4.4 THE SURVIVAL FUNCTION

The reliability of a system is expressed by means of the survival function. Hence, estimating the survival of a system or a component in a system has attained a crucial place in the reliability study for improvement of productivity or other related areas. The Jeffrey's prior of the survival function of the exponential model and the corresponding posterior distribution have been derived in the next section. Furthermore, Bayesian estimators have also been derived relative to the squared and LINEX loss functions.

4.4.1 The Jeffreys' prior and posterior analysis

Suppose now, using (2.8), the survival function of the exponential model at any time point t, is given by

$$S(t) = \exp[-\theta t] = \nu.$$

Hence, solving for θ and substituting the result, the corresponding exponential model in terms of the survival function is given by

$$f(x|\nu) = \frac{(-\log[\nu])}{t} \exp\left[\frac{x}{t}\log[\nu]\right]. \tag{4.93}$$

Therefore, for a sequence of lifetimes, the corresponding likelihood function in terms of the survival function ν, is given by

$$\begin{aligned} \ell(data|\nu) &\propto \left(\frac{(-\log[\nu])}{t}\right)^r \exp\left[-\sum_{i=1}^{r}\frac{x_i}{t}\log[\nu]\right] \\ &= \left(\frac{(-\log[\nu])}{t}\right)^r \nu^{\frac{T}{t}}, \end{aligned}$$

with

$$T = \sum_{i=1}^{n} x_i.$$

The Jeffreys' prior for the survival function, ν,can be obtained from (4.93) as follows. Suppose

$$\log\left[f(x|\nu)\right] = \log\left[(-\log[\nu])\right] - \log[t] + \frac{x}{t}\log[\nu] \equiv K.$$

Now,

$$\frac{\partial K}{\partial \nu} = \frac{1}{\nu\log[\nu]} + \frac{x}{t\nu}$$

and

$$\frac{\partial^2 K}{\partial \nu^2} = -\frac{(1+\log[\nu])}{\nu^2\log^2[\nu]} - \frac{x}{t\nu^2}.$$

Hence,

$$I_{11}(\theta) = -E_\theta\left[\frac{\partial^2\{\log[f\,(x|\nu)]\}}{\partial\nu^2}\right] = \frac{(1+\log[\nu])}{\nu^2\log^2[\nu]} + E_\nu\left[\frac{x}{t\nu^2}\right],$$

the latter expression is given by

$$E_\nu\left[\frac{x}{t\nu^2}\right] = \int_0^\infty \frac{x}{t\nu^2}\frac{(-\log[\nu])}{t}\exp\left[\frac{x}{t}\log[\nu]\right]dx.$$

Letting

$$u = \frac{x}{t}(-\log[\nu])$$

results in

$$E_\nu\left[\frac{x}{t\nu^2}\right] = \frac{1}{\nu^2(-\log[\nu])}\int_0^\infty u\exp[-u]du$$
$$= \frac{1}{\nu^2(-\log[\nu])}.$$

Therefore, the Fisher information matrix is given by

$$I_{11}(\theta) = \frac{(1+\log[\nu])}{\nu^2\log^2[\nu]} - \frac{1}{\nu^2\log[\nu]} = \frac{1}{\nu^2(-\log[\nu])^2},$$

from which the Jeffreys' prior for the survival function is given by

$$\pi(\nu) \propto \frac{1}{\nu(-\log[\nu])}. \tag{4.94}$$

The corresponding posterior distribution of ν is obtained as

$$p(\nu|data) \propto (-\log[\nu])^{r-1}\,\nu^{\frac{T}{t}-1}. \tag{4.95}$$

The normalising constant in (4.95) (using transformation $u = -\log[\nu]$) is equal to

$$\int_0^1 (-\log[\nu])^{r-1}\,\nu^{\frac{T}{t}-1}d\nu = \int_0^\infty u^{r-1}\exp[-uw]du = \frac{\Gamma(r)}{w^r},$$

where

$$w = \frac{T}{t}.$$

Hence, the posterior (4.95) becomes

$$p(\nu|data) = \frac{(-\log[\nu])^{r-1}\,\nu^{w-1}w^r}{\Gamma(r)}. \tag{4.96}$$

It can be easily shown that

$$E_{post}[\nu^k] = \left(\frac{w}{k+w}\right)^r. \tag{4.97}$$

Once the posterior distribution of the survival function, ν, is obtained, it is now possible to obtain the Bayesian estimators based on the results given in chapter 2. Hence, using (4.97) where $k = 1$, the Bayesian estimator relative to the squared error loss function, when defined in terms of the type I estimation error, is

$$\widehat{\nu}_{S1} = \left(\frac{w}{1+w}\right)^r.$$

The corresponding Bayesian estimator relative to the squared error loss function, when it is defined in terms of the type II estimation error, is

$$\widehat{\nu}_{S2} = \frac{E_{post}[\nu^2]}{E_{post}[\nu]} = \left(\frac{1+w}{2+w}\right)^r.$$

Similarly, using (2.30) and (4.96), the Bayesian estimator relative to the LINEX loss function, when it is defined in terms of the type I estimation error, is

$$\widehat{\nu}_{L1} = -\frac{1}{a}\log[E_{post}[\exp[-a\nu]]]. \tag{4.98}$$

Now, using the Taylor series expansion about zero for $\exp[-a\nu]$ and applying the result in (4.97), (4.98) reduces to

$$\widehat{\nu}_{L1} \approx -\frac{1}{a}\log\left[\sum_{i=0}^{\infty}(-a)^i\left(\frac{w}{i+w}\right)^r\right].$$

The Bayesian estimator relative to the LINEX loss function, when defined in terms of the type II estimation error is not obtainable.

4.4.2 Posterior risk

The posterior risk relative to the squared error loss function when the type I estimation error is used, is always equal to the variance of the posterior whenever it exists. Hence in this case, it is given by

$$\begin{aligned} R_1^S(\widehat{\nu}_{S1}) &= E_{post}\left[\nu^2\right] - E_{post}^2[\nu] \\ &= \left(\frac{w}{2+w}\right)^r - \left(\frac{w}{1+w}\right)^{2r} \end{aligned}$$

in the range where it attains non-negative. The risk functions, however, do not exist.

4.5 CONCLUSION

The Jeffreys' prior θ and the survival functions ν of the exponential model, as well as their respective posterior distributions were derived. Furthermore, the conjugate prior distribution was used in the derivation of the posterior and subsequent analyses. The Bayesian estimators of the parameter and the survival function of the exponential model were derived, relative to the squared error loss and the LINEX loss functions. Moreover, two forms of the estimation error were used in defining the loss functions. Risk measures were obtained to establish the robustness of estimators used under different loss functions.

Relative to the squared error loss function, none of the estimators uniformly dominates the other with respect to the risk function. There are, however, situations where the correct estimator dominates the other and vice versa. In such situations, some sufficient conditions of inadmissibility or admissibility might be of importance.

Relative to the LINEX loss function, however, it was observed that the correct estimators are superior to the other alternatives, with respect to the risk function, when the Jeffreys' prior is considered. However, when the conjugate prior is considered, there are situations where none of the estimators uniformly dominates the other. Furthermore, even though not proved analytically, for different values considered, the correct estimator $\widehat{\theta}_{L2}$ performed better than $\widehat{\theta}_{S1}$. From the numerical computations performed, it could also be observed that more often, a wrongly chosen alternative leads to a greater risk and this is quite plausible, since at a risk analysis stage, one expects to use the same estimator that is obtained relative to the loss function under consideration.

Similar pair-wise performance comparisons were done using posterior risks. In all these comparisons, it was shown that the correct Bayesian estimator outperforms the alternative.

CHAPTER 5

THE WEIBULL MODEL

5.1 INTRODUCTION

In the previous chapter, the Bayesian analyses of the univariate exponential model was performed. It is known that the exponential distribution is a special case of the two-parameter Weibull distribution, where the shape parameter is taken as unity. In this chapter, the same analyses are done for a general Weibull model. Two situations are considered where, only the scale parameter is unknown and secondly, where both the scale and shape parameters are unknown.

In many practical problems of reliability, such as in electrical components and complex mechanisms, the Weibull distribution is known to sufficiently represent a lifetime distribution (Soland, 1969). The Weibull model is useful mainly in situations where components change stochastically and data is observed from repairable systems (Ananda & Ravindra, 1998). Quite often, the parameters underlying the Weibull model cannot be known beforehand especially when insufficient data is available. A Bayesian approach can ease this problem at hand. Many authors performed a Bayesian analysis where the shape parameter is known. Yao and William (2002) described the life-test plan for type II censored data. In their analysis,a conjugate prior was used. Evans and Nigm (1980) also performed an approximate Bayesian approach to obtain predictive bounds for failure times. They devised methods of approximating the remaining components of a sample, whose lifetimes follow a two parameter Weibull pdf and the predictive bounds of a similar, but unused component. Dellaportas and Wright (1991) also carried out a numerical approach of evaluating a posterior expectation using the Weibull pdf.

The remainder of the chapter is schemed as follows. In section 5.2, the Jeffreys' prior for the scale parameter is derived treating the shape parameter as known. The Bayesian estimators and their respective risk measures are also derived. Section 5.3 is devoted to the analogous derivation of the Jeffreys' prior, as well as the reference prior for the Weibull model when both parameters are unknown. Moreover, the computational difficulties of the posterior analysis for the general case of the Weibull model is also discussed. Section 5.4 deals with the derivation of the Jeffreys' prior of the survival function where the shape parameter is treated as a nuisance parameter, as well as in the case when it is the parameter of interest. Section 5.5 concludes the chapter.

5.2 SCALE PARAMETER UNKNOWN

In many practical reliability problems, the Weibull model with only the scale parameter unknown is employed for analysis (Soland, 1968). The shape parameter is often considered as less important and thus may be approximated and treated as known. Another reason for assuming a known shape parameter is, since in some reliability analyses, the shape parameter might be known from previous test experience or sometimes, decision-makers might as a matter of policy, treat it as a constant.

Given its wide application in the reliability theory, the posterior analysis of the Weibull model with only the scale parameter unknown is considered in this chapter. The two functional forms of the Weibull model are considered to derive the Bayesian estimators, posterior distribution, risk measures and thereby used for performance evaluations.

The first stage in the posterior analysis is to derive the Jeffreys' prior for the parameter(s) of interest. In this case, the scale parameter θ, is considered as the parameter of interest while the shape parameter is treated as a nuisance parameter or known. Therefore, the Jeffrey's prior with λ known, using (2.11) and (2.14) is obtained from

$$I_{11}(\theta) = -E_\theta\left[\frac{\partial^2\{\log[f(x,\lambda|\theta)]\}}{\partial\theta^2}\right] = -E_\theta\left[-\frac{1}{\theta^2}\right] = \frac{1}{\theta^2}.$$

Similarly, the Jeffreys' prior for the Weibull model of the form (2.12), is derived from

$$I_{11}(\theta) = -E_\theta\left[\frac{\partial^2\{\log[f(x,\lambda|\theta)]\}}{\partial\theta^2}\right] = -E_\theta\left[-\frac{\lambda}{\theta^2} - \lambda(\lambda-1)x^\lambda\theta^{\lambda-2}\right]. \qquad (5.1)$$

Since

$$E\left[x^{\lambda}\right]=\int_{0}^{\infty}(\theta x)^{\lambda}\lambda x^{\lambda-1}\exp\left[-(\theta x)^{\lambda}\right]dx=\frac{1}{\theta^{\lambda}}, \tag{5.2}$$

(5.2) is equal to

$$I_{11}(\theta)=-\frac{\lambda}{\theta^{2}}-\lambda(\lambda-1)\frac{\theta^{\lambda-2}}{\theta^{\lambda}}\propto\frac{1}{\theta^{2}}.$$

Therefore, using (2.15), the Jeffreys' prior for both Weibull models, (2.11) and (2.12) with λ known, is given by

$$\pi(\theta|\lambda)\propto\frac{1}{\theta}. \tag{5.3}$$

Hence, using the corresponding likelihood functions of (2.11), the posterior distribution under the Jeffreys' prior, is given by

$$p(\theta|data)\propto\theta^{r-1}\exp[-T\theta], \tag{5.4}$$

with a normalising constant of $\frac{T^{r}}{\Gamma(r)}$, which is $\mathcal{G}(r,T)$.

Since λ is known, the one-to-one transformation of $\gamma=\theta^{\lambda}$ results in an equivalent Weibull model of (2.12), which is given by

$$f(x,\lambda|\gamma)=\lambda\gamma x^{\lambda-1}\exp[-(x^{\lambda}\gamma)],\qquad\lambda,\ \gamma>0. \tag{5.5}$$

Therefore, the corresponding Jeffreys' prior is given by

$$\pi(\gamma|\lambda)=\frac{1}{\gamma},$$

which by using the corresponding likelihood function, yields the posterior distribution

$$p(\gamma|data)=\frac{T^{r}\gamma^{r-1}\exp[-T\gamma]}{\Gamma(r)}, \tag{5.6}$$

where

$$T=\sum_{i=1}^{n}x_{i}^{\lambda}.$$

In both cases, the two Weibull models yield equivalent posterior distribution when λ is known. It should also be noted that when λ is known, estimating θ^{λ} and θ and consequently their risk measures, it gives the same results. It can be easily observed that since λ is regarded as a constant, the posterior distribution (4.2) and (5.4), as well as (5.6) are equivalent, except the form of the statistic, T. As a matter of fact, every subsequent analysis regarding the Bayesian estimators and their risk measures are equivalent. Therefore, any risk measure obtained in chapter 4 can analogously be used for comparisons of performance using the Weibull model. It was mentioned in chapter 2 that when the shape parameter $\lambda = 2$, the Weibull model reduces to the Rayleigh pdf. Hence, the analysis using the Rayleigh model can be done by adjusting the consequent term of the statistic T. Similarly, if data pertaining a Rayleigh pdf is observed, a one-to-one transformation of $\gamma = \theta^2$ reduces to (5.6).

5.3 SHAPE AND SCALE PARAMETERS UNKNOWN

In this section, the analysis is extended to the case where both the scale and shape parameters are unknown. The analysis of the two-parameter unknown Weibull model is computationally complicated, both in classical and Bayesian approaches, since there is no simple sufficient statistic for the two parameters. In fact, inferences and predictions are not obtainable in closed-form and hence a numerical procedure is usually necessary. To alleviate the mathematical complexity in the Bayesian analysis, Soland (1969) introduced a method by discretising the shape parameter, while keeping a continuous prior distribution for the shape parameter. Soland (1968) also presents a complete Bayesian analysis of prior, posterior and pre-posterior distributions. Some of the papers on the Bayesian analysis of the two-parameter Weibull model are cited by Dellaportas and Wright (1991). Due to the constraints mentioned in the previous section, the Jeffreys' prior and the reference prior are derived and no further analysis is carried out.

5.3.1 Posterior analysis using the Jeffreys' prior

Consider the Weibull model (2.11) and the Fisher information matrix given by (2.14). The logarithm of the pdf is

$$\log[f(x|\lambda,\theta)] = \log[\lambda] + \log[\theta] + \lambda\log[x] - \log[x] - \theta x^{\lambda} \equiv K. \tag{5.7}$$

From (5.7), the elements of the Fisher information matrix are obtained from

$$\begin{aligned} I_{11} &= -\frac{1}{\theta^2}, \\ I_{12} &= -x^{\lambda}\log[x] \\ I_{22} &= -\frac{1}{\lambda^2} - \theta x^{\lambda}(\log[x])^2. \end{aligned}$$

Hence, applying (2.14) for every element yields,

$$-E[I_{12}] = -E[I_{21}] = \int_0^{\infty} x^{\lambda}\log[x]\lambda\theta x^{\lambda-1}\exp\left[-\theta x^{\lambda}\right]dx. \tag{5.8}$$

Letting

$$u = \theta x^{\lambda},$$

the integral (5.8) reduces to

$$\begin{aligned} -E[I_{12}] &= \frac{1}{\lambda\theta}\int_0^\infty u\log[u]\exp[-u]du - \frac{\log[\theta]}{\lambda\theta}\int_0^\infty u\exp[-u]du \\ &= \frac{\gamma_1}{\lambda\theta} - \frac{\log[\theta]}{\lambda\theta}, \end{aligned}$$

where

$$\gamma_1 = \int_0^\infty u\log[u]\exp[-u]du.$$

Similarly,

$$\begin{aligned} -E[I_{22}] &= E\left[\frac{1}{\lambda^2} + \theta\int_0^\infty x^{\lambda}(\log[x])^2\right] \\ &= \frac{1}{\lambda^2} + \theta\int_0^\infty x^{\lambda}(\log[x])^2\lambda\theta x^{\lambda-1}\exp\left[-\theta x^{\lambda}\right]dx, \end{aligned}$$

and with the same substitution as above, simplifies to

$$\begin{aligned} &\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\int_0^\infty u(\log[u] - \log[\theta])^2\exp[-u]du \\ = &\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\left(\int_0^\infty u^2\log^2[u]\exp[-u]\right)\times \\ &\frac{1}{\lambda^2}\left(-2\log[\theta]\int_0^\infty u\log[u]\exp[-u]du + \log^2[\theta]\int_0^\infty u\exp[-u]du\right) \\ = &\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\left(\gamma_2 - 2\gamma_1\log[\theta] + \log^2[\theta]\right), \end{aligned} \tag{5.9}$$

where

$$\gamma_2 = \int_0^\infty u^2\log^2[u]\exp[-u].$$

From (5.8) and (5.9), the Fisher information matrix is given as

$$I = -E\begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{\theta^2} & \frac{\gamma_1}{\lambda\theta} - \frac{\log[\theta]}{\lambda\theta} \\ \frac{\gamma_1}{\lambda\theta} - \frac{\log[\theta]}{\lambda\theta} & \frac{1}{\lambda^2}\left(1 + \gamma_2 - 2\log[\theta]\gamma_1 + \log^2[\theta]\right) \end{bmatrix},$$

from which the Jeffreys' prior is

$$\pi(\lambda,\theta) \propto (\det[I(\lambda,\theta)])^{\frac{1}{2}} = \frac{1}{\lambda\theta}. \tag{5.10}$$

This implies that when λ and θ are assumed to be independent, the individual non-informative priors could be the reciprocal of the parameters, i.e., $\frac{1}{\lambda}$ and $\frac{1}{\theta}$, respectively. Using (5.10), the joint posterior pdf of λ and θ is given as

$$p(\lambda,\theta|data) \propto \lambda^{r-1}\theta^{r-1}(\prod_{i=1}^{r} x_i^{\lambda-1})\exp[-T\theta]. \tag{5.11}$$

From (5.11), the posterior distribution is given by

$$p(\lambda,\theta|data) = \frac{(\lambda\theta)^{r-1}(\prod_{i=1}^{r} x_i^{\lambda-1})\exp[-T\theta]}{\int_0^\infty\int_0^\infty (\lambda\theta)^{r-1}(\prod_{i=1}^{r} x_i^{\lambda-1})\exp[-T\theta]d\theta d\lambda}. \tag{5.12}$$

The Jeffreys' prior for the Weibull model of the form (2.12) is as follows. Given

$$f(x|\lambda,\theta) = \lambda\theta^\lambda x^{\lambda-1}\exp\left[-(\theta x)^\lambda\right]$$

and

$$\log[f(x|\lambda,\theta)] = \log[\lambda] + \lambda\log[\theta] + \lambda\log[x] - \log[x] - \theta^\lambda x^\lambda \equiv K.$$

The elements are

$$\begin{aligned}
I_{11} &= \frac{\partial^2 K}{\partial\theta^2} = -\frac{\lambda}{\theta^2} - \frac{\lambda(\lambda-1)}{\theta^2}\theta^\lambda x^\lambda \\
I_{12} &= \frac{\partial^2 K}{\partial\lambda\partial\theta} = \frac{1}{\theta} - \lambda\theta^{\lambda-1}x^\lambda\log[x] - \theta^{\lambda-1}x^\lambda - \lambda\theta^{\lambda-1}\log[\theta]x^\lambda = I_{21} \\
I_{22} &= \frac{\partial^2 K}{\partial\lambda^2} = -\frac{1}{\lambda^2} - \log[\theta](\theta x)^\lambda\log[\theta x] - (\theta x)^\lambda\log[x]\log[\theta x].
\end{aligned}$$

The elements of the Fisher information matrix are now obtained as

$$-E[I_{11}] = \frac{\lambda}{\theta^2} + \frac{\lambda(\lambda-1)}{\theta^2}E[\theta x]^\lambda = \frac{\lambda}{\theta^2} + \frac{\lambda(\lambda-1)}{\theta^2}\int_0^\infty(\theta x)^\lambda\lambda\theta^\lambda x^{\lambda-1}\exp\left[-(\theta x)^\lambda\right]dx,$$

and when

$$u = (\theta x)^\lambda$$

yields

$$\begin{aligned}
&\frac{\lambda}{\theta^2} + \frac{\lambda(\lambda-1)}{\theta^2}\int_0^\infty u\lambda u\frac{\theta}{u^{\frac{1}{\lambda}}}\frac{1}{\lambda\theta}u^{\frac{1}{\lambda}-1}\exp[-u]du \\
=\ &\frac{\lambda}{\theta^2} + \frac{\lambda(\lambda-1)}{\theta^2} = \frac{\lambda^2}{\theta^2}.
\end{aligned}$$

Similarly,

$$\begin{aligned} -E[I_{12}] &= -\frac{1}{\theta} + E\left[\lambda\theta^{\lambda-1}x^{\lambda}\log[x] + \theta^{\lambda-1}x^{\lambda} + \lambda\theta^{\lambda-1}\log[\theta]x^{\lambda}\right] \\ &= -\frac{1}{\theta} + \frac{1}{\theta}E\left[(\theta x)^{\lambda}\right] + \frac{\lambda}{\theta}E\left[(\theta x)^{\lambda}\log[\theta x]\right] \\ &= \frac{\lambda}{\theta}E\left[(\theta x)^{\lambda}\log[\theta x]\right] = \frac{\lambda}{\theta}\int_0^{\infty}(\theta x)^{\lambda}\log[\theta x]\lambda(\theta x)^{\lambda}x^{-1}\exp\left[-(\theta x)^{\lambda}\right]dx, \end{aligned}$$

using the same substitution as above, with

$$\gamma_1 = \int_0^{\infty} u\log[u]\exp[-u]du,$$

yields

$$\frac{\gamma_1}{\theta}.$$

Lastly,

$$\begin{aligned} -E[I_{22}] &= \frac{1}{\lambda^2} + E\left[(\theta x)^{\lambda}\log^2[\theta x]\right] \\ &= \frac{1}{\lambda^2} + \int_0^{\infty}(\theta x)^{\lambda}\log^2[\theta x]\lambda\theta^{\lambda}x^{x-1}\exp\left[-(\theta x)^{\lambda}\right]dx, \end{aligned}$$

using the same substitution as before yields

$$\begin{aligned} -E[I_{22}] &= \frac{1}{\lambda^2} + \frac{1}{\lambda^2}\int_0^{\infty}u\log^2[u]\exp[-u]du \\ &= \frac{1+\gamma_3}{\lambda^2}, \end{aligned}$$

with

$$\gamma_3 = \int_0^{\infty} u\log^2[u]\exp[-u]du.$$

Therefore, the corresponding Fisher information matrix is given by

$$I = -E\begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} = \begin{bmatrix} \frac{\lambda^2}{\theta^2} & \frac{\gamma_1}{\theta} \\ \frac{\gamma_1}{\theta} & \frac{1+\gamma_3}{\lambda^2} \end{bmatrix},$$

from which Jeffreys' prior is

$$\pi(\lambda,\theta) \propto (\det[I(\lambda,\theta)])^{\frac{1}{2}} = \frac{1}{\theta}. \tag{5.13}$$

The posterior distribution corresponding to the Jeffreys' prior (5.13) is given as

$$p(\lambda,\theta|data) \propto \lambda^{r-1}\theta^{r\lambda-1}\left(\prod_{i=1}^{r}x_i^{\lambda-1}\right)\exp[-T\theta^{\lambda}],$$

with normalising equivalent to

$$p(\lambda, \theta|data) = \frac{\lambda^{r-1}\theta^{r\lambda-1}\left(\prod_{i=1}^{r} x_i^{\lambda-1}\right)\exp[-T\theta^{\lambda}]}{\int_0^{\infty}\int_0^{\infty}\lambda^{r-1}\theta^{r\lambda-1}\left(\prod_{i=1}^{r} x_i^{\lambda-1}\right)\exp[-T\theta^{\lambda}]d\theta d\lambda}. \tag{5.14}$$

It can be observed from (5.10) and (5.13) that the two Jeffreys' priors are not equivalent. This difference stems from the fact that the two Weibull models have different forms of likelihood functions. The integrals (5.12) and (5.14) cannot be obtained explicitly in closed-form, since in both cases the statistic T depends on the parameter λ as well. This problem persists even when the Jeffreys' prior (5.10) is considered, assuming λ and θ to be independent. Hence, one of the computer intensive Bayesian computational techniques discussed in chapter 3 might be employed to obtain the denominator in (5.12) and (5.14). The method devised by Soland (1969) is another alternative to handle this problem of obtaining the posterior distribution and consequent posterior moments of interest.

5.3.2 Posterior analysis using the reference prior

As was pointed out in chapter 2, that the reference prior can be obtained by utilising the Jeffreys' prior (5.3). The reference prior for the Weibull model in (2.12) will be derived, according to the steps mentioned in chapter 2.

Consider the sequence of subsets given by $\Delta_1 \subset \Delta_2 \subset \Delta_3 ...$ where $\cup_i \Delta_i = \Theta$ are the rectangles, whose width are given by the pair of numbers (b_{1i}, b_{2i}) and corresponding length of (a_{1i}, a_{2i}).

For this embedded subsets, if $i \to \infty$, then a_{1i} and $b_{1i} \to 0$ while a_{2i} and $b_{2i} \to \infty$.

Step 1: The first step has already been performed, which assigns an usual reference prior, that is obtained by assuming λ to be known (in this case, the Jeffreys' prior).

Step 2: Choosing the appropriate rectangles on the parameter space of the two parameters, λ and θ, the prior $\pi(\theta|\lambda) = \frac{1}{\theta}$ has finite mass on

$$\Omega_{i,\lambda} = \{\theta : (\theta, \lambda) \in \Delta_i = (a_{1i}, a_{2i}) \times (b_{1i}, b_{2i})\},$$

for any λ. Normalising $\pi(\theta|\lambda)$ on each of the $\Omega_{i,\lambda}$ yields

$$\begin{aligned}\pi_i(\theta|\lambda) &= K_i(\lambda)\pi(\theta|\lambda)1_{\Omega_{i,\lambda}}(\theta)\\ &= K_i(\lambda)\left(\frac{1}{\theta}\right)1_{\Omega_{i,\lambda}}(\theta),\end{aligned}$$

where $1_{\Omega_{i,\lambda}}(\theta)$ is an indicator function on Ω and

$$K_i(\lambda) = \frac{1}{\int\limits_{\Omega_{i,\lambda}} \pi(\theta|\lambda)d\theta}.$$

Step 3: The marginal reference prior for θ, with respect to $\pi_i(\theta|\lambda)$, is obtained by

$$\pi_i(\lambda) = \exp\left[\frac{1}{2}\int_{\Omega_{i,\lambda}} \pi_i(\theta|\lambda) \times \log\left[\frac{\det[I(\lambda,\theta)]}{\det[I_{11}(\lambda,\theta)]}\right] d\theta\right]. \tag{5.15}$$

First obtaining

$$\begin{aligned}\pi_i(\theta|\lambda) &= \frac{1}{\theta}\frac{I_{a_{1i},a_{2i}}(\theta)}{\int\limits_{a_{1i}}^{a_{2i}} \frac{1}{\theta}d\theta}\\ &= \frac{I_{a_{1i},a_{2i}}(\theta)}{\theta(\log[a_{2i}] - \log[a_{1i}])}.\end{aligned} \tag{5.16}$$

Substitute (5.16) in (5.15) yields the corresponding marginal reference prior for θ, which is given by

$$\begin{aligned}\pi_i(\lambda) &\propto \exp\left[\frac{1}{2}\int_{a_{1i}}^{a_{2i}} \frac{1}{\theta(\log[a_{2i}] - \log[a_{1i}])}\log\left[\frac{\frac{1+\gamma_3-\gamma_1^2}{\theta^2}}{\frac{\lambda^2}{\theta^2}}\right]\right] d\theta\\ &= \exp\left[\frac{1}{2(\log[a_{2i}] - \log[a_{1i}])}\log\left[\frac{1+\gamma_3-\gamma_1^2}{\lambda^2}\right](\log[a_{2i}] - \log[a_{1i}])\right]\\ &= \exp\left[\frac{1}{2}\log\left[\frac{1+\gamma_3-\gamma_1^2}{\lambda^2}\right]\right]\\ &= \frac{\sqrt{1+\gamma_3-\gamma_1^2}}{\lambda}.\end{aligned}$$

Step 4: Using (2.18) and letting $\lambda_0 = 1$, the corresponding reference prior is

$$\begin{aligned}\pi(\lambda,\theta) &= \lim_{i\to\infty}\left[\frac{(\log[a_{2i}] - \log[a_{1i}])\frac{\sqrt{1+\gamma_3-\gamma_1^2}}{\lambda}}{(\log[a_{2i}] - \log[a_{1i}])\frac{\sqrt{1+\gamma_3-\gamma_1^2}}{1}}\right] \times \frac{1}{\theta}\\ &= \frac{1}{\lambda\theta},\end{aligned}$$

which is equivalent to the Jeffreys' prior obtained for the Weibull model given in (2.11). The reference prior for the Weibull model of the form (2.11) is however not obtainable.

5.4 THE SURVIVAL FUNCTION OF THE WEIBULL MODEL

Many problems of reliability improvement, such as the space shuttle reaction control system are modelled as a two-parameter Weibull pdf. This model can be used in situations where there are two explanatory parameters dictating a system or when one parameter mainly determines the system, while the second is considered a nuisance parameter.

In this section, the Jeffreys' prior for the survival function of the Weibull model has been derived for shape parameter known and for both parameters unknown. It should be noted that when the shape parameter is known, the Bayesian estimators of the survival function will be similar to the derivations of chapter 4.

In a similar fashion, for the Weibull model, (2.11), the survival function at any point t, is given as

$$S(t) = \exp\left[-t^{\lambda}\theta\right]$$

equating it to υ, results in

$$\theta = \frac{(-\log[\upsilon])}{t^{\lambda}}.$$

The corresponding Weibull model in terms of υ is give as

$$f(x|\lambda,\upsilon) = \frac{\lambda(-\log[\upsilon])}{t^{\lambda}} x^{\lambda-1} \exp\left[-\left(\frac{x}{t}\right)^{\lambda}(-\log[\upsilon])\right]. \tag{5.17}$$

The Jeffreys' prior for the survival function can now be obtained using the reparameterised Weibull model (5.17). Rewriting (5.17) yields

$$\log\left[f(x,\lambda|\upsilon)\right] = \log[\lambda] + \log[(-\log[\upsilon])] - \lambda\log[t] + \lambda\log[x] - \log[x] - \left(\frac{x}{t}\right)^{\lambda}(-\log[\upsilon]) \equiv K.$$

Differentiating K twice with respect to the parameter of interest υ, gives

$$\frac{\partial K}{\partial \upsilon} = \frac{1}{\upsilon\log[\upsilon]} + \left(\frac{x}{t}\right)^{\lambda}\frac{1}{\upsilon} \quad \text{and} \quad \frac{\partial^2 K}{\partial \upsilon^2} = -\frac{1+\log[\upsilon]}{(\upsilon\log[\upsilon])^2} - \left(\frac{x}{t}\right)^{\lambda}\frac{1}{\upsilon^2}.$$

The elements of the Fisher information matrix are

$$I_{11}(\theta) = -E_{\theta}\left[\frac{\partial^2\left\{\log\left[f(x,\lambda|\upsilon)\right]\right\}}{\partial \upsilon^2}\right] = \frac{1+\log[\upsilon]}{\upsilon^2\log^2[\upsilon]} + E_{\upsilon}\left[\left(\frac{x}{t}\right)^{\lambda}\frac{1}{\upsilon^2}\right], \tag{5.18}$$

but

$$E_v\left[\left(\frac{x}{t}\right)^\lambda \frac{1}{v^2}\right] = \int_0^\infty \frac{\lambda \log[v]}{t^\lambda} x^{\lambda-1}\left(\frac{x}{t}\right)^\lambda \frac{1}{v^2}\exp\left[\left(\frac{x}{t}\right)^\lambda \log[v]\right] dx. \tag{5.19}$$

Letting

$$u = \left(\frac{x}{t}\right)^\lambda (-\log[v])$$

reduces to

$$\frac{1}{v^2(-\log[v])}$$

Therefore, substituting this into (5.18) results in the appropriate element for the Fisher information matrix

$$I_{11}(\theta) = \frac{(1+\log[v])}{v^2\log^2[v]} - \frac{1}{v^2(\log[v])} = \frac{1}{v^2(-\log[v])^2}.$$

Hence, using (2.15) provides the appropriate Jeffreys' prior for the survival function v, which is given by

$$\pi(v|\lambda) \propto \frac{1}{v(-\log[v])}$$

and is the same as the prior of the survival function in the case of the exponential model. It can also be noted by substituting the survival function for the Weibull model given by (2.12) reduces the same Weibull model as (5.17), which implies that the Jeffreys' prior in the second model and the subsequent analyses are all equivalent.

The theory is now extended to the case where both λ and v are unknown. As before, K is now rewritten as

$$\log\left[f(x|\lambda, v)\right] = \log[\lambda] + \log[(-\log[v])] - \lambda\log[t] + \lambda\log[x] - \log[x] + \left(\frac{x}{t}\right)^\lambda \log[v]. \tag{5.20}$$

The elements for the Fisher information matrix are

$$\begin{aligned}
I_{11} &= \frac{1}{v^2(-\log[v])^2},\\
I_{12} &= I_{21} = -E\left[\frac{1}{v}\left(\frac{x}{t}\right)^\lambda \log\left[\frac{x}{t}\right]\right]\\
&= \int_0^\infty \frac{\lambda\log[v]}{t^\lambda}x^{\lambda-1}\left(\frac{x}{t}\right)^\lambda \frac{1}{v}\log\left[\frac{x}{t}\right]\exp\left[\left(\frac{x}{t}\right)^\lambda \log[v]\right] dx. \qquad (5.21)
\end{aligned}$$

If

$$u = \left(\frac{x}{t}\right)^\lambda (-\log[v]),$$

then (5.21) reduces to

$$I_{12} = \frac{\gamma_4}{\lambda \upsilon(-\log[\upsilon])},$$

where

$$\gamma_4 = \int_0^\infty u \exp[-u] \log[u] du - \int_0^\infty u \exp[-u] \log[(-\log[u])] du$$

and

$$I_{22} = \frac{1}{\lambda^2}.$$

The Fisher information matrix is therefore

$$I = \begin{bmatrix} \frac{1}{\upsilon^2(-\log[\upsilon])^2} & \frac{\gamma_4}{\lambda \upsilon(-\log[\upsilon])} \\ \frac{\gamma_4}{\lambda \upsilon(-\log[\upsilon])} & \frac{1}{\lambda^2} \end{bmatrix}$$

Hence, using (2.15), the corresponding Jeffreys' prior is

$$\pi(\lambda, \upsilon) \propto \frac{1}{\upsilon \lambda(-\log[\upsilon])}. \tag{5.22}$$

The posteriors using both Jeffreys' and reference priors cannot be expressed in closed-form and hence explicit expressions of point estimators and their corresponding risk measures are complicated.

5.5 CONCLUSION

The two-parameter Weibull model is one of the widely applied models in the Bayesian paradigm to decision-making. The two commonly used Weibull model representations have been considered in this chapter. The Jeffreys' prior for the scale parameter of the Weibull model is the same as Jeffreys' prior of the parameter of the exponential model obtained in chapter 4. The Bayesian estimators and their corresponding posterior risks and risk functions for the Weibull model in this case correspond to that of the exponential model, by simply adjusting the statistic T.

The posterior distribution of the two-parameter Weibull model, when using the Jeffreys' prior is not explicitly obtainable in its closed-form mainly due to the fact that T depends on the unknown shape parameter. Moreover, there are no continuous joint pdfs, which are closed under sampling that can be considered as a conjugate. An alternative option is to use numerical techniques to simulate the relevant posterior distributions and hence, the risk measures.

CHAPTER 6

CONCLUSION

The Bayesian approach to decision-making process requires a choice of the prior distribution, that reflects the state of nature for the parameter(s) of interest. The Jeffreys' prior has been widely used in Bayesian analyses, especially for vague information about one parameter. In problems with more than one parameter however, the general reference prior is preferable. Increasing the number of the parameter(s) in a model, on the other hand, complicates the derivation of the reference prior. If a more mathematically amenable analysis is required, under certain conditions, the conjugate priors can be used as alternative. In general, the choice of an appropriate prior, that reflects the true state of nature, is by itself a decision-making process.

The objective in this thesis was to obtain various alternative Bayesian estimators, using the squared error loss and LINEX loss functions and compare their performance. The choice of these two loss functions stems from the fact that they are widely used many analyses within the Bayesian paradigm. Practically speaking, there are problems pertaining to either symmetrical or asymmetrical loss functions. Two different forms of the scalar estimation error were used, to define the loss functions. Defining the loss functions with different forms of the estimation error gives a decision-maker flexibility to choose among many alternatives. However, when their risk measures are of the crossing-over type, it might be onerous to choose the best among alternatives.

In decision-making, sometimes one might use a certain estimator, which is obtained relative to an "incorrect loss function", where in fact another more appropriate loss function can be used to measure the penalty or risk. The performance of the estimators were compared for acceptability in a decision-making framework. This can be seen as type of procedure that addresses the robustness of an estimator relative to a chosen loss function.

Relative to the risk function measure under the squared error loss function, the correct and wrongly chosen estimators are of a crossing-over type. Therefore, under the squared error loss function, both estimators can be inadmissible. Relative to the LINEX loss function, however, the LINEX estimator outperforms the other alternative, when the Jeffreys' prior is considered. Nonetheless, when the conjugate prior is considered, the risk functions are of a crossing-over type, portraying the fact that none of the estimators are uniformly dominating.

The sensitivity of the deviation with respect to the LINEX parameter might be an important point worth analysing, even though the choice of the LINEX parameter is often made subjectively.

It can be easily noticed that the performance of the estimators depends both on the model choice, the prior and the loss function. Thus, it can in general be recommended that one should use the appropriate loss functions reflecting the specific situation. If that is not the case, a choice of other loss functions might be at the cost of incurring a greater risk. Even though there are crossing-over of risks, when pair-wise comparisons were made, in most cases the wrongly chosen alternative results in a greater risk. In addition to this, an appropriate prior distribution should also be incorporated in the analysis, reflecting the reality as much as possible. A further analytical study in the behavior of the posterior risks, risk functions and integrated risks is important to establish the sufficient or necessary conditions of inadmissibility or admissibility.

In general, care must be taken especially when many alternatives of estimators are available. The challenge for the decision-maker escalates, when confronted with a scenario where the correct estimator results in a greater risk, especially for crossing-over risk measures. Moreover, in all situations, a decision-maker must foresee all possible scenarios, needless to mention, however, that the decision made at the end, should not contradict the reality.

REFERENCES

Aitchison, J. & Dunsmore, I. R. 1975. *Statistical Prediction Analysis*, Cambridge University Press, London.

Ananda, S. & Ravindra, K. 1998. On estimating the current intensity of failure for the power-law process. *Journal of Statistical Planning and Inference,* vol. 74, p. 253-272.

Aven, T. & Kvalóy, J. T. 2002. Implementing the Bayesian paradigm in risk analysis. *Reliability Engineering and Systems Safety*, vol. 78, Issue 2. p. 195-201.

Aven, T. & Pörn, K. 1998. Expressing and interpreting the results of quantitative risk analyses. Review and Discussion. *Reliability Engineering and Safety*, vol. 61, p. 3-10.

Basu, A. P. & Ebrahimi, N. 1991. Bayesian approach to life testing and reliability estimation using asymmetric loss function, *Journal of Statistical Planning and Inference*, vol. 29, p. 21-31.

Basu, A. P. & Ebrahimi, N. 1992. Bayesian approach to some problems in life testing and reliability estimation.in *Bayesian analysis in Statistics and Economics.* Springer-Verdag, New York.

Bayes, T. 1958. A biographical note. *Biometrika*, vol. 45, p. 293-315.

Bekker, A., Roux, J. J. J. & Mostert, P. J. 2000. A generalization of the compound Rayleigh distribution: using a Bayesian method on cancer survival times. *Communication in Statistics: Theory and Methods*, vol. 29, no. 7, p. 1419-1433.

Berger, J. O. 1985. *Statistical Decision Theory: Foundations, Concepts and Methods*, Springer-Verlag, New York.

Berger, J. O. & Bernardo, J. M. 1989. Estimating a Product of Means: Bayesian Analysis With Reference Priors. *Journal of the American Statistical Association*, vol. 84, no. 405. p. 200-207.

Bernard, G. A. 1954. Sampling inspection and statistical decisions. *Journal Royal Statistical Society*, vol. 16, series B, p. 151-174.

Bernardo, J. M. 1979. Reference posterior distributions for Bayesian inference (with discussion). *Journal Royal Statistical Society*, vol. 41, Series B, p. 113-147.

Bernardo, J. M. 1997. Non-informative priors do not exist: A dialogue with Jose, M. Bernardo. *Journal of Statistical Planning and inference*, vol. 65, p. 159-189.

Bernardo, J. M & Ramon, J. M. 1998. An introduction to Bayesian inference analysis: inference on the ratio of multinomial parameters. *The Statistician*, vol. 47, Part 1, p.101-135.

Bhattacharya, S. K. 1967. Bayesian Approach to life testing and reliability estimation using a loss function. *Journal of the American Statistical Association*, vol. 62, p. 48-62.

Bhattacharya, D., Samaneigo, F. J. & Vestrup, E. M. 2002. On the comparative performance of Bayesian and classical point estimators under Asymmetric loss. *The Indian Journal of Statistics*, vol. 64 Series B, Pt. 3, p. 239-266.

Bolfarine, H. 1989. A note on finite population prediction under asymmetric loss functions. *Communication in Statistics: Theory and Methods,* vol. 18, p. 1863-1869.

Box, G. E. P. & Tiao, G. C. 1973. *Bayesian inference in Statistical Analysis*. Addison-Wesley, Reading, Massachusetts.

Casella, G. & George, E. I. 1992. Explaining the Gibbs Sampler. *The American Statistician,* vol. 46, no. 3, p.167–174.

Chib, C. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, vol. 90, p. 1313-1321.

Chaturvedi, A., Bhatti, M. I. & Kumar, K. 2000. Bayesian analysis of disturbances variance in the linear regression model under asymmetric loss functions. *Applied Mathematics and Computation*, vol. 114, p. 149-153.

Damien, P., Wakefield, J. & Walker, S. 1999. Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables, *Royal Statistical Society*. Series B, vol. 61, no. 1, p. 331-344.

Dellaportas, P. & Wright, D. E. 1991. Numerical prediction for the two-parameter Weibull distribution. *The Statistician*, vol. 44, p. 365-372.

El-Sayyad, G. M. 1967. *Estimation of the parameter of an exponential distribution*, University College of Wales, Aberystwyth.

Enis, P. & Geisser, S. 1971. Estimation of the probability that $Y < X$. *Journal of the American Statistical Association*, vol. 66, p. 167-168.

Evans, I. G. & Nigm, A. M. 1980. Bayesian Prediction for 2-parameter Weibull Lifetime Models. *Communication in Statistics: Theory and Methods*, vol. A9, no. 6, p. 649-658.

Evans, I. G. & Nigm, A. H. M. 1980. Bayesian 1-Sample Prediction for 2-parameter Weibull Distribution. *IEEE Transactions on Reliability*, vol. R-29, no. 5, p. 410-412.

Farsipour, N. S. & Asgharzadeh, A. 2002. On the Admissibility of Estimators of the common mean of two normal populations under symmetric and asymmetric loss functions. *South African Statistical Journal*, vol. 36, p. 39-54.

Ferguson, T. S. 1967. *Mathematical Statistics: A decision Theoretic Approach*, Academic Press, New York.

Feynman, R. P. 1987. Mr. Feynman goes to Washington. *Engineering and science.* California Institute of Technology, Pasadena, California.

Gauss, C. F. 1810. *Methode des Moindres Carres. Memoire sur la Combination des Observations*. Transl. J. Bertrand (1955). Mallet-Bachelier, Paris.

Gelfand, A. E. & Smith, A. F. M. 1990. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, vol. 85, no. 410, p. 398-409.

Gelfand, A. E., Smith, A. F. M. & Lee, T-M. 1992. Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling. *Journal of the American Statistical Association*, vol. 87, no. 418, p. 523-532.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. & Smith, A. F. M. 1990. Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, vol. 85, no. 412, p. 972-985.

Geman, S. & Geman, D. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern analysis and Machine intelligence,* vol. 6, p. 721-740.

Geweke, J. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, vol. 57, p. 1317-1340.

Geyer, C. 1992. Practical Markov chain Monte Carlo. *Statistical Science*, vol.7, p. 473-482.

Girshick, M. A. & Savage, L. J. 1951. Bayes and minimax estimates for quadratic loss functions. *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability.* vol. 1, p. 53-73.

Gross, A. J. & Clark, V. A. 1975. *Survival distributions: Reliability Applications in Biomedical Sciences*, Wiley, New York.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their application. *Biometrika,* vol. 57, p. 907-919.

Hodges, J. L. & Lehmann, E. L. 1951. Some applications of the Cramer-Rao inequality, *Proceedings of the. 2nd Berkeley Symposium on Mathematical Statistics and Probability.* vol. 1, p. 13-22.

Holton, G. 1997. *Subjective value at risk.* Financial Engineering News: August.

Jaheen, Z. F. 2003. A Bayesian analysis of record statistics from the Gompertz model. *Applied Mathematics and Computation*, vol. 145, p. 307-320.

Jeffrey, H. D. 1997. *Bayesian Economics through Numerical Methods: A Guide to Econometrics and Decision-Making with Prior Information.* Springer-Verlag, New York.

Jeffreys, H. 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London*, vol. 186, Series B, p. 453-461.

Jeffreys, H. 1961. *Theory of probability* (third edition). Oxford University Press, Oxford.

Kotz, S. & Johnson, N. L. 1982. *Encyclopedia of Statistical Sciences*, vol. 1, p. 25-28.

Kotz, S. & Johnson, N. L. 1983. *Encyclopedia of Statistical Sciences*, vol. 4, p. 26.

Kuo, L. & Dey, D. 1990. On the admissibility of the linear estimates of the Poisson mean using LINEX loss function. *Statistical decisions*, vol. 8, p. 201-210.

Kuo, L. & Smith, A. F. M. 1992. Bayesian computations in survival models via the Gibbs sampler. Klein, J. P. & Goel, P. K. (eds.). *Survival Analysis: State of Art*, p. 11-24.

Kvam, P. H., Singh, H. & Tiwari, R. C. 1999. Nonparametric estimation of the survival function based on censored data with additional observations from the residual life distribution. *Statistica Sinica*, vol. 9, p. 229-246.

Laplace, P.S. 1773. *Mémoire sur la probabilité des causes par les événements. Mémoires de l'Academie Royale des sciences présentés pardivers savans*, vol. 6, p. 621-656. [Reprinted in Laplace (1878) vol. 8, p. 27-65.]

Lee, E. T. 1980. *Statistical Methods for Survival Data Analysis*, Lifetime Learning Publications, Inc., Belmount.

Lehmer, D. H. 1949. Mathematical methods in large-scale computing units. *Proceedings of the 2nd Symposium on Large-Scale Digital Calculating Machinery*, Massachusetts: Harvard University Press, Cambridge, p. 141-146.

Lindley, D. V. 1980. Approximate Bayesian methods: in *Bayesian statistics*, eds. Bernardo, J. M. DeGroot, M. H., Lindley, D. V. & Smith, A. F. M. Valenticia., Spain: University press, p. 223-245.

Martz, H. F. & Waller, R. A. 1982. *Bayesian Reliability Analysis*. John Wiley & Sons, New York.

Metropolis, N. Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, vol. 21. p. 1087-1092.

Metropolis, N. & Ulam, S. 1949. The Monte Carlo method. *Journal of the American Statistical Association,* vol. 44, p. 335-341.

Morris, M. 1994. *Introduction to Game Theory*. Springer-Verlag, New York.

Mosteller, F. & Wallace, D. L. 1964. *Inference and disputed Authorship: The Federalist Papers Reading*. Addison-Wesley, MA.

Mostert, P. J., Bekker, A. & Roux, J. J. J. 1998. Bayesian analysis of survival data using the Rayleigh model and linex loss. *South African Statistical Journal*, vol. 32, no. 1, p. 19-42.

Mostert, P. J., Roux, J. J. J. & Bekker, A. 1999. Bayes estimators of the lifetime parameters using the compound Rayleigh model. *South African Statistical Journal*, vol. 33, p. 117-138.

Nayak. T. K. & Kundu, S. 2001. Calculating and Describing Uncertainty in Risk Assessment: *The Bayesian Approach: Human and Ecological Risk Assessment*, vol. 7, Issue 2, p. 307-328.

Naylor, J. C. & Smith, A. F. M. 1982. Application of a Method for the efficient Computation of Posterior Distributions, *Applied Statistics*, vol. 31. p. 214-225.

Nigm, A. M. 1989. An informative Bayesian prediction for the Weibull Lifetime Distribution. *Communication in Statistics: Theory and Methods*, vol.18, no. 3, p. 897-911.

Nordman, D. J. & Meeker, W. Q. 2002. Weibull Prediction Intervals for a Future Number of Failures. *Technometrics*, vol. 44, p. 15-23.

Pandy, B. N., Singh, B. P. & Mishara, C. S. 1996. Bayes Estimation of Shape Parameter of Classical Pareto Distribution under LINEX loss function. *Communication in Statistics: Theory and Methods*, vol. 25, no. 12, p. 3125-3145.

Parsian, A. & Farsipour, S. N. 1993. On the Admissibility and Inadmissibility of estimators of scale parameter using an asymmetric loss function, *Communications in Statistics: Theory and Methods*, vol. 22 no. 10, p. 2877-2901.

Parsian, A. 1990b. Bayes estimation using LINEX loss function. *Journal of science*, IROI, vol. 1, p. 305-307.

Raiffa, H. & Schlaifer, R. 1961. *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University.

Raiffa, H. & Schlaifer, R. 2000. *Applied Statistical Decision Theory*. Wiley, New York.

Rayleigh, J. W. S. 1919. *Philosophy Magazine*, 6th Series, vol. 37, p. 321-347.

Rice, J. A. 1995. *Mathematical Statistics and data Analysis*, Second edition, Wadsworth, Inc., California.

Robert, C. P. 2001. *The Bayesian Choice*, Springer Texts in Statistics, second edition, New York.

Rojo, J. 1987. *On the Admissibility of with respect to the LINEX loss function*, Mathematical Sciences Department, University of Texas, El Paso, Texas.

Samaniego, F. J. & Reneuau, D. 1994. Towards reconciliation of the Bayesian and frequentist approaches to point estimation. *Journal of the American Statistical Association*, vol. 89, p. 947-959.

Shalaby, O. A. 1993. Bayesian inference in truncated and censored exponential distribution and reliability estimation. *Communication in Statistics: Theory and Methods*, vol. 22, no.1, p. 57-79.

Singh, U., Gupta, P. K. & Upadhyay, S. K. 2004. Estimation of parameters for exponentiated-Weibull family under type-II censoring scheme. *Computational Statistics and Data Analysis*, (in press).

Sinha, S. K & Guttman, I. 1976. Bayesian inferences about the reliability function of the exponential distribution. *Communication in Statistics: Theory and Methods,* vol. 5, series A5, p. 471-479.

Sinha, S. K. & Kale, B. K. 1980. *Life testing and reliability estimation*. John Wiley & Sons, New York.

Siu, T..K & Yang, H. 1999. Subjective risk measures: Bayesian predictive scenarios analysis. Insurance: *Mathematics and Economics*, vol. 25, p. 157-169.

Soland, R. M. 1968. Bayesian Analysis of the Weibull Process with Unknown Scale Parameter and Its Application to Acceptance Sampling. *IEEE Transactions on Reliability*, vol. R-17, no. 2, p. 84-90.

Soland, R. M. 1969. Bayesian Analysis of the Weibull Process With Unknown Scale and Shape Parameters. *IEEE Transactions on Reliability*, vol. R-18, no.4, p. 181-184.

Stein, C. 1964. Inadmissibility of the usual estimator for the variance of the normal distribution with unknown mean. *Annals of the Institute for Statistics and Mathematics*, vol. 16, p. 155-160.

Tanner, M. A.& Wong, W. H. 1987. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, vol. 82, no. 398, p. 528-550.

Tierney, L. & Kadane, J. B. 1986. Accurate Approximations for Posterior Moments and Marginal densities. *Journal of the American Statistical Association*, vol. 81, p. 82-86.

Tierney, L., Kass, R. E. & Kadane, J. B. 1989. Fully exponential Laplace approximations Expectations and Variance Non-positive Functions. *Journal of the American Statistical Association*, vol. 84, p. 710-716.

Varian, H. R. 1975. A Bayesian approach to real estate assessment. In: Feinberg, S. E. & Zellner, A., Eds., *Studies in Bayesian Econometrics and Statistics in Honor of L. J. Savage*. Amsterdam, North-Holland, p. 195-208.

Von Neumann, J. 1951. Various techniques used in connection with random digits. *J. Resources of the National Bureau of Standards-Applied Mathematics,* Series 12, p. 36-38.

Wakefield, J. C., Gelfand, A. E. & Smith, A. F. M. 1991. Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing*, vol. 1, p. 129-133.

Wald, A. 1950. *Statistical decision Functions*. Wiley, New York.

Wetherill, G. B. 1961. Bayesian sequential analysis. *Biometrika*, vol. 48, p. 281-292.

Zellner, A. & Geisel , M. S. 1968. Sensitivity of Control to Uncertainty and Form of the Criterion Function in *The future of Statistics*, ed. Donald G. Watts, Academic Press, New York. p. 269-289.

Zellner, A. 1980. *Bayesian Analysis in Econometrics and Statistics*: *Studies in Bayesian econometrics. Essays in honor of Harold Jeffreys*. Amsterdam, North-Holland.

Zellner, A. 1986. Bayes estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, vol. 81, p. 446-451.

Zhang, Y. & Meeker, W. Q. 2002. *Bayesian Life Test Planning for the Weibull Distribution with Given Shape Parameter*, Department of Statistics, Iowa State University, Ames,IA.

Zou, G. 1997. Admissible estimation for finite population under the Linex Loss function. *Journal of Statistical Planning and inference*, vol. 61, p. 373-384.

[Online] Bayesian Computational Methods in Most Bayesian Inference Problems, [cited 2004, April], Available at:
http://www.biostat.harvard.edu/courses/individual/ibrahim/bio249/bayesnotes10.ps

[Online] *Weibull Analysis*.[cited 2004, March], Available at:
http://www.mathpages.com/home/kmath122/kmath122.htm.

[Online] Ghosal, S. 1997. *Non-informative Priors*, [cited 2004, May], Available at:
http://www.cs.vu.nl/~ghosal/noninf.html.

[Online] Yang, G. 2002. A Monte Carlo Method of Integration.[cited 2004, May], Available at:
http://unicast.org/enclosures/text.pdf