Loss functions, utility functions and Bayesian sample size determination.

Islam, A. F. M. Saiful

LOSS FUNCTIONS, UTILITY FUNCTIONS AND BAYESIAN SAMPLE SIZE DETERMINATION

by

A. F. M. Saiful Islam, M.Sc.

Thesis submitted for the degree of Doctor of Philosophy

in Queen Mary, University of London

February, 2011

Dedicated to my parents.

Declaration

I declare that the work presented in this thesis is my own research.

Acknowledgements

I would like to thank Dr Lawrence I. Pettit, my supervisor, for his invaluable suggestion and constant encouragement which helped me to complete this thesis. I would also like to thank Professor Steven G. Gilmour, Professor Rosemary A. Bailey and Dr Roger Sugden of Queen Mary, University of London for their advice at different stages of my research.

I am also grateful to the School of Mathematical Science, Queen Mary, University of London which assisted me with the necessary financial support in order to complete my research.

I would like to thank the computer technical support department who provided me with computer programs for my research.

I am grateful to my family especially to my parents, who have always encouraged me to continue my research. Finally, I would like to give special thanks to my wife whose love and support enabled me to complete this work.

Abbreviations and notations

AE : Absolute error

APR : Average posterior risk

BLINEX : Bounded linear exponential

LINEX : linear exponential

$l(d, \theta)$: Loss function

$p(x|\theta)$: Probability density function

$p(\underline{x}|\theta)$: Likelihood of the data

$p(\theta)$: Prior density

$p(\theta|\underline{x})$: Posterior density

PR : Posterior risk

APR : Average posterior risk

SE : Squared error

SSD : Sample Size Determination

$TC(n)$: Total cost for an optimal sample of size n

$E[TC(n)]$: Expected total cost for an optimal sample of size n

$u(d, \theta)$: Utility function

Abstract

This thesis consists of two parts. The purpose of the first part of the research is to obtain Bayesian sample size determination (SSD) using loss or utility function with a linear cost function. A number of researchers have studied the Bayesian SSD problem. One group has considered utility (loss) functions and cost functions in the SSD problem and others not. Among the former most of the SSD problems are based on a symmetrical squared error (SE) loss function. On the other hand, in a situation when underestimation is more serious than overestimation or vice-versa, then an asymmetric loss function should be used. For such a loss function how many observations do we need to take to estimate the parameter under study? We consider different types of asymmetric loss functions and a linear cost function for sample size determination. For the purposes of comparison, firstly we discuss the SSD for a symmetric squared error loss function. Then we consider the SSD under different types of asymmetric loss functions found in the literature. We also introduce a new bounded asymmetric loss function and obtain SSD under this loss function. In addition, to estimate a parameter following a particular model, we present some theoretical results for the optimum SSD problem under a particular choice of loss function. We also develop computer programs to obtain the optimum SSD where the analytic results are not possible.

In the two parameter exponential family it is difficult to estimate the parameters when both are unknown. The aim of the second part is to obtain an optimum decision for the two parameter exponential family under the two parameter conjugate utility function. In this case we discuss Lindley's (1976) optimum decision for one

parameter exponential family under the conjugate utility function for the one parameter exponential family and then extend the results to the two parameter exponential family. We propose a two parameter conjugate utility function and then lay out the approximation procedure to make decisions on the two parameters. We also offer a few examples, normal distribution, trinomial distribution and inverse Gaussian distribution and provide the optimum decisions on both parameters of these distributions under the two parameter conjugate utility function.

Contents

List of Figures

List of Tables

Chapter 1

Literature Review

1.1 Introduction

In this thesis we will focus on loss functions, utility functions and Bayesian sample size determination (SSD). For point estimation we need to consider either a loss or a utility function. The decision depends on the pattern (shape) of loss or utility function adopted in a particular situation. This shape could either be a symmetric or an asymmetric form. Apart from the squared error loss function, very little attention has been paid to obtaining optimum sample size under specific loss functions. We mainly consider asymmetric loss functions and a linear cost function to determine an optimum sample size. We will also investigate estimation of the parameters of bivariate exponential family under a bivariate conjugate utility function.

The aim of this chapter is to review the literature related to loss functions,

Bayesian sample size determination and Lindley's (1976) paper on the conjugate utility function. In the first section we discuss different types of loss functions found in the literature. Then we review the literature related to the Bayesian SSD. Finally we review Lindley's (1976) paper on conjugate utility functions.

A loss function $l(d, \theta)$ represents losses incurred when we estimate the parameter θ by d. There is always some difference observed between the estimate and the parameter. Let d be an estimate of parameter θ. Loss is generally measured as some function of the difference $d-\theta$ or the ratio $\frac{d}{\theta}$. For one parameter, if we fix θ we might get different values of d as an estimate of θ. If $d = \theta$ there is no loss, if $d < \theta$ we call it underestimation, on the other hand if $d > \theta$ then we call it overestimation. Now we can define a loss function as a function of d and θ and denote it by $l(d, \theta)$. The loss function, $l(d, \theta)$ is defined to be a real valued function satisfying i) $l(d, \theta) \geq 0$ for all possible estimates d and all θ under the chosen population ii) $l(d, \theta) = 0$ for $d = \theta$.

In the Bayesian approach to estimate any parameter we consider some prior distribution representing our beliefs about θ with density $p(\theta)$. We collect a sample x_1,x_2,$\ldots x_n$ of size n from the probability density function $p(x|\theta)$ and the likelihood of the sample is $p(\underline{x}|\theta)$. Now combining the prior and the likelihood and using Bayes theorem we have an updated information about the parameter of interest θ represented by the posterior distribution of θ with density $p(\theta|\underline{x})$. We obtain a Bayes estimate, $\hat{d}$ of the parameter θ by choosing a particular form of loss function, $l(d, \theta)$. To obtain the Bayes estimate first we need to find the posterior expected loss $E[l(d, \theta)]$ by $\int l(d, \theta)p(\theta|\underline{x})d\theta$ which is also known as the posterior risk for θ. Then we minimize it with respect to d. For example, the Bayes estimate under a squared error

loss function $(d-\theta)^2$ will be the mean of the posterior distribution and the Bayes estimate under an absolute error loss function $|d-\theta|$, will be the posterior median. It is to be noted that the different loss functions will give different Bayes estimates of θ and in any estimation procedure we can model these losses either in symmetric or in asymmetric form. Now we will discuss symmetric loss functions.

1.2 Symmetric loss functions

If the amount of loss assigned by a loss function to a positive error is equal to the negative error of the same magnitude, then the loss function is called a symmetric loss function. Mathematically, a loss function $l(d,\theta)$ is said to be symmetric if it is a function of $|d-\theta|$. Some possible symmetric loss functions given by Mood, Graybill and Boes (1974) are as follows. If we want to estimate θ by d then the squared error loss function is defined as,

$$l_1(d,\theta) = a_0(d-\theta)^2 \tag{1.1}$$

where $a_0 > 0$ is the scale of the loss function. This scale a_0 is often taken equal to one. If we are only interested in finding the Bayes estimate under the loss function then this has no effect on the estimate. In SSD problems the scale does make a difference. Some more symmetric loss functions are:

$$l_2(d,\theta) = |d-\theta|,$$

$$l_3(d,\theta) = \begin{cases} A, & \text{if}, |d-\theta| > \epsilon \\ 0, & \text{if }, |d-\theta| \leq \epsilon, \text{ where } A > 0 \end{cases}$$

$$l_4(d,\theta) = g(\theta)|d-\theta|^r, \text{ for } g(\theta) \geq 0 \text{ and } r > 0.$$

l_2 is called absolute error loss function. Note that both l_1 and l_2 increase as the error $d-\theta$ increases in magnitude. l_3 says that we lose nothing if the estimate $\hat{d}$ is within ϵ units of d and otherwise we lose an amount A. l_4 is a general loss function that includes both l_1 and l_2 as special cases. Here are some more symmetric loss functions for the functions of the parameter.

If $g(\theta)$ denotes a function of θ, the absolute difference loss function is

$$l_{ad}(\hat{g}(\theta), g(\theta)) = \mid \hat{g}(\theta) - g(\theta) \mid \tag{1.2}$$

Under this loss function the Bayes estimate of $g(\theta)$ will be the median of the posterior distribution of $g(\theta)$. The well known squared error loss function is

$$l_{se}(\hat{g}(\theta), g(\theta)) = (\hat{g}(\theta) - g(\theta))^2. \tag{1.3}$$

Under this loss function the Bayes estimate of $g(\theta)$ will be the mean of the posterior distribution of $g(\theta)$. Another loss function named as the logarithmic loss function is given by Brown (1968) as

$$l_{ln}(\hat{g}(\theta), g(\theta)) = \mid \ln \frac{\hat{g}(\theta)}{g(\theta)} \mid . \tag{1.4}$$

The Bayes estimate of the function $g(\theta)$ under the logarithmic loss function (1.4) is

$$\hat{g}(\theta) = \exp\left[\int \ln g(\theta) p(\theta|\underline{x}) d\theta\right] = \exp\left[E_\theta\{\ln g(\theta)\}\right].$$

Now we will discuss asymmetric loss functions.

1.3 Asymmetric loss functions

The use of symmetric loss function is based on the assumption that the loss is the same in any direction. However, this assumption may not be valid in many practical situations and the use of the symmetric loss function may be inappropriate. In such cases a given positive error may be more serious than a given negative error of the same magnitude or vice-versa. Now we will define the asymmetric loss function. Let $p(\theta|\underline{x})$ be the posterior distribution of a continuous random variable θ for which it is required to find an optimal decision d, associated with the error, $d-\theta$ (which could be some other function of d and θ). The loss function, $l(d,\theta)$, can be written in the form

$$l(d,\theta) = \begin{cases} l_1(d,\theta), & d \geq \theta \\ \\ l_2(d,\theta), & d < \theta, \end{cases}$$

where, $l_1'(d,\theta) > 0$ for $d-\theta > 0$ and $l_2'(d,\theta) < 0$ for $d-\theta < 0$. It is required to minimize the expected loss $\int l(d,\theta)p(\theta|\underline{x})d\theta$ and an optimal decision $\hat{d}$ is a solution of the following equation

$$\int_d^{\infty} l_1'(d,\theta)p(\theta|\underline{x})d\theta + \int_{-\infty}^{d} l_2'(d,\theta)p(\theta|\underline{x})d\theta = 0. \tag{1.5}$$

1.3.1 Asymmetric linear loss function

Granger (1969) obtained an optimum decision under the asymmetric linear loss (cost) function of the form

$$l(d,\theta) = \begin{cases} a(d-\theta), & d \geq \theta, \\ \\ -b(d-\theta), & d < \theta, \end{cases}$$

where, $a > 0$, $b > 0$ and the solution of (1.5) is $d^* = F^{-1}(\frac{a}{a+b})$, where F is the posterior cumulative distribution function of θ. If the loss function is symmetric then $a = b$ and the posterior median is optimal.

1.3.2 Loss function from DeGroot (1970)

DeGroot (1970) discussed different types of loss functions and obtained the Bayes estimates under these loss functions. Here is one example of the asymmetric loss function defined for the positive values of the parameter. If d is an estimate of θ then the loss function $l(d, \theta)$ will be,

$$l(d, \theta) = \left(\frac{\theta - d}{d}\right)^2. \tag{1.6}$$

1.3.3 Power and exponential loss

Britney and Winkler (1974) studied some loss functions other than linear and quadratic form. They investigate the Bayesian point estimate under two special forms of loss function namely, the power and the exponential loss functions. In the following we will discuss these loss functions. Let

$$l(d, \theta) = \begin{cases} l_0(d, \theta), & d - \theta \geq 0, \\ l_u(d, \theta), & d - \theta < 0. \end{cases}.$$

If $l_0(d, \theta) = k_0(d-\theta)^r$ and $l_u(d, \theta) = k_u(\theta-d)^s$, then the loss function is called a power loss function. On the other hand if $l_0(d, \theta) = k_0|e^{r(d-\theta)}|$ and $l_u(d, \theta) = k_u|e^{s(\theta-d)}|$ then the loss function $l(d, \theta)$ is called an exponential loss function. For both the cases,

k_u,k_0, r and s are positive and they are strictly convex functions if $r > 1$ and $s > 1$. They also obtained an optimum Bayes estimate for a standard normal distribution under both the power loss function and the exponential loss function respectively. Finally, they suggested that in any Bayesian point estimation problem for an optimum estimate it is important to specify the form of loss function carefully.

1.3.4 Linex loss function

Varian (1975) introduced a very useful asymmetric loss function. He explained that in a particular area of the United States real estate assessors are required by law to estimate the current market value of all taxable properties in their district each year. This burden has been somewhat alleviated by the application of regression techniques based on previous sales using an ordinary least square procedure. He worked entirely with the appraisal of single family houses given their characteristics such as total living area, number of bedrooms, etc. If the assessors' estimate is denoted by d and the actual value is θ, certain losses are incurred, denoted by $l(d, \theta)$. Now if the assessors' office underestimate the value of a house, the loss is equal to the amount of underestimate. If the office overestimates the value of a house the California homeowner has two courses of action; he can 1) complain to the assessors' office, confer with them and attempt to convince them that his home was over-assessed 2) present an appeal to the Assessment Board of Equalization, which evaluates evidence presented by both sides so as to correct the assessment. Both of these are lengthy procedures which incur considerable expense for both parties. In such a situation the usual quadratic loss function, (1.1) seems inappropriate because it assigns the same loss to overesti-

mates as to underestimates and it is clear that the assessor's loss function should be asymmetric. He mentioned the following features:

1) The loss function should be linear for large negative errors.

2) The loss function should be increasing for positive errors at a greater than linear rate.

3) Even if the court costs are constant, the probability of a complaint being submitted increases with the magnitude of the overestimate, and thus the loss function should increase monotonically for positive errors. Finally, he proposed the following asymmetric loss function as,

$$l(d,\theta) = a\exp[b(d-\theta)] - c(d-\theta) - a; \tag{1.7}$$

which is called a linex (linear exponential) loss function with properties

1) for $c = 1$, $a = b^{-1}$ it has a minimum loss of 0 for $d = \theta$;

2) it has exponential losses for large overestimates and linear losses for large underestimates;

3) parameters can be chosen to provide a variety of asymmetric effects;

4) it combines readily with a normal probability distribution.

Given a sample of previous sales of houses and their characteristics, he formulated the following regression model,

$$y = X\beta + \varepsilon,$$

where y is an m vector of selling prices, X is an $m \times k$ matrix of the observed characteristics of the houses, β is a k vector of unknown parameters, and ε is an m vector of error terms, assumed to be distributed $N(0, \sigma^2 I)$. Under the assumption of normality and using the proposed loss function he obtained the loss minimizing

estimator of β for a univariate case as

$$d = \mu - \frac{b\sigma^2}{2} + b^{-1} \ln \frac{c}{ab},$$

where, μ and σ^2 is the mean and variance of β respectively following a normal distribution. Finally, he compares his estimates obtained from a squared error loss function with the linex loss function under different choices of priors.

1.3.5 Bounded loss function

Smith (1980) studied Bayes estimators under the bounded loss functions. He explained that the Bayes estimates should be made under a bounded loss function rather than convex one, because in practice we can never lose an infinite amount. At first he proposed the following loss functions called the step loss function with guage u.

$$l(d, \theta) = \begin{cases} 0, & (|d - \theta| < u), \\ 1, & \text{otherwise} \end{cases}$$

where d is an estimate of a parameter θ. The expected loss function will be,

$$E[l(d, \theta)] = 1 - F(d + u) + F(d - u),$$

where $F(\theta)$ is continuous. He also showed that, for $d \in (k_1, d_2)$ then $E[l(d, \theta)]$ is strictly decreasing and for $d \in (d_2, k_2)$, then $E[l(d, \theta)]$ is strictly increasing and the Bayes decision must lie in the interval (d_1, d_2). Then he obtained the Bayes estimates under the step loss function for some standard distributions. He also introduced another kind of loss function giving an additional information about the loss $l(d, \theta)$

namely, the ramp loss function as follows.

$$l(d,\theta) = \begin{cases} B^{-1}|d-\theta|, & (|d-\theta| < B) \\ 1, & \text{otherwise} \end{cases}$$

where, $B > 0$. Finally he formalizes a Bayes decision interval under the family of loss functions as,

$$l_b(d,\theta) = \begin{cases} \kappa(u)l^*(\theta-d), & l^*(\theta-d) < c(u)) \\ \kappa(u)c(u), & \text{otherwise} \end{cases}$$

where, $\kappa(u) > 0$, $c(u)$ is strictly increasing in u and $l^*(\theta-d)$ is some symmetric loss function.

1.3.6 Linex loss function, Zellner (1986)

Zellner (1986) re-expressed the Varian's (1975) loss function (1.7) with $l(d,\theta) = 0$ for a minimum to exist at $d-\theta = 0$, then we must have $ab = c$ and the linex loss function reduced as,

$$l(d,\theta) = a[\exp(b(d-\theta)) - b(d-\theta) - 1], \tag{1.8}$$

with two parameters $a > 0, b \neq 0$, where, a is the scale of the loss function and b determines its shape. Zellner (1986) also studied the properties of this loss function showing for $b > 1$ the function is quite asymmetric with overestimation more costly than underestimation. On the other hand when $b < 0$, the function rises almost exponentially when $w < 0$ and almost linearly when $w > 0$, where $w = d-\theta$. In figure (1.1) we present the shape of the linex loss function for a different combinations of parameters. For small values of b the function is almost symmetric and not far from a

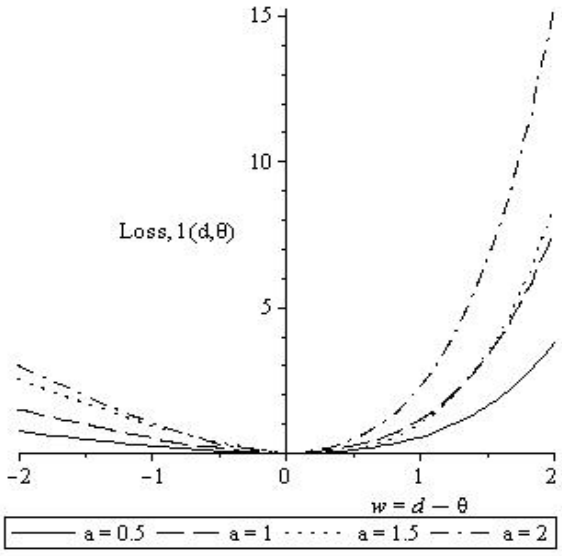

Figure 1.1: Shape of the linex loss function for different values of the scale parameter a when $b = 1.2$ is kept fixed.

squared error loss function, that is, on expanding $e^{bw} \approx 1+bw+b^2w^2/2, l(w) \approx b^2w^2/2$, a squared error loss function. Zellner obtained the Bayes estimator under the linex loss function (1.8) by minimizing the posterior expected loss or posterior risk (PR) (we will discuss it in the section 1.4) as,

$$\hat{d}_{lin} = -\frac{1}{b}\ln[E_\theta(e^{-b\theta})]. \tag{1.9}$$

He used this estimate in an univariate prediction problem and extended the result for multivariate cases. He showed the Bayes risk and the risk function for alternative estimators of normal mean with prior as a $N(0, \tau^2)$ distribution. He also studied the scalar parameters of an exponential distribution, a multiple regression model parameter and multi-parameter problems under a linex loss function.

1.3.7 Modified linex loss function

Despite the flexibility of the linex loss function (1.8) for the estimation of a location parameter, it appears not to be suitable for the estimation of scale parameters and other quantities. For these reasons Basu and Ibrahimi (1991) defined a modified linex loss function as follows. If d is an estimate of θ, then the loss function is defined as,

$$l_m(d,\theta) \propto \exp\left[b\left(\frac{d}{\theta}-1\right)\right] - b\left(\frac{d}{\theta}-1\right) - 1, \tag{1.10}$$

where the estimation error is expressed by $\frac{d}{\theta}$ and the shape parameter is $b \neq 0$. Such a modification does not change the main characteristics of Zellner's (1986) linex loss function described above. The posterior expectation of (1.10) is,

$$E_\theta\left[l_m(d,\theta)\right] \propto \exp(-b)E_\theta\left\{\exp\left[b\left(\frac{d}{\theta}\right)\right]\right\} - bE_\theta\left(\frac{d}{\theta}\right) + b - 1.$$

The value of d that minimizes $E_\theta\left[l_m(d,\theta)\right]$, say, θ_m, is the solution of

$$E_\theta\left[\theta^{-1}\exp\left\{b\left(\frac{\theta_m}{\theta}\right)\right\}\right] = \exp\left[E_\theta\left(\frac{1}{\theta}\right)\right],$$

provided that all expectations are finite.

1.3.8 Asymmetric quadratic loss function

To estimate θ by d Cain (1991) considered an asymmetric quadratic loss function of the following form

$$l(d,\theta) = \begin{cases} a(d-\theta)^2, & d \geq \theta \\ b(d-\theta)^2, & d < \theta, \end{cases}$$

where, $a > 0$ and $b > 0$. Now to obtain an optimum decision $\hat{d}$ under this loss function we need to solve the equation (1.5). Clearly, if the loss function is symmetric then the expected posterior loss is $aVar(\theta|\underline{x})$.

1.3.9 Entropy loss function

Calabria and Pulcini (1994) proposed another alternative to the modified linex loss function named general entropy loss function and defined it as,

$$l_e(d,\theta) \propto \left(\frac{d}{\theta}\right)^p - p\ln\left(\frac{d}{\theta}\right) - 1, \tag{1.11}$$

which has a minimum at $d = \theta$. This loss is a generalization of the entropy loss function used by several authors taking the shape parameter $p = 1$. The more general version (1.11) allows different shapes of loss function when $p > 0$ and for $d > \theta$, i.e. a positive error causes more serious consequences than a negative error. The Bayes estimator of θ under the general entropy loss will be,

$$\hat{d} = \left[E_\theta\{\theta^{-p}\}\right]^{-\frac{1}{p}}, \tag{1.12}$$

provided that, $E_\theta\left[\theta^{-p}\right]$ exists and is finite. They also detailed the properties of a general entropy loss function.

a) When $p = 1$ the Bayes estimate (1.12) coincides with the Bayes estimate under the weighted squared error loss function $\frac{(d-\theta)^2}{\theta}$.

b) When $p = -1$ the Bayes estimate (1.12) coincides with the Bayes estimate under a squared error loss function (1.1).

1.3.10 Bounded linex loss

Wen and Levy (2001) proposed a new parametric family of bounded and asymmetric loss function named the blinex loss function. They explained, due to the nature of many decision problems, such as reliability, the analysis requires the use of asymmetric losses and Zellner's (1986) linex loss can be applied but the utility theory and other practical arguments suggest that the bounded loss functions are more appropriate than the unbounded ones. They also noticed that, while the linex loss function has been extensively explored in the literature and found to be quite useful, it is limited in application because the expected linex loss does not exist under some densities, for example, the student-t. The expected blinex loss does not suffer this disadvantage. This shortcoming prevents linex loss from wider usage as in predictive analysis under a normal likelihood. They proposed the following bounded linex loss function based on the linex loss function.

If d is an estimate of θ and w represent the scalar error resulting from a decision (estimate) d which is denoted as $w = d - \theta$, then the linex loss function is defined as

$$l(w) = c(e^{bw} - bw - 1);$$

where $b \neq 0$ and $c > 0$. Linex is not bounded and depending on the sign of the parameter b, the linex loss is approximately exponential on one side and linear on the other side. Wen and Levy define the new parametric family of bounded asymmetric loss function, denoted by $l_b(w)$, based on the linex loss function $l(w)$ as,

$$\begin{aligned} l_b(w) = \frac{l(w)}{1 + \lambda l(w)} &= \frac{c(e^{bw} - bw - 1)}{1 + \lambda c(e^{bw} - bw - 1)} \\ &= \frac{1}{\lambda}\left[1 - \frac{1}{1 + \lambda c(e^{bw} - bw - 1)}\right]. \end{aligned}$$

where, $b \neq 0$, $c > 0$ and $\lambda > 0$. Setting $a = \lambda c$, the function becomes,

$$l_b(w) = \frac{1}{\lambda}\left[1 - \frac{1}{1 + a(e^{bw} - bw - 1)}\right]. \tag{1.13}$$

They call this function a blinex loss, since it is bounded and is derived from a linex loss function. For the mathematical properties of the blinex loss function, see Wen and Levy (2001). In their paper they also proved that the Bayes estimate of mean based on a normal posterior distribution under the blinex loss function exists and is unique. Wen and Levy (2004) also developed an algorithm for fitting the parameter of the blinex loss function.

We have discussed different types of symmetric and asymmetric loss function above. Now we will outline the posterior expected loss or posterior risk function for the Bayes estimate θ as we need this later for an optimum SSD.

1.4 Posterior expected loss or posterior risk function

If $l(d, \theta)$ is any loss function and $p(\theta|\underline{x})$ is any posterior distribution of θ then the expected loss, which is called the posterior risk's(PR) for the Bayes estimate θ, is defined as

$$PR = \int l(d, \theta)p(\theta|\underline{x})d\theta. \tag{1.14}$$

Now for the squared error loss function (1.1) the PR's for the Bayes estimate θ will be,

$$PR = \int l_1(d, \theta)p(\theta|\underline{x})d\theta = a_0 \int (d - \theta)^2 p(\theta|\underline{x})d\theta, \tag{1.15}$$

which is clearly $a_0 \times$ posterior variance. The scale a_0, which often takes the value 1, does not affect the Bayes estimate but has some effects on the optimum sample size. The PR's for the Bayes estimate θ under the loss function (1.6) is,

$$\begin{aligned} PR &= \int \left(\frac{\theta - d}{d}\right)^2 p(\theta|\underline{x})d\theta \\ &= \frac{Var(\theta|\underline{x})}{E(\theta^2|\underline{x})}, \end{aligned} \tag{1.16}$$

provided $E(\theta^2|\underline{x}) > 0$. The PR's for the Bayes estimate θ under the linex loss function (1.8) is

$$\begin{aligned} PR &= a\int [\exp(b(d-\theta)) - b(d-\theta) - 1]p(\theta|\underline{x})d\theta \\ &= a[\exp(bd)E\{\exp(-b\theta|\underline{x})\} - bd + bE(\theta|\underline{x}) - 1] \\ &= a[\exp(bd)\exp(-bd) - bd + bm - 1] \\ &= ab(m - \hat{d}_{lin}), \end{aligned} \tag{1.17}$$

where a is the scale parameter, b is the shape parameter of the linex loss, m is the posterior mean and $\hat{d}_{lin}$ is the Bayes estimate under the linex loss function defined in (1.9). The PR's for the Bayes estimate θ under the bounded linex loss function (1.13) is,

$$PR = \frac{1}{\gamma}\int \left[1 - \frac{1}{1 + a(e^{b(d-\theta)} - b(d-\theta) - 1)}\right] p(\theta|\underline{x})d\theta. \tag{1.18}$$

It is difficult to solve (1.18) analytically. So we will solve this integral numerically using R for an optimum sample size. Now we will review the literature related to the asymmetric loss functions.

1.5 Application of asymmetric loss function

We have defined different types of symmetric and asymmetric loss functions above. Now we shall review the literature where these loss functions may be used. Rojo (1987) proposed a general class of linear estimators of the form $c\bar{x}+d$ where $\bar{x}$ and $\bar{x}-\frac{a\sigma^2}{2n}$ are special cases of this general form. He determined the range of c and d for which $c\bar{x}+d$ is admissible under a linex loss function. He proved that $c\bar{x}+d$ is admissible if $0 \leq c < 1$ or $c = 1$ and $d = -\frac{a\sigma^2}{2n}$ and otherwise is inadmissible. Later Sadooghi-Alvadani and Nematollahi (1989) in a short paper corrected the proof of the inadmissibility of $c\bar{x}+d$ when $c < 0$ given by Rojo (1987) and showed that his result is true with a different choice of dominating estimators. Bolfarine (1989) studied the prediction of a population total using a linex loss function under a simple regression model through the origin. He also studied two stage sampling and a binomial super population model for an optimal prediction under a linex loss function.

Basu and Ibrahimi (1991) considered the exponential life testing model as,

$$p(x|\theta) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), x > 0, \theta > 0,$$

where θ is a random variable and obtained the Bayes estimator of θ under their proposed loss function (1.10) using three different types of priors such as:

1. The uniform prior with no prior information about θ or more generally the quasi-density of the form

$$p_1(\theta) = 1/\theta^a, 0 < \theta < \infty, a \geq 0.$$

2. The inverted gamma with parameters $\alpha, \beta > 0$ with the density function given by

$$p_2(\theta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{\theta})^{\alpha+1}\exp(-\frac{\beta}{\theta}), & \theta > 0 \\ 0, & \text{otherwise .} \end{cases}$$

3. If the life tester knows in advance that the probable values of θ lie over a finite range (α, β) but he does not have any strong opinion about any subset of values over the range $(\alpha\ , \beta)$ then a uniform distribution over $(\alpha\ , \beta)$ may be a good approximation. So the prior distribution will be,

$$p_3(\theta) = \begin{cases} \frac{1}{(\beta-\alpha)}, & 0 < \alpha \leq \theta \leq \beta, \\ 0, & \text{elsewhere.} \end{cases}$$

They considered the reliability function, $\gamma = R(t)$ as the probability that a system will survive at a specified mission time t.

4. For a situation where the experimenter has no prior information about γ, they used the following non-informative prior distribution as

$$p_4(\gamma) = \frac{1}{\gamma \log \gamma}, 0 < \gamma \leq 1,$$

which is a special case of $p_1(\theta)$ giving $a = 1$.

5. They considered a prior distribution for γ which is a beta distribution with parameter α and $\beta > 0$ as follows:

$$p_5(\gamma) = \frac{1}{B(\alpha, \beta)}(\gamma)^{\alpha-1}(1-\gamma)^{\beta-1}, 0 < \gamma \leq 1.$$

They also obtained the Bayes estimator of γ using this prior $p_5(\gamma)$ considering the

loss function given by Canfield (1970) as,

$$l(\hat{\gamma},\gamma)=\begin{cases} k_1(\frac{\hat{\gamma}}{\gamma}-1)^2, & \text{if } \hat{\gamma}\leq\gamma, \\ k_1(\frac{\hat{\gamma}}{\gamma}-1)^2+k_2(\frac{\hat{\gamma}}{\gamma}-1), & \text{if } \hat{\gamma}\geq\gamma. \end{cases}$$

Canfield (1970) concludes that when $\alpha=\beta=1$ (implying p_5 is uniform) and if $k_2=0$ (symmetric loss function), the resulting estimator is the minimum variance unbiased estimator of the reliability function. They obtained the estimators of the exponential parameter θ under the Varian's (1975) linex loss , the modified linex loss (1.10) and the squared error loss function considering the priors $p_1(\theta)$, $p_2(\theta)$, $p_3(\theta)$ respectively and compared the estimates by the posterior risk functions. Finally, they obtained the Bayes estimator of the reliability, $R(t)=\exp(-t/\theta)$ for the Varian (1975) loss function using non-informative priors $p_4(\gamma)$ and $p_5(\gamma)$ respectively.

Sadooghi-Alvadandi and Parsian (1992) studied the estimation of a binomial parameters n,p with known p and unknown $n\in\{0,1,...\}$ using the linex loss function.

Pandey and Rai (1992) studied the Bayesian estimation of the mean and the square of the mean of a normal distribution with mean (μ) and variance σ^2 (known) using the linex loss function. They compared the estimates relative to the risk functions and Bayes risk under the linex loss and the squared error loss function with the alternative estimators e.g. UMVUE.

Parsian and Farsipour (1993) studied the admissibility and inadmissibility of a scale parameter using an asymmetric loss function which is similar to (1.10). They defined Pitman (1939) estimator to estimate a parameter, θ. Let x_1,x_2,...,x_n be a random sample of size n taken from a density $\frac{1}{\theta}p(\frac{x}{\theta})$, then the estimate of θ under the

loss function $l(d,\theta) = \left(\frac{d}{\theta} - 1\right)^2$ is given by,

$$\hat{d} = \frac{\int_0^\infty \theta^{-n} \prod_{i=1}^n p(\frac{x}{\theta}) d\theta}{\int_0^\infty \theta^{-n+1} \prod_{i=1}^n p(\frac{x}{\theta}) d\theta},$$

which is also called the Pitman estimator of θ. Use of this quadratic loss function in scale parameter estimation heavily penalizes overestimation.

Calabria and Pulcini (1994) considered m censored observations out of a sample of size n. First m observations are failed at t_1,t_2,...,t_m and the remaining $(n-m)$ items are functioning at t_{m+1},...t_n. These observations are assumed to have a two-parameter Weibull distribution given by,

$$p(t|\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{t}{\beta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]; \alpha,\beta > 0, t > 0,$$

where α and β are the scale and shape parameter respectively. They obtained the mean lifetime of the underlying population, $E_T = \alpha\Gamma\left(\frac{1}{\beta+1}\right)$, the reliability of the item for a given mission time t, say, $R_t = \exp\left[-t\left(\frac{t}{\alpha}\right)^{\beta}\right]$, the reliability life of a given reliability level, say $R_g = \alpha\left[\ln\left(\frac{1}{R}\right)\right]^{\frac{1}{\beta}}$ and the hazard rate at a given magnitude of time, say, $H_t = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$ under the linex loss (1.8) and their proposed entropy loss (1.11) using different types of prior distributions. Finally, they compared the estimates obtained from two different loss functions.

Parsian and Nematollahi (1994) studied the estimation of a scale parameter under the entropy loss function ,

$$l(d,\theta) = \nu\left(\frac{\theta}{d} - \ln\frac{\theta}{d} - 1\right).$$

Which is a special case of (1.11) with $p = -1$. This loss function is strictly convex in $\frac{\theta}{d}$ but it is not convex in d, and as a function of d it has a unique minimum at $d = \theta$.

They obtained a Pitman type generalized Bayes estimator of $\theta = \tau^r$ with respect to a noninformative prior, $p(\theta) = \frac{1}{\tau}$ under the entropy loss considering a gamma, an inverse Gaussian, a Normal with zero mean, and an exponential distribution. They also considered the admissibility and inadmissibility of a class of linear estimator of the form $cT + d$ under this loss function. Later on, Cain and Janssen (1995) studied the real estate price prediction under the asymmetric loss functions. They used three different types of asymmetric loss functions namely asymmetric linear loss function by Granger (1969), asymmetric quadratic loss function by Cain (1991) and the asymmetric linex loss function Zellner (1986) to predict the selling price of houses through a real estate agents when the distribution of prediction error is normal. They compared their results and obtained the adjustment factor for the predictive mean under these three loss functions.

Calabria and Pulcini (1996) studied point estimation for left truncated exponential samples and used the linex loss function of Zellner (1986), modified linex loss function of Basu and Ibrahimi (1991) and general entropy loss function of Calabria and Pulcini (1994). They compared the general entropy and the modified linex loss function graphically. Moreover they obtained the Bayes estimate using a non-informative and a conjugate prior under these asymmetric loss functions and compared their result with that of ML estimators.

Cain and Janssen (1999) studied the market price of individual town house units where there is an asymmetry in the loss function. The selling price of single family dwellings is estimated by means of a linear model applied to a sample of comparable properties with explanatory variables including the number of rooms, and bathrooms,

the floor area, number of fireplaces, year of construction, garage, parking space, neighborhood location, freehold or leasehold etc. The estimated market price of an actual 16-unit town house is then adjusted for an asymmetry in the loss function and they observed that the overpricing is more serious than underpricing with optimal adjustment factors derived under the linex loss function.

Moore and Papadopoulos (2000) studied the Burr-XII distribution as a failure model under various loss functions. The two parameter Burr type XII distribution with pdf and distribution function in the respective forms

$$p(x|\theta,\beta) = \theta.\beta.x^{\beta-1}(1+x^{\beta})^{-(\theta+1)}, \theta, \beta > 0, \tag{1.19}$$

$$F(x|\theta,\beta) = 1-(1+x^{\beta})^{-\theta}, 0 \leq x, \theta, \beta > 0,$$

where θ and β are the parameters. Assuming β is known they obtained the Bayes estimate of the parameter θ and the reliability $R(t)$ under the loss functions (1.2),(1.3) and (1.4). Moreover, they considered the following priors for the parameter θ to obtain the Bayes estimate of $g(\theta)$.

1. gamma prior, $p_1(\theta) = \frac{m^{v+1}}{\Gamma(v+1)}\theta^v e^{-m\theta}$.
2. Jeffreys' improper prior, $p_2(\theta) \propto \frac{1}{\theta}$.

Finally, they compared their estimates obtained from three different loss functions using two different priors through simulation.

Soliman (2001) studied Bayes estimates under the linex and the quadratic error loss functions for the Pareto model. The density, reliability and failure rate of the two-parameter Pareto distribution with parameter (α, β) are given respectively by,

$$p(x|\alpha,\beta) = \alpha\beta^{\alpha}(x+\beta)^{-(\alpha+1)}; x > 0; (\alpha > 0, \beta > 0)$$

$$R(t) = \beta^{\alpha}(t+\beta)^{-\alpha}; t > 0.$$

$$H(t) = \alpha(t+\beta)^{-1}; t > 0,$$

where α is the shape parameter, β is the scale parameter, $R(t)$ is the reliability function and $H(t)$ is the failure rate. He considered the squared error and the linex loss function and obtained the Bayes estimate of α, $R(t)$ and $H(t)$ respectively when β is known. Then he considered the Bayesian estimation of the parameters α, β when both are unknown taking some suitable priors for α and β. In this situation he also obtained the estimates of $R(t)$ and $H(t)$. In both cases he used Lindley's (1980) approximation to obtain the estimates of the parameters. Finally, he compared the estimates obtained from a squared error loss function and a linex loss function with the usual ML estimators through simulation.

Soliman (2002) studied reliability estimation in a generalized life model using Burr-XII distribution given in (1.19). He considered some suitable prior distributions for (θ, β) and used the squared error loss, linex loss and general entropy loss functions (Calabria and Pulcini, 1996) to obtain the approximate Bayes estimate of the reliability function $R(t)$ through Lindley's (1980) approximation. Finally, he compared the estimate of the reliability function using a simulation study.

Marsh and Zellner (2004), studied the Bayesian solutions to the graduate admissions to achieve a targeted number of acceptances for its entering in any class who accepts an offer. Firstly, they obtained the point predictions of the numbers enrolling given the number of offers relative to the SE loss function and the linex loss function for a homogeneous set of data. They also did the same for the heterogeneous data like probabilities of enrolment associated with cost for n individuals.

Demetrescu (2006), extended the Gauss-Newton algorithm under an asymmetric loss function. Xiao, Takada and Shi (2005) obtained a minimax confidence bound of the normal mean under the linex loss function. Shi, Gao and Shi (2005) studied the convergence rate of the empirical Bayes estimation for two dimensional truncation parameters under a linex loss function. Chai et al.,(2007), studied the empirical Bayes estimators using a linex loss function under type-II censored samples. Farsipour and Zakerzadeh (2006) studied the estimation of generalized variance under an asymmetric loss function named the squared log error loss function. Moreover, Sengupta (2008) and Chattopadhyay (2000) used the linex loss function in sequential estimation problems, to determine the optimum process parameters by Cheng, Hung (2007), to estimate the failure rate of integrated circuit by Cain and Janssen (1995), the estimation of the market price of individual town house by Cain and Janssen (1999). Linex loss function is also used in obtaining failure rates by Schabe (1991, 1996).

We have discussed both the symmetric loss and asymmetric loss functions indicating their applications in different fields. As our plan is to obtain Bayesian SSD under symmetric and asymmetric loss functions so we will now review the literature related to the Bayesian SSD.

1.6 Literature review on Bayesian Sample Size Determination (SSD)

The problem of determining the size of a sample has received much attention in the literature. Frequentist methods are generally based on a power calculation. There are two main strands in the Bayesian literature. We will outline these two approaches and then give some more detail about each. The first is to look at measures which are analogous to power. For example Adcock (1988) obtained a closed form solution for determining sample size for the normal distribution with known and unknown variances by the average coverage criteria (ACC) method. Joseph, Wolfson, and Du Berger (1995) obtained the sample size for a binomial proportion considering the highest posterior density (HPD) interval. In this paper they also proposed a related method to ACC called the average length criteria (ALC) method where the coverage probability $1-\alpha$ is fixed and the length of HPD interval is equal to l on average. In the same paper they outlined another Bayesian method of determining sample size called worst outcome criteria (WOC) where both l and α are fixed in advance. Weiss (1997) obtained the sample size for a normal distribution with known variance using a Bayes factor. Adcock (1997) reviewed the SSD problem including frequentist and Bayesian methods. Pham-Gia and Turkkan (1992) obtained the sample size for a binomial proportion. Pham-Gia (1997) described a method of matching the ALC and maximization of expected utility (MEU). Joseph and Belisle (1997) obtained Bayesian sample size for a normal mean and difference between two normal means using all three ACC, ALC and WOC methods. Joseph, Du Berger and Belisle (1997) determine the sample size based on length and coverage of the posterior credible interval

to estimate the difference between two binomial proportions using the Bayesian and mixed Bayesian criteria. Stuger (2006) obtained the optimum sample size using an asymmetric loss function considering ACC and ALC method for a binomial distribution. Sahu and Smith (2006) discussed sample size determination with applications in clinical trials and in financial audits. They considered a bounded loss function but not any cost function in their study. Santis (2006) discussed SSD for a robust Bayesian analysis. Santis (2007) used historical data for Bayesian sample size determination. Wang and Gelfand (2002) discussed sample size considering a sampling prior and a fitting prior to get the posterior distribution for different models through simulation. All these methods discussed above are based on the probability coverage α and the length, l of the interval containing the parameter.

A second strand uses ideas taken from decision theory to combine the cost of taking a sample with the posterior expected loss and choose the sample size which minimizes the (pre-posterior) expected cost. There are choices to be made about the cost function and loss function. The first Bayesian account of this sample size determination method has been given by Raiffa and Schlaifer (1961) considering binomial and normal distributions. DeGroot (1970) obtained the Bayesian sample size considering squared error and absolute error loss function and a linear cost function. Lindley (1972) obtained the sample size using a squared error loss function for a normal distribution with known variance. Lindley (1997a) gave a clear discussion of the SSD problem through Maximization of Expected Utility (MEU). In the same paper he compared his result with the ALC method. There has been much debate between advocates of the ACC method and the MEU method about whether the ACC method is fully Bayesian or not. For a discussion paper see Adcock (1997a)

and Lindley (1997b). Muller and Parmigiani (1995) considered utility functions and optimized the expected utility by curve fitting of Monte Carlo samples for sample size determination. They also considered a cost function in their study. In the following we will discuss the optimum sample size based on power like calculations.

1.6.1 SSD based on power calculation

The Average Coverage Criteria (ACC) method was introduced by Adcock (1988). If the vector $\underline{x}$ denotes a sample of size n and θ is the parameter under study then conditional on $\underline{x}$, θ is contained in a specified interval $R(x)$ with probability $1-\alpha$ on an average over all possible samples. So defining $T(\underline{x}) = Pr[\theta \in R(\underline{x})|\underline{x}]$, we now determine n so that

$$E[T(\underline{x})] = \int \left\{ \int_{R(\underline{x})} p(\theta|\underline{x}, n) d\theta \right\} p(\underline{x}) d\underline{x} = 1 - \alpha.$$

Here $R(\underline{x})$ is a symmetric interval about the posterior mean $E[\theta|\underline{x}]$ of the form $R(\underline{x}) = E[\theta|\underline{x}] \pm l/2$ where l is the specified length. For a non-symmetric posterior distributions, Joseph et al.(1995) proposed a region of the form $R(\underline{x}) = [a, a+l]$ where l is given and a is determined so that $R(\underline{x})$ is the highest posterior density (HPD) interval. Joseph, Wolfson, and Du Berger, (1995) determined the binomial sample size via the highest posterior density (HPD) interval. A related method to ACC is called the average length criteria (ALC) method where the coverage probability $1-\alpha$ is fixed and the length of HPD interval is equal to l on average. In the same paper they also mentioned another Bayesian method of determining sample size called the worst outcome criteria (WOC) based on averaging over the marginal

distribution of x. The minimum sample size is obtained such that

$$inf_x \left\{ \int_{a(\underline{x},n)}^{a(\underline{x},n)+l} p(\theta|\underline{x},n)d\theta \right\} \geq 1-\alpha.$$

Where, $a(\underline{x},n)$ is the lower limit of the HPD credible set of length l for the posterior density $p(\theta|\underline{x},n)$ and α and l both are fixed in advance. However, Lindley mentioned that it is not a Bayesian rule as it didn't include the expected value operation over the marginal distribution of $\underline{x}$.

Weiss (1997) obtained the sample size for a normal distribution with known variance using Bayes factor, i.e.

$$B = \ln[p(\underline{x}|H_0)/p(\underline{x}|H_1)].$$

He obtained the sample size by fixing α and β and then solving a pair of equations:

$$Pr[B > B_{cut}|H_0] = 1-\alpha,$$

$$Pr[B \leq B_{cut}|H_1] = 1-\beta,$$

for given H_0 and H_1 to determine both the sample size n and the cut-off point B_{cut}.

1.6.2 Simulation based approach

Weng and Gelfand (2002) proposed a simulation based approach to Bayesian SSD for performance under a given model and for separating models. They discussed different models including a survival model with censoring and a logistic regression model. To obtain the SSD they introduced the idea of a sampling prior and fitting priors to get the posterior distributions for different models where they suggested

the sampling prior would be informative and the fitting prior is non-informative. Sahu and Smith (2006) used this approach and they distinguished between these two priors. The sampling prior is used to generate the parameter values which are then conditioned to generate the data from $p(x|\theta)$ in substantive experiments, i.e. data $\underline{x}$ are generated from the joint hierarchical model $\pi^{(s)}$. Once data are available we would like to pretend that the informative prior distribution which generated the data is unknown to us, and we would like to make an inference with assumption of relatively non-informative prior distribution. They also considered a bounded type loss function for an optimum sample size.

1.6.3 SSD using utility functions

Lindley (1997a) gave a clear direction of SSD problem through Maximization of Expected Utility (MEU) considering a linear cost function described below. Let, x_1,x_2,...,x_n be a sample of size n taken from the density $p(x|\theta)$ with unknown parameter θ. If $p(\underline{x}|\theta, n)$ is the likelihood of the data and $p(\theta)$ is the prior density of θ then the posterior density of θ will be,

$$p(\theta|\underline{x}, n) = \frac{p(\underline{x}|\theta, n)p(\theta)}{p(\underline{x}|n)}. \tag{1.20}$$

Now consider a utility function $u(n, \underline{x}, d, \theta)$ so the sample size will be found first taking expectations of utility over random quantities θ and x and then maximizing over deterministic quantities d and n. From (1.20) we can find $p(\theta|d, \underline{x}, n)$, where the density $p(\theta|d, \underline{x}, n)$ is the same as $p(\theta|\underline{x}, n)$ since the choice of d does not influence the distribution of θ, given $\underline{x}$ and n. Then we maximize the expected utility over d to select the terminal decision. Going backward in time, the expectation over $\underline{x}$ of

this maximized value can be found by using $p(\underline{x}|n)$. Finally, this expectation can be maximized over n to give the optimum sample size. It is the solution of

$$\max_n \left[\sum_x \max_d \left\{ \int u(n, \underline{x}, d, \theta) p(\theta|\underline{x}, n) d\theta \right\} p(\underline{x}|n) \right]. \tag{1.21}$$

In the continuous case the summation will be replaced by an integration. Now by Raiffa and Schlaifer (1961), the utility does not depend on x and is additive and linear in n as follows,

$$u(n, \underline{x}, d, \theta) = u(d, \theta) - cn - c_0, \tag{1.22}$$

where c is the cost in utiles of each observation, c_0 is the sampling set-up cost. It is usual for each observation to cost the same, so justifying the linearity in n. Now using special form of utility, (1.22), (1.21) becomes,

$$\max_n \left[\sum_x \max_d \left\{ \int u(d, \theta) p(\theta|\underline{x}, n) d\theta \right\} p(\underline{x}|n) - cn - c_0 \right]. \tag{1.23}$$

Now using the result of equation (1.20) in (1.23) the final expression for getting an optimum sample of size n will be,

$$\max_n \left[\sum_x \max_d \left\{ \int u(d, \theta) p(\underline{x}|\theta, n) p(\theta) d\theta \right\} - cn - c_0 \right]. \tag{1.24}$$

This method of determining sample size termed as maximization of expected utility (MEU). In this method MEU is used twice, first over θ and d then over x and n. In the same paper Lindley compared his result with the ALC method with discussions. In our case, we will use a loss function rather than a utility function, $u(n, \underline{x}, d, \theta)$ where the form of loss will be,

$$-u(n, \underline{x}, d, \theta) = l(d, \theta) + cn + c_0 \tag{1.25}$$

and the final form to get a minimum sample size will be,

$$\min_n \left[\sum_x \max_d \left\{ \int l(d, \theta) p(\underline{x}|\theta, n) p(\theta) d\theta \right\} + cn + c_0 \right]. \tag{1.26}$$

Muller and Parmigiani (1995) discussed the Bayesian optimal design via curve fitting of Monte Carlo experiments. They described a class of optimization algorithms designed to gain efficiency in such situations, by exploring smoothness of the expected utility surface and borrowing information from neighbouring design points. They considered the different forms of utility functions and the cost function in different situations. Now we discuss the pre posterior analysis as we need it later for an optimum SSD.

1.7 Pre-posterior analysis

Suppose, $\underline{x} = (x_1, \ldots, x_n)$ is a random sample from the distribution with pdf $p(x_i|n, \theta)$ and the prior for θ is $p(\theta)$. If the likelihood of the sample is denoted by $p(\underline{x}|n, \theta)$ then the pre-posterior distribution of $\underline{x}|n$ will be,

$$p(\underline{x}|n) = \int p(\underline{x}|\theta, n)p(\theta)d\theta.$$

In determining the optimum sample size we will consider the posterior risk (PR) defined earlier which may sometimes depend on $\underline{x}$. Then we need to average over $\underline{x}$ knowing the pre-posterior distribution of $\underline{x}|n$. So to make the PR independent of $\underline{x}$ we need to find an average posterior risk (APR) by

$$\int \text{PR } p(\underline{x}|n)d\underline{x}. \tag{1.27}$$

Then we will add a linear cost function to the risk and minimize them together to get the optimum sample size. Here are the details of the cost functions we will be using for an optimum sample size.

1.8 Cost function

Sampling cost per unit is one of the very important ingredients to consider to get an optimum sample size. In this study we will consider a linear cost function given by Lindley (1972) as,

$$C(n) = c_0 + cn, \qquad n > 0 \tag{1.28}$$

and

$$C(0) = 0.$$

Here c_0 is the sampling set-up cost or any other related cost involve in sampling and c is the sampling cost per unit. If there is no plan to sample at all, obviously there is no set-up cost or sampling related cost. If the posterior risk function for the Bayes estimates does not depend on $\underline{x}$ then we will add it to the cost function (1.28) to obtain the total cost. That is we use

$$TC(n) = C(n) + PR.$$

On the other hand if it does depend on $\underline{x}$ then we find the APR using (1.27) and then we add it to the cost function (1.28) for the expected total cost. that is in this situation we use

$$E[TC(n)] = C(n) + APR.$$

The problem that costs of sampling and APR are measured in different units will be discussed in chapter 8, section 8.3. If we want to draw samples from two populations then we need to consider the following cost function. Let c_1 be the cost of recruiting units of size n_1 from one population and c_2 be the cost of recruiting units of size n_2 from another population. For example, to compare the difference between two

treatment means, we consider a cost function,

$$C(n) = c_0 + c_1 n_1 + c_2 n_2. \tag{1.29}$$

If we want to have an equal number of units then consider $n_1 = n_2$, if not then consider $n_1 \neq n_2$. The sampling cost per unit could be the same for the samples drawn from both populations so take, $c_1 = c_2$ if different, then consider $c_1 \neq c_2$. Note that, in finding a minimum cost, we usually treat n as a real quantity so are able to differentiate $TC(n)$ with respect to it to find a turning point n^*. In practice, of course, n is a positive integer and we must check whether $\llcorner n^* \lrcorner$ or $\ulcorner n^* \urcorner$ gives the smaller value.

1.9 Lindley's conjugate utility

Alongside the Bayesian optimum sample size determination we will obtain an approximate optimum decisions under the Lindley's (1976) conjugate utility function for one parameter exponential family. In the same paper Lindley noted that it is possible to extend the idea for several parameters. To do this first we propose a two parameter conjugate utility function which fits nicely with the bivariate exponential family. Then we obtain the approximate optimum decisions for both parameters of the bivariate exponential family under the proposed conjugate utility function when both parameters are unknown. In the following we will review Lindley's (1976) conjugate utility functions.

If x is a random variable depending on a single parameter θ for a suitable param-

eterizations the density of x, given θ, is

$$p(x|\theta) = \exp(x\theta)H(x)G(\theta), \tag{1.30}$$

where $H(x)$ is some non negative function and $G(\theta)^{-1} = \int \exp(x\theta)H(x)dx$ for all θ for which the integral is finite. The natural conjugate family for θ by Raiffa and Schlaifer (1961) is proportional to

$$p(\theta) \propto \exp(x_0\theta)G(\theta)^{n_0}$$

for suitable x_0 and n_0. Defining

$$K(n_0, x_0)^{-1} = \int \exp(x_0\theta)G(\theta)^{n_0}d\theta, \tag{1.31}$$

the integral being over the relevant θ-values, the conjugate density of θ, given n_0 and x_0 is

$$p(\theta|n_0, x_0) = \exp(x_0\theta)G(\theta)^{n_0}K(n_0, x_0), \tag{1.32}$$

where $K(n_0, x_0)$ is defined in (1.31). Now if we draw a random sample of size n from (1.30), and if the distribution of θ prior to sample is given by (1.32), then the distribution after the sample is

$$p(\theta|\underline{x}, n) \propto \exp(\theta \sum x_i)G(\theta)^{n+n_0}, \tag{1.33}$$

where the summations are from zero (not one) to n. Besides a distribution of θ we need to introduce a utility function $u(d, \theta)$. A convenient utility function which uses the distribution of θ and fits nicely with x is

$$u(d, \theta) = \exp\{x(d)\theta\}G(\theta)^{n(d)}F(d) \tag{1.34}$$

where $x(d)$, $n(d)$ and $F(d)$ are suitable functions of d whose form we will discuss below. Consider the maximum of (1.34) for a fixed d. Let, $g(\theta) = \log G(\theta)$ so we

have,

$$\log u(d,\theta) = x(d)\theta + n(d)g(\theta) + \log F(d)$$

Then take,

$$\frac{\partial \log U(d,\theta)}{\partial \theta} = x(d) + n(d)g'(\theta).$$

We see that the logarithmic derivative of (1.34) vanishes if

$$x(d) + n(d)g'(\theta) = 0. \tag{1.35}$$

It would be natural in most applications for such a maximum to occur at $\theta = d$, that is for the decision being taken, d is the best decision for θ. So from (1.35) we have,

$$x(d) = -n(d)g'(d) \tag{1.36}$$

referring to this condition C_1, and using it to eliminate $x(d)$. Using this condition in (1.34), we have the maximum utility will be,

$$u(d,d) = \exp[n(d)g(d) - g'(d)d]F(d), \tag{1.37}$$

Taking the maximum utility as 1, let $f(d) = \log F(d)$ and define it as

$$f(d) = n(d)[g'(d)d - g(d)], \tag{1.38}$$

referring to this condition as C_2. Now using the condition (1.38) in (1.37) we have the final expression of Lindley's utility function as follows,

$$u(d,\theta) = \exp[n(d)\{g(\theta) - g(d) - g'(d)(\theta - d)\}] \tag{1.39}$$

and only $n(d)$ is free to be selected. Finally, he gave an outline of finding $n(d)$ where he considered $n(d)^{-1} = -kg''(d)$ for some constant k and noted it as as condition C_3. He then maximized the the expected utility (1.39) under some approximations which we will discuss in chapter 7. Then we will extend these results to the two parameter exponential family.

1.10 Conclusion

In this chapter we have reviewed the literature related to the loss functions, utility function and the Bayesian sample size determination. We have discussed both the symmetric and the asymmetric loss functions. We found most of the SSD has been done under the symmetric squared error (SE) loss function but we haven't seen any SSD under any of the asymmetric loss functions. So we will review SSD under the symmetric SE loss function in Chapter 2, then we will deal with the Bayesian SSD problem under the asymmetric loss functions in the Chapters 3-4. In chapter 5 we will discuss the optimum SSD under the loss function from DeGroot(1970). In Chapter 6, we will also deal with SSD problem but under our proposed utility (or loss) function. Since the maximization expected utility (MEU) function is equivalent to the minimization expected loss function. We will deal with the Bayesian SSD problem through the minimization of expected loss function. This method includes the following elements.

i) Prior distribution

ii) Current data following a distribution

iii) Marginal or Pre-posterior distribution

iv) Posterior distribution

v) Loss function

vi) Posterior risk function

vii) Cost function

Finally, adding vi) and vii),we will minimize both the costs the posterior risks together to get the optimum sample of size n. In Chapter 7, we will discuss and extend

Lindley's (1976) paper on conjugate utility function for one parameter. We will present how to estimate the parameters of the bivariate exponential family when both parameters are unknown under a two parameter conjugate utility function. Finally in Chapter 8 we will discuss practical implications together with the limitations of the research done and provide some suggestions for further research.

Chapter 2

Bayesian SSD using squared error loss

2.1 Introduction

Sample Size Determination (SSD) is an important issue to consider when estimating any parameter. A number of researchers have studied the Bayesian SSD problem. One group have considered utility (or loss) functions and cost functions in their SSD problems and the others have considered power calculations which we discussed in the previous chapter. A very common approach is to consider a squared (SE) error loss function for an optimum sample size. In this chapter we will discuss the optimum sample size under this commonly used SE loss function. We will explore the normal, exponential, Poisson and binomial distributions. For the normal distribution we

consider the optimum sample size for different situations. To obtain the optimum sample size our objective is to minimize the posterior risk function and a linear cost function (1.28) described in previous chapter. We will also discuss the situation where it is not worth sampling due to a high sampling cost or strong prior information. We will present all the figures and tables at the end of the chapter.

2.2 SSD to estimate a normal mean when precision is known (Lindley, 1972)

Suppose x_1, x_2,. . . ,x_n is a random sample of size n taken from a normal distribution with mean θ and known precision ξ. So the likelihood of the sample will be,

$$p(\underline{x}|\theta) = \left(\frac{\xi}{2\pi}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\xi\Sigma_{i=1}^{n}(x_i-\theta)^2\right].$$

We can equivalently write,

$$p(\underline{x}|\theta) \propto \exp(n\overline{x}\xi\theta)\exp\left(-\frac{1}{2}n\xi\theta^2\right). \tag{2.1}$$

Let us take a conjugate prior of θ with mean μ_0 and precision $n_0\xi$ as,

$$p(\theta) \propto \exp(n_0\mu_0\xi\theta)\exp\left(-\frac{1}{2}n_0\xi\theta^2\right). \tag{2.2}$$

Now combining (2.1) and (2.2) the posterior distribution will be,

$$p(\theta|\underline{x}) \propto \exp\left\{(n\overline{x}+n_0\mu_0)\xi\theta\right\}\exp\left\{-\frac{1}{2}(n+n_0)\xi\theta^2\right\}; \tag{2.3}$$

which is normal with mean $\frac{n\overline{x}+n_0\mu_0}{n+n_0}$ and precision $(n+n_0)\xi$. So under the SE loss (1.1), from (1.15) we have the posterior risk function as,

$$PR = \frac{a_0}{(n+n_0)\xi}. \tag{2.4}$$

Now adding the cost function (1.28) to the risk function (2.4) we have the total cost as,

$$\begin{aligned} TC(n) &= c_0 + cn + PR \\ &= c_0 + cn + \frac{a_0}{(n+n_0)\xi}, \end{aligned} \tag{2.5}$$

which is clearly independent of $\underline{x}$. It is to be noted that c_0, c and a_0 have to be chosen so that losses and costs are measured in the same units (we have discussed on this point in chapter 8 section 8.3). Now consider n is a real positive number. To have a minimum sample of size n differentiate $TC(n)$ w.r.t. n and setting equal to zero , i.e., $\frac{\partial TC(n)}{\partial n} = 0$ gives

$$n^*_{se} = \sqrt{\frac{a_0}{c\xi}} - n_0. \tag{2.6}$$

If there is little prior information, that is $n_0 \to 0$ in (2.6), then the optimum sample size under the SE loss function will be,

$$n^*_{se} = \sigma\sqrt{\frac{a_0}{c}}, \tag{2.7}$$

where a_0 is the scale of the loss function. It is also clear that the sample size is proportional to the standard deviation of the data but inversely proportional to the sampling cost per unit. If the data variability increases the size of the sample also increases but if the sampling cost increases then the optimum sample size decreases.

2.2.1 No sampling situation

Now if we do not sample the total cost is given by

$$TC(0) = \frac{a_0}{n_0\xi}.$$

If we take a sample of size n^* then the total cost of sampling (giving $n = n^*_{se}$ in (2.5)) will be,

$$TC(n^*) = c_0 - cn_0 + 2\sqrt{\frac{ca_0}{\xi}}.$$

Therefore the overall optimum sample size is

$$n^* = \max\left\{0,\ \sqrt{\frac{a_0}{c\xi}} - n_0\right\}$$

The truncation at 0 matters and we automatically choose not to sample when $\sqrt{\frac{a_0}{c\xi}} \leq n_0$ in addition to the situation when $\sqrt{\frac{a_0}{c\xi}} > n_0$ and $TC(0)$ is exceeded by $TC(n^*)$. This implies the decision not to sample occurs when

$$\left[\frac{a_0}{n_0\xi} < c_0 - cn_0 + 2\sqrt{\frac{ca_0}{\xi}} \text{and } \sqrt{\frac{a_0}{c\xi}} > n_0\right] \text{ or } \sqrt{\frac{a_0}{c\xi}} \leq n_0.$$

Equivalently the decision to sample occurs when

$$\frac{a_0}{n_0\xi} > c_0 - cn_0 + 2\sqrt{\frac{ca_0}{\xi}} \text{ and } \sqrt{\frac{a_0}{c\xi}} > n_0. \tag{2.8}$$

Solving for $y = \sqrt{c}$ we find that we sample when

$$n_0y^2 - 2\sqrt{\frac{a_0}{\xi}}y + \frac{a_0}{n_0\xi} - c_0 > 0 \text{ and } y < \sqrt{\frac{a_0}{n_0^2\xi}}.$$

So that

$$\left[y < \sqrt{\frac{a_0}{n_0^2\xi}} - \sqrt{\frac{c_0}{n_0}} \text{ or } y > \sqrt{\frac{a_0}{n_0^2\xi}} + \sqrt{\frac{c_0}{n_0}}\right] \text{ and } y < \sqrt{\frac{a_0}{n_0^2\xi}},$$

which implies that

$$\sqrt{c} < \sqrt{\frac{a_0}{n_0^2\xi}} - \sqrt{\frac{c_0}{n_0}}.$$

So choose not to sample when

$$c > \left[\sqrt{\frac{a_0}{n_0^2\xi}} - \sqrt{\frac{c_0}{n_0}}\right]^2.$$

That is for known a_0, c_0, ξ, and n_0 if the sampling cost per unit, c exceeds $\left(\frac{1}{n_0}\sqrt{\frac{a_0}{\xi}} - \sqrt{\frac{c_0}{n_0}}\right)^2$ then it is not worth sampling at all due to high sampling cost per unit.

Now we will obtain the range of n_0 so that the decision not to sample occurs. From (2.8) the decision to sample occurs when

$$cn_0^2 - \left\{c_0 + 2\sqrt{\frac{ca_0}{\xi}}\right\} n_0 + \frac{a_0}{\xi} > 0 \text{ and } n_0 < \sqrt{\frac{a_0}{c\xi}}.$$

So that

$$\left[n_0 < \frac{c_0}{2c} + \sqrt{\frac{a_0}{c\xi}} - \frac{1}{2c}\sqrt{c_0^2 + 4c_0\sqrt{\frac{ca_0}{\xi}}} \text{ or } n_0 > \frac{c_0}{2c} + \sqrt{\frac{a_0}{c\xi}} + \frac{1}{2c}\sqrt{c_0^2 + 4c_0\sqrt{\frac{ca_0}{\xi}}}\right]$$

and

$$n_0 < \sqrt{\frac{a_0}{c\xi}}.$$

This implies that the optimal decision is to sample when

$$n_0 < \frac{c_0}{2c} + \sqrt{\frac{a_0}{c\xi}} - \frac{1}{2c}\sqrt{c_0^2 + 4c_0\sqrt{\frac{ca_0}{\xi}}}.$$

So we choose not to sample when

$$n_0 > \frac{c_0}{2c} + \sqrt{\frac{a_0}{c\xi}} - \frac{1}{2c}\sqrt{c_0^2 + 4c_0\sqrt{\frac{ca_0}{\xi}}},$$

where a_0, c_0, c and ξ are known.

2.2.2 Numerical study

From figure 2.1, we can see that if the sampling cost per unit c goes up, then the optimum sample size goes down. It is also clear that if the sampling cost per unit

is too high then it is not worth sampling at all. When the precision goes up this implies that the variability within data is reduced so that the optimum sample size goes down. Moreover, if we have bigger prior size of samples then it is not worth sampling to estimate a normal mean under the SE loss function.

2.3 SSD to estimate a normal precision when mean is known

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean μ_0 and precision ξ. So the likelihood of the sample will be,

$$p(\underline{x}|\xi) \propto \xi^{n/2} \exp\left[-\frac{1}{2}\xi \sum (x_i - \mu_0)^2\right]. \tag{2.9}$$

Let us take a gamma conjugate prior as,

$$p(\xi) \propto \xi^{\alpha-1} \exp(-\beta\xi) \tag{2.10}$$

Now combining (2.9) and (2.10) the posterior distribution will be,

$$p(\xi|\underline{x}) \propto \xi^{\frac{n}{2}+\alpha-1} \exp\left\{-\frac{1}{2}\xi\left[\sum (x_i - \mu_0)^2 + 2\beta\right]\right\}. \tag{2.11}$$

Let, $t = \sum (x_i - \mu_0)^2$, so the posterior distribution of $\xi|t$ has a $Gamma(\alpha + \frac{n}{2}, \beta + \frac{t}{2})$ distribution with mean, $\frac{\alpha+\frac{n}{2}}{\beta+\frac{t}{2}}$ and variance, $\frac{\alpha+\frac{n}{2}}{(\beta+\frac{t}{2})^2}$. Now we want to estimate ξ by d. So following (1.1) the squared error loss function becomes,

$$l(d, \xi) = a_0(d - \xi)^2.$$

Now from (1.15) the posterior risk will be,

$$PR = \frac{a_0\left(\alpha + \frac{n}{2}\right)}{\left(\beta + \frac{t}{2}\right)^2}, \tag{2.12}$$

Note that t is a sufficient statistic for ξ and the pre-posterior distribution of t follows a Gamma-Gamma distribution with parameters α, 2β and $\frac{n}{2}$ (Bernardo and Smith 1994). We have,

$$p(t) = \frac{ct^{\frac{n}{2}-1}}{(2\beta+t)^{\alpha+\frac{n}{2}}}, t > 0; \tag{2.13}$$

where,

$$c = \frac{(2\beta)^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma\left(\alpha+\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Consider the cost function (1.28) and adding this cost with the risk function (2.12), we have the total cost as,

$$TC(n) = c_0 + cn + PR = c_0 + cn + \frac{a_0\left(\alpha+\frac{n}{2}\right)}{\left(\beta+\frac{t}{2}\right)^2}.$$

Now we will find the APR by taking expectation over (2.12) w.r.t. t using the pre-posterior distribution (2.13). That is,

$$E_t[TC(n)] = c_0 + cn + a_0 E_t\left[\frac{\alpha+\frac{n}{2}}{(\beta+\frac{t}{2})^2}\right]. \tag{2.14}$$

Now find

$$\begin{aligned}
E_t\left[\frac{\alpha+\frac{n}{2}}{(\beta+\frac{t}{2})^2}\right] &= \frac{(2\beta)^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\frac{n}{2})}{\Gamma(\frac{n}{2})} \int \frac{(\alpha+\frac{n}{2})}{(\beta+\frac{t}{2})^2} \times \frac{t^{\frac{n}{2}-1}}{(2\beta+t)^{\alpha+\frac{n}{2}}} dt \\
&= \frac{(2\beta)^{\alpha}}{\Gamma(\alpha)} \frac{4(\alpha+\frac{n}{2})\Gamma(\alpha+\frac{n}{2})}{\Gamma(\frac{n}{2})} \int \frac{t^{\frac{n}{2}-1}}{(2\beta+t)^{\alpha+2+\frac{n}{2}}} dt \\
&= \frac{(2\beta)^{\alpha}}{\Gamma(\alpha)} \frac{4(\alpha+\frac{n}{2})\Gamma(\alpha+\frac{n}{2})}{\Gamma(\frac{n}{2})} \frac{\Gamma(\alpha+2)\Gamma(\frac{n}{2})}{(2\beta)^{\alpha+2}\Gamma(\alpha+2+\frac{n}{2})} \\
&= \frac{\alpha(\alpha+1)}{\beta^2\left(\alpha+1+\frac{n}{2}\right)}.
\end{aligned} \tag{2.15}$$

Using this result in (2.14), we have the expected total cost as,

$$E_t[TC(n)] = c_0 + cn + \frac{a_0\alpha(\alpha+1)}{\beta^2\left(\alpha+\frac{n}{2}+1\right)}. \tag{2.16}$$

To obtain a minimum sample of size n setting $\frac{\partial E_t[TC(n)]}{\partial n} = 0$ we have,

$$n^*_{se} = 2\left[\sqrt{\frac{a_0\alpha(\alpha+1)}{2\beta^2 c}} - (\alpha+1)\right], \tag{2.17}$$

which is the optimum sample of size n to estimate a normal precision when mean is known under the SE loss function.

2.3.1 No sampling situation

From (2.16), the expected total cost of sampling at $n = 0$ is

$$E_t[TC(0)] = \frac{a_0\alpha}{\beta^2}.$$

Again from (2.16) the expected total cost of sampling at $n = n^*_{se}$ is

$$E_t[TC(n^*)] = c_0 - 2c(\alpha+1) + \frac{2\sqrt{2a_0c\alpha(\alpha+1)}}{\beta}.$$

Now the optimum sample size (2.17) should be,

$$n^* = \max\left\{0,\ 2\sqrt{\frac{a_0\alpha(\alpha+1)}{2\beta^2 c}} - 2(\alpha+1)\right\}.$$

Choose not to sample when $\sqrt{\frac{a_0\alpha(\alpha+1)}{2\beta^2 c}} \leq \alpha+1$ i.e.,$\sqrt{\frac{a_0\alpha}{2\beta^2(\alpha+1)}} \leq \sqrt{c}$ in addition to the situation when $\sqrt{\frac{a_0\alpha(\alpha+1)}{2\beta^2 c}} > \alpha+1$ i.e., $\sqrt{\frac{a_0\alpha}{2\beta^2(\alpha+1)}} > \sqrt{c}$ and $E_t[TC(0)]$ is exceeded by $E_t[TC(n^*)]$. This implies that the decision not to sample occurs when

$$\left[\frac{a_0\alpha}{\beta^2} < c_0 - 2c(\alpha+1) + \frac{2\sqrt{2a_0c\alpha(\alpha+1)}}{\beta} \text{ and } \sqrt{\frac{a_0\alpha}{2\beta^2(\alpha+1)}} > \sqrt{c}\right]$$

or

$$\sqrt{\frac{a_0\alpha}{2\beta^2(\alpha+1)}} \leq \sqrt{c}.$$

Equivalently the decision to sample occurs when

$$\frac{a_0\alpha}{\beta^2} > c_0 - 2c(\alpha+1) + \frac{2\sqrt{2a_0c\alpha(\alpha+1)}}{\beta} \text{ and } \sqrt{\frac{a_0\alpha}{2\beta^2(\alpha+1)}} > \sqrt{c}.$$

Now solving for $y = \sqrt{c}$ we obtain that we sample when

$$2c(\alpha+1)y^2 - \frac{2y\sqrt{2a_0\alpha(\alpha+1)}}{\beta} + \frac{a_0\alpha}{\beta^2} - c_0 > 0 \text{ and } y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}}$$

So that

$$\left[y > \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}} + \sqrt{\frac{c_0}{2(\alpha+1)}} \text{ or } y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{2(\alpha+1)}}\right]$$

and

$$y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}}.$$

Which implies that

$$\sqrt{c} < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{2(\alpha+1)}}.$$

So choose not to sample when

$$c > \left[\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{2(\alpha+1)}}\right]^2.$$

That is for known a_0, c_0, α and β if the sampling cost per unit exceeds the amount $\left[\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{2(\alpha+1)}}\right]^2$ then it is not worth sampling due to the high sampling cost to estimate a normal precision when mean is known under the SE loss function.

2.3.2 Numerical study

From figure 2.2, we observe that if the sampling cost per unit goes up the optimum sample size goes down whatever the value of α or β. From graph it is also clear that for $\alpha > \beta$ the sample size is higher than $\alpha = \beta$. On the other hand for $\alpha < \beta$ the optimum sample size is lower than the case when $\alpha = \beta$. From figure 2.3 it is clearly seen that if the value of prior parameter β goes up, then the optimum sample size goes down for fixed values of c and α. When we kept c and β fixed and increased α then we see that the bigger the value of α gives bigger the sample size compared to the smaller values of α.

2.4 SSD to estimate the difference between two normal means

Suppose, $\underline{x} = x_1, x_2,....,x_{n_1}$ is a random sample of size n_1 with mean μ_1 and precision ξ_1 and $\underline{y} = y_1, y_2,....,y_{n_2}$ is another random sample of size n_2 with mean μ_2 and precision ξ_2 taken from a normal distribution. Consider prior distributions for $\mu_1|\xi_1$ as $N(\mu_{01}, n_{01}\xi_1)$ and for $\mu_2|\xi_2$ as $N(\mu_{02}, n_{02}\xi_2)$. We will determine the sample size to estimate the difference between two means, θ, where $\theta = \mu_1 - \mu_2$, under the squared error loss function. First of all we will assume the two populations have an equal known precision $\xi_1 = \xi_2 = \xi$ (say). Secondly, we will consider known but unequal precisions. In this situation we will consider $n_1 \neq n_2$.

2.4.1 SSD when precision is known and common

The posterior density of $\theta = \mu_1 - \mu_2$ having observed data X and Y to estimate $\theta = \mu_1 - \mu_2$ given n_1 and n_2 is

$$\theta|\underline{x}, \underline{y} \sim N\left\{m_2 - m_1, \frac{\xi(n_{01}+n_1)(n_{02}+n_2)}{n_1+n_2+n_{01}+n_{02}}\right\}, \tag{2.18}$$

where m_1 and m_2 are the estimates of μ_1 and μ_2 respectively and ξ is a common, known precision. Let m be the posterior mean where, $m = m_2 - m_1$, $m_1 = \frac{n_{01}\mu_{01}+n_1\overline{x}}{n_{01}+n_1}$, $m_2 = \frac{n_{02}\mu_{02}+n_2\overline{y}}{n_{02}+n_2}$ and the posterior variance, $v^2 = \frac{n_1+n_2+n_{01}+n_{02}}{\xi(n_{01}+n_1)(n_{02}+n_2)}$. Now from (1.15) the posterior risk under the SE loss function (1.1) will be,

$$PR = \frac{a_0(n_1+n_2+n_{01}+n_{02})}{\xi(n_{01}+n_1)(n_{02}+n_2)}.$$

For simplicity let us take $a_0 = 1$ and adding this posterior risk with the cost function (1.29) we have the total cost as,

$$TC(n) = c_0 + c_1 n_1 + c_2 n_2 + \frac{n_1+n_2+n_{01}+n_{02}}{\xi(n_{01}+n_1)(n_{02}+n_2)}. \tag{2.19}$$

Now consider the case that both the sample sizes are the same and the prior sample sizes are also the same but the sampling costs per unit are different when we draw the samples from two populations. That is $n_1 = n_2 = n$ (say), $n_{01} = n_{02} = n_0$ (say), We have from (2.19),

$$TC(n) = c_0 + (c_1+c_2)n + \frac{2}{\xi(n_0+n)}. \tag{2.20}$$

To get the optimum sample size n differentiate (2.20) w.r.t. n and setting equal zero we have,

$$n^* = \sqrt{\frac{2}{\xi(c_1+c_2)}} - n_0; \tag{2.21}$$

which is the optimum sample of size n per group when the equal size of the prior and the equal size of the posterior samples are taken from both populations and the sampling cost per unit are the same within the population but different between populations. If the sampling costs per unit are the same within and between populations then we have $c_1 = c_2 = c$ (say) and from equation (2.21), we can work out the optimum sample size as,

$$n^* = \frac{1}{\sqrt{c\xi}} - n_0. \tag{2.22}$$

Now we consider the optimum sample size when the equal number of samples will be taken from both populations but the prior sample sizes are unequal with a common precision ξ. That is $n_1 = n_2 = n$, (say) $n_{01} \neq n_{02}$ and $c_1 \neq c_2$. So from (2.19) we have,

$$TC(n) = c_0 + (c_1 + c_2)n + \frac{2n + n_{01} + n_{02}}{\xi(n_{01} + n)(n_{02} + n)} \tag{2.23}$$

Now differentiating (2.23) w.r.t. n and setting equal zero we have,

$$A_1 n^4 + B_1 n^3 + C_1 n^2 + D_1 n + E_1 = 0; \tag{2.24}$$

where,

$A_1 = (c_1 + c_2)$, $B_1 = 2A_1 N_0$, $C_1 = A_1 N_0^2 + 2A_1 N_0' - 2k_1$, $D_1 = 2A_1 N_0 N_0' - 2k_1 N_0$, $E_1 = A_1 N_0'^2 + 2k_1 N_0' - N_0{}^2$, $N_0 = n_{01} + n_{02}$, $N_0' = n_{01} n_{02}$, and $k_1 = \frac{1}{\xi}$.

We can easily solve the equation (2.24) for given, c_1, c_2 , n_{01}, n_{02} , and ξ to determine the sample of size n. Clearly, here the sampling cost per unit within the population is the same but between the populations is different and the prior sample sizes are different as well. To solve the equation numerically we assumed, $n_{01} = n_{02} = 10$, $\xi = 1$. We solved this polynomial equation using Maple 13 assuming $c_1 + c_2 = c$. We found four roots of n and out of these four roots, three roots are negative and only one

root is positive. We have considered that positive root as the optimum sample size for a given value of c' . Clearly from figure 2.4, we see that if the sampling cost goes up, the optimum sample size goes down. We observe that it is not worth sampling when sampling cost per unit $c > 0.01$.

Now we will find the optimum sample sizes (n_1, n_2) jointly when $n_1 \neq n_2, n_{01} \neq n_{02}$, $c_1 \neq c_2$ but $\xi_1 = \xi_2 = \xi$. To do so let us differentiate (2.19) w.r.t. n_1 first and setting equal zero we have,

$$n_1^* = \frac{1}{\sqrt{c_1 \xi}} - n_{01}. \tag{2.25}$$

Again differentiate (2.19) w.r.t. n_2, we have,

$$n_2^* = \frac{1}{\sqrt{c_2 \xi}} - n_{02}. \tag{2.26}$$

We will have a pair of the optimum sample of size (n_1^*, n_2^*) from (2.25) and (2.26) which jointly minimizes the total cost of (2.19).

2.4.2 SSD when precision is known but unequal

Now we will determine the optimum sample size to estimate the difference between two normal means by θ, where $\theta = \mu_1 - \mu_2$, under the squared error loss function (1.1) considering $\xi_1 \neq \xi_2$. For this case the posterior density observing the data $\underline{x}$ and $\underline{y}$ to estimate $\theta = \mu_1 - \mu_2$ will be,

$$\theta | \underline{x}, \underline{y} \sim N \left\{ m_2 - m_1, \frac{\xi_1 \xi_2 (n_{01} + n_1)(n_{02} + n_2)}{\xi_1 (n_1 + n_{01}) + \xi_2 (n_2 + n_{02})} \right\}, \tag{2.27}$$

where, $\xi_1 \neq \xi_2$ but known, $m_1 = \frac{n_{01}\mu_{01} + n_1 \overline{x}}{n_{01} + n_1}$, $m_2 = \frac{n_{02}\mu_{02} + n_2 \overline{y}}{n_{02} + n_2}$. Here the posterior variance is, $\frac{\xi_1 (n_1 + n_{01}) + \xi_2 (n_2 + n_{02})}{\xi_1 \xi_2 (n_{01} + n_1)(n_{02} + n_2)}$ which is the posterior risk under the SE loss function.

Now adding the cost function (1.28) with the variance we have the total cost,

$$TC(n) = c_0 + c_1 n_1 + c_2 n_2 + \frac{\xi_1(n_1 + n_{01}) + \xi_2(n_2 + n_{02})}{\xi_1 \xi_2 (n_{01} + n_1)(n_{02} + n_2)}. \tag{2.28}$$

Let choosing samples from population X are independent of choosing samples from population Y. So to find the optimum sample of sizes (n_1^*, n_2^*), differentiate, (2.28) w.r.t. n_1 and n_2 separately and setting each equation equation equal zero we have,

$$n_1^* = \frac{1}{\sqrt{c_1 \xi_1}} - n_{01} \tag{2.29}$$

and

$$n_2^* = \frac{1}{\sqrt{c_2 \xi_2}} - n_{02}. \tag{2.30}$$

For given, c_1, c_2 , n_{01}, n_{02} , ξ_1 and ξ_2 we will have a pair of optimum sample of size (n_1^*, n_2^*) from (2.29) and (2.30) which jointly minimizes the equation (2.28).

In the following section we will discuss the optimum sample size to estimate an exponential parameter.

2.5 SSD to estimate an exponential parameter

Suppose x_1, x_2,....,x_n is a random sample of size n taken from an exponential distribution with parameter λ with density,

$$p(x|\lambda) = \lambda \exp(-\lambda x); x > 0, \lambda > 0.$$

Let $t = \sum x_i$, so the likelihood of the sample is,

$$p(\underline{x}|\lambda) = \lambda^n \exp(-\lambda t). \tag{2.31}$$

Consider the conjugate gamma prior for λ as,

$$p(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda).$$

So the posterior density is,

$$p(\lambda|\underline{x}) \propto \lambda^{\alpha+n-1}\exp\{-(\beta+t)\lambda)\}; \tag{2.32}$$

which is a $Gamma(\alpha+n, \beta+t)$ distribution with mean, $\frac{\alpha+n}{\beta+t}$ and variance, $\frac{\alpha+n}{(\beta+t)^2}$. Now under the SE loss function (1.1), following the equation (1.15), the posterior risk will be,

$$PR = \frac{a_0(\alpha+n)}{(\beta+t)^2}.$$

Now inserting the cost function (1.28) into the risk PR, we have the total cost as,

$$\begin{aligned} TC(n) &= c_0 + cn + PR \\ &= c_0 + cn + \frac{a_0(\alpha+n)}{(\beta+t)^2}; \end{aligned}$$

which depends on t. At this stage we will take expectation over $TC(n)$ w.r.t t where t follows a Gamma-Gamma distribution with parameters α, β and n (Bernardo and Smith, 1994). So we have,

$$\begin{aligned} E_t[TC(n)] &= \int \left[c_0 + cn + \frac{a_0(\alpha+n)}{(\beta+t)^2}\right] p(t)dt \\ &= c_0 + cn + \frac{a_0\beta^{\alpha}(\alpha+n)}{\Gamma(\alpha)}\frac{\Gamma(\alpha+n)}{\Gamma(n)}\int \frac{t^{n-1}}{(\beta+t)^{\alpha+2+n}}dt \\ &= c_0 + cn + \frac{a_0\beta^{\alpha}(\alpha+n)}{\Gamma\alpha}\frac{\Gamma(\alpha+n)}{\Gamma(n)}\frac{\Gamma(\alpha+2)}{\beta^{\alpha+2}}\frac{\Gamma(n)}{\Gamma(\alpha+n+2)} \\ &= c_0 + cn + \frac{a_0\alpha(\alpha+1)}{\beta^2(\alpha+n+1)}. \end{aligned} \tag{2.33}$$

To obtain the optimum sample of size n setting $\frac{\partial E_t[TC(n)]}{\partial n} = 0$ gives,

$$n^*_{se} = \sqrt{\frac{a_0\alpha(\alpha+1)}{\beta^2 c}} - (\alpha+1), \tag{2.34}$$

which is the optimum sample size to estimate an exponential parameter.

2.5.1 No sampling situation

If we have no plan to sample then the expected total cost is (giving $n = 0$ in (2.33)),

$$E_t[TC(0)] = \frac{a_0\alpha}{\beta^2}.$$

Also the expected total cost for an optimum sample of size $n = n^*_{se}$ will be,

$$E_t[TC(n^*)] = c_0 - c(\alpha+1) + \frac{2}{\beta}\sqrt{a_0 c\alpha(\alpha+1)}.$$

Now the optimum sample size (2.34) should be,

$$n^* = \max\left\{0,\ \sqrt{\frac{a_0\alpha(\alpha+1)}{\beta^2 c}} - (\alpha+1)\right\}.$$

Now choose not to sample when $\sqrt{\frac{a_0\alpha(\alpha+1)}{\beta^2 c}} \leq (\alpha+1)$ i.e.,$\sqrt{\frac{a_0\alpha}{\beta^2(\alpha+1)}} \leq \sqrt{c}$ in addition to the situation when $\sqrt{\frac{a_0\alpha(\alpha+1)}{\beta^2 c}} > (\alpha+1)$ i.e.,$\sqrt{\frac{a_0\alpha}{\beta^2(\alpha+1)}} > \sqrt{c}$ and $E_t[TC(0)]$ is exceeded by $E_t[TC(n^*)]$. This implies the decision not to sample occurs when

$$\left[\frac{a_0\alpha}{\beta^2} < c_0 - c(\alpha+1) + \frac{2}{\beta}\sqrt{a_0 c\alpha(\alpha+1)} \text{ and } \sqrt{\frac{a_0\alpha}{\beta^2(\alpha+1)}} > \sqrt{c}\right]$$

or

$$\sqrt{\frac{a_0\alpha}{\beta^2(\alpha+1)}} \leq \sqrt{c}.$$

Equivalently the decision to sample occurs when

$$\frac{a_0\alpha}{\beta^2} > c_0 - c(\alpha+1) + \frac{2}{\beta}\sqrt{a_0 c\alpha(\alpha+1)} \text{ and } \sqrt{\frac{a_0\alpha}{\beta^2(\alpha+1)}} > \sqrt{c}.$$

Solving for $y = \sqrt{c}$ we obtain that we sample when

$$(\alpha+1)y^2 - \frac{2y}{\beta}\sqrt{a_0\alpha(\alpha+1)} - \left(c_0 - \frac{a_0\alpha}{\beta^2}\right) > 0 \text{ and } y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}}.$$

So that

$$\left[y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{\alpha+1)}} \text{ or } y > \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}} + \sqrt{\frac{c_0}{\alpha+1}}\right]$$

and

$$y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}},$$

which implies that

$$\sqrt{c} < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{\alpha+1}}.$$

So choose not to sample when

$$c > \left[\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{\alpha+1}}\right]^2.$$

That is for known a_0, c_0, ξ, and n_0 if the sampling cost per unit, c exceeds $\left[\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{\alpha+1}}\right]^2$ then it is not worth sampling at all due to high sampling cost.

2.5.2 Numerical study

From figure 2.5, we found that to estimate an exponential parameter, if the sampling cost per unit c goes up then the optimum sample size goes down for all values of α or β. Then we fixed the sampling cost per unit c and changed the values of the prior parameters α and β to see the effect of the prior parameters on the optimum sample size. From the same figure it is clear that if $\alpha > \beta$ then the sample size is higher than the optimum sample with $\alpha = \beta$. On the other hand for $\alpha < \beta$ the optimum sample size is smaller than the optimum sample size when the prior parameters, $\alpha = \beta$. From figure 2.6 it is evident that the optimum sample size goes up for an increasing β when c and α are kept fixed.

2.6 SSD for a Poisson parameter (DeGroot, 1970)

Let x follow a Poisson distribution with an unknown parameter θ. Then the probability density function of x will be,

$$p(x|\theta) = \frac{\exp(-\theta)\theta^x}{x!}; x = 0, 1, 2, \ldots. \tag{2.35}$$

Let x_1, x_2,...,x_n be a sample of size n from a Poisson distribution given in (2.35). Also let $t = \sum_{i=1}^{n} x_i$. So the likelihood of the sample is,

$$p(\underline{x}|\theta) = \frac{\exp(-n\theta)\theta^t}{\prod x!}. \tag{2.36}$$

Let us take the prior distribution of θ as,

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta). \tag{2.37}$$

Then the posterior distribution of $\theta|\underline{x}$ is,

$$p(\theta|\underline{x}) = \frac{(n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)}\theta^{t+\alpha-1}\exp\{-(\beta+n)\}d\theta, \tag{2.38}$$

which is a $Gamma(t+\alpha, n+\beta)$ distribution. Now under the SE loss (1.1) following the equation (1.15), the posterior risk will be,

$$PR = \frac{a_0(t+\alpha)}{(n+\beta)^2}. \tag{2.39}$$

Now add a linear cost function (1.28) with (2.39) we have the total cost as,

$$TC(n) = c_0 + cn + \frac{a_0(t+\alpha)}{(n+\beta)^2};$$

which clearly depends on t. So to have a minimum n we need to take expectation over $TC(n)$ w.r.t t as follows.

$$E_t[TC(n)] = c_0 + cn + E_t(PR) = c_0 + cn + \frac{a_0(\alpha + E(t|n))}{(n+\beta)^2}. \tag{2.40}$$

At this stage, we need to find $E[t|n]$, where t is a sufficient statistics for θ. Since x_i has a Poisson (θ) distribution, so t has a Poisson $(n\theta)$ distribution. We have,

$$\begin{aligned} E(t|n) &= E[E(t|\theta, n)] \\ &= nE[\theta|n] \\ &= \frac{n\alpha}{\beta}. \end{aligned} \tag{2.41}$$

Using this result in (2.40) we have,

$$E_t[TC(n)] = c_0 + cn + \frac{a_0\alpha}{\beta(n+\beta)}. \tag{2.42}$$

To determine the minimum sample of size n^*, differentiating, $E_t[TC(n)]$ w.r.t. n and setting equal to zero we have,

$$n^*_{se} = \sqrt{\frac{a_0\alpha}{c\beta}} - \beta. \tag{2.43}$$

Clearly, the optimum sample size under the SE loss function n^*_{se} depends on the shape parameter of the loss function a_0, prior parameters α, β and the sampling cost per unit, c.

2.6.1 No sampling situation

If we have no plan to sample then the expected total cost is (giving $n = 0$ in (2.42)),

$$E_t[TC(0)] = \frac{a_0\alpha}{\beta^2}.$$

Also the expected total cost for an optimum sample of size $n = n^*_{se}$ will be,

$$E_t[TC(n^*)] = c_0 - c\beta + 2\sqrt{\frac{a_0 c\alpha}{\beta}}.$$

Now the optimum sample size (2.43) should be,

$$n^* = \max\left\{0,\ \sqrt{\frac{a_0\alpha}{c\beta}} - \beta\right\}.$$

Now we automatically choose not to sample when $\sqrt{\frac{a_0\alpha}{c\beta}} \leq \beta$ i.e., $\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} \leq \sqrt{c}$ in addition to the situation when $\sqrt{\frac{a_0\alpha}{c\beta}} > \beta$ i.e., $\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} > \sqrt{c}$ and $E_t[TC(0)]$ is exceeded by $E_t[TC(n^*)]$. This implies the decision not to sample occurs when

$$\left[\frac{a_0\alpha}{\beta^2} < c_0 - c\beta + 2\sqrt{\frac{a_0c\alpha}{\beta}} \text{ and } \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} > \sqrt{c}\right] \text{ or } \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} \leq \sqrt{c}.$$

Equivalently the decision to sample occurs when

$$\frac{a_0\alpha}{\beta^2} > c_0 - c\beta + 2\sqrt{\frac{a_0c\alpha}{\beta}} \text{ and } \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} > \sqrt{c}.$$

Solving for $y = \sqrt{c}$ we obtain that we sample when

$$\beta y^2 - 2y\sqrt{\frac{a_0\alpha}{\beta}} - \left(c_0 - \frac{a_0\alpha}{\beta}\right) > 0 \text{ and } y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}}.$$

So that

$$\left[y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} - \frac{1}{\beta}\sqrt{a_0\alpha\left(\frac{1}{\beta} - 1\right) + c_0\beta} \text{ or } y > \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} + \frac{1}{\beta}\sqrt{a_0\alpha\left(\frac{1}{\beta} - 1\right) + c_0\beta}\right]$$

and

$$y < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}}.$$

Which implies that

$$\sqrt{c} < \frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} - \frac{1}{\beta}\sqrt{a_0\alpha\left(\frac{1}{\beta} - 1\right) + c_0\beta}.$$

So choose not to sample when

$$c > \left[\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} - \frac{1}{\beta}\sqrt{a_0\alpha\left(\frac{1}{\beta} - 1\right) + c_0\beta}\right]^2.$$

That is for known a_0, c_0, ξ, and n_0 if the sampling cost per unit, c exceeds $\left[\frac{1}{\beta}\sqrt{\frac{a_0\alpha}{\beta}} - \frac{1}{\beta}\sqrt{a_0\alpha\left(\frac{1}{\beta}-1\right)+c_0\beta}\right]^2$ then it is not worth sampling at all due to high sampling cost.

2.6.2 Numerical study

From figure 2.7, we observe that to estimate a Poisson parameter, if the sampling cost per unit goes up then the optimum sample size goes down for the prior parameters α or β. We also fixed the sampling cost per unit and then changed the values of α and β to see the effect of the prior parameters on the optimal sample size. From the graph it is clear that if $\alpha > \beta$ then the sample size is higher than the optimum sample when the prior parameters, $\alpha = \beta$. On the other hand for $\alpha < \beta$ the optimum sample sizes are less than the optimum size of samples when prior parameters, $\alpha = \beta$. From figure 2.8 we can see that if the prior parameter β goes up, then the optimum sample sizes goes down. On the other hand for a fixed value of β, when the value of α increases, then the optimum sample sizes also increases.

2.7 SSD for a Binomial distribution

Suppose a discrete random variable X has a binomial distribution with parameters θ and n and its probability function is,

$$p(x|\theta) = n_{C_x}\theta^x(1-\theta)^{n-x}, 0 \leq \theta \leq 1, x = 0, 1, \ldots, n. \tag{2.44}$$

Let us take the prior distribution for the proportion θ as a beta distribution,

$$p(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \alpha,\beta > 0, 0 \leq \theta \leq 1. \tag{2.45}$$

($B(\alpha,\beta)$ is the beta function with parameters α and β). The posterior distribution of $\theta|\underline{x}$ will be,

$$p(\theta|x) = \frac{1}{B(\alpha+x,\beta+n-x)}\theta^{x+\alpha-1}(1-\theta)^{n+\beta-x-1}, \alpha,\beta > 0, 0 \leq \theta \leq 1. \tag{2.46}$$

Now using (2.45) and (2.46), we have the marginal or pre-posterior distribution of $x|n$ is,

$$\begin{aligned} p(x|n) &= \int_0^1 p(\theta)p(\theta|x)d\theta \\ &= \frac{n_{C_x}}{B(\alpha,\beta)}\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n+\beta-x-1}d\theta \\ &= \frac{n!B(\alpha+x,n+\beta-x)}{x!(n-x)!B(\alpha,\beta)}, x = 0,1,\ldots,n. \end{aligned} \tag{2.47}$$

Here, x follows a beta-binomial distribution with parameter α, β and n. Now under the squared error loss function (1.1), from the equation (1.15) the posterior risk will be,

$$PR = \frac{a_0(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}. \tag{2.48}$$

Now adding the cost function (1.28) with the risk function (2.48) we have the total cost as,

$$TC(n) = c_0 + cn + \frac{a_0(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)},$$

which depends on x. For simplicity let us take $a_0 = 1$. So to find the minimum n we first take the expectation over $TC(n)$ w.r.t the pre posterior distribution of x given

in (2.47). We have,

$$\begin{aligned} E_x[TC(n)] &= c_0 + cn + E_x[PR] \\ &= c_0 + cn + \frac{\alpha(n+\beta) + (\beta + n - \alpha)E(x|n) - E(x^2|n)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \\ &= c_0 + cn + \frac{a(n+\beta) + (\beta + n - \alpha)E(x|n) - Var(x|n) - [E(x|n)]^2}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}. \end{aligned}$$

At this stage, we need to find $E(x|n)$ and $Var(x|n)$ following a beta-binomial distribution (2.47), from Bernardo and Smith (1994) we have,

$$E(x|n) = \frac{n\alpha}{\alpha+\beta}.$$

and variance,

$$Var(x|n) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Using these results in $E_x[TC(n)]$ we have,

$$\begin{aligned} E_x[TC(n)] &= c_0 + cn + \frac{\alpha(n+\beta)(\alpha+\beta+1)(\alpha+\beta)^2 + n\alpha(n+\beta-\alpha)(\alpha+\beta)(\alpha+\beta+1)}{(\alpha+\beta)^2(n+\alpha+\beta)^2(\alpha+\beta+1)(\alpha+\beta+n+1)} \\ &\quad - \frac{n\alpha\beta(\alpha+\beta+n) + n^2\alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(n+\alpha+\beta)^2(\alpha+\beta+1)(\alpha+\beta+n+1)}. \end{aligned}$$

We can re-write this equation as,

$$E_x[TC(n)] = c_0 + cn + \frac{n^2k_1 + nk_2 + k_3}{k_4(n+\alpha+\beta)^2(\alpha+\beta+n+1)} \tag{2.49}$$

where, $k_1 = \alpha\beta$, $k_2 = \alpha(\alpha+\beta)(\alpha+\beta+1) + \alpha(\beta-\alpha)(\alpha+\beta+1) - \alpha\beta$, $k_3 = \alpha\beta(\alpha+\beta)(\alpha+\beta+1)$, $k_4 = \frac{k_3}{\alpha\beta}$ so that k_1, k_2, k_3 and k_4 are independent of n. To obtain the optimum sample of size n, differentiating (2.49) w.r.t. n and setting equal zero we have,

$ck_4(\alpha+\beta+n)^3(\alpha+\beta+n+1)^2 + (\alpha+\beta+n)(\alpha+\beta+n+1)(2nk_1+k_2)$

$-(\alpha+\beta+n)(n^2k_1+nk_2+k_3) - 2(n^2k_1+nk_2+k_3)(\alpha+\beta+n+1) = 0.$

To simplify this equation let us assume, $\alpha+\beta+n+1 \simeq \alpha+\beta+n$ for a large n. Then we have more simplified equation as follows,

$$ck_4(\alpha+\beta+n)^4+(\alpha+\beta+n)(2nk_1+k_2)-3(n^2k_1+nk_2+k_3)=0,$$

which can be re-written as,

$$ck_4(\alpha+\beta+n)^4-n^2k_1+2n[(\alpha+\beta)k_1-k_2]+(\alpha+\beta)k_2-3k_3=0. \tag{2.50}$$

2.7.1 Numerical study

We can easily solve the equation (2.50) to obtain the optimum sample size of n using maple 13. For the given prior parameters α, β and c we obtained four roots of n and out of these four, three roots are negative, so the only positive solution of n is considered as we need the optimum sample size as a positive whole number to perform our study. From figure 2.9 we can see that if the sampling cost per unit increases, then the optimum sample size decreases. It can also be clearly observed from the graph that it is not worth sampling if the sampling cost per unit is $c > 0.04$.

2.8 Summary

In this chapter we have determined the optimum sample size of two continuous and two discrete distributions under the symmetric SE loss function. These are the normal, the exponential, the Poisson and the binomial distribution. In the case of normal distribution first we reviewed the SSD problem of Lindley (1972), where he obtained

the sample size to estimate a normal mean when precision is known. We extended the result obtaining the optimum sample size to estimate a normal precision when mean is known. We also determined the sample size considering a number of situations, such as, estimating the difference between two normal means for a known equal and unequal precision subsequently. Next, we reviewed the DeGroot (1970) SSD problem for a Poisson distribution under the SE loss function. Then we obtained the sample size for an exponential parameter and a Binomial proportion under the SE loss function. We observed that the sample size depends on the shape parameter of the loss function, prior parameter(s) and the sampling cost per unit. In all the situations we showed the effect of the sampling cost and the prior parameters on the optimum sample size graphically. We also noted the situation of not worth sampling for the higher sampling cost (generally c is bigger than some function of the prior parameters) or a strong prior information for all distributions.

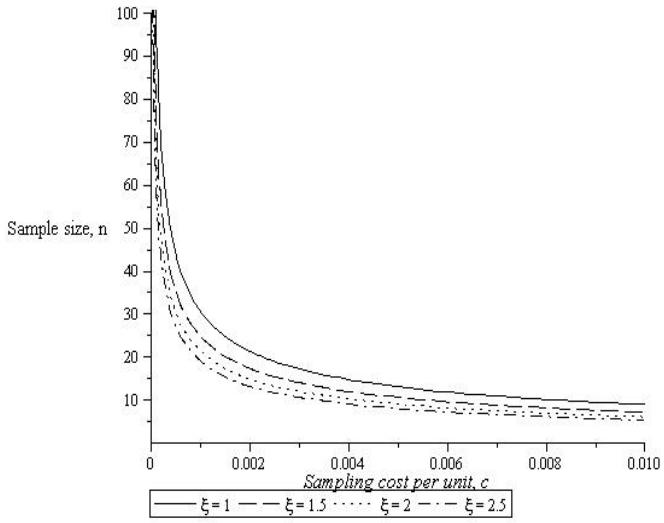

Figure 2.1: Optimum sample size to estimate a normal mean as a function of the sampling cost c for different values of ξ and $n_0 = 1$ under the SE loss function.

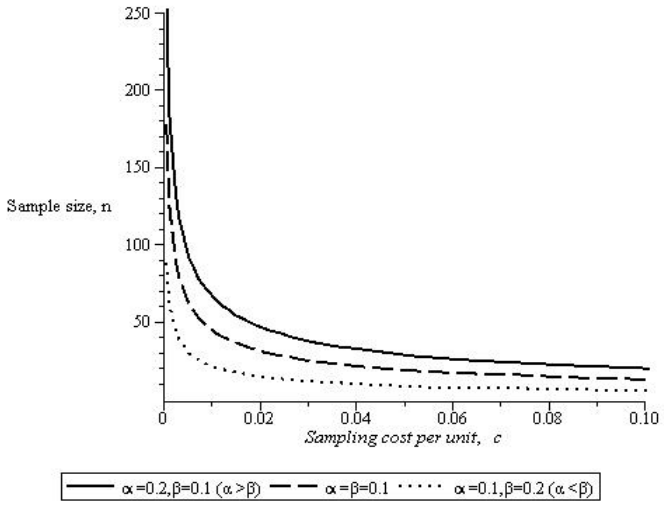

Figure 2.2: Optimum sample size to estimate a normal precision when mean is known as a function of the sampling cost per unit c for different values of α, β under the SE loss function.

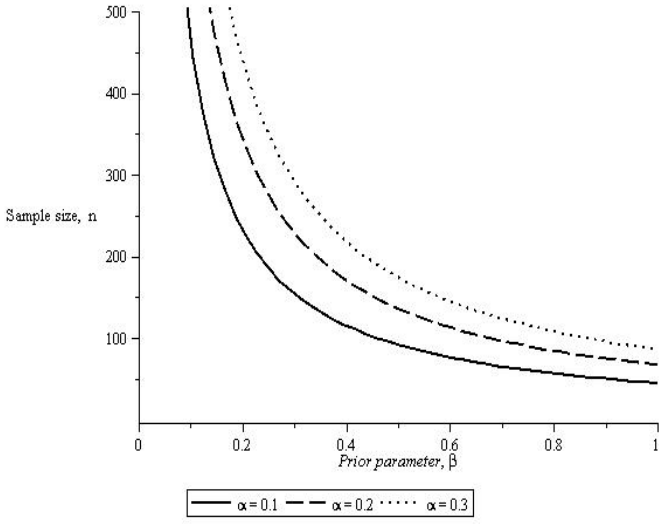

Figure 2.3: Optimum sample size to estimate a normal precision as a function of the prior parameter β when the sampling cost per unit $c = 0.0001$ for different values of α.

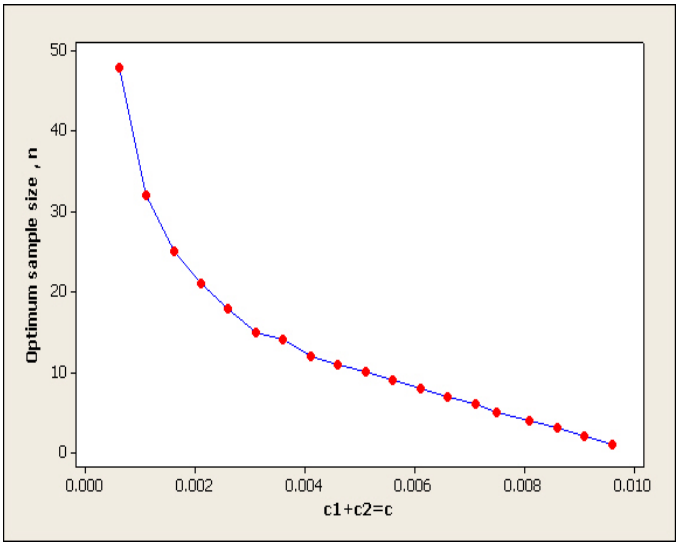

Figure 2.4: Optimum sample size to estimate the difference between two normal means as a function of the sampling cost per unit c (assumed $c_1 + c_2 = c$) when the precision is known and common.

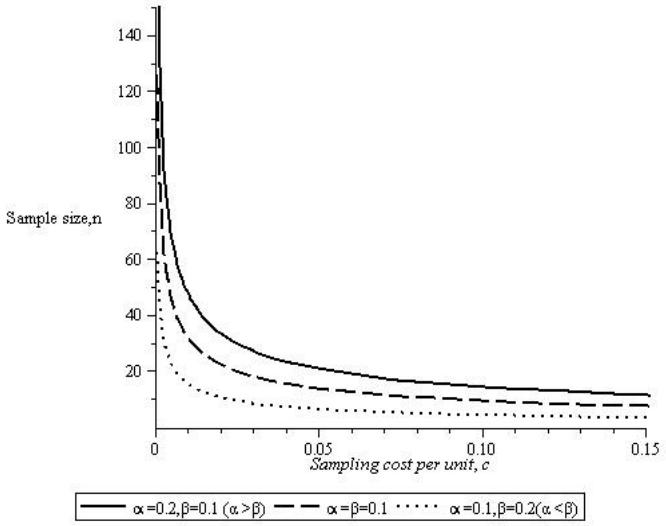

Figure 2.5: Optimum sample size to estimate an exponential parameter as a function of the sampling cost per unit c for different values of α, β under the SE loss function.

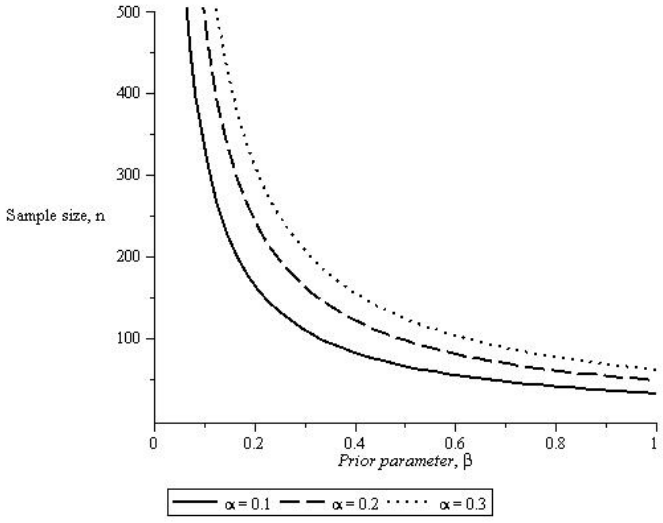

Figure 2.6: Optimum sample size to estimate an exponential parameter as a function of the prior parameter β when the sampling cost per unit $c = 0.0001$ and for a fixed prior parameter α.

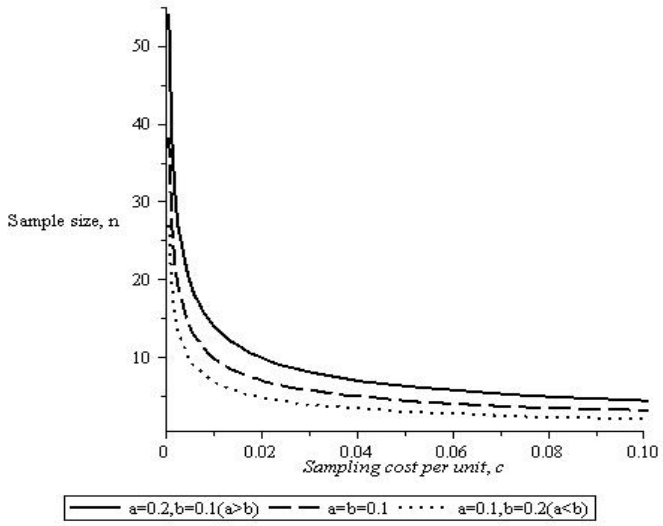

Figure 2.7: Optimum sample size to estimate a Poisson parameter as a function of the sampling cost per unit c for different values of α, β under the SE loss function.

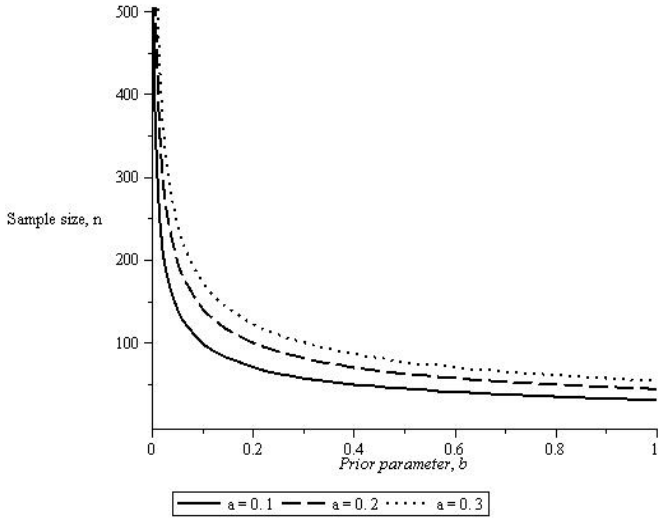

Figure 2.8: Optimum sample size to estimate a Poisson parameter as a function of the prior parameter β when the sampling cost per unit $c = 0.0001$ and the prior parameter α fixed.

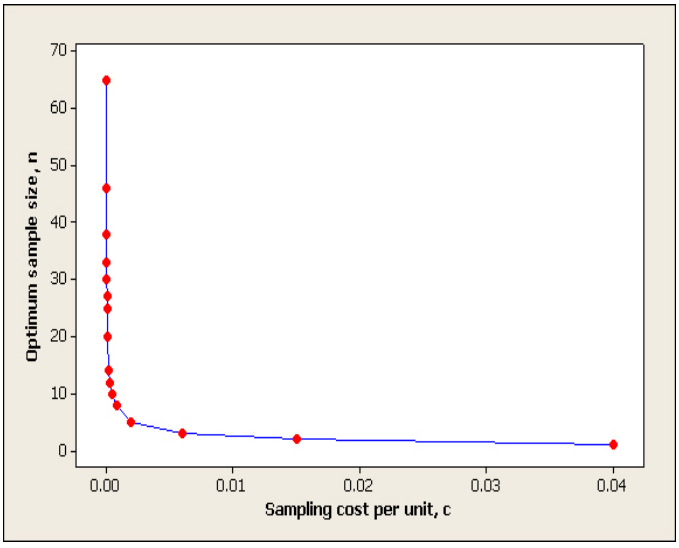

Figure 2.9: Optimum sample size to estimate a Binomial proportion θ as a function of the sampling cost per unit c when $\alpha = \beta = 0.1$ under the SE loss function.

Chapter 3

Bayesian SSD using linex loss

3.1 Introduction

In the previous chapter we discussed the optimum sample size using the squared error (SE) loss function. In a situation where underestimation is more serious than overestimation or vice-versa, then an asymmetric loss function should be used. In such a situation how many samples do we need to take to estimate the parameter under study? In this chapter, we consider the sample size using an asymmetric (linex) loss function and a linear cost function for various distributions. We compare the sample size obtained from this asymmetric loss function with the sample size from the symmetric SE loss function. We also discuss the situation where it is not worth sampling due to high sampling cost or strong prior information.

The plan of this chapter is to obtain the optimum sample size under the linex

loss function for a number of situations. First we discuss SSD for different cases of normal distribution. Then we obtain SSD for an exponential distribution. Finally we consider SSD for a Poisson distribution.

3.2 SSD to estimate a normal mean when precision is known

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean θ and known precision ξ. Now consider a conjugate prior of θ as $N(\mu_0, n_0\xi)$. From the previous chapter, section 2.2, we have the posterior distribution of $\theta|\underline{x}$ is,

$$p(\theta|\underline{x}) \propto \exp\left\{(n\overline{x} + n_0\mu_0)\xi\theta\right\} \exp\left\{-\frac{1}{2}(n+n_0)\xi\theta^2\right\} \tag{3.1}$$

We have,

$$E\{\exp(-b\theta)|\underline{x}\} = \exp\left(-bm + \frac{ab^2v^2}{2}\right),$$

where m is the posterior mean and v^2 is the posterior variance. Using this result in (1.9) we have the Bayes estimate of θ is,

$$\hat{d}_{lin} = m - \frac{abv^2}{2}. \tag{3.2}$$

Now following the equation (1.17) the posterior risk function under the linex loss function (1.8) will be,

$$PR = \frac{ab^2v^2}{2} \tag{3.3}$$

Now adding the cost function (1.28) with the posterior risk (3.3) the total cost will be,

$$\begin{aligned} TC(n) &= c_0 + cn + \frac{ab^2v^2}{2} \\ &= c_0 + cn + \frac{ab^2}{2\xi(n+n_0)}. \end{aligned} \tag{3.4}$$

Since $TC(n)$ depends on n and the precision (known) only, the optimum sample size will be the solution of $\frac{\partial TC(n)}{\partial n} = 0$. This gives,

$$n^*_{lin} = b\sqrt{\frac{a}{2c\xi}} - n_0. \tag{3.5}$$

If the scale parameter for the SE loss and the linex loss are equal to 1, that is, $a_0 = a = 1$, then for $b = \sqrt{2}$, we have the sample size under the linex loss given in (3.5) and under the SE loss functions given in the chapter-2, equation (2.5) are the same. Clearly n depends on the data precision, the scale and the shape parameter of the loss function and the sampling costs per unit c. If the prior information is not informative, giving $n_0 = 0$ in (3.5), we have the optimum sample size under the linex loss function as,

$$n^*_{lin} = \sigma b\sqrt{\frac{a}{2c}}. \tag{3.6}$$

That is the optimum sample size under the linex loss function is directly proportional to the shape parameter of the loss function and variability of the data but inversely proportional to the square root of the sampling cost per unit when there is no prior information available.

3.2.1 No sampling situation

The total cost if we do not sample is

$$TC(0) = \frac{ab^2}{2n_0\xi}.$$

The total cost of a sample for size n^* is

$$TC(n^*) = c_0 - n_0c + b\sqrt{\frac{2ac}{\xi}}.$$

Now the optimum sample size (3.5) is,

$$n^* = \max\left\{0,\ b\sqrt{\frac{a}{2c\xi}} - n_0\right\}$$

The truncation at 0 matters, so we choose not to sample when $b\sqrt{\frac{a}{2c\xi}} \leq n_0$ in addition to the situation when $b\sqrt{\frac{a}{2c\xi}} > n_0$ and $TC(0)$ is exceeded by $TC(n^*)$. This implies the decision not to sample occurs when

$$\left[\frac{ab^2}{2n_0\xi} < c_0 - n_0c + b\sqrt{\frac{2ac}{\xi}} \text{ and } b\sqrt{\frac{a}{2c\xi}} > n_0\right] \text{ or } b\sqrt{\frac{a}{2c\xi}} \leq n_0.$$

Equivalently the decision to sample occurs when

$$\frac{ab^2}{2n_0\xi} > c_0 - n_0c + b\sqrt{\frac{2ac}{\xi}} \text{ and } b\sqrt{\frac{a}{2c\xi}} > n_0. \tag{3.7}$$

Solving for $y = \sqrt{c}$ we obtain that we sample when

$$n_0y^2 - by\sqrt{\frac{2a}{\xi}} - \left(c_0 - \frac{ab^2}{2n_0\xi}\right) > 0 \text{ and } y < \frac{b}{n_0}\sqrt{\frac{a}{2\xi}}.$$

So that

$$\left[y < \frac{b}{n_0}\sqrt{\frac{a}{2\xi}} - \sqrt{\frac{c_0}{n_0}} \text{ or } y > \frac{b}{n_0}\sqrt{\frac{a}{2\xi}} + \sqrt{\frac{c_0}{n_0}}\right] \text{ and } y < \frac{b}{n_0}\sqrt{\frac{a}{2\xi}}.$$

Which implies that

$$c^{1/2} < \frac{b}{n_0}\sqrt{\frac{a}{2\xi}} - \sqrt{\frac{c_0}{n_0}}.$$

So choose not to sample when

$$c > \left[\frac{b}{n_0}\sqrt{\frac{a}{2\xi}} - \sqrt{\frac{c_0}{n_0}}\right]^2.$$

That is for known a_0, c_0, ξ, and n_0 if the sampling cost per unit, c exceeds $\left[\frac{b}{n_0}\sqrt{\frac{a}{2\xi}} - \sqrt{\frac{c_0}{n_0}}\right]^2$ then it is not worth sampling at all.

Now we will obtain the range of n_0 so that the decision not to sample occurs. From (3.7) the decision to sample occurs when

$$cn_0^2 - \left\{c_0 + b\sqrt{\frac{2ac}{\xi}}\right\} n_0 + \frac{ab^2}{2\xi} > 0 \text{ and } n_0 < b\sqrt{\frac{a}{2c\xi}}$$

so that

$$\left[n_0 < b\sqrt{\tfrac{a}{2c\xi}} - \tfrac{1}{2c}\left\{\sqrt{c_0^2 + 2bc_0\sqrt{\tfrac{2ac}{\xi}}} - c_0\right\} \text{ or } n_0 > b\sqrt{\tfrac{a}{2c\xi}} + \tfrac{1}{2c}\left\{\sqrt{c_0^2 + 2bc_0\sqrt{\tfrac{2ac}{\xi}}} - c_0\right\}\right]$$

and

$$n_0 < b\sqrt{\frac{a}{2c\xi}}.$$

Which implies that the decision to sample when

$$n_0 < b\sqrt{\frac{a}{2c\xi}} - \frac{1}{2c}\left\{\sqrt{c_0^2 + 2bc_0\sqrt{\frac{2ac}{\xi}}} - c_0\right\}.$$

So we can choose not to sample when

$$n_0 > b\sqrt{\frac{a}{2c\xi}} - \frac{1}{2c}\left\{\sqrt{c_0^2 + 2bc_0\sqrt{\frac{2ac}{\xi}}} - c_0\right\},$$

where a, b, c, c_0 and ξ are known.

3.2.2 Numerical study

In figure 3.1, we kept $a = 1$, $\sigma = 0.1$, n_0 fixed and we can see that as the sampling cost per unit c increases, the optimum sample size decreases. We observe that due to the high sampling cost $c \geq 0.0001$ (figure 3.1) it is not worth sampling at all. For $a_0 = a = 1$ and $b = \sqrt{2}$, then the optimum sample size under the SE loss and under the linex loss function are the same. If $b > \sqrt{2}$ the sample size under the SE loss is less than the sample size under the linex loss function. If $b < \sqrt{2}$ the sample size under the SE loss is more than the sample size under the linex loss function. We also observe from the equation (3.5) that, if the prior sample size n_0 increases then the optimum sample size decreases for the fixed values of a, b, c and ξ. On the other hand, the optimum sample size increases either the shape parameter, b or the scale parameter, a increases.

3.3 SSD to estimate a normal precision when mean is known

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean μ_0 and precision ξ. Now take a conjugate prior as a $Gamma(\alpha, \beta)$ distribution, so from chapter 2, equation (2.11) we have the posterior distribution of $\xi|\underline{x}$ also has a $Gamma\left(\alpha + \frac{n}{2}, \beta + \frac{\sum(x_i-\mu_0)^2}{2}\right)$ distribution. This time we want to estimate the normal precision ξ by d. We have,

$$E\left\{\exp(-b\xi)|\underline{x}\right\} = \left(1 + \frac{b}{\beta + \frac{t}{2}}\right)^{-(\alpha+\frac{n}{2})}. \tag{3.8}$$

Using (3.8) in (1.9) the Bayes estimate under the linex loss function (1.8) will be,

$$\hat{d} = \frac{\alpha + \frac{n}{2}}{b} \ln\left(1 + \frac{b}{\beta + \frac{t}{2}}\right). \tag{3.9}$$

Now from (1.17) the posterior risk for the linex loss function (1.9) will be,

$$PR = ab\left[\frac{\alpha + \frac{n}{2}}{\beta + \frac{t}{2}} - \frac{\alpha + \frac{n}{2}}{b} \ln\left(1 + \frac{b}{\beta + \frac{t}{2}}\right)\right].$$

Now adding the cost function (1.28) with the PR we have the total cost as,

$$\begin{aligned} TC(n) &= c_0 + cn + PR \\ &= c_0 + cn + ab\left[\frac{\alpha + \frac{n}{2}}{\beta + \frac{t}{2}} - \frac{\alpha + \frac{n}{2}}{b} \ln\left(1 + \frac{b}{\beta + \frac{t}{2}}\right)\right]. \end{aligned}$$

Let, $z = \frac{b}{\beta + \frac{t}{2}}$. Now expanding $\ln(1+z)$, since $b < \beta + \frac{t}{2}$ is small, neglecting 3rd and higher powers of z we have,

$$TC(n) \approx c_0 + cn + \frac{ab^2}{2} \frac{(\alpha + \frac{n}{2})}{(\beta + \frac{t}{2})^2}. \tag{3.10}$$

For a minimum n first take the expectation over $TC(n)$ w.r.t. t following the pre-posterior distribution (2.13) and using the result of (2.15) we have,

$$E_t[TC(n)] = c_0 + cn + \frac{ab^2\alpha(\alpha+1)}{2\beta^2\left(\alpha + \frac{n}{2} + 1\right)}. \tag{3.11}$$

To obtain a minimum sample of size n setting $\frac{\partial E_t[C(n)]}{\partial n} = 0$ and we have,

$$n^*_{lin} = \frac{b}{\beta}\left[\sqrt{\frac{a\alpha(\alpha+1)}{c}}\right] - 2(\alpha+1). \tag{3.12}$$

Which is the required optimum sample size to estimate a normal precision when mean is known under the linex loss function.

3.3.1 No sampling Situation

The expected total cost of sampling under the linex loss function is

$$E_t[TC(n)] = c_0 + cn + \frac{ab^2\alpha(\alpha+1)}{2\beta^2\left(\alpha+\frac{n}{2}+1\right)}.$$

Now find the expected total cost for choosing no sample and the optimum sample of size n^* giving $n = 0$ and $n = n^*_{lin}$ in $E_t[TC(n)]$ respectively. We have,

$$E_t[TC(0)] = \frac{ab^2\alpha}{2\beta^2}$$

and

$$E_t[TC(n^*)] = c_0 + \frac{2b}{\beta}\sqrt{ac\alpha(\alpha+1)} - 2c(\alpha+1).$$

Now the optimum sample size (3.12) should be

$$n^* = \max\left\{0,\ \frac{b}{\beta}\left[\sqrt{\frac{a\alpha(\alpha+1)}{c}}\right] - 2(\alpha+1)\right\}.$$

Now we choose not to sample when $\frac{b}{\beta}\sqrt{\frac{a\alpha(\alpha+1)}{c}} \leq 2(\alpha+1)$ i.e., $\frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} \leq \sqrt{c}$ in addition to the situation when $\frac{b}{\beta}\sqrt{\frac{a\alpha(\alpha+1)}{c}} > 2(\alpha+1)$ i.e., $\frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} \geq \sqrt{c}$ and $TC(0)$ is exceeded by $TC(n^*)$. This implies the decision not to sample occurs when

$$\left[\frac{ab^2\alpha}{2\beta^2} < c_0 + \frac{2b}{\beta}\sqrt{ac\alpha(\alpha+1)} - 2c(\alpha+1) \text{ and } \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} > \sqrt{c}\right] \text{ or } \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} \leq \sqrt{c}.$$

Equivalently the decision to sample occurs when

$$\frac{ab^2\alpha}{2\beta^2} > c_0 + \frac{2b}{\beta}\sqrt{ac\alpha(\alpha+1)} - 2c(\alpha+1) \text{ and } \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} > \sqrt{c}.$$

Solving for $y = \sqrt{c}$ we obtain that we sample when

$$2y^2(\alpha+1) - \frac{2b}{\beta}\sqrt{a\alpha(\alpha+1)}y - \left(c_0 - \frac{ab^2\alpha}{2\beta^2}\right) > 0 \text{ and } y < \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}}.$$

So that

$$\left[y < \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{2(\alpha+1)}} \text{ or } y > \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} + \sqrt{\frac{c_0}{2(\alpha+1)}}\right]$$

and

$$y < \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}}.$$

Which implies that

$$\sqrt{c} < \frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{2(\alpha+1)}}.$$

So choose not to sample when

$$c > \left[\frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{2(\alpha+1)}}\right]^2.$$

That is for known a, b, c_0, α, β and n_0 if the sampling cost per unit, c exceeds $\left[\frac{b}{2\beta}\sqrt{\frac{a\alpha}{\alpha+1}} - \sqrt{\frac{c_0}{2(\alpha+1)}}\right]^2$ then it is not worth sampling due to high sampling cost per unit.

3.3.2 Numerical study

From figure 3.2 we found that, if the sampling cost per unit goes up then the optimum sample size goes down for both the SE and the linex loss functions assuming both the scale parameter of the SE loss a_0 and the shape parameter of the linex loss a are equal to 1. Again for $a_0 = a = 1$, if the shape parameter of the linex loss, $b > \sqrt{2}$, then the optimum sample size is higher than the SE loss function. The reverse is also true. From figure 3.3 we see that, if the prior parameter β goes up, then the optimum sample sizes are goes down for any c. On the other hand, looking at the equation

(3.12), the scale parameter a, the shape parameter b and the prior parameter, α has the similar type of effects on the optimum sample size. If any one of them increases knowing the others, then the optimum sample size also increases.

3.4 SSD to estimate the difference between two normal means

Suppose, $\underline{x} = x_1, x_2,....,x_{n_1}$ is a random sample of size n_1 with mean μ_1 and precision ξ_1 and $\underline{y} = y_1, y_2,....,y_{n_2}$ is another random sample of size n_2 with mean μ_2 and precision ξ_2 taken from a normal distribution. Consider the prior distributions for $\mu_1|\xi_1$ as $N(\mu_{01}, n_{01}\xi_1)$ and for $\mu_2|\xi_2$ as $N(\mu_{02}, n_{02}\xi_2)$. We will determine the sample size to estimate the difference between two means by θ where $\theta = \mu_1 - \mu_2$ under the linex loss function. At first, we assume the two populations have an equal and known precisions $\xi_1 = \xi_2 = \xi$ (say) and in this case we consider the equal size of sample could be drawn from both populations, that is $n_1 = n_2 = n$. Secondly, we consider the known but unequal precisions where we consider $n_1 \neq n_2$. In this situation we obtain the joint minimum sample of sizes n_1 and n_2 under the linex loss functions.

3.4.1 SSD when precision is known and common

Let us take the estimates of the posterior means μ_1 and μ_2 are m_1 and m_2 respectively. Now consider the estimate of the difference between two means is θ where $\theta = \mu_1 - \mu_2$. So from chapter 2 equation (2.18), the posterior density of θ observing the data $\underline{x}$

and $\underline{y}$ which is a normal distribution with mean, $m = m_2 - m_1$ and variance, $v^2 = \frac{n_1+n_2+n_{01}+n_{02}}{\xi(n_{01}+n_1)(n_{02}+n_2)}$. Where ξ is both a common and known precision, $m_1 = \frac{n_{01}\mu_{01}+n_1\overline{x}}{n_{01}+n_1}$ and $m_2 = \frac{n_{02}\mu_{02}+n_2\overline{y}}{n_{02}+n_2}$. Using the posterior density provided in the equation (2.18) we have,

$$E\{\exp(-b\theta)|\underline{x}\} = \exp\left(-bm + \frac{b^2v^2}{2}\right).$$

Now using this result in (1.9) the Bayes estimate under the linex loss function (1.8) will be,

$$\hat{d_{lin}} = m - \frac{bv^2}{2}. \tag{3.13}$$

So using this result in the equation (1.17) the posterior risk will be,

$$PR = \frac{ab^2v^2}{2},$$

where a is the scale , b is the shape parameter of the linex loss function and v^2 is the posterior variance. Now adding a linear cost function (1.28) to the risk function PR we have the total cost as,

$$TC(n) = c_0 + c_1n_1 + c_2n_2 + \frac{ab^2}{2\xi}\frac{n_1+n_2+n_{01}+n_{02}}{(n_{01}+n_1)(n_{02}+n_2)}. \tag{3.14}$$

Consider $n_1 = n_2 = n(say), n_{01} = n_{02} = n_0(say)$ in (3.14) and for a minimum n differentiate it w.r.t n then setting equal zero we have,

$$n^* = b\sqrt{\frac{a}{\xi(c_1+c_2)}} - n_0. \tag{3.15}$$

If $c_1 = c_2 = c$ we have the optimum sample size,

$$n^* = b\sqrt{\frac{a}{2\xi c}} - n_0. \tag{3.16}$$

If $n_1 = n_2 = n$ but $n_{01} \neq n_{02}$, $c_1 \neq c_2$ then (3.14) becomes,

$$TC(n) = c_0 + (c_1+c_2)n + \frac{ab^2}{2\xi}\frac{2n+n_{01}+n_{02}}{(n_{01}+n)(n_{02}+n)}. \tag{3.17}$$

For an optimum sample size, differentiating (3.17) w.r.t. n and setting equal to zero we have,

$$An^4 + Bn^3 + Cn^2 + Dn + E = 0, \tag{3.18}$$

where, $A = c_1 + c_2$, $B = 2AN_0$, $C = AN_0^2 + 2AN_0' - 2k_2$, $D = 2AN_0N_0' - 2k_2N_0$, $E = AN_0'^2 + 2k_2N_0' - {N_0}^2$, $N_0 = n_{01} + n_{02}$, $N_0' = n_{01}n_{02}$, $k_2 = \frac{ab^2}{2\xi}$. For the given, c_1, c_2 , n_{01}, n_{02} , ξ and b we can easily solve the equation (3.18) for an optimum sample of size n.

Consider a situation where, $n_1 \neq n_2$, $n_{01} \neq n_{02}$ and $c_1 \neq c_2$. If the sampling from X is independent of sampling from Y then we can determine the optimum sample sizes for both n_1 and n_2 separately. Let us differentiate (3.14) w.r.t. n_1 we have,

$$n_1^* = \frac{b\sqrt{a}}{\sqrt{2c_1\xi}} - n_{01}. \tag{3.19}$$

Again differentiate (3.14) w.r.t. n_2 we have,

$$n_2^* = \frac{b\sqrt{a}}{\sqrt{2c_2\xi}} - n_{02}. \tag{3.20}$$

For the given, c_1, c_2 , n_{01}, n_{02} , ξ, a and b we can find n_1 and n_2 from equation (3.19) and (3.20) respectively which jointly minimizes the equation (3.14).

3.4.2 Numerical study

We have solved the equation (3.18) using Maple and found that out of four roots three roots are negative. So the only positive root is considered as an optimum sample size and from figure 3.4, we observe that if the sampling cost per unit goes up, then the optimum sample size goes down to estimate the difference between two

normal means when $n_{01} = 8$, $n_{02} = 12$, $\xi = 1$, $a = 1$ and $b = 0.1$. From figure 3.5, we observe that if the shape parameter, b increases then the optimum sample size decreases when $n_{01} = 8$, $n_{02} = 12$, $\xi = 1$, $c_1 + c_2 = c = 0.001$ are kept fixed. Also from the equations (3.15), (3.16), (3.19), and (3.20) we have the optimum sample per group is proportional to the shape parameter of the linex loss function b but inversely proportional to the sampling cost per unit c.

In the above section we obtained the optimum sample size for the difference between two normal means when the precisions of two populations are assumed to be known and equal. In the following section we will obtain the optimum sample of size considering the known but unequal precisions of two populations.

3.4.3 SSD when the precision is known but unequal

From (1.17) the posterior risk function for the Bayes estimate $\theta = \mu_1 - \mu_2$ under the linex the loss function is,

$$PR = \frac{ab^2}{2} \times \frac{\xi_1(n_1 + n_{01}) + \xi_2(n_2 + n_{02})}{\xi_1\xi_2(n_{01} + n_1)(n_{02} + n_2)}. \tag{3.21}$$

Now adding the cost function (1.28) with the risk function (3.21) we have the total cost as,

$$TC(n) = c_0 + c_1 n_1 + c_2 n_2 + \frac{ab^2}{2}\left[\frac{\xi_1(n_1 + n_{01}) + \xi_2(n_2 + n_{02})}{\xi_1\xi_2(n_{01} + n_1)(n_{02} + n_2)}\right]. \tag{3.22}$$

The sample of sizes n_1 and n_2 may be drawn independently from two different populations X and Y respectively under the same experiment. For example if we want to compare the difference between two treatments to cure the same disease, then we

might allocate patients to the treatments randomly. Or even if the samples of patients are taken from the same hospital, we might need to draw samples from two different groups of patients who are taking two different treatments to treat the same disease. So the samples we want to draw from two populations may be independent. Let us differentiate (3.22) w.r.t. n_1 we have,

$$n_1^* = b\sqrt{\frac{a}{2c_1\xi_1}} - n_{01}. \tag{3.23}$$

Again differentiate (3.22) w.r.t. n_2 we have,

$$n_2^* = b\sqrt{\frac{a}{2c_2\xi_2}} - n_{02}. \tag{3.24}$$

For given, c_1, c_2, n_{01}, n_{02} , ξ_1, ξ_2, a and b we can find n_1 and n_2 from equation (3.23) and (3.24) respectively which jointly minimizes the total cost (3.22).

In the following section we will obtain the optimum sample to estimate an exponential parameter under the linex loss function.

3.5 SSD to estimate an exponential parameter

Suppose x_1,x_2,...,x_n is a random sample of size n taken from an exponential distribution with parameter λ. From Chapter 2, equation (2.32), we have the posterior density of $\lambda|\underline{x}$ is a $Gamma(\alpha+n, \beta+t)$ distribution with mean $\frac{\alpha+n}{\beta+t}$ and variance $\frac{\alpha+n}{(\beta+t)^2}$. Now using this posterior distribution we have,

$$E\{\exp(-b\lambda)|\underline{x}\} = \left[1 + \frac{b}{\beta+t}\right]^{-(\alpha+n)}.$$

Using this result in (1.9), the Bayes estimate of λ under the linex loss function (1.8) is,

$$\hat{d} = \frac{\alpha+n}{b}\left[\ln\left\{1+\frac{b}{\beta+t}\right\}\right].$$

Now using this result in (1.17) the posterior risk will be,

$$PR = \frac{ab(\alpha+n)}{\beta+t} - a(\alpha+n)\left[\log\left\{1+\frac{b}{\beta+t}\right\}\right].$$

Now inserting the linear cost function (1.28) into the risk function we have the total cost as,

$$\begin{aligned} TC(n) &= c_0 + cn + \frac{ab(\alpha+n)}{\beta+t} - a(\alpha+n)\left[\log\left\{1+\frac{b}{\beta+t}\right\}\right] \\ &\approx c_0 + cn + \frac{ab(\alpha+n)}{\beta+t} - a(\alpha+n)\left[\frac{b}{\beta+t} - \frac{b^2}{2(\beta+t)^2}\right] \\ &\approx c_0 + cn + \frac{a(\alpha+n)b^2}{2(\beta+t)^2}. \end{aligned}$$

Here we have expanded the log term and neglected 3rd and higher powers of $\frac{b}{\beta+t}$ as $b < \beta + t$. Clearly, $TC(n)$ depends on the data t. So at this stage take the expectation over $TC(n)$ w.r.t. t that is find $E_t[TC(n)]$ where, t follows a Gamma-Gamma distribution with parameters α, β and n (Bernardo and Smith,1994). We have,

$$\begin{aligned} E_t[TC(n)] &= c + cn + \frac{ab^2\beta^\alpha(\alpha+n)}{2\Gamma(\alpha)}\frac{\Gamma(\alpha+n)}{\Gamma(n)}\int\frac{t^{n-1}}{(\beta+t)^{\alpha+2+n}}dt \\ &= c_0 + cn + \frac{ab^2\beta^\alpha(\alpha+n)}{2\Gamma(\alpha)}\frac{\Gamma(\alpha+n)}{\Gamma(n)}\frac{\Gamma(\alpha+2)}{\beta^{\alpha+2}}\frac{\Gamma(n)}{\Gamma(\alpha+n+2)} \\ &= c_0 + cn + \frac{ab^2\alpha(\alpha+1)}{2\beta^2(\alpha+n+1)}. \end{aligned}$$

To obtain an optimum sample of size n setting $\frac{\partial E_t[TC(n)]}{\partial n} = 0$ gives,

$$n^*_{lin} = \frac{b}{\beta}\sqrt{\frac{a\alpha(\alpha+1)}{2c}} - (\alpha+1). \tag{3.25}$$

Which is the optimum sample size to estimate an exponential parameter under the linex loss function.

3.5.1 No sampling situation

The total expected cost to estimate an exponential parameter under the linex loss function is

$$E_t[TC(n)] = c_0 + cn + \frac{ab^2\alpha(\alpha+1)}{2\beta^2(\alpha+n+1)}.$$

Now find the expected total cost for $n = 0$ and $n = n^*_{lin}$ respectively we have,

$$E_t[TC(0)] = \frac{ab^2\alpha}{2\beta^2}$$

and

$$E_t[TC(n^*)] = c_0 + \frac{b}{\beta}\sqrt{2ac\alpha(\alpha+1)} - c(\alpha+1).$$

The optimum sample size (3.25) should be

$$n^* = \max\left\{0,\ \frac{b}{\beta}\sqrt{\frac{a\alpha(\alpha+1)}{2c}} - (\alpha+1)\right\}.$$

So we choose not to sample when $\frac{b}{\beta}\sqrt{\frac{a\alpha(\alpha+1)}{2c}} \leq (\alpha+1)$ i.e., $\frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} \leq \sqrt{c}$ in addition to the situation when $\frac{b}{\beta}\sqrt{\frac{a\alpha(\alpha+1)}{2c}} > (\alpha+1)$ i.e., $\frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} > \sqrt{c}$ and $TC(0)$ is exceeded by $TC(n^*)$. This implies the decision not to sample occurs when

$$\left[\frac{ab^2\alpha}{2\beta^2} < c_0 + \frac{b}{\beta}\sqrt{2ac\alpha(\alpha+1)} - c(\alpha+1) \text{ and } \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} > \sqrt{c}\right]$$

or

$$\frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} \leq \sqrt{c}.$$

Equivalently the decision to sample occurs when

$$\frac{ab^2\alpha}{2\beta^2} > c_0 + \frac{b}{\beta}\sqrt{2ac\alpha(\alpha+1)} - c(\alpha+1) \text{ and } \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} > \sqrt{c}.$$

Solving for $y = c^{1/2}$ we obtain that we sample when

$$y^2(\alpha+1) - \frac{b}{\beta}\sqrt{2a\alpha(\alpha+1)}y - \left(c_0 - \frac{ab^2\alpha}{2\beta^2}\right) > 0 \text{ and } y < \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}}.$$

So that

$$\left[y < \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{\alpha+1}} \text{ or } y > \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} + \sqrt{\frac{c_0}{\alpha+1}}\right]$$

and

$$y < \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}}.$$

Which implies that

$$\sqrt{c} < \frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{\alpha+1}}.$$

So choose not to sample when

$$c > \left[\frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{\alpha+1}}\right]^2.$$

That is for known a, b, c_0, α, β and n_0 if the sampling cost per unit, c exceeds $\left[\frac{b}{\beta}\sqrt{\frac{a\alpha}{2(\alpha+1)}} - \sqrt{\frac{c_0}{\alpha+1}}\right]^2$ then it is not worth sampling due to high sampling cost per unit.

3.5.2 Numerical study

From figure 3.6 we observe that if the sampling cost per unit c goes up, then the optimum sample size goes down for both SE and linex loss function when $a_0 = a = 1$.

Moreover, if the shape parameter of the linex loss $b > 2$ then the optimum sample size is bigger than the SE loss. The reverse is also true. Interestingly, for $a_0 = a = 1$, when $b = 2$ then the optimum sample sizes are equal under both linex and SE loss function. From figure 3.7, the optimum sample sizes goes up if the prior parameter β goes down. On the other hand, from the equation (3.25), we see that, if the prior parameter α goes up then the optimum sample sizes also goes up. Also if the shape parameter b of linex loss function goes up then the optimum sample size also goes up. That is the prior parameter, α and the shape parameter b has the similar type of effect on the optimum sample size n^*.

3.6 SSD to estimate a Poisson parameter

Let x_1, x_2,. . . ,x_n be a random sample of size n taken from a Poisson distribution given in (2.35). Now consider prior distribution of θ as a $Gamma(\alpha, \beta)$ distribution and from Chapter 2 equation (2.38) we have the posterior distribution of $\theta|\underline{x}$ which is also a $Gamma\,(t + \alpha, n + \beta)$ distribution. Using this posterior distribution we have,

$$E\{\exp(-b\theta)|\underline{x}\} = \left[\frac{n+\beta}{n+\beta+b}\right]^{\alpha+t}. \tag{3.26}$$

So using the result of (3.26) in (1.9), the Bayes estimate under the linex loss function (1.8) will be,

$$\hat{d}_{lin} = -\frac{\alpha+t}{b}\ln\left(\frac{n+\beta}{n+\beta+b}\right). \tag{3.27}$$

From the equation (1.17), the posterior risk under the linex loss function will be,

$$\begin{aligned} PR &= ab\left[\frac{t+\alpha}{\beta+n}+\frac{\alpha+t}{b}\ln\left(\frac{n+\beta}{n+\beta+b}\right)\right] \\ &= a\left[\frac{b}{\beta+n}+(t+\alpha)\ln\left(\frac{n+\beta}{n+\beta+b}\right)\right]. \end{aligned}$$

Now adding the cost function (1.28) with the PR we have the total cost as,

$$TC(n) = c_0 + cn + \frac{ab}{\beta+n} + a(t+\alpha)\ln\left(\frac{n+\beta}{n+\beta+b}\right),$$

which depends on t. Now taking the expectation over $TC(n)$ w.r.t. t we have,

$$E_t[TC(n)] = c_0 + cn + a\frac{b}{\beta+n} + (\alpha + E(t|n))\ln\left(\frac{\beta+n}{n+\beta+b}\right). \qquad (3.28)$$

From previous chapter equation (2.41) we have,

$$E(t|n) = \frac{\alpha n}{\beta}.$$

Using this result in (3.28) we have,

$$E_t[TC(n)] = c_0 + cn + a\left[\frac{b\alpha}{\beta}+\frac{\alpha(n+\beta)}{\beta}\ln\left(\frac{n+\beta}{n+\beta+b}\right)\right] \qquad (3.29)$$

To have a minimum sample of size n differentiate (3.29) w.r.t. n and setting equal to zero we have,

$$\frac{c}{a}+\frac{b\alpha}{\beta(n+\beta+b)}+\frac{\alpha}{\beta}\ln\left(\frac{n+\beta}{n+\beta+b}\right)=0$$

$$\frac{c\beta}{a\alpha}+\frac{b}{(n+\beta+b)}+\ln\left(\frac{n+\beta}{n+\beta+b}\right)=0$$

Let $z = n+\beta+b$ gives,

$$\frac{c\beta}{a\alpha}+\frac{b}{z}+\ln\left(1-\frac{b}{z}\right)=0$$

Expanding $\log(1-\frac{b}{z})$, since $b < z$, neglecting 3rd and higher powers of $\frac{b}{z}$ we have,

$$\frac{\beta c}{a\alpha}+\frac{b}{z}-\frac{b}{z}-\frac{b^2}{2z^2}=0$$

$$\frac{c\beta}{a\alpha} - \frac{b^2}{2z^2} = 0$$

Since, $z = n + \beta + b$, after some simplification the approximate sample size under the linex loss function will be,

$$n^*_{lin} = b\sqrt{\frac{a\alpha}{2c\beta}} - (\beta + b). \tag{3.30}$$

Here the optimum sample size depends on the prior parameters α and β, the cost of sampling per unit c, the scale and shape parameter a, b of the linex loss function respectively.

3.6.1 No sampling situation

If $z = n + \beta + b$ we can re-write the equation (3.29) as,

$$E_t[TC(n)] = c_0 + cn + a\left[\frac{b\alpha}{\beta} + \frac{\alpha(n+\beta)}{\beta}\ln\left(1 - \frac{b}{z}\right)\right].$$

Expanding the log term, since $b < n + \beta + b$, neglecting the power of 2 and more we have,

$$\begin{aligned} E_t[TC(n)] &\approx c_0 + cn + a\left[\frac{b\alpha}{\beta} - \frac{b\alpha}{\beta}\left(1 - \frac{b}{n+b+\beta}\right)\right] \\ &= c_0 + cn + \frac{a\alpha b^2}{\beta(n+\beta+b)}. \end{aligned}$$

Now we find the expected total cost when there is no plan to sample giving $n = 0$ in $E_t[TC(n)]$ we have,

$$E_t[TC(0)] = \frac{a\alpha b^2}{\beta(n+\beta+b)}.$$

Again obtain the expected total cost for an optimum sample of size n^* giving $n = n^*_{lin}$ in $E_t[TC(n)]$ we have,

$$\begin{aligned} E_t[TC(n^*)] &= c_0 + c\left(b\sqrt{\frac{a\alpha}{2c\beta}} - b - \beta\right) + \frac{b\sqrt{2ac\alpha\beta}}{\beta} \\ &= c_0 + 3b\sqrt{\frac{a\alpha c}{2\beta}} - c(\beta + b). \end{aligned}$$

The optimum sample size (3.30) should be

$$n^* = \max\left\{0,\ b\sqrt{\frac{a\alpha}{2c\beta}} - (\beta + b)\right\}.$$

So we choose not to sample when $b\sqrt{\frac{a\alpha}{2c\beta}} \leq (\beta + b)$ i.e., $\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} \leq \sqrt{c}$ in addition to the situation when $b\sqrt{\frac{a\alpha}{2c\beta}} > \beta + b$ i.e., $\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} > \sqrt{c}$ and $TC(0)$ is exceeded by $TC(n^*)$. This implies the decision not to sample occurs when

$$\left[\frac{a\alpha b^2}{\beta(n + b + \beta)} < c_0 + 3b\sqrt{\frac{a\alpha c}{2\beta}} - c(\beta + b) \text{ and } \frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} > \sqrt{c}\right] \text{ or } \frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} \leq \sqrt{c}.$$

Equivalently the decision to sample occurs when

$$\frac{a\alpha b^2}{\beta(n + b + \beta)} > c_0 + 3b\sqrt{\frac{a\alpha c}{2\beta}} - c(\beta + b) \text{ and } \frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} > \sqrt{c}.$$

Solving for $y = \sqrt{c}$ we obtain that we sample when

$$(\beta + b)y^2 - 3b\sqrt{\frac{a\alpha}{2\beta}}y - \left(c_0 - \frac{a\alpha b^2}{\beta(n + b + \beta)}\right) > 0 \text{ and } y < \frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}}.$$

So that

$$\left[y < \frac{3\beta}{2(\beta + b)}\left\{\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} - \frac{v}{3\beta}\right\} \text{ or } y > \frac{3\beta}{2(\beta + b)}\left\{\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} + \frac{v}{3\beta}\right\}\right]$$

and

$$y < \frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}}.$$

Which implies that

$$\sqrt{c} < \frac{3\beta}{2(\beta+b)}\left\{\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} - \frac{v}{3\beta}\right\},$$

where $v = \sqrt{\frac{9a\alpha b^2}{2\beta} + 4(\beta+b)\left(c_0 - \frac{a\alpha b^2}{\beta(n+\beta+b)}\right)}$. So choose not to sample when

$$c > \frac{9\beta^2}{4(\beta+b)^2}\left[\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} - \frac{v}{3\beta}\right]^2.$$

That is for known a, b, α, β and c_0 if the sampling cost per unit, c exceeds $\frac{9\beta^2}{4(\beta+b)^2}\left[\frac{b}{\beta}\sqrt{\frac{a\alpha}{2\beta}} - \frac{v}{3\beta}\right]^2$ then it is not worth sampling due to high sampling cost per unit to estimate a Poisson parameter under the linex loss function. Since getting sample is expensive, so if we look at the sampling cost per unit, then it is possible to take an initial decision according to our budget that the sampling is possible or not in such a situation. We also need to consider the overall cost of sampling c_0 because if the sampling set-up cost is too high then it may also not worth sampling at all.

3.6.2 Numerical study

From figure 3.8 we observe that if the sampling cost per unit goes up then the optimum sample size goes down for both the SE and linex loss functions. From figure 3.9 it is clear that the optimum sample size goes up when the prior parameter β goes down. So the prior parameter, β and the sampling cost per unit c has the the similar type of effect on the optimum sample size. On the other hand from the equation (3.30), we see that as the prior parameter, α increases then the optimum sample size also increases for any c. Also if the shape parameter b goes up then the optimum sample size also goes up for any value of c. Again we observe from (3.30) that if the shape

parameter b goes up then the optimum sample sizes are also goes up for any values of the prior parameter α. We conclude that the optimum sample size is proportional to the prior parameter α, the scale parameter a, the shape parameter b but inversely proportional to the square root of the prior parameter β and the sampling cost per unit c.

3.7 Summary

In this chapter we have shown how a Bayesian decision theoretic approach to determine the optimum sample size can be extended to the asymmetric linex loss function. There are a number of situations where the losses involved are asymmetric. Varian was concerned with estimating the value of real estate of taxation. In medicine it could be more serious to underestimate temperature than overestimate it. In section 3.2 we obtained the optimum sample size to estimate a normal mean when the precision is known. Section 3.3 contains the optimum sample size determination to estimate a normal precision when mean is known. Then, in section 3.4, we obtained the optimum sample size to estimate the difference between the two normal means under the linex loss function which we could use to compare the two treatment means when both treatment follows a normal distribution. Next, in section 3.5 we explored optimum SSD to estimate an exponential parameter. Finally, in section 3.6 we detailed an optimum SSD to estimate a Poisson parameter. In most cases the optimum sample size depends on the prior parameters, data variability, sampling cost per unit, c and the scale and shape parameter of the loss function. For most cases we have

also used graph to plot the sampling cost per unit against the optimum sample size n^* and observed that if the sampling cost per unit goes up, the optimum sample size goes down. We have also noted cases where, due to high sampling cost of collecting data or strong prior information, it is not worth sampling.

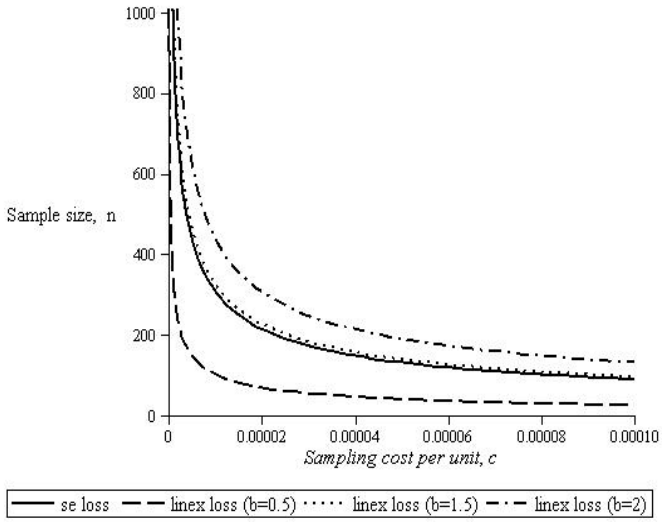

Figure 3.1: Optimum sample size to estimate a normal mean μ as a function of the sampling cost per unit c.

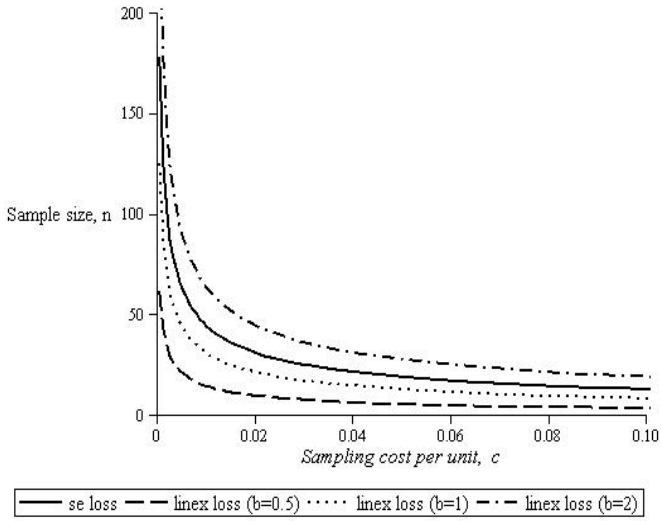

Figure 3.2: Optimum sample size to estimate a normal precision ξ as a function of the sampling cost per unit c when $\alpha = \beta = 0.1$ under the SE and the linex loss function.

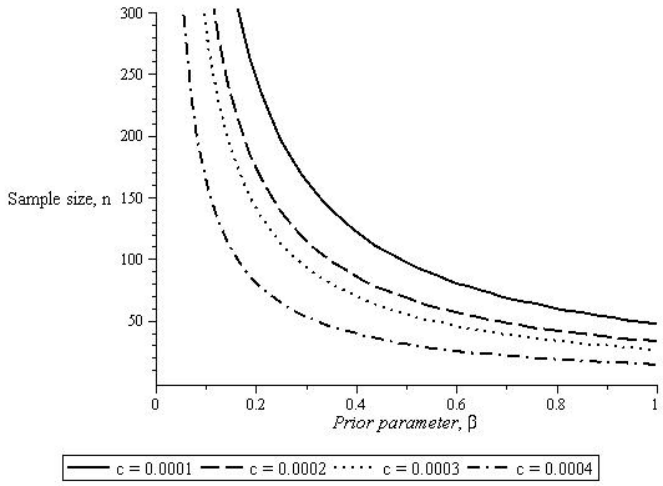

Figure 3.3: Optimum sample size to estimate a normal precision ξ as a function of the prior parameter β (when prior parameter $\alpha = 0.1$, shape parameter $b = 1.5$ kept fixed) for different values of the sampling cost per unit c under the linex loss function.

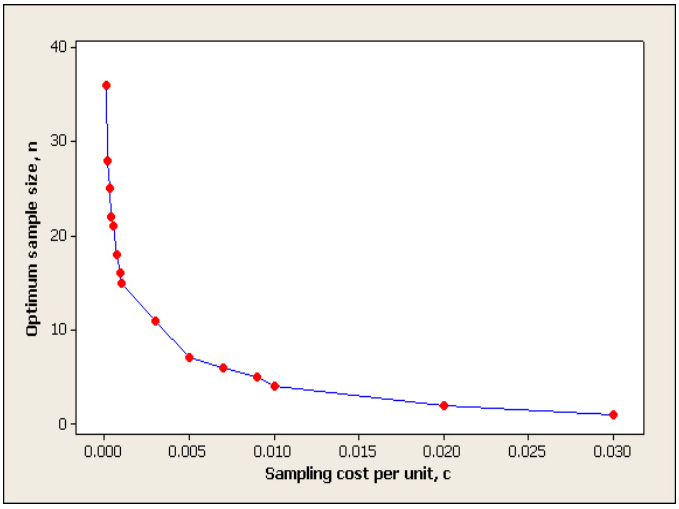

Figure 3.4: Optimum sample size to estimate the difference between two normal means as a function of the sampling cost per unit c when the precision $\xi = 1$, the prior sample sizes, $n_{01} = 8$, $n_{02} = 12$ and the shape parameter $b = 0.1$ of the linex loss function.

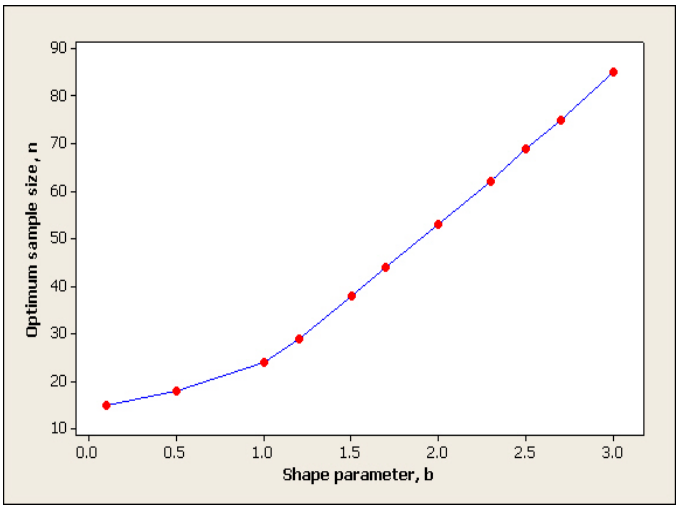

Figure 3.5: Optimum sample size to estimate the difference between two normal means as a function of the shape parameter b when the precision $\xi = 1$, prior sample sizes, $n_{01} = 8$, $n_{02} = 12$ and the sampling cost per unit $c = 0.001$ under the linex loss function.

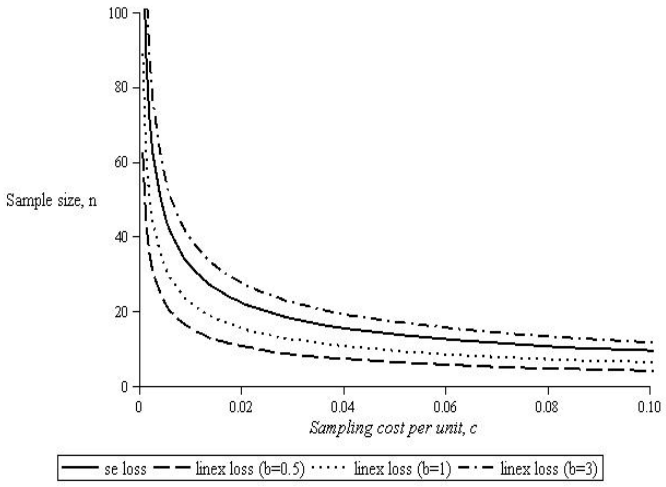

Figure 3.6: Optimum sample size to estimate an exponential parameter θ as a function of the sampling cost c when $\alpha = \beta = 0.1$ and for different values of b of the linex loss function when $a = 1$.

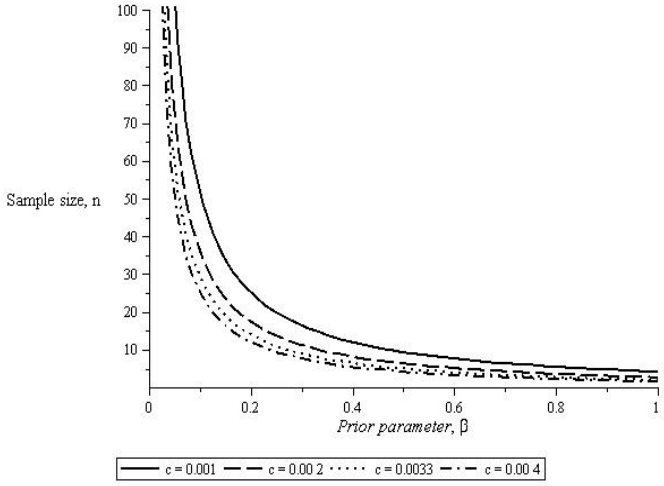

Figure 3.7: Optimum sample size to estimate an exponential parameter θ as a function of the prior parameter β when $a = 1$, $\alpha = 0.1$, $b = 0.5$ and for different values of c.

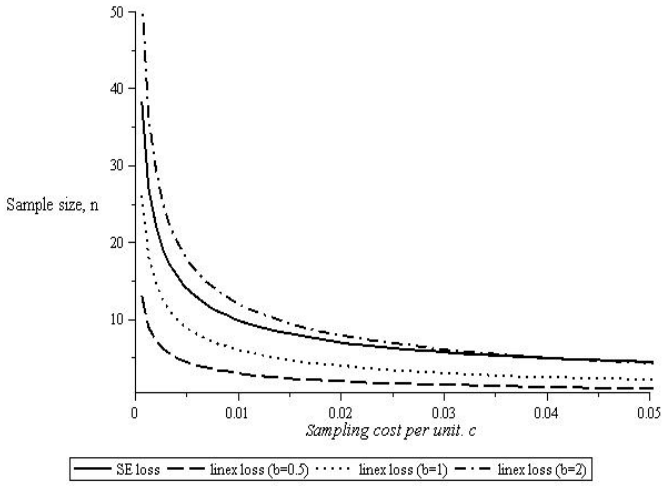

Figure 3.8: Optimum sample size to estimate a Poisson parameter θ as a function of the sampling cost per unit c when the scale parameter $a = 1$, prior parameters $\alpha = \beta = 0.1$ under the SE and different values of the shape parameter, b of the linex loss function.

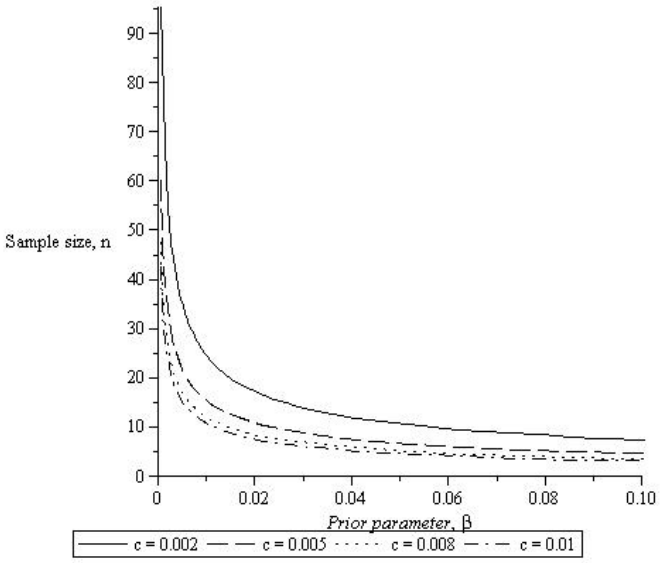

Figure 3.9: Optimum sample size to estimate a Poisson parameter θ as a function of the prior parameter β when $\alpha = 0.1$, $b = 0.5$ for different values of the sampling cost per unit c under the linex loss function.

Chapter 4

Bayesian SSD under the bounded linex loss function

4.1 Introduction

In this chapter we will obtain the optimum sample size for a bounded linex loss function given by Wen and Levy (2001). We have reviewed this bounded asymmetric loss function in the section 1.3.10 of Chapter 1. If $p(\theta|\underline{x})$ is any posterior density then we also defined the posterior risk in the equation (1.18), section 1.4 of Chapter 1. Since it is difficult to solve the integral (1.18) analytically, we will find the optimum sample size for different distributions under this blinex loss function numerically using the program R. In our simulation study, first we will average the posterior risk of (1.18) to obtain the average posterior risk (APR), then adding a linear cost function

(1.28) with this APR we obtain the expected total cost as,

$$E(TC) = c_0 + cn + APR. \tag{4.1}$$

Then find $E(TC)$ for different n and plot $E(TC)$ vs n. Finally, find the optimum sample size n^* giving a minimum cost.

The algorithm to obtain an optimum sample size is as follows.

1. Generate a sample, x^* of size n from the pre-posterior distribution, $p(\underline{x}|n)$.
2. Find the posterior distribution for the parameter of interest $p(\theta|x^*, n)$
3. Generate θ^* from the posterior distribution $p(\theta|x^*, n)$. We considered 10,000 θ^* at this stage.
4. Find the minimum value of the posterior risk.
5. Repeat 1 to 4 for different samples and find the average posterior risk (APR).
6. Add a linear cost function $c_0 + cn$ to the APR to get $E(TC)$.
7. Repeat 1 to 6 for different n.
8. Plot $E(TC)$ vs n.
9. Find the sample size giving the minimum cost which is our required optimum sample size.

Now following the steps described above we will determine the optimum sample size under the bounded linex loss function (1.18) for different distributions. We shall present all the tables and figures at the end of this chapter.

4.2 SSD to estimate a Normal mean when precision is known

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean θ and known precision ξ. Let us take a conjugate prior as normal with mean μ_0 and precision $n_0\xi$. The posterior distribution, $p(\theta|\underline{x})$ given in chapter 2, equation (2.3) which is also a normal distribution with mean, $\hat{\mu} = \frac{n\overline{x}+n_0\mu_0}{n+n_0}$ and precision, $\hat{\xi} = (n + n_0)\xi$. Under the blinex loss given in (1.13) the posterior risk function is given in the equation (1.18), where the posterior density $p(\theta|\underline{x})$ is now $N(\hat{\mu}, \hat{\xi})$. It is difficult to solve the integral (1.18) analytically, so we will use R to get the optimum sample size following the steps described in the previous section.

4.2.1 Numerical study

In table 4.1, we have presented the optimum sample size by varying the bounding parameter γ and the shape parameter b when the scale parameter $a = 0.5$ is kept fixed to estimate a normal mean. We also found from the table that, if the shape and scale parameter b and a are fixed but the bounding parameter γ increases then the optimum sample size decreases. On the other hand if γ fixed then for an increasing shape parameter b, the optimum sample size also increases when the scale parameter a is fixed. In figure 4.1 we found that for the given set of parameters the optimum sample size is around 24 as it gives the minimum cost of sampling which is given in bold type in table 1 as R output.

4.3 SSD to estimate a normal precision when mean is known.

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean μ_0 and precision ξ. The posterior distribution of $\xi|\underline{x}$ is same as Chapter 2, equation (2.11). This time we want to estimate ξ by d. Under the blinex loss given in (1.13) the posterior risk function is given in the equation (1.18), where the posterior density $p(\theta|\underline{x})$ is given in (2.11). It is difficult to solve the integral (1.18) analytically, so we will run the program R to get the APR then we add it with the cost function (1.28) for the expected total cost. In step 1 of the R program we have simulated x^* from a t distribution (Bernardo and Smith 1994) as the pre-posterior distribution. Then we followed all the steps in order for an optimum sample size described earlier. In table 4.2, we have presented the optimum sample size under the blinex loss function for different bounding parameter γ and shape parameter b. We obtained each of the sample sizes from the figure (one is shown in figure 4.2) by comparing the sample size against the total cost and the sample with the minimum cost is chosen as the optimum sample size.

4.3.1 Numerical study

In table 4.2 we found that for a fixed shape parameter b if the bounding parameter γ goes up then the optimum sample size goes down. On the other hand for a fixed bounding parameter γ if we increase the shape parameter b then the optimum sample

size also increase. The R output is presented in the figure 4.2 and we found that the optimum sample size is around 45 (shown as bold type in the table 4.2) to estimate a normal precision under the blinex loss function when $b = \gamma = 0.1$.

4.4 SSD to estimate a normal mean when precision is unknown

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean μ and precision ξ. So the likelihood of the sample will be,

$$p(\underline{x}|\mu,\xi) \propto \xi^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\xi \sum (x_i - \mu)^2\right\}. \tag{4.2}$$

Now specify a prior for μ and ξ, $p(\mu,\xi)$ in the form $p(\mu|\xi)p(\xi)$. Since for known ξ the normal is self conjugate we use a normal prior for $\mu|\xi$ and we also use a gamma prior for ξ. So take $\mu|\xi$ is a $N(\mu_0, n_0\xi)$ distribution and ξ is a $Gamma(\alpha/2, \beta/2)$ distribution. That is,

$$p(\mu|\xi) = \sqrt{\frac{n_0\xi}{2\pi}} \exp\left\{-\frac{(n_0\xi)}{2} \sum (\mu - \mu_0)^2\right\} \tag{4.3}$$

and

$$p(\xi) = \{\Gamma(\alpha/2)\}^{-1}(\beta/\alpha)^{\alpha/2}\xi^{\alpha/2-1} \exp(-(1/2)\beta\xi). \tag{4.4}$$

Now using the result of (4.2), (4.3) and (4.4) we have the joint distribution of (μ, ξ) will be,

$$\begin{aligned} p(\mu, \xi|\underline{x}) &= p(\underline{x}|\mu, \xi) \times p(\mu|\xi) \times p(\xi) \\ &\propto \xi^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\xi \sum (x_i - \mu)^2\right\} \times \xi^{\frac{1}{2}} \exp\{-\frac{n_0\xi}{2}(\mu - \mu_0)^2\} \\ &\quad \times \xi^{\frac{\alpha}{2}-1} \exp\left[-\frac{\beta\xi}{2}\right] \\ &\propto \xi^{\frac{\alpha+n+1}{2}-1} \exp\left[-\frac{\xi}{2}\{\beta + \sum (x_i - \mu)^2 + n_0(\mu - \mu_0)^2\}\right]. \end{aligned} \tag{4.5}$$

Now we want $p(\mu)$ by

$$p(\mu) = \int p(\mu, \xi|\underline{x}) d\xi = \int p(\underline{x}|\mu, \xi) p(\mu|\xi) p(\xi) d\xi. \tag{4.6}$$

Now integrating (4.6) and after some calculations we have the posterior unconditional distribution of μ has a t distribution with $\alpha + n$ degrees of freedom with location $\mu^* = \frac{n_0\mu_0 + n\overline{x}}{n_0 + n}$ and precision $\frac{\alpha^*}{\beta^*}$ where,

$$\alpha^* = (\alpha + n)(n_0 + n)$$

and

$$\beta^* = \beta + \frac{1}{n + n_0}\left[n \sum (x_i - \mu)^2 + n_0(x_i - \mu_0)^2\right].$$

That is, $\mu|x$ has a $t_{\alpha+n}\left(\mu^*, \frac{\alpha^*}{\beta^*}\right)$ distribution. The posterior risk is the blinex loss averaged over this t distribution. So we will run R program to obtain the optimum sample size. We assumed the scale parameter $a = 0.5$ throughout the simulation study. In our usual optimization procedure we obtained the average posterior risk (APR) of μ through the simulation study using R, then minimized the expected total cost for an optimum sample size following the steps presented in the section 4.1.

4.4.1 Numerical study

From table 4.3 we observe that for a fixed shape parameter b if the bounding parameter γ goes up then the optimum sample size goes down. On the other hand for a fixed bounding parameter γ if we increase the shape parameter b then the optimum sample sizes are also increases. Figure 4.3 is one example of the R output where we found that for $\alpha = 1$, $\beta = 2$, $n_0 = 1$, $\mu_0 = 1$, $c_0 = 1$, $c = 0.001$, $\gamma = 0.3$, $b = 0.4$ and $a = 0.5$ the optimum sample size is around 18 (shown in bold type in the table 4.3). We considered prior mean of the precision is 2 so that $sd = \frac{1}{\sqrt{2}}$. This means results are directly comparable to the known precision case. We found that the optimum sample size to estimate a normal mean when precision is unknown is bigger than the optimum sample size to estimate a normal mean when precision is known (table 4.1) because of the extra uncertainty about the precision.

4.5 SSD for an exponential distribution

Suppose x_1, x_2,....,x_n is a random sample of size n taken from an exponential distribution with parameter λ with density,

$$p(x|\lambda) = \lambda \exp(-\lambda x); x > 0, \lambda > 0.$$

The posterior density of $\lambda|\underline{x}$ is same as of Chapter 2 equation (2.32). This time we want to estimate λ by d, hence under the blinex loss (1.13) we can easily get a mathematical form of the posterior risk function using (1.18) of Chapter 1, which is difficult to obtain analytically. So to obtain the APR we will use the steps involve in R

and to do that in the step-1, we have simulated x^* from the pre-posterior distribution which is a gamma-gamma distribution. Then for the expected total cost we have added APR with the cost function (1.28) with $c_0 = 1$ and $c = 0.0001$. Finally plot $E(TC)$ against n to get the optimum sample size. In table 4.4 we have presented the optimum sample size for different combination of the bounding parameter γ and the shape parameter b, keeping the scale parameter a is fixed.

4.5.1 Numerical study

From table 4.4 we found that for a fixed shape parameter b if the bounding parameter goes up then the optimum sample sizes goes down. On the other hand for a fixed bounding parameter γ if we increase the shape parameter b then the optimum sample sizes are also increases. In figure 4.4 we have presented the R output where we found that for $\alpha = \beta = 1$, $a = 0.5$, $\gamma = 0.3$ and $b = 0.3$ the optimum sample size is around 50 (shown in bold type in table 4.4).

4.6 SSD for a Poisson distribution

Let x_1, x_2,....,x_n be a sample of size n taken from a Poisson distribution. The posterior distribution of $\theta|\underline{x}$ is presented in Chapter 2, equation (2.38). Now if we want to estimate θ by d, then the mathematical form of the posterior risk function under the blinex loss (1.13) can easily obtain from (1.18) which is difficult to find analytically. The optimum sample size is obtained by minimizing (4.1). This time in step-1, to

draw x^*, we have considered a negative binomial distribution with parameter α and $\frac{\beta}{\beta+n}$ where α is a positive integer. In table 4.5, we have presented the optimum sample size for different combination of γ and b, keeping the scale parameter $a = 0.5$ is fixed. We have considered the prior parameters, $\alpha = \beta = 2$, $c_0 = 1$ and $c = 0.0001$ in our simulation study.

4.6.1 Numerical study

From table 4.5 we observe that for a fixed shape parameter b if the bounding parameter γ goes up then the optimum sample size goes down. On the other hand for a fixed bounding parameter γ if we increase the shape parameter b then the optimum sample sizes also increases. In figure 4.5, we have showed how we obtained the optimum sample size against the minimum cost for a given set of parameters using the program R. We found that the optimum sample size is around 30 under the blinex loss function when the shape parameter, $b = 0.4$ and the bounding parameter $\gamma = 0.5$, which is shown in bold type in table 4.5.

4.7 Conclusion

In this chapter we obtained the optimum sample size for various distributions under the blinex loss function. Because of the complicated form of the posterior risk function, we have used R program to minimize the total expected cost. First we considered three different cases of normal distribution. We considered SSD to estimate a

normal mean when the precision is known, SSD to estimate a normal precision when mean is known and SSD to estimate a normal mean when the precision is unknown. In the last case we have simulated the pre-posterior samples from a t distribution. In this situation we are not able to use the linex loss function because the mgf of t distribution does't exists. As we discussed in Chapter 1, only distributions which has the mgf can give an optimum Bayes decision under the linex loss function. Hence the blinex loss overcomes this weakness of the linex loss function and we have obtained the optimum sample size for this loss function. Then we considered SSD to estimate an exponential parameter. Finally, we considered SSD to estimate a Poisson parameter. We found that the optimum sample size depends on the prior parameters, the scale, shape and bounding parameter of the blinex loss function. In all the cases for a fixed scale, shape and prior parameters, if the bounding parameter of the blinex loss function increases then the optimum sample sizes are decreases. On the other hand for a fixed scale, bounding and prior parameters, if the shape parameter of the blinex loss function increases then the optimum sample size also increases.

Table 4.1: Optimum sample size to estimate a normal mean for different values of γ and b when $\theta = 1, sd = \sqrt{2}, n_0 = 1, \mu_0 = 1$, $c_0 = 1$ and $c = 0.001$.

Sample size	b				
γ	0.1	0.2	0.3	0.4	0.5
0.1	10	15	18	22	**24**
0.2	7	10	12	15	20
0.3	5	8	10	12	16
0.4	3	6	8	11	13
0.5	1	3	4	6	8

Table 4.2: Optimum sample size to estimate a normal precision (mean known) when $\xi = 1$, $n_0 = 10$, $\mu_0 = 1$, $c = 0.0001$, $\alpha = \beta = 1$.

Sample size	b				
γ	0.1	0.2	0.3	0.4	0.5
0.1	**45**	62	74	96	110
0.2	35	48	60	72	85
0.3	30	40	50	63	70
0.4	26	34	44	55	56
0.5	20	30	40	46	48

Table 4.3: Optimum sample size to estimate a normal mean (precision unknown) when $\alpha = 1$, $\beta = 2$ $n_0 = 1$, $\mu_0 = 1$, $c = 0.001$.

Sample size	b				
γ	0.1	0.2	0.3	0.4	0.5
0.1	16	19	23	27	33
0.2	13	17	22	25	28
0.3	10	12	15	**18**	21
0.4	7	11	13	16	19
0.5	3	6	9	13	17

Table 4.4: Optimum sample size to estimate an exponential parameter λ for different values of γ and b when $\alpha = \beta = 1$ and $a = 0.5$.

Sample size	b				
γ	0.1	0.2	0.3	0.4	0.5
0.1	45	68	81	97	110
0.2	35	46	69	81	92
0.3	25	36	**50**	68	77
0.4	15	27	34	45	63
0.5	10	18	30	36	47

Table 4.5: Optimum sample size to estimate a Poisson parameter θ for different values of γ and b when $\alpha = \beta = 2$ and $a = 0.5$.

Sample size	b				
γ	0.1	0.2	0.3	0.4	0.5
0.1	34	45	56	65	72
0.2	25	36	40	49	58
0.3	20	31	36	42	49
0.4	15	24	32	35	43
0.5	12	21	26	**30**	36

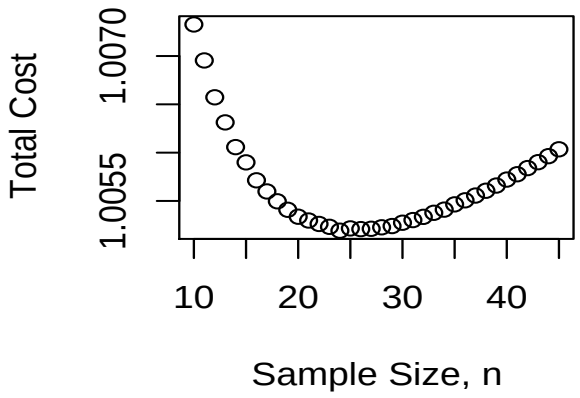

Figure 4.1: SSD to estimate a normal mean (precision is known) under the blinex loss for $\theta = 1$, $sd = \sqrt{2}$, $n_0 = 1$, $\mu_0 = 1$, $c = 0.001$, $\gamma = 0.1$, $b = 0.5$ and $a = 0.5$.

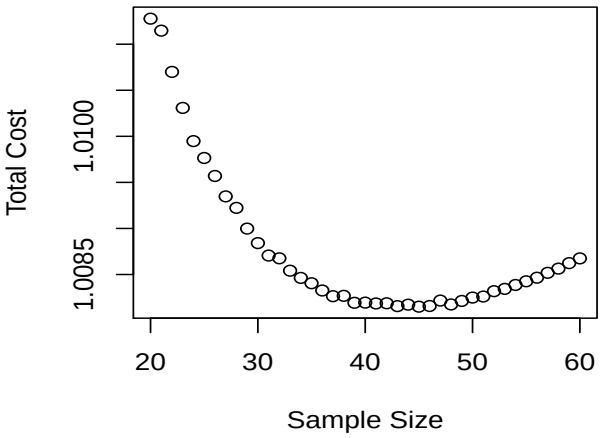

Figure 4.2: SSD to estimate a normal precision (mean is known) for $\xi = 1$, $n_0 = 10$, $\mu_0 = 1$, $c = 0.0001$, $\gamma = b = 0.1$, $\alpha = \beta = 1$.

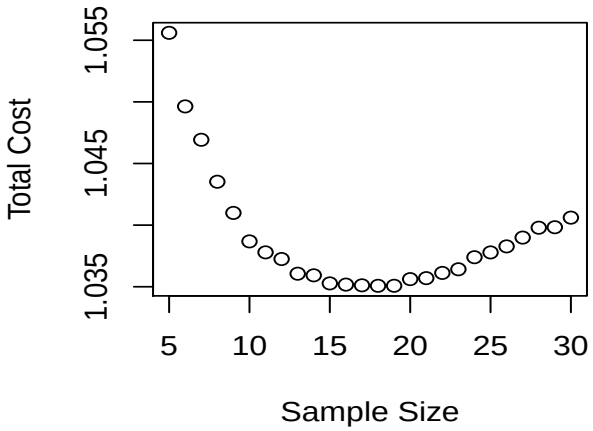

Figure 4.3: SSD to estimate a normal mean (precision unknown) when $\alpha = 1$, $\beta = 2$, $n_0 = 1$, $\mu_0 = 1$, $c = 0.001$, $c_0 = 1$, $\gamma = 0.3$, $b = 0.4$ and $a = 0.5$.

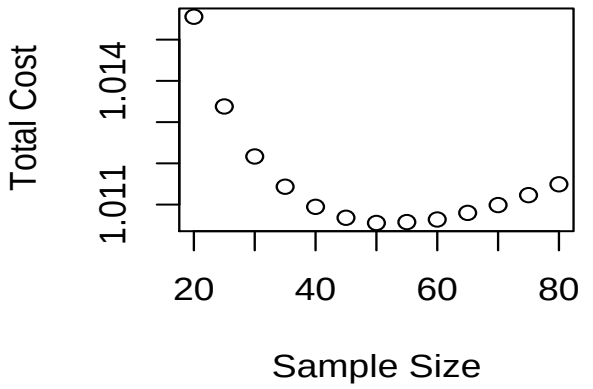

Figure 4.4: SSD to estimate an exponential parameter when $\lambda = 1.5$, $\alpha = \beta = 1$, $c = 0.0001$, $\gamma = 0.3$, $b = 0.3$ and $a = 0.5$.

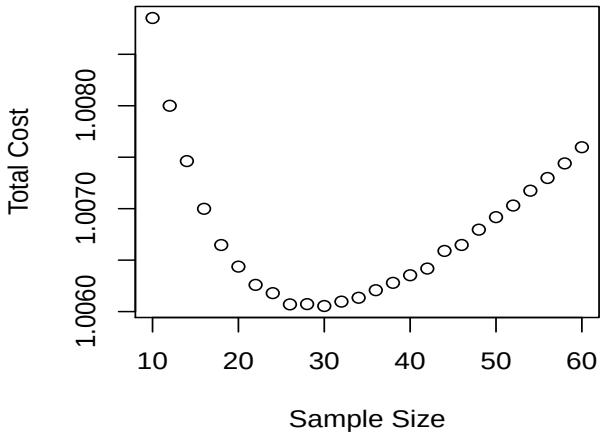

Figure 4.5: SSD to estimate a Poisson parameter for $\lambda = 1$, $\alpha = \beta = 2$, $c_0 = 1$, $c = 0.0001$, $b = 0.4$ and $\gamma = 0.5$.

Chapter 5

SSD under the loss function by DeGroot (1970)

5.1 Introduction

In Chapters 3 and 4 we obtained the optimum sample size under an asymmetric linex loss function and an asymmetric bounded linex loss function respectively. But there are some other asymmetric loss functions which may be considered when under estimation is more serious than overestimation or vice-versa. Besides the linex loss function we also observed some other asymmetric loss functions which are to be found in the literature and it is possible to obtain the optimum sample size for these asymmetric loss functions. In this chapter, we obtain the optimum sample size under an asymmetric loss function due to DeGroot (1970) with a linear cost function given

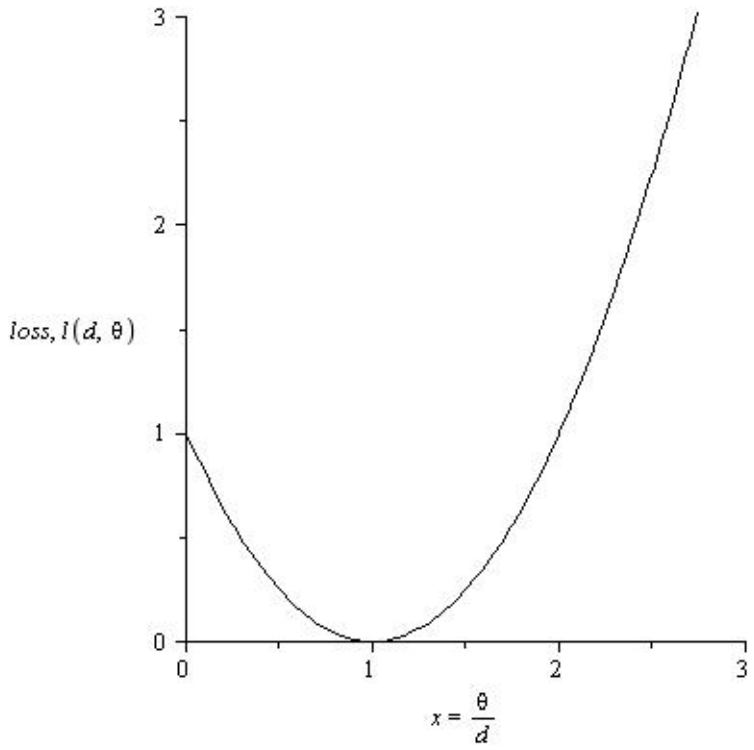

Figure 5.1: Shape of the loss function where, $x = \frac{\theta}{d}$.

in (1.28) for various distributions. In a situation where we are unable to get analytic results we will use the R program to find the optimum sample size. If d is an estimate of θ then recall the asymmetric loss function is defined by DeGroot (1970) from Chapter 1 as,

$$l(d,\theta) = \left(\frac{\theta - d}{d}\right)^2, \tag{5.1}$$

where $\theta > 0$. The form of the posterior risk function for the Bayes estimate θ has been given in Chapter 1, equation (1.16).

Properties: i) The estimation error is $x - 1$, where $x = \frac{\theta}{d}$.

ii) For $0 \leq x \leq 2$, the loss function $l(d,\theta)$ is symmetric.

iii) At the point $x = \frac{\theta}{d} = 1$ that is for $\theta = d$, the loss is zero.

iv) For $x > 2$ the shape of the loss function is increasing.

In the following we will provide the optimum SSD for a number of cases under the loss function (5.1) and present all the tables and figures at the end of the chapter.

5.2 SSD to estimate a normal mean when precision is known.

Suppose x_1, x_2,...., x_n is a random sample of size n taken from a normal distribution with mean θ and known precision ξ. Let us take a conjugate prior for θ also as a normal with mean μ_0 and precision $n_0\xi$. Now from chapter 2 equation (2.3) we have the posterior distribution of $\theta|\underline{x}$ is also a normal distribution with mean $\frac{n\bar{x}+n_0\mu_0}{n+n_0}$ and precision $(n+n_0)\xi$. Using the equation (1.16) the posterior risk for the loss function (5.1) will be,

$$PR = \left[1 + \frac{\xi(n\bar{x} + n_0\mu_0)^2}{n+n_0}\right]^{-1}. \tag{5.2}$$

Now adding the cost function (1.28) with the posterior risk (5.2) the total cost will be,

$$TC(n) = c_0 + cn + \left[1 + \frac{\xi(n\bar{x} + n_0\mu_0)^2}{n+n_0}\right]^{-1}; \tag{5.3}$$

which depends on x. Now we need to take expectation over (5.3) w.r.t the pre-posterior distribution of $\bar{x}$, we have

$$E_{\bar{x}}[TC(n)] \;=\; c_0 + cn + \int \left[1 + \frac{\xi(n\bar{x} + n_0\mu_0)^2}{n+n_0}\right]^{-1} p(\bar{x})d\bar{x}.$$

It is not possible to obtain $E(TC)$, so we will use the R program to find an optimum sample size. We will follow the steps as in Chapter 4. This time in step

1, we have simulated the data, x^* from the pre-posterior distribution of x which is a $N\left(\mu_0, \frac{1}{\frac{1}{\xi}+\frac{1}{n_0\xi}}\right)$ distribution, then for an optimum SSD we followed all the steps described in the chapter 4 section (4.1). In our study we have considered a linear cost function (1.28) with $c_0 = 1$ and $c = 0.0001$. In table 5.1, we shall present the optimum sample sizes to estimate a normal mean, θ when the precision $\xi = 10$.

5.2.1 Numerical study

From table 5.1, we observed that for the fixed prior mean μ_0 if the prior sample size n_0 increases then the optimum sample size decreases. Again for the fixed prior sample size n_0 if the prior mean μ_0 increases, then the optimum sample size decreases. The notation '-' shows that it is not worth sampling as for these combination of parameters the values of n are less than zero. Moreover, in figure 5.2 we have presented the R output where the optimum sample is around 28 for $\mu_0 = 0.5$ and $n_0 = 5$ which is shown as bold type in the table 5.1.

5.3 SSD to a estimate a normal precision when mean is known.

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a normal distribution with mean μ_0 and precision ξ. The posterior distribution of $\xi|\underline{x}$ is given in Chapter 2 equation (2.11). Under the asymmetric loss function (5.1), following the equation

(1.16) the posterior risk is

$$PR = \frac{1}{1+\alpha+\frac{n}{2}}.$$

Adding the linear cost function (1.28) with the posterior risk PR, the total cost will be,

$$TC(n) = c_0 + cn + \frac{1}{1+\alpha+\frac{n}{2}}. \tag{5.4}$$

To get the optimum sample size differentiate (5.4) w.r.t. n and setting equal to zero we have,

$$n^* = \sqrt{\frac{2}{c}} - 2(\alpha+1). \tag{5.5}$$

The optimum sample size is independent of β because we found the posterior risk function (PR) is independent of β.

5.3.1 Numerical study

From figure 5.3 we see that if the sampling cost per unit c goes up then the optimum sample size goes down for any prior parameter, α. We also see from the equation (5.5) that if the prior parameter α increases then the optimum sample size decreases for any fixed sampling cost per unit, c.

5.4 SSD for an exponential parameter

Suppose x_1, x_2,....,x_n is a random sample of size n taken from an exponential distribution with parameter λ. The posterior distribution of $\lambda|\underline{x}$ is presented in Chapter

2 equation (2.32). Now following equation (1.16) the posterior risk under the loss function (5.1) will be,

$$PR = \frac{1}{1+\alpha+n}.$$

Now the total cost will be,

$$TC(n) = c_0 + cn + \frac{1}{1+\alpha+\frac{n}{2}}. \tag{5.6}$$

So for the optimum sample of size n differentiate (5.6) w.r.t. n and setting equal to zero we have,

$$n^* = \frac{1}{\sqrt{c}} - (\alpha+1), \tag{5.7}$$

which is the optimum sample size to estimate an exponential parameter under the loss function (5.1).

5.4.1 Numerical study

From figure 5.4 we can see that if the sampling cost per unit c goes up then the optimum sample sizes are goes down for any prior parameter, α. We also observe from the equation (5.7) that if the prior parameter α increases then the optimum sample size decreases for the fixed sampling cost per unit, c.

5.5 SSD for a Poisson distribution

Let x_1, x_2,....,x_n be a sample of size n from a Poisson distribution with parameter θ. The posterior distribution of $\theta|\underline{x}$ is given in Chapter 2 equation (2.38). Now following

the equation (1.16) the posterior risk under the loss function (5.1) will be,

$$PR = \frac{1}{1+t+n}.$$

So the total cost will be,

$$TC(n) = c_0 + cn + \frac{1}{1+t+n};$$

which depends on data vector t. Now take an expectation over $TC(n)$ w.r.t. t we have,

$$E_t[TC(n)] = c_0 + cn + E_t\left[\frac{1}{1+t+n}\right].$$

It is not possible to obtain an analytic result of $E_t(TC)$ so we will use the R program for an optimum sample size n^*. We will follow the steps in R given in the previous chapter and in step 1 we simulated the data, x^* from the pre-posterior distribution which is a negative binomial distribution with parameter α and $\frac{\beta}{\beta+n}$. Then we followed all the steps from 2-9 in order for an optimum sample size. In this study we have considered a linear cost function (1.28) with $c_0 = 1$ and $c = 0.001$. In table 5.2, we present the optimum sample size to estimate a Poisson parameter, λ. In figure 5.5 we have presented the R output for the prior parameters $\alpha = 1$, $\beta = 2$.

5.5.1 Numerical study

From table 5.2 we observe that if the prior parameter α goes up, then the optimum sample size goes down for a fixed β. On the other hand if the prior parameter, β goes up, then the optimum sample sizes also go up for a fixed value of α. From figure 5.5 we found that the optimum sample size is around 15 for the prior parameter $\alpha = 1, \beta = 2$ which is presented in bold type in table 5.2 as R output.

5.6 Conclusion

In this chapter we obtained the optimum sample size for a normal, an exponential and a Poisson distributions under an asymmetric loss function due to DeGroot (1970). We have defined this loss function in the equation (5.1) and the form of the posterior risk function is provided in the equation (1.16). To estimate a normal mean and a Poisson parameter we are unable to obtain the optimum sample size analytically because of the complicated form of the posterior risk function. We then developed the programs in R to obtain an optimum sample size in these cases. We also obtained the optimum sample size to estimate a normal precision when mean is known and SSD to estimate an exponential parameter. In these two situations we obtained the optimum sample size with closed and simple form formulas under the asymmetric loss function (5.1) without any approximation.

Table 5.1: Optimum sample size to estimate a normal mean for different values of the prior parameters μ_0 and n_0, $c_0 = 1$ and $c = 0.0001$.

μ_0	n_0	n^*	n_0	μ_0	n^*
0.5	5	**28**	5	0.5	28
	10	22		1	15
	15	16		1.5	12
	20	11		2	8
1	5	15	10	0.5	22
	10	12		1	12
	15	8		1.5	7
	20	5		2	3
1.5	5	12	15	0.5	15
	10	7		1	8
	15	3		1.5	5
	20	-		2	2

Table 5.2: Optimum sample size to estimate a Poisson parameter for different values of the prior parameters α and β when $c_0 = 1$ and $c = 0.001$.

α	β	n^*	β	α	n^*
1	1	12	1	1	12
	2	**15**		2	10
	3	18		3	8
	4	20		4	6
2	1	10	2	1	15
	2	12		2	12
	3	14		3	10
	4	16		4	8
3	1	8	3	1	18
	2	10		2	14
	3	12		3	12
	4	14		4	10

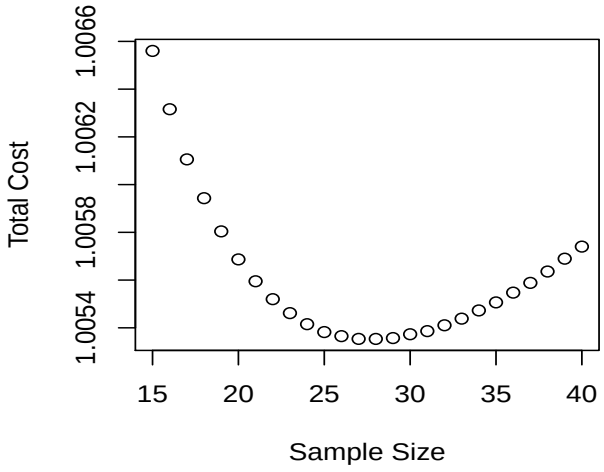

Figure 5.2: SSD to estimate a normal mean θ for $\mu_0 = 0.5$, $n_0 = 5$, $c_0 = 1$ and $c = 0.0001$.

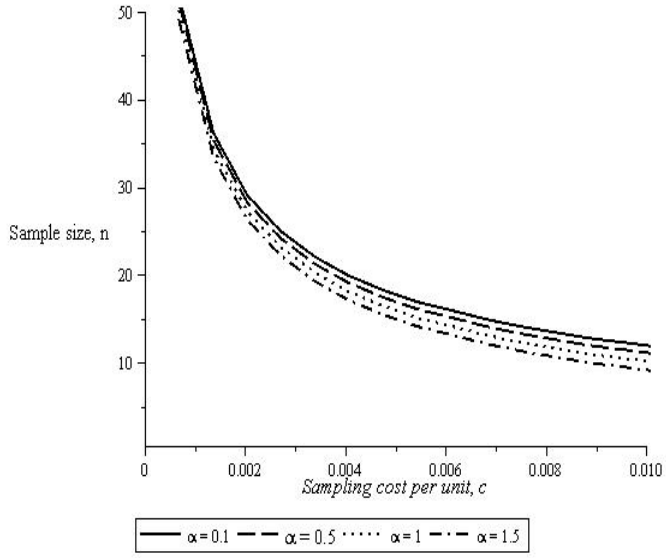

Figure 5.3: Optimum sample size to estimate a normal precision (mean known) as a function of the sampling cost per unit, c for different values of α.

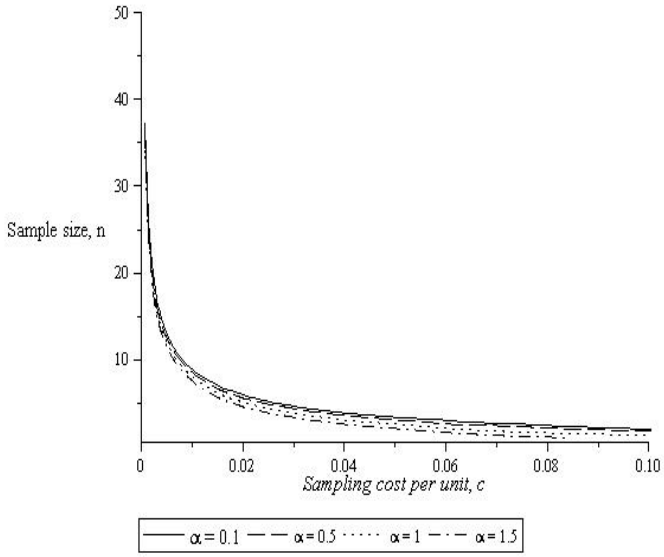

Figure 5.4: Optimum sample size to estimate an exponential parameter θ as a function of the sampling cost per unit, c for different values of α.

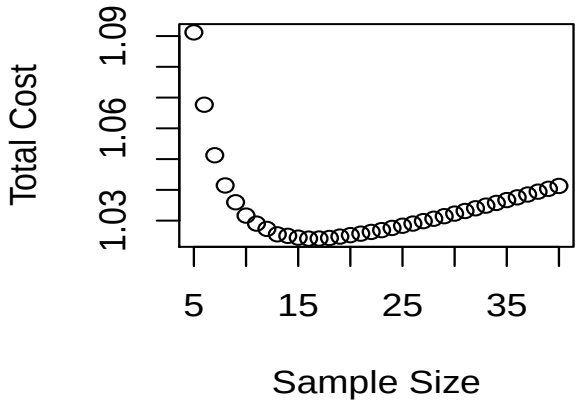

Figure 5.5: SSD to estimate a Poisson parameter λ for $\alpha = 1$, $\beta = 2$, $c_0 = 1$ and $c = 0.001$.

Chapter 6

SSD under a scaled exponential utility function

6.1 Introduction

In this chapter we propose a utility function and obtain the Bayes estimate and the optimum sample size under this utility function. This utility function is designed especially to obtain the Bayes estimate when the posterior follows a gamma distribution. If d is an estimate of θ then the loss or utility function should be functions of d and θ which incorporate either the difference of d and θ or the ratio of d and θ or both. But there is an inverse relation between loss and utility functions. If we define a loss function as $l(d, \theta)$ and an utility function as $u(d, \theta)$, then the relation between them is $l(d, \theta) = \max_d u(d, \theta) - u(d, \theta)$. This means for a unique Bayes estimate, mini-

mizing expected loss is equivalent to the maximizing expected utility. In practice, the experimenter wants to have some closed form estimates for their decisions without approximation considering either utility or loss functions. In most cases, they are happy to work with the posterior mean. So if the posterior mean is the estimate to choose then what will be the form of the utility or loss function(s)? The simple answer is a squared error loss function. Is this always true? Can we have any other utility or loss function which is asymmetric in nature and gives the posterior mean as a Bayes estimate. To answer this question we will define a new utility function to estimate the parameter $\in R_+$. We will also obtain the expected utility for this utility function. Then we will present an equivalent form of loss function. Finally we will obtain the optimum sample size by minimizing both the posterior expected loss and a linear cost function defined in Chapter 1 equation (1.28). As some of the posterior risk functions are in complicated form, we are unable to differentiate it for an optimum sample size. In those situations we use the R program to obtain the optimum sample size. We consider a Normal with known mean, a Pareto, an Exponential and a Poisson distribution for an optimum sample size under the proposed utility function. Now we will define this utility function.

6.2 Scaled exponential utility function

If d is an estimate of θ, then define a utility function as,

$$u(d,\theta) = \frac{\theta}{d}\exp\left[\left(1-\frac{\theta}{d}\right)\right]. \tag{6.1}$$

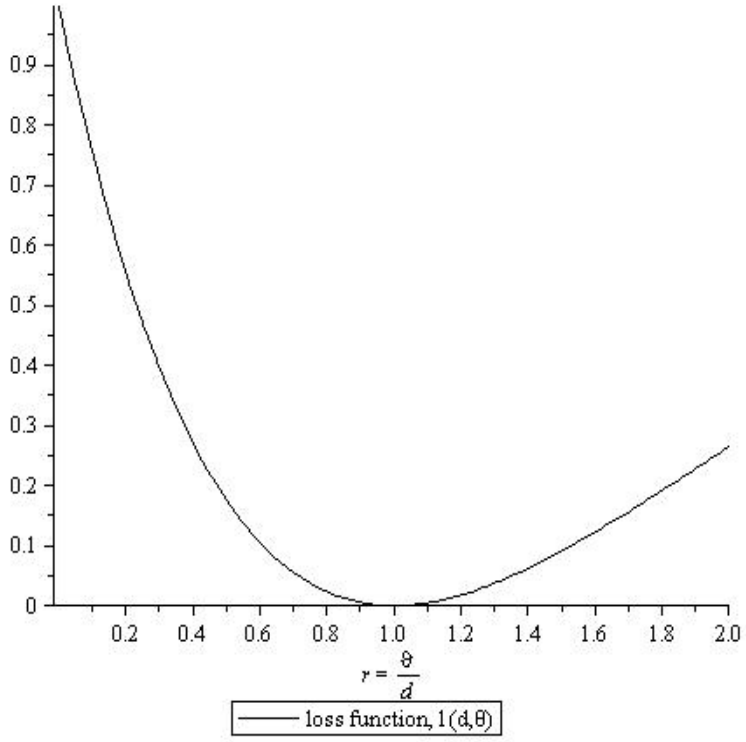

Figure 6.1: Shape of the scaled exponential loss function for, $r = \frac{\theta}{d}$.

Where, $\theta \in R_+$ and the utility will be maximum at the point $\theta = d$ that is $u(d, d) = 1$. Alternatively we can consider the above utility function as a loss function below.

$$l(d, \theta) = 1 - r \exp(1 - r), \tag{6.2}$$

where, $r = \frac{\theta}{d}$ and for $r = 1$ we have $d = \theta$ means there is no loss implies $l(d, d) = 0$.

Properties:

i) This loss function is bounded as the maximum loss could be 1 and the minimum loss is zero.

ii) The loss is decreasing if, $0 < r \leq 1$.

iii) If $r = 1$ that is $\theta = d$ then the loss is zero as the estimation error is $r - 1$.

iv) The loss is increasing for $r > 1$.

This is an asymmetric loss function because the loss, $l(d,\theta)$ decreases sharply when r takes the value from 0 to 1 and then it will be 0 for $r = 1$. Finally, the shape of the loss increases less steeply for $r > 1$. Since the exponential form of utility is scaled by $\frac{\theta}{d}$ hence the name of the proposed utility function is a scaled exponential utility function. In the following theorem we will obtain the Bayes estimate for this utility function.

Theorem 1 *Suppose x_1, x_2,....,x_n is a random sample of size n taken from any distribution $p(x|\theta)$, where θ is the parameter under study $\in R_+$. Let $p(\theta)$ be any prior distribution of θ that gives the posterior distribution $p(\theta|\underline{x})$ as a Gamma distribution with parameters (say) α and β. Then the Bayes estimate under the utility function $u(d,\theta) = \frac{\theta}{d}\exp\left[(1-\frac{\theta}{d})\right]$ will be the posterior mean $\frac{\alpha}{\beta}$.*

Proof

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a density $p(x|\theta)$. If $p(\theta)$ is any prior for θ gives the posterior $p(\theta|\underline{x})$ as a gamma distribution with parameter α and β. Consider the utility function $u(d,\theta) = \frac{\theta}{d}\exp\left[(1-\frac{\theta}{d})\right]$. So the expected utility w.r.t. the posterior density $p(\theta|\underline{x})$ will be,

$$E[u(d,\theta)] = \int \frac{\theta}{d}\exp\left(1-\frac{\theta}{d}\right)p(\theta|\underline{x})d\theta. \tag{6.3}$$

Now differentiate (6.3) w.r.t. d we have,

$$\begin{aligned}
\frac{\partial E[u(d,\theta)]}{\partial d} &= \frac{\beta^{\alpha} e}{\Gamma(\alpha)} \int \theta^{\alpha+1-1} \exp\{-(\beta\theta)\} \\
&\quad \times \left[\frac{1}{d}\left(\frac{\theta}{d^2}\right) \exp\left\{-\frac{\theta}{d}\right\} - \frac{1}{d^2} \exp\left\{-\frac{\theta}{d}\right\}\right] d\theta \\
&= \frac{\beta^{\alpha} e}{\Gamma(\alpha)} [\frac{1}{d^3} \int \theta^{\alpha+2-1} \exp\left\{-\left(\beta + \frac{1}{d}\right)\theta\right\} d\theta \\
&\quad -\frac{1}{d^2} \int \theta^{\alpha+1-1} \exp\left\{-\left(\beta + \frac{1}{d}\right)\theta\right\} d\theta] \\
&= \frac{\beta^{\alpha} e}{\Gamma(\alpha)} \left[\frac{\Gamma(\alpha+2)}{d^3(\beta+\frac{1}{d})^{\alpha+2}} - \frac{\Gamma(\alpha+1)}{d^2(\beta+\frac{1}{d})^{\alpha+1}}\right] \\
&= \frac{\beta^{\alpha} e}{d^2(\beta+\frac{1}{d})^{\alpha+1}} \left[\frac{\alpha(\alpha+1)}{d(\beta+\frac{1}{d})} - \alpha\right] \\
&= \frac{\alpha\beta^{\alpha} e^k}{d^2(\beta+\frac{1}{d})^{\alpha+1}} \left[\frac{(\alpha+1)}{d(\beta+\frac{1}{d})} - 1\right].
\end{aligned}$$

To obtain the Bayes estimate now setting $\frac{\partial Eu(d,\theta)}{\partial d} = 0$ we have,

$$\frac{(\alpha+1)}{d(\beta+\frac{1}{d})} - 1 = 0.$$

Gives,

$$\hat{d} = \frac{\alpha}{\beta},$$

which is clearly the posterior mean.

In the following theorem we will present the expected utility for the utility function defined in the equation (6.1).

Theorem 2 *Suppose x_1, x_2,...,x_n is a random sample of size n taken from any distribution $p(x|\theta)$, where θ is the parameter under study $\in R_+$. Let, $p(\theta)$ is any prior distribution of θ that gives the posterior distribution $p(\theta|\underline{x})$ as a Gamma distribution with parameters (say) α and β. Then the maximum expected utility under the utility function $u(d,\theta) = \frac{\theta}{d} \exp\left[(1-\frac{\theta}{d})\right]$ will be, $e\left(1+\frac{1}{\alpha}\right)^{-(\alpha+1)}$.*

Proof

Suppose $x_1,\ x_2,\ldots,x_n$ is a random sample of size n taken from a density $p(x|\theta)$. If $p(\theta)$ is any prior for θ gives the posterior $p(\theta|\underline{x})$ as a gamma distribution with parameter α and β. From theorem 1, the Bayes estimate under the utility function $u(d,\theta) = \frac{\theta}{d}\exp\left[(1-\frac{\theta}{d})\right]$ is,

$$\hat{d} = \frac{\alpha}{\beta},$$

which is the posterior mean. So the expected utility will be,

$$E[u(d,\theta)] = \int \frac{\theta}{d}\exp\left\{\left(1-\frac{\theta}{d}\right)\right\} p(\theta|\underline{x})d\theta. \tag{6.4}$$

The expected utility (6.4) will be maximum at $d = \frac{\alpha}{\beta}$. We have,

$$\begin{aligned}
E[u(d,\theta)] &= \int \frac{\theta}{d}\exp\left(1-\frac{\theta}{d}\right) p(\theta|\underline{x})d\theta \\
&= \frac{\beta^{\alpha}e}{d\Gamma(\alpha)} \times \int \theta^{\alpha+1-1}\exp\left\{-\left(\beta+\frac{1}{d}\right)\theta\right\} d\theta \\
&= \frac{\beta^{\alpha}e}{d\Gamma(\alpha)} \times \frac{\Gamma(\alpha+1)}{(\beta+\frac{1}{d})^{\alpha+1}} \\
&= \frac{\beta^{\alpha+1}e}{\alpha} \times \frac{\alpha}{(\beta+\frac{\beta}{\alpha})^{\alpha+1}} \\
&= \frac{\beta^{\alpha+1}e}{\beta^{\alpha+1}} \times \frac{1}{(1+\frac{1}{\alpha})^{\alpha+1}} \\
&= e\left(1+\frac{1}{\alpha}\right)^{-(\alpha+1)}.
\end{aligned}$$

Clearly the expected utility is independent of the prior parameter β. Now the equivalent form of the posterior risk under the loss function (6.2) will be,

$$PR = 1 - e\left(1+\frac{1}{\alpha}\right)^{-(\alpha+1)}. \tag{6.5}$$

In the following we will consider different distributions to obtain the optimum sample size under the loss function (6.2) using the results of the theorem 1 and theorem 2 and present all the tables and figures at the end of the chapter.

6.3 SSD to estimate a normal precision when mean is known

Suppose x_1, x_2,...,x_n is a random sample of size n taken from a normal distribution with mean μ_0 and precision ξ. The posterior distribution of $\xi|\underline{x}$ is given in chapter 2, equation (2.11). Now by theorem 1, Bayes estimate under the loss function (6.2) will be,

$$\hat{d} = \frac{\alpha + \frac{n}{2}}{\beta + \frac{t}{2}}. \tag{6.6}$$

From theorem 2 the expected utility will be,

$$E[u(\xi, \theta)] = e \times \left(1 + \frac{1}{\alpha + \frac{n}{2}}\right)^{-(\alpha + \frac{n}{2} + 1)}. \tag{6.7}$$

So from (6.5) the posterior risk function will be,

$$\begin{aligned} PR &= 1 - e \times \left(1 + \frac{1}{\alpha + \frac{n}{2}}\right)^{-(\alpha + \frac{n}{2} + 1)} \\ &= 1 - \exp\left\{1 - \left(\alpha + \frac{n}{2} + 1\right) \log\left(1 + \frac{1}{\alpha + \frac{n}{2}}\right)\right\} \\ &= 1 - \exp\left\{1 + \left(\alpha + \frac{n}{2} + 1\right) \log\left(\frac{\alpha + \frac{n}{2}}{\alpha + \frac{n}{2} + 1}\right)\right\} \end{aligned}$$

Let, $z = \alpha + \frac{n}{2} + 1$, so we have,

$$PR = 1 - \exp\left\{1 + z \log\left(1 - \frac{1}{z}\right)\right\}.$$

Now expanding the log term and neglecting 3rd and the higher powers of z (as if the power of z increases then the value of $\frac{1}{z}$ decreases), we have,

$$PR \approx 1 - \exp\left(-\frac{1}{z}\right),$$

substituting the value of z in PR we have,

$$PR \approx 1 - \exp\left(-\frac{1}{\alpha + \frac{n}{2} + 1}\right). \tag{6.8}$$

Now adding the cost function (1.28) with the posterior risk (6.8) we have the total cost as,

$$TC(n) \approx c_0 + cn + 1 - \exp\left(-\frac{1}{\alpha + \frac{n}{2} + 1}\right). \tag{6.9}$$

To obtain the optimum sample of size n, differentiate (6.9) w.r.t n, then setting equal zero we have,

$$4c\left(\alpha + \frac{n}{2} + 1\right)^2 - \exp\left\{-\frac{1}{2\left(\alpha + \frac{n}{2} + 1\right)^2}\right\} = 0. \tag{6.10}$$

Now we will use Maple 13 to solve the equation (6.10) for n. To solve for n we see that out of the three roots there is only one root that is positive and we shall consider it as the optimum sample size for the given values of c and α. In table 6.1 we present the optimum sample size for different values of c keeping α fixed and in table 6.2 we present the optimum sample size to estimate a normal precision for different values of α keeping the sampling cost per unit c fixed.

6.4 SSD to estimate an exponential parameter

Suppose x_1, x_2,....,x_n is a random sample of size n taken from an exponential distribution with parameter λ. From Chapter 2 equation (2.32) we have the posterior

density $\lambda|\underline{x}$ has a $Gamma(\alpha+n, \beta+s)$ distribution. Now by theorem 1, the Bayes estimate under the loss function (6.2) will be,

$$\hat{d} = \frac{\alpha+n}{\beta+s}$$

and by (6.5) the posterior risk (PR) will be,

$$\begin{aligned} PR &= 1 - e \times \left(1 + \frac{1}{\alpha+n}\right)^{-(\alpha+n+1)} \\ &= 1 - \exp\left\{1 - (\alpha+n+1)\log\left(1+\frac{1}{\alpha+n}\right)\right\} \\ &= 1 - \exp\left\{1 + (\alpha+n+1)\log\left(\frac{\alpha+n}{\alpha+n+1}\right)\right\} \end{aligned}$$

Let, $z = \alpha+n+1$, so the PR becomes,

$$PR = 1 - \exp\left\{1 + z\log\left(1-\frac{1}{z}\right)\right\}.$$

Now expanding the log term and neglecting 3rd and the higher powers of z (as if the power of z increases then the value of $\frac{1}{z}$ decreases), we have,

$$PR \approx 1 - \exp\left\{-\frac{1}{z}\right\},$$

substituting the value of z in PR, we have,

$$PR \approx 1 - \exp\left\{-\frac{1}{\alpha+n+1}\right\}. \tag{6.11}$$

Now adding the cost function (1.28) with the posterior risk (6.11) we have the total cost as,

$$TC(n) \approx c_0 + cn + 1 - \exp\left\{-\frac{1}{\alpha+n+1}\right\} \tag{6.12}$$

To obtain an optimum sample size n differentiate (6.12) w.r.t n, then setting equal to zero we have,

$$2c(\alpha+n+1)^2 - \exp\left\{-\frac{1}{2(\alpha+n+1)^2}\right\} = 0. \tag{6.13}$$

Now we will use Maple 13 to solve the equation (6.13) for n. To solve for n we see that out of the three roots only one root is positive and consider it as the optimum sample size for the given values of c and α. In table 6.3 we will present the optimum sample size to estimate an exponential parameter for different values of c keeping α is fixed and in table 6.4 we will provide the optimum sample size when the prior parameter α varies but the sampling cost per unit c kept fixed.

6.5 SSD to estimate the shape parameter of a Pareto distribution

Suppose x_1, x_2,....,x_n be a random sample of size n taken from a Pareto distribution with parameter θ. The density function will be,

$$p(x|\theta) = \theta\alpha_0^{\theta}x^{-(\theta+1)}; x > \alpha_0, \theta > 0.$$

We can re-write it as,

$$p(x|\theta) = \theta \exp\left\{-\theta \log\left(\frac{x}{\alpha_0}\right)\right\}\frac{1}{x}. \tag{6.14}$$

So the likelihood of the sample is,

$$p(\underline{x}|\theta) = \theta^n \exp\left\{-\theta\sum_{i=1}^{n}\log\left(\frac{x_i}{\alpha_0}\right)\right\}\prod\frac{1}{x_i}. \tag{6.15}$$

Let us consider the prior for θ as,

$$p(\theta) \propto \theta^{\alpha-1}\exp(-\beta\theta). \tag{6.16}$$

Now combining (6.15) and (6.16) the posterior distribution of $\theta|\underline{x}$ is,

$$p(\theta|\underline{x}) \propto \theta^{n+\alpha-1}\exp\{-(\beta + \sum\log x_i - n\log\alpha_0)\theta\}. \tag{6.17}$$

Which follows a gamma distribution with parameter $n+\alpha$ and $\beta+\sum \log x_i - n \log \alpha_0$. Now by theorem 1, the Bayes estimate under the loss function (6.2) is

$$\hat{d} = \frac{n+\alpha}{\beta + \sum \log x_i - n \log \alpha_0}.$$

Following the equation (6.5) we can obtain the posterior risk as,

$$PR = 1 - e \left[1 + \frac{1}{n+\alpha}\right]^{-(n+\alpha+1)}, \tag{6.18}$$

which is exactly the same as the posterior risk to estimate an exponential parameter obtained in previous section. This is because the posterior risk under the loss function (6.2) is independent of the prior parameter β. So in this situation we will have exactly the same cost function as of (6.12) and of course for an optimum sample size we need to solve the same equation as (6.13) which will lead us to the similar numerical results obtained from an exponential distribution.

6.5.1 Numerical study

For all the optimum sample sizes (to estimate a particular parameter discussed above) increases when the sampling cost per unit decreases for the fixed values of the prior parameters (refer to the table 6.1 and the table 6.3). Again we observe that if the prior parameter, α increases then the optimum sample size decreases for a fixed value of the sampling cost per unit c (refer to the table 6.2 and the table 6.4). We also observe that for the large sampling cost per unit, c and the bigger prior parameter value it might not worth sampling due to the high sampling cost or enough prior information. In table 6.2, to estimate a normal precision when mean is known for $\alpha = 10$ and $c = 0.002$, it is not worth sampling due to big sampling cost c and

large α. Similarly, in the table 6.4, it not worth sampling to estimate an exponential parameter for $(c = 0.0002, \alpha = 50)$, $(c = 0.0005, \alpha = 45)$ and $(c = 0.0005, \alpha = 50)$ because of the high sampling cost or enough prior information.

6.6 SSD to estimate a Poisson parameter

Let x follows a Poisson distribution with parameter θ. From chapter 2 equation (2.38), we have the posterior distribution of $\theta|\underline{x}$ is a gamma distribution with parameter $t+\beta$ and $n+\alpha$. By theorem 1, the Bayes estimate under the loss function (6.2) will be,

$$\hat{d} = \frac{t+\beta}{n+\alpha} \tag{6.19}$$

and by theorem 2, the corresponding posterior risk function will be,

$$PR = 1 - e\left(1 + \frac{1}{t+\beta}\right)^{-(t+\beta+1)}. \tag{6.20}$$

Adding the linear cost function (1.28) with (6.20) we have the total cost as,

$$TC = c_0 + cn + 1 - e\left(1 + \frac{1}{t+\beta}\right)^{-(t+\beta+1)}, \tag{6.21}$$

which depends on the data t. To obtain the optimum sample size we need find $E_t(TC)$. Because of the complicated form of the equation (6.21) we are unable to solve it analytically and use the R program to get the optimum sample size following the steps described in chapter 4. Here we have considered a negative binomial as the pre posterior distribution to draw samples in step-1. In table 6.5, we have presented the optimum sample size to estimate a Poisson parameter, θ for different values of the prior parameters α and β.

6.6.1 Numerical study

From table 6.5 we observe that if the prior parameter α goes up then the optimum sample size goes down for a fixed β. On the other hand, we found that if the prior parameter β goes up then the optimum sample sizes also go up for a fixed α. In figure 6.2 we have presented the R output where we can see that the optimum sample size is around 15 which gives the minimum cost for the prior parameters, $\alpha = 1, \beta = 4$ (which is shown in bold type in table 6.5).

6.7 Conclusion

Normally we know the Bayes estimate under the symmetric squared error utility (or loss) function will be the posterior mean. In this chapter we proposed a new asymmetric utility function which also gives the posterior mean as a Bayes estimate for the parameter $\in R_+$. The advantage of using this utility function is that it is bounded and it gives the posterior mean as a Bayes estimate for the gamma posterior distribution which is presented in theorem 1. Looking at the practical importance Smith (1980) obtained Bayes estimates under different form of bounded loss functions. We have reviewed his loss functions in Chapter 1. We obtained the expected utility function with a conjugate prior set up for this asymmetric form of utility function in theorem 2. Then we obtained the optimum sample size using this utility function. We considered a normal distribution to estimate the precision, a Pareto distribution, an exponential distribution and a Poisson distribution. We have presented the analytic results to obtain SSD for the normal, an exponential and a Pareto distribution. For

these distributions we found from the numerical study that if the prior parameter goes up then the optimum sample size goes down. We have obtained the optimum sample size for a Poisson distribution using the R program because of the complicated form of the posterior risk function. In this case if the prior parameter α goes up when β is fixed then optimum sample size goes down, on the other hand if α is fixed and β goes up then the optimum sample size also goes up. This is possibly because we have drawn the pre-posterior samples from a negative binomial distribution with the parameter α which is an integer and the parameter β is a function of n. We can easily obtain an optimum sample size for the scale parameter of Maxwell and Rayleigh distributions following the same approach as the one we discussed in this chapter.

Table 6.1: Optimum sample size to estimate a normal precision when mean is known for different values of the sampling cost per unit, c.

Optimum sample size	n^*		
c	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
0.0005	40	32	28
0.0010	26	22	18
0.0015	20	16	12
0.0020	17	13	9
0.0025	15	11	7
0.0030	13	9	5
0.0035	12	8	4
0.0040	11	7	3
0.0045	10	6	3
0.0050	9	5	1

Table 6.2: Optimum sample size to estimate a normal precision when mean is known for different values of the prior parameter, α.

Optimum sample size	n^*		
α	$c = 0.001$	$c = 0.0015$	$c = 0.002$
1	27	21	18
2	25	19	16
3	23	17	14
4	21	15	12
5	19	13	10
6	17	11	8
7	15	9	6
8	13	7	4
9	11	5	2
10	9	3	-

Table 6.3: Optimum sample size to estimate an exponential parameter for different values of the sampling cost per unit, c.

Optimum sample size	n^*		
c	$\alpha = 1$	$\alpha = 3$	$\alpha = 5$
0.0001	69	67	65
0.00015	56	54	52
0.0002	48	46	44
0.0025	43	41	39
0.0003	39	37	35
0.00035	36	34	32
0.0004	32	31	30
0.00045	31	30	28
0.0005	30	28	26
0.00055	28	26	24
0.0006	27	25	23
0.00065	24	26	22
0.0007	25	23	21

Table 6.4: Optimum sample size to estimate an exponential parameter for different values of the prior parameter, α.

Optimum sample size	n^*		
α	$c = 0.0001$	$c = 0.0002$	$c = 0.0005$
5	65	44	39
10	60	39	34
15	55	34	29
20	50	29	24
25	45	24	19
30	40	19	14
35	35	14	9
40	30	9	4
45	25	4	-
50	20	-	-

Table 6.5: Optimum sample size to estimate a Poisson parameter for different values of the prior parameters α and β when $c_0 = 1$ and $c = 0.001$.

α	β	n^*	β	α	n^*
1	1	10	1	1	10
	2	12		2	6
	3	14		3	4
	4	**15**		4	2
2	1	6	2	1	14
	2	8		2	8
	3	10		3	6
	4	12		4	4
3	1	4	3	1	16
	2	6		2	10
	3	8		3	8
	4	10		4	6

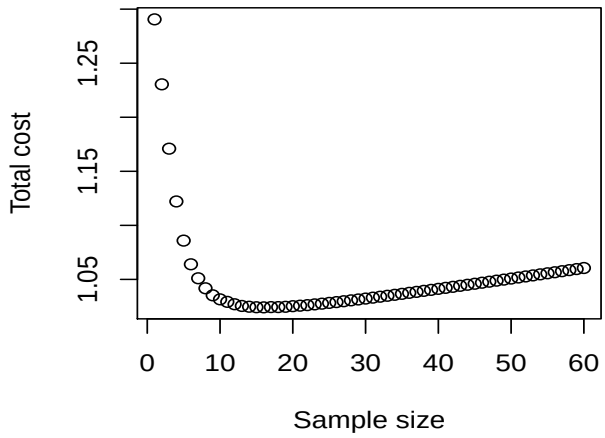

Figure 6.2: SSD to estimate a Poisson parameter, λ when prior parameters $\alpha = 1, \beta = 3$, $c_0 = 1$, $c = 0.001$.

Chapter 7

Optimum decisions under Lindley's conjugate utility function

7.1 Introduction

In this chapter we will obtain the optimum decisions under the conjugate utility function described by Lindley(1976). First he proposed a conjugate utility function for the one parameter exponential family. He solved the ratio of two integrals of the expected utility through approximations. Then he considered a normal and a binomial distribution and approximated a large sample with some conditions to solve the equations for the optimum decisions under the conjugate utility function. In this chapter first we will consider his conjugate utility function and some other distributions namely, exponential, Poisson and Maxwell distributions to obtain the optimum decisions us-

ing his approximation. In the first chapter, section 5.1 we reviewed the literature on Lindley's conjugate utility function. In the first half of this chapter we will discuss the maximization of the expected utility for the one parameter exponential family. We will then obtain the approximate optimum decision for the Poisson parameter, exponential parameter, Pareto parameter and the parameter of Maxwell distribution based on Lindley's conjugate utility for one parameter exponential family. We will also sketch the shape of the utility function for each distribution at the end of the chapter.

In many situations we have seen that the estimation of the parameters of the two parameter exponential family are difficult when both parameters are unknown. For such an estimation we need to define a conjugate utility for a two parameter exponential family. In Lindley's (1976) paper, he noted that it is possible to extend the results of the one parameter exponential family to the two parameter exponential family but gave no details. Note the technique of expanding the posterior distribution as a Taylor series and taking the terms up to the second order is essentially the Laplace method for integrals although Lindley (1976) does not mention this point. For details of the use of Laplace's method for approximate calculation of posterior moments and deriving marginal distributions see Tierney and Kadane (1986). They described approximations to the posterior means and variances of positive functions of a real and vector-valued parameter, and to the marginal posterior densities of an arbitrary parameter. In our case we will maximize the expected utility for a two parameter exponential family and to approximate posterior integrals we will use a Taylor expansion in all cases. So the second half of this chapter contains the whole procedure for maximizing the expected utility for the two parameter exponential family when

both parameters are unknown. First we will define the posterior distribution for the two parameter exponential family. Then we will propose a conjugate utility function for two parameters. Next we will lay out the maximization procedure of the expected utility with some approximations which will result in two theorems. Finally, we will estimate the parameters of the normal, trinomial, inverse Gaussian distribution using the results of the theorems when both parameters are unknown. We will also sketch the two parameter utility functions for these distributions at the end of the chapter.

In the following we will describe the maximization of expected utility for the one parameter exponential family given by Lindley.

7.2 Optimum decisions for one parameter exponential family

We have presented the posterior distribution of θ in the the equation (1.33) and the utility function in (1.34). Let $x_m = \sum_{i=0}^{n} x_i$ and $N = n + n_0$. The expected utility of d with respect to the posterior density (1.33) will be,

$$\begin{aligned} U(d) &= \int \exp[(x_m + x(d))\theta] G(\theta)^{N+n(d)} K(N, x_m) F(d) d\theta \\ &= \frac{K(N, x_m)}{K(N + n(d), x_m + x(d))} \end{aligned} \tag{7.1}$$

in terms of known functions. But it is difficult to obtain analytic results from (7.1). So Lindley developed an approximation for a larger N as he noted it will be useful when either the sample size is larger or the prior knowledge is substantial.

If $N \to \infty$, $x_m = \sum_{i=0}^{n} x_i$ will also increase and we have $x_m = N\overline{x}$, follows

$$\overline{x} = \frac{x_m}{N} = \frac{x_0 + \sum_{i=1}^{n} x_i}{n_0 + n} \tag{7.2}$$

It is to be noted that $\overline{x}$ is the only sample mean when $x_0 = n_0 = 0$, otherwise it is modified in a familiar way by the prior knowledge. Then he gave the following lemma.

Lemma 1 *For large N, $K[N + n(d), N\overline{x} + x(d)]^{-1}$ is asymptotically*

$$\left[\frac{-2\pi}{Ng''(\theta_0)}\right]^{\frac{1}{2}} \left[h(\theta_0) - \frac{h''(\theta_0)}{2Ng''(\theta_0)}\right] \exp\{N[\overline{x}\theta_0 + g(\theta_0)]\},$$

where θ_0 is the root of the equation $\overline{x} + g'(\theta_0) = 0$ and $h(\theta) = \exp[x(d)\theta + n(d)g(\theta)]$.

Proof

Following (1.31) let,

$$\begin{aligned} I &= [K(N + n(d), x_m + x(d)]^{-1} \\ &= \int \exp[N\overline{x} + x(d)\theta] G(\theta)^{N+n(d)} d\theta \\ &= \int \exp[N\overline{x} + x(d)\theta + (N + n(d))g(\theta)] d\theta \\ &= \int \exp[N(\overline{x} + g(\theta))] \exp[x(d)\theta + n(d)g(\theta)] d\theta \\ &= \int \exp[Nf(\theta)] h(\theta) d\theta, \end{aligned} \tag{7.3}$$

where $f(\theta) = \overline{x} + g(\theta)$ and $h(\theta) = \exp[x(d)\theta + n(d)g(\theta)]$. Now expanding both $f(\theta)$ and $h(\theta)$ of (7.3) in a Taylor series about θ_0 and give the root $f'(\theta_0) = 0$, implies

$g'(\theta_0) = 0$, so the integral I becomes,

$$\begin{aligned} I &= \int [\{h(\theta_0) + (\theta - \theta_0)h'(\theta_0) + \frac{1}{2}(\theta - \theta_0)^2 h''(\theta_0)\} \\ &\times \exp\{N[f(\theta_0) + \frac{1}{2}(\theta - \theta_0)^2 f''(\theta_0)]\}]d\theta \\ &= \exp[Nf(\theta_0)] \int \left\{ h(\theta_0) + \theta h'(\theta_0) - \theta_0 h'(\theta_0) + \frac{h''(\theta_0)}{2}(\theta - \theta_0)^2 \right\} \\ &\times \exp\left(\frac{N}{2}(\theta - \theta_0)^2 f''(\theta_0)\right) d\theta \end{aligned}$$

For large N if θ is distributed as normal with mean θ_0 and precision $Ng''(\theta_0)$ (where $f''(\theta_0) = g''(\theta_0) < 0$) we have, $\left[\frac{Ng''(\theta_0)}{-2\pi}\right]^{\frac{1}{2}} \int \exp\left(\frac{N}{2}(\theta - \theta_0)^2 g''(\theta_0)\right) d\theta = 1$. This implies the mean of θ, $E(\theta|x) = \theta_0$ and variance of θ, $var(\theta) = -\frac{1}{Ng''(\theta_0)}$. Using these results in the integral I we have,

$$\begin{aligned} I &= \left[\frac{-2\pi}{Ng''(\theta_0)}\right] \left[h(\theta_0) + \frac{h''(\theta_0)}{2} var(\theta|x)\right] \exp[Nf(\theta_0)] \\ &= \left[\frac{-2\pi}{Ng''(\theta_0)}\right] \left[h(\theta_0) - \frac{h''(\theta_0)}{2Ng''(\theta_0)}\right] \exp[N\{\overline{x} + g(\theta_0)\}] \end{aligned} \tag{7.4}$$

proved.

The approximate result for the denominator of (7.1) is given in (7.4). Since the numerator is a special case of the denominator with $n(d) = x(d) = 0$ which implies $h(\theta) = 1$, we have,

$$K(N, x_m)^{-1} = \left[\frac{-2\pi}{Ng''(\theta_0)}\right] \exp[N\{\overline{x} + g(\theta_0)\}]. \tag{7.5}$$

Now using the result of (7.4) and (7.5) in (7.1), $U(d)$ is asymptotically,

$$F(d) \left[h(\theta_0) - \frac{h''(\theta_0)}{2Ng''(\theta_0}\right]. \tag{7.6}$$

First consider if the term $O(N^{-1})$ is omitted, then the maximum expected utility will only be $F(d)h(\theta_0)$ and using conditions C_1 and C_2 it becomes,

$$\exp[-n(d)g'(d)\theta_0 + n(d)g(\theta_0) + n(d)\{g'(d)d - g(d)\}].$$

After few simplifications this expression will be the same as the utility function (1.39) which clearly has the maximum at $\hat{d} = \theta_0$. In the following theorem Lindley proves an approximate optimum decision for θ for this utility function.

Theorem 3 *Under conditions C_1 and C_2 the optimum decision for large N is given by the root, θ_0, of the equation $\overline{x} + g'(\theta) = 0$.*

Clearly the above theorem is for an optimum approximate decision for θ when $O(N^{-1})$ is omitted from the equation (7.6). Lindley then retained the term $O(N^{-1})$ as it is in the equation (7.6) and maximized it for an improved estimate of θ stating the following theorem.

Theorem 4 *Under conditions C_1 and C_2 the optimum decision, to $O(N^{-1})$ is*

$$\theta_0 - \frac{\frac{1}{2}n'(\theta_0)}{Nn(\theta_0)g''(\theta_0)}.$$

In equation (7.6) the term $n(d)$ is free to select but Lindley also gave an outline how we can choose it. For θ near d we may expand the expression in braces of (1.39) about d which is approximately,

$$\exp\left[\frac{1}{2}n(d)g''(d)(\theta - d)^2\right],$$

so that near the θ best for that d, the utility behaves like a normal density with spread, $[-n(d)g''(d)]^{-\frac{1}{2}}]$ where $g''(d) < 0$. So $n(d)$ measures how near θ has to be to d for decision to be good: large $n(d)$ says it has to be very near, small $n(d)$ means that it is not critical. A special case would be where the departure is the same for all

d. Finally Lindley gave $n(d)$ as

$$n(d)^{-1} = -\kappa g''(d)$$

for some constant κ and referred to this as condition C_3. Lindley (1961) considered this condition C_3 along with condition C_1 and C_2 to obtain an optimum decision for θ which is,

$$\theta_0 + \frac{\frac{1}{2}g^{(3)}(\theta_0)}{N\{g''(\theta_0)\}^2}. \tag{7.7}$$

For an optimum decision it gives similar results given in theorem 4. In the paper Lindley (1976) gave an optimum decision for the Normal mean with known variance (equal to one). So we have,

$$p(x|\theta) = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\theta)^2\right].$$

This is the form of (1.30) with $G(\theta) = \exp[-\frac{1}{2}\theta^2]$, so $g(\theta) = -\frac{1}{2}\theta^2$. Theorem 1 gives the optimum decision (under C_1 and C_2) to be $\overline{x}$. Then he considered the case for normal variance, known mean taken to be zero. So the probability density will be,

$$p(y|\phi) = (2\pi\phi)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}y^2\phi\right],$$

where ϕ is the variance. Now let, $x = -\frac{1}{2}y^2$ so that $x < 0$, and $\theta = \phi^{-1}$, the precision. Then $G(\theta) = \theta^{\frac{1}{2}}$ and $g(\theta) = \frac{1}{2}\log\theta$ for $\theta > 0$. Theorem1 gives the large sample decision as $\theta_0^{-1} = -2\overline{x}$. Turning back to the original data $\{y_i\}$ and the variance ϕ, the decision for ϕ is $\frac{\sum_{i=1}^{n} y_i^2 + y_0^2}{n+n_0}$. Then he considered the approximate optimal decision for the Bernoulli sample. In each situation he also discussed the corrected estimates of the parameters using theorem 4. But we haven't seen the shape of the conjugate utility function for the parameter of the distributions he studied. In the following section we will obtain the optimum decisions for an exponential, a Poisson, a Pareto

and a Maxwell distribution. We will use the conjugate utility structure (1.31) and for an optimum decision will follow theorem 3 and theorem 4. We will also sketch the shape of the conjugate utility functions and present them at the end of the chapter.

7.2.1 Optimum decision for a Poisson parameter

Let x follow a Poisson distribution with parameter θ. The likelihood of the sample is given in Chapter 2 equation (2.36). Let, $\psi = \log\theta$ and $e^{\psi} = \theta$. So the likelihood is now

$$p(\underline{x}|\psi, n) \propto \exp[-ne^{\psi}] \exp[t\psi] \tag{7.8}$$

Let us consider a natural conjugate prior for ψ as,

$$p(\psi|x_0, n_0) \propto \exp\left[-n_0 e^{\psi}\right] \exp\left[x_0\psi\right] \tag{7.9}$$

Combining (7.8) and (7.9) we have the posterior distribution as,

$$p(\psi|x_0, n_0) \propto \exp\left[-(n+n_0)e^{\psi}\right] \exp\left[(t+x_0)\psi\right] \tag{7.10}$$

Now from (7.2) we have,

$$\overline{x} = \frac{x_m}{N}, \tag{7.11}$$

where $x_m = t + x_0 = \sum_{i=0}^{n} x_i$ and $N = (n + n_0)$. From (7.2) we have, $G(\psi) = \exp[-e^{\psi}]$, $g(\psi) = -e^{\psi}$, $g(d) = -e^{d}$, $g'(d) = -e^{d}$, $g''(d) = -e^{d}$, $n(d)^{-1} = -kg''(d)$, so, $n(d) = \frac{1}{ke^{d}}$. Now by theorem 3, the optimum decision for ψ will be $\psi_0 = \log\overline{x}$, so that, $\theta_0 = \overline{x}$. By theorem 4, the decision for ψ is $\psi_0 = \log\overline{x} - \frac{1}{2N\overline{x}}$ gives the decision for $\theta = \overline{x}\exp[-\frac{1}{2N\overline{x}}]$.

Shape of the utility function

If d is an estimate of ψ then the conjugate utility for ψ under conditions C_1 and C_2 is,

$$u(d,\psi) = \exp\{n(d)[g(\psi) - g(d) - g'(d)(\psi - d)]\}$$

Now under condition C_3 the utility function will be,

$$u(d,\psi) = \exp\left\{\frac{1}{ke^d}[-e^{\psi} + e^d + e^d(\psi - d)]\right\}. \tag{7.12}$$

In figure 7.1 we have plotted $u(d)$ against d for various ϕ and observe that at the point $d = \phi$ the utility reach a maximum at 1. We also found that if the parameter value goes up then underestimation is more serious than overestimation for the given set of parameters.

7.2.2 Optimum decision for an exponential parameter

Suppose x_1, x_2,....,x_n is a random sample of size n taken from an exponential distribution with parameter λ. The likelihood of the sample is given in chapter 2 equation (2.31). Consider a conjugate prior for λ as,

$$p(\lambda) \propto \lambda^{n_0} \exp(-x_0\lambda) \tag{7.13}$$

So combining (2.31) and (7.13) the posterior density is,

$$p(\theta|x) \propto \lambda^{n_0+n} \exp\{-(x_0 + t)\lambda)\}. \tag{7.14}$$

Now comparing with (7.2) we have,

$$\overline{x} = \frac{x_m}{N} \tag{7.15}$$

Where, $x_m = -(t + x_0) = \sum_{i=0}^{n} x_i$ and $N = n + n_0$. We also have, $G(\lambda) = \lambda$, $g(\lambda) = \log \lambda$, $g'(\lambda) = \frac{1}{\lambda}$, so , $g(d) = \log d$, $g'(d) = \frac{1}{d}$,$g''(d) = -\frac{1}{d^2}$, $n(d)^{-1} = -kg''(d) = \frac{k}{d^2}$, so, $n(d) = \frac{d^2}{k}$. By theorem 3, for large N the decision for λ under the conjugate utility (7.16) will be, $\lambda_0 = \frac{N}{\sum_{i=0}^{n} x_i}$. By theorem 4, the decision for λ will be $\lambda_0 \left(1 + \frac{1}{N}\right)$.

Shape of the utility function

If d is an estimate of θ, under conditions C_1 and C_2 the conjugate utility for λ will be,

$$u(d, \lambda) = \exp\left[n(d)\left\{n \log\left(\frac{\lambda}{d}\right) - \frac{n}{d}(\lambda - d)\right\}\right]; \tag{7.16}$$

where $n(d)$ can be chosen independently. Using condition C_3 we have the utility function as,

$$u(\lambda, d) = \exp\left[\frac{d^2}{k}\left\{n \log\left(\frac{\lambda}{d}\right) - \frac{n}{d}(\lambda - d)\right\}\right]; \tag{7.17}$$

From figure 7.2 we observe that for $\lambda = d$, the maximum utility is 1. If the values of λ goes up then the shape of the utility shifts to the right for fixed range of d. We also observe from figure 7.3 that if the sample size goes up, the decision range goes down. The reverse is also true. From both shapes we can see that, after reaching the maximum at $\lambda = d$, the utility comes down to the right faster than to its left which implies that if over estimation is more serious than under estimation then the utility $u(d, \lambda)$ is more appropriate.

7.2.3 Optimum decision for a Pareto parameter

Suppose x_1, x_2,....,x_n is a random sample of size n taken from a Pareto distribution with parameter θ. The likelihood is given in Chapter 6 equation (6.15). Consider the conjugate prior of θ as,

$$p(\theta) \propto \theta^{n_0} \exp\left[-\theta \log\left(\frac{x_0}{\alpha_0}\right)\right]. \tag{7.18}$$

Now combining, (6.15) and (7.18) we have,

$$p(\theta|\underline{x}) \propto \theta^{n+n_0} \exp\left[-\theta \sum_{i=0}^{n} \log\left(\frac{x_i}{\alpha_0}\right)\right]. \tag{7.19}$$

Now comparing with (7.2) we have,

$$\overline{x} = \frac{x_m}{N} \tag{7.20}$$

Where, $x_m = -\sum_{i=0}^{n} \log\left(\frac{x_i}{\alpha_0}\right)$ and $N = n + n_0$, $G(\theta) = \theta$, $g(\theta) = \log\theta$, $g'(\theta) = \frac{1}{\theta}$, so , $g(d) = \log d$, $g'(d) = \frac{1}{d}$,$g''(d) = -\frac{1}{d^2}$, $n(d)^{-1} = -kg''(d) = \frac{k}{d^2}$, so, $n(d) = \frac{d^2}{k}$. By theorem 3, for large N the decision for θ under the conjugate utility (7.21) will be, $\theta_0 = \frac{n+n_0}{\exp\left[\sum_{i=0}^{n} \log\left(\frac{x_i}{\alpha_0}\right)\right]}$. By theorem 4, the decision for θ will be $\theta_0\left(1 + \frac{1}{N}\right)$.

Shape of the utility function

If d is an estimate of θ, under conditions C_1 and C_2 the conjugate utility for θ will be,

$$u(\theta, d) = \exp\left[n(d)\left\{\log\theta - \log d - \frac{1}{d}(\theta - d)\right\}\right]; \tag{7.21}$$

where $n(d)$ can be chosen independently. Using condition C_3 we have the utility function as,

$$u(\theta, d) = \exp\left[\frac{d^2}{k}\left\{\log\theta - \log d - \frac{1}{d}(\theta - d)\right\}\right]; \tag{7.22}$$

From figure 7.4 we observe that for $\theta = d$ the maximum utility is 1. The shape of the utility function increases up to its maximum then decreases. If the values of θ goes up then the shape of the utility shifts to the right for a fixed range of d. From both figures 7.4 and 7.5 we can see that, after reaching the maximum at $\theta = d$ the utility comes down to the right faster than to its left which implies that if over estimation is more serious than under estimation then this utility $u(d, \theta)$ seems to be more appropriate.

7.2.4 Optimum decision for a Maxwell distribution

The Maxwell distribution is used to model the distribution of speeds of molecules in thermal equilibrium as given by statistical mechanics. Defining $\sigma = \sqrt{\frac{KT}{m}}$, where K is the Stephan- Boltzmann constant, T is the temperature, m is the mass of molecule, and x denotes the speed of a molecule, the probability distribution over the range $x \in [0, \infty)$ is

$$p(x|\sigma) = \sqrt{\frac{2}{\pi}} \frac{x^2 \exp(-\frac{x^2}{2\sigma^2})}{\sigma^3}, \tag{7.23}$$

where, $\sigma > 0$ is the scale parameter. In the study of the velocity distributions of molecules in R^3 Maxwell assumed that in every cartesian coordinate system the three components of the velocity are mutually independent random variables with zero expectation. It can be shown that the three components are normally distributed with the same variance. Let us take $\theta = \frac{1}{\sigma^2}$. So the distribution (7.49) will be,

$$p(x|\theta) \propto \theta^{\frac{3}{2}} \exp\left(-\frac{\theta x^2}{2}\right) \tag{7.24}$$

Let x_1, x_2,..., x_n be a sample of size n taken from Maxwell distribution. So the likelihood of the sample is,

$$p(\underline{x}|\theta) \propto \theta^{\frac{3n}{2}} \exp\left(-\frac{\theta \sum x_i^2}{2}\right) \tag{7.25}$$

Let us take the conjugate family prior as,

$$p(\theta) \propto \exp\left(-\frac{1}{2}x_0^2\theta\right) \theta^{\frac{3n_0}{2}} \tag{7.26}$$

Now combining (7.50) and (7.26) the posterior distribution will be,

$$p(\theta|\underline{x}) \propto \exp\left[-\left(\frac{\sum x_i^2}{2} + \frac{x_0^2}{2}\right)\theta\right] \theta^{\frac{3}{2}(n+n_0)}$$

We can equivalently write,

$$p(\theta|\underline{x}) \propto G(\theta)^N \exp(x\theta)$$

where, $N = \frac{3}{2}(n+n_0)$, $x_m = -\frac{1}{2}(\sum_{i=0}^{n} x_i^2)$ and $G(\theta) = \theta$. Now comparing with (7.2) we have,

$$\overline{x} = \frac{x_m}{N} \tag{7.27}$$

If d is an estimate of θ, under conditions C_1 and C_2 the conjugate utility for θ will be,

$$u(d,\theta) = \exp\left[n(d)\left\{\log\theta - \log d - \frac{1}{d}(\theta - d)\right\}\right]; \tag{7.28}$$

where $n(d)$ can be chosen independently. Using condition C_3 we have the utility function as

$$u(d,\theta) = \exp\left[\frac{d^2}{k}\left\{\log\theta - \log d - \frac{1}{d}(\theta - d)\right\}\right]; \tag{7.29}$$

where, $G(\theta) = \theta$, $g(\theta) = \log\theta$, $g'(\theta) = \frac{1}{\theta}$, so , $g(d) = \log d$, $g'(d) = \frac{1}{d}$,$g''(d) = -\frac{1}{d^2}$, $n(d)^{-1} = -kg''(d) = \frac{k}{d^2}$, so, $n(d) = \frac{d^2}{k}$. By theorem 3, for large N the decision for

θ under the conjugate utility (7.28) will be, $\theta_0 = \frac{3(n+n_0)}{2\sum_{i=0}^{n} x_i^2}$. By theorem 4, retaining $O(N^{-1})$, the decision for θ will be $\theta_0 \left(1 + \frac{1}{N}\right)$, where, $\theta_0 = \frac{3(n+n_0)}{2\sum_{i=0}^{n} x_i^2}$.

We have discussed optimization procedure of the estimates for one parameter exponential family under Lindley's one parameter conjugate utility function. In the following we will first propose a conjugate utility function for the two parameter exponential family. Then we will discuss the procedure of maximization of this utility function for the two parameter exponential family. At the end, we will consider the estimation procedure of the parameters of the normal, trinomial and inverse Gaussian distributions when both parameters are unknown.

7.3 Optimum decisions for the two parameter exponential family

Lindley (1976) suggested that his results could be extended to multivariate situations but gave no details. Here we develop his method for the bivariate exponential family. If $p(x|\underline{\theta})$ depends on $\underline{\theta} = (\psi_1(\underline{\theta}), \psi_2(\underline{\theta}))$ and $\underline{t}(x) = (t_1(x), t_2(x))$ are jointly sufficient for $(\psi_1(\underline{\theta}), \psi_2(\underline{\theta}))$ and if we are able to write the density $p(x|\underline{\theta})$ in form

$$p(x|\underline{\theta}) = \exp\{t_1(x)\psi_1(\underline{\theta}) + t_2(x)\psi_2(\underline{\theta})\}G(\underline{\theta})H(x), \tag{7.30}$$

then the density belongs to two parameter exponential family. Where,

$$G(\underline{\theta})^{-1} = \int \exp\{t_1(x)\psi_1(\underline{\theta}) + t_2(x)\psi_2(\underline{\theta})\}H(x)dx.$$

If we draw a sample of size n from (7.30) then the likelihood of the sample will be

$$p(\underline{x}|\underline{\theta}) = \exp\left\{\sum_{i=1}^{n} t_1(x_i)\psi_1(\underline{\theta}) + \sum_{i=1}^{n} t_2(x_i)\psi_2(\underline{\theta})\right\} G(\underline{\theta})^n \prod_{i=1}^{n} H(x_i). \tag{7.31}$$

Now consider the natural conjugate prior family for $\underline{\theta}$ as,

$$p(\underline{\theta}|n_0, \underline{t}(x_0)) \propto \exp\left\{t_1(x_0)\psi_1(\underline{\theta}) + t_2(x_0)\psi_2(\underline{\theta})\right\} G(\underline{\theta})^{n_0} \tag{7.32}$$

Also define,

$$K(n_0, \underline{t}(x_0))^{-1} = \int \exp\left\{t_1(x_0)\psi_1(\underline{\theta}) + t_2(x_0)\psi_2(\underline{\theta})\right\} G(\underline{\theta})^{n_0} d\underline{\theta} \tag{7.33}$$

So the conjugate prior density of $\underline{\theta}$ given n_0 and x_0 is,

$$p(\underline{\theta}|n_0, \underline{t}(x_0)) = \exp\left\{t_1(x_0)\psi_1(\underline{\theta}) + t_2(x_0)\psi_2(\underline{\theta})\right\} G(\underline{\theta})^{n_0} K(n_0, \underline{t}(x_0)). \tag{7.34}$$

Combining (7.31) and (7.34) the posterior density will be

$$\begin{aligned} p(\underline{\theta}|n_0, \underline{t}(x)) \quad &\propto \quad \exp\left\{\sum_{i=0}^{n} t_1(x_i)\psi_1(\underline{\theta}) + \sum_{i=0}^{n} t_2(x_i)\psi_2(\underline{\theta})\right\} \\ &\quad \times G(\underline{\theta})^{n+n_0} \prod_{i=1}^{n} H(x_i) K(N, \sum \underline{t}(x_i)). \end{aligned} \tag{7.35}$$

Clearly,

$$K(N, \sum \underline{t}(x_i))^{-1} = \int \exp\left\{\sum_{i=0}^{n} t_1(x_i)\psi_1(\underline{\theta}) + \sum_{i=0}^{n} t_2(x_i)\psi_2(\underline{\theta})\right\} G(\underline{\theta})^N d\underline{\theta}, \tag{7.36}$$

where, $N = n + n_0$.

In any multiple decision problem we have to take decisions $\underline{d} = (d_1, d_2, \ldots, d_p)$ for the parameter values $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ respectively. Now we will define a conjugate utility function for such a problem. Let us denote, $u(\underline{d}, \underline{\theta})$ as the utility function which represents multiple decisions for multiple parameters and define as follows:

$$u(\underline{d}, \underline{\theta}) = \exp\{x_1(\underline{d})\psi_1(\underline{\theta}) + x_2(\underline{d})\psi_2(\underline{\theta}) G(\underline{\theta})^{n(\underline{d})} F(\underline{d}), \tag{7.37}$$

where, $x_1(\underline{d})$, $x_2(\underline{d})$ and $F(\underline{d})$ are some suitable functions of $\underline{d}$. It is not necessary to impose normalizing constraints on a utility function as the function $F(\underline{d})$ does not have to integrate to 1. The only restriction on $F(\underline{d})$ is that it must be positive. We want to maximize the expected utility (7.37) with respect to the posterior, (7.35) by taking expectation on $U(\underline{d}, \underline{\theta})$. So the expected utility of $\underline{d}$ will be,

$$\begin{aligned} u(\underline{d}) &= \int \exp\left\{ (\sum_{i=0}^{n} t_1(x_i) + x_1(\underline{d}))\psi_1(\underline{\theta}) + (\sum_{i=0}^{n} t_2(x_i) + x_2(\underline{d}))\psi_2(\underline{\theta}) \right\} \\ &\quad \times G(\underline{\theta})^{n+n_0+n(\underline{d})} K(N, \sum \underline{t}(x_i)) F(\underline{d}) d\underline{\theta} \\ &= \frac{K(N, \sum \underline{t}(x_i)) F(\underline{d})}{K(N + n(\underline{d}), \sum \underline{t}(x_i) + \underline{x}(\underline{d}))}. \end{aligned} \tag{7.38}$$

Clearly, $K(N, \sum \underline{t}(x))$ is defined in (7.36) and in a similar fashion the denominator will be,

$$\begin{aligned} K(N + n(\underline{d}), \sum \underline{t}(x) + \underline{x}(\underline{d}))^{-1} &= \int \exp\{ (\sum_{i=0}^{n} t_1(x_i) + x_1(\underline{d}))\psi_1(\underline{\theta}) \\ &\quad + (\sum_{i=0}^{n} t_2(x_i) + x_2(\underline{d}))\psi_2(\underline{\theta}) \} \\ &\quad \times G(\underline{\theta})^{N+n(\underline{d})} d\underline{\theta}. \end{aligned} \tag{7.39}$$

Now maximization over $\underline{d}$ will give us the optimum decisions for $\underline{\theta}$. We also see that the expected utility and the probability distribution belong to the same closed family. The maximization over $\underline{d}$ is the solution of the ratio of two K functions. In the following section we will discuss the choice of the function of $\underline{d}$ in (7.37). Then we will discuss the method of maximizing the ratio of two K functions given in (7.38).

7.3.1 Form of the two parameter conjugate utility function

For the two parameter case ($p = 2$), consider $\psi_1(\underline{\theta}) = \psi_1$ and $\psi_2(\underline{\theta}) = \psi_2$, and $g(\psi_1, \psi_2) = \log G(\psi_1, \psi_2)$. Obviously, ψ_1 and ψ_2 can be obtained after some suitable re-parametrization. So (7.37) becomes,

$$u(d_1, d_2, \psi_1, \psi_2) = \exp\{x_1(d_1, d_2)\psi_1 + x_2(d_1, d_2)\psi_2\}G(\psi_1, \psi_2)^{n(d_1,d_2)}F(d_1, d_2), \quad (7.40)$$

where, $x_1(d_1, d_2)$, $x_2(d_1, d_2)$ and $F(d_1, d_2)$ are some suitable functions of d_1 and d_2. Since we want to maximize d_1, d_2 of (7.40), so the logarithmic derivative of (7.40) vanishes for both parameters. So we can write,

$$\log U(d_1, d_2, \psi_1, \psi_2) = x_1(d_1, d_2)\psi_1 + x_2(d_1, d_2)\psi_2 + n(d_1, d_2)g(\psi_1, \psi_2) + \log F(d_1, d_2). \quad (7.41)$$

Now differentiate (7.41) w.r.t. ψ_1 and ψ_2 separately and setting each equation equal to zero we have,

$$x_1(d_1, d_2) = -n(d_1, d_2)g_1'(\psi_1, \psi_2) \quad (7.42)$$

and

$$x_2(d_1, d_2) = -n(d_1, d_2)g_2'(\psi_1, \psi_2), \quad (7.43)$$

where, $g_1'(\psi_1, \psi_2) = \frac{\partial g(\psi_1,\psi_2)}{\partial \psi_1}$ and $g_2'(\psi_1, \psi_2) = \frac{\partial g(\psi_1,\psi_2)}{\partial \psi_2}$. Using (7.42) and (7.43) in (7.40) the utility function will be,

$$u(d_1, d_2, \psi_1, \psi_2) = \exp\left[n(d_1, d_2)\{g(\psi_1, \psi_2) - \psi_1 g_1'(\psi_1, \psi_2) - \psi_2 g_2'(\psi_1, \psi_2)\}\right] F(d_1, d_2) \quad (7.44)$$

At the points $d_1 = \psi_1$ and $d_2 = \psi_2$, the utility function will be maximum. That is the best possible decision has been made. So

$$u(\underline{d}, \underline{d}) = \exp\left[n(d_1, d_2)\{g(d_1, d_2) - d_1 g_1'(d_1, d_2) - d_2 g_2'(d_1, d_2)\}\right] F(d_1, d_2). \quad (7.45)$$

Now we need to define $F(d_1, d_2)$. To make $u(\underline{d}, \underline{d})$ to a common value to 1. Let,

$$f(d_1, d_2) = \log F(d_1, d_2) = n(d_1, d_2)[d_1 g_1'(d_1, d_2) + d_2 g_2'(d_1, d_2) - g(d_1, d_2)]. \quad (7.46)$$

Now using (7.46) in (7.44) we have the utility function as,

$$\begin{aligned} u(d_1, d_2, \psi_1, \psi_2) &= \exp[n(d_1, d_2)\{g(\psi_1, \psi_2) - g_1'(\psi_1, \psi_2)(\psi_1 - d_1) \\ &\quad - g_2'(\psi_1, \psi_2)(\psi_2 - d_2) - g(d_1, d_2)\}] \end{aligned} \quad (7.47)$$

For $d_1 = \psi_1$ and $d_2 = \psi_2$ the maximum utility will be 1. The advantage of this kind of utility is it is bounded compared to unbounded squared error or linex loss function. Here $n(d_1, d_2)$ can be chosen independently as some suitable function of d_1 and d_2. The utility structure doesn't affect the value of $n(d_1, d_2)$, as it is simply the scale of the utility function. Now we will discuss how to chose $n(\underline{d})$.

7.3.2 Outline of choosing $n(\underline{d})$

If we expand $g(\psi_1, \psi_2)$ of (7.47) by Taylor series near d_1 , d_2 up to second order terms we have,

$$\begin{aligned} u(d_1, d_2, \psi_1, \psi_2) &= \exp[n(d_1, d_2)\{g(d_1, d_2) + \hat{g}_1(\psi_1 - d_1) + \hat{g}_2(\psi_2 - d_2) \\ &+ \frac{1}{2}\hat{g}_{11}(\psi_1 - d_1)^2 + \frac{1}{2}\hat{g}_{22}(\psi_2 - d_2)^2 + \hat{g}_{12}(\psi_1 - d_1)(\psi_2 - d_2) \\ &- (\hat{g}_1 + (\psi_1 - d_1)\hat{g}_{11})(\psi_1 - d_1) - (\hat{g}_2 + (\psi_2 - d_2)\hat{g}_{22})(\psi_2 - d_2) \\ &- g(d_1, d_2)\}], \end{aligned}$$

where, $\hat{g}_1 = \frac{\partial g(\psi_1,\psi_2)}{\partial \psi_1}|_{\psi_1=d_1,\psi_2=d_2}$, $\hat{g}_2 = \frac{\partial g(\psi_1,\psi_2)}{\partial \psi_2}|_{\psi_1=d_1,\psi_2=d_2}$, $\hat{g}_{11} = \frac{\partial^2 g(\psi_1,\psi_2)}{\partial^2 \psi_1}|_{\psi_1=d_1,\psi_2=d_2}$, $\hat{g}_{22} = \frac{\partial^2 g(\psi_1,\psi_2)}{\partial^2 \psi_2}|_{\psi_1=d_1,\psi_2=d_2}$ and $\hat{g}_{12} = \frac{\partial g(\psi_1,\psi_2)}{\partial \psi_1,\partial \psi_2}|_{\psi_1=d_1,\psi_2=d_2}$. After simplification we have,

$$\begin{aligned} u(d_1,d_2,\psi_1,\psi_2) &= \exp[-\frac{n(d_1,d_2)}{2}\{\hat{g}_{11}(\psi_1-d_1)^2+\hat{g}_{22}(\psi_2-d_2)^2 \\ &\quad -2\hat{g}_{12}(\psi_1-d_1)(\psi_2-d_2)\}] \\ &= K_1 \times \frac{n(d_1,d_2)\sqrt{\hat{g}_{11}\hat{g}_{22}}}{2\pi\sqrt{\hat{g}_{11}\hat{g}_{22}-g_{12}^2}}\exp[-\frac{n(d_1,d_2)}{2}\{\hat{g}_{11}(\psi_1-d_1)^2 \\ &\quad +\hat{g}_{22}(\psi_2-d_2)^2-2\hat{g}_{12}(\psi_1-d_1)(\psi_2-d_2)\}], \end{aligned}$$

where $K_1 = \frac{2\pi\sqrt{\hat{g}_{11}\hat{g}_{22}-\hat{g}_{12}^2}}{n(d_1,d_2)\sqrt{\hat{g}_{11}\hat{g}_{22}}}$. Clearly, (ψ_1,ψ_2) has an asymptotic bivariate normal distribution where the precision of ψ_1 is $n(d_1,d_2)\hat{g}_{11}$, the precision of ψ_2 is $n(d_1,d_2)\hat{g}_{22}$ and the correlation between ψ_1 and ψ_2 is $\rho = \frac{\hat{g}_{12}}{\sqrt{\hat{g}_{11}\hat{g}_{22}}}$. Now assume the parameters are independent so (ψ_1,ψ_2) asymptotically follows a bivariate normal distribution with precision of ψ_1 is $n(d_1,d_2)\hat{g}_{11}$ and the precision of ψ_2 is $n(d_1,d_2)\hat{g}_{22}$. So the precision of ψ_1 and ψ_2 depends on $n(d_1,d_2)$ and it measures the closeness between (d_1,ψ_1) and (d_2,ψ_2). One possibility could be, the departure between d_1 and ψ_1, d_2 and ψ_2 are expected to be the same for all d_1 and d_2; so we define

$$n(d_1,d_2)^{-1} = -\kappa\hat{g}_{12}, \tag{7.48}$$

for some constant κ can be chosen independently as the utility is not affected by a scale change.

7.3.3 Maximization of the expected utility

At this stage we need to maximize $u(\underline{d})$ which is the ratio of two K functions given in (7.38). We can obtain this ratio numerically if $K(N+n(\underline{d}),\sum \underline{t}(x)+\underline{x}(\underline{d}))$ is available

but it is difficult to obtain analytic results. Now we will do some approximations which will help us to make optimum decision at least about the parameters of interest. Let, $N = n_0 + n$. If $N \to \infty$ then $\sum_{i=0}^{n} t_1(x_i)$ and $\sum_{i=0}^{n} t_2(x_i)$ will also increase. So define,

$$\overline{x_1} = \frac{\sum_{i=0}^{n} t_1(x_i)}{N} \tag{7.49}$$

and

$$\overline{x_2} = \frac{\sum_{i=0}^{n} t_2(x_i)}{N} \tag{7.50}$$

Now we need to re-write (7.39) for the two parameter case, considering $\psi_1(\underline{\theta}) = \psi_1$, $\psi_2(\underline{\theta}) = \psi_2$, $g(\psi_1, \psi_2) = \log G(\psi_1, \psi_2)$. So we have an integral, say I as,

$$\begin{aligned}
I &= K[N + n(\underline{d}), \sum \underline{t}(x) + \underline{x}(\underline{d})]^{-1} \\
&= \int\int \exp\{[\sum_{i=0}^{n} t_1(x_i) + x_1(d_1, d_2)]\psi_1 \\
&\quad +[\sum_{i=0}^{n} t_2(x_i) + x_2(d_1, d_2)]\psi_2\} G(\psi_1, \psi_2)^{N+n(d_1,d_2)} d\psi_1 d\psi_2 \\
&= \int \exp\{(N\overline{x_1} + x_1(d_1, d_2))\psi_1 + (N\overline{x_2} + x_2(d_2, d_2))\psi_2 \\
&\quad +(N + n(d_1, d_2))g(\psi_1, \psi_2)\} d\psi_1 d\psi_2 \\
&= \int\int \exp[N\{\overline{x_1}\psi_1 + \overline{x_2}\psi_2 + g(\psi_1, \psi_2)\}] \\
&\quad \times \exp[\psi_1 x_1(d_1, d_2) + \psi_2 x_2(d_1, d_2) + n(d_1, d_2)g(\psi_1, \psi_2)] d\psi_1 d\psi_2 \\
&= \int\int \exp[Nf(\psi_1, \psi_2)] \times h(\psi_1, \psi_2) d\psi_1 d\psi_2. \qquad (7.51)
\end{aligned}$$

Where, $f(\psi_1, \psi_2) = \overline{x_1}\psi_1 + \overline{x_2}\psi_2 + g(\psi_1, \psi_2)$, $h(\psi_1, \psi_2) = \exp[\psi_1 x_1(d_1, d_2) + \psi_2 x_2(d_1, d_2) + n(d_1, d_2)g(\psi_1, \psi_2)]$, $N = n + n_0$. Now expanding $f(\psi_1, \psi_2)$ and $h(\psi_1, \psi_2)$ of (7.51) by

Taylor series for two variables ψ_1 and ψ_2 up to second order about (ψ_{01}, ψ_{02}) we have,

$$\begin{aligned}
\exp[Nf(\psi_1,\psi_2)] &= \exp[N\{f(\psi_{01},\psi_{02}) + \frac{\partial f(\psi_1,\psi_2)}{\partial\psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01}) \\
&\quad + \frac{\partial f(\psi_1,\psi_2)}{\partial\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_2-\psi_{02}) \\
&\quad + \frac{1}{2!}\{\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})^2 \\
&\quad + 2\frac{\partial^2 f(\psi_1,\psi_2)}{\partial\psi_1\partial\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) \\
&\quad + \frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_2-\psi_{02})^2\}\}].
\end{aligned}$$

In the expansion of $\exp[Nf(\psi_1,\psi_2)]$ consider the root,

$$f_1'(\psi_1,\psi_2) = \frac{\partial f(\psi_1,\psi_2)}{\partial\psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}} = 0 \tag{7.52}$$

and

$$f_2'(\psi_1,\psi_2) = \frac{\partial f(\psi_1,\psi_2)}{\partial\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}} = 0 \tag{7.53}$$

Now $\exp[Nf(\psi_1,\psi_2)]$ becomes,

$$\begin{aligned}
\exp[Nf(\psi_1,\psi_2)] &= \exp[N\{f(\psi_{01},\psi_{02}) + \frac{1}{2}\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})^2 \\
&\quad + \frac{\partial^2 f(\psi_1,\psi_2)}{\partial\psi_1\partial\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) \\
&\quad + \frac{1}{2}\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_2-\psi_{02})^2\}] \\
&= \exp[Nf(\psi_{01},\psi_{02})]\exp[-\frac{1}{2}\{-N\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})^2 \\
&\quad - 2N\frac{\partial^2 f(\psi_1,\psi_2)}{\partial\psi_1\partial\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) \\
&\quad - N\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_2-\psi_{02})^2\}]
\end{aligned}$$

Now by definition of $f(\psi_1,\psi_2)$,

$$\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}} = \hat{g}_{11},$$

$$\frac{\partial^2 f(\psi_1,\psi_2)}{\partial\psi_1\partial\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}} = \hat{g}_{12}$$

and

$$\frac{\partial^2 f(\psi_1,\psi_2)}{\partial^2\psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}} = \hat{g}_{22}.$$

So,

$$\begin{aligned}
\exp[Nf(\psi_1,\psi_2)] &= \exp[Nf(\psi_{01},\psi_{02})]\exp[-\frac{1}{2}\{-N\hat{g}_{11}(\psi_1-\psi_{01})^2 \\
&\quad -2N\hat{g}_{12}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) - N\hat{g}_{22}(\psi_2-\psi_{02})^2\}] \\
&= \exp[Nf(\psi_{01},\psi_{02})]\times C_1 \times \frac{N\sqrt{\hat{g}_{11}-\frac{\hat{g}_{12}^2}{\hat{g}_{22}}}\sqrt{\hat{g}_{22}-\frac{\hat{g}_{12}^2}{\hat{g}_{11}}}}{2\pi[1-\frac{\hat{g}_{12}^2}{\hat{g}_{11}\hat{g}_{22}}]} \\
&\quad \times\exp[-\frac{1}{2}\{-N\hat{g}_{11}(\psi_1-\psi_{01})^2 \\
&\quad -2N\hat{g}_{12}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) - N\hat{g}_{22}(\psi_2-\psi_{02})^2\}].
\end{aligned}$$

Where, $C_1 = \frac{2\pi[1-\frac{\hat{g}_{12}^2}{\hat{g}_{11}\hat{g}_{22}}]}{N\sqrt{\hat{g}_{11}-\frac{\hat{g}_{12}^2}{\hat{g}_{22}}}\sqrt{\hat{g}_{22}-\frac{\hat{g}_{12}^2}{\hat{g}_{11}}}}$. In the above expansion the last exponent approximately follows bivariate normal distribution with precision of ψ_1 is $-N\left(\hat{g}_{11}-\frac{\hat{g}_{12}^2}{\hat{g}_{22}}\right)$, precision of ψ_2 is $-N\left(\hat{g}_{22}-\frac{\hat{g}_{12}^2}{\hat{g}_{11}}\right)$ and the correlation between ψ_1 and ψ_2 is $\rho = \frac{\hat{g}_{12}}{\sqrt{\hat{g}_{11}\hat{g}_{22}}}$. Clearly, both $\hat{g}_{11}$ and $\hat{g}_{22}$ are always negative. If the variables are independent then, $\rho = 0$ implies $\hat{g}_{12} = 0$, so $\exp[Nf(\psi_1,\psi_2)]$ becomes

$$\begin{aligned}
\exp[Nf(\psi_1,\psi_2)] &= \exp[Nf(\psi_{01},\psi_{02})]\times C_2 \times \frac{N\sqrt{\hat{g}_{11}\hat{g}_{22}}}{2\pi} \\
&\quad \times\exp[-\frac{N}{2}\{\hat{g}_{11}(\psi_1-\psi_{01})^2 + \hat{g}_{22}(\psi_2-\psi_{02})^2\}],
\end{aligned}$$

where, $C_2 = \frac{2\pi}{N\sqrt{\hat{g}_{11}\hat{g}_{22}}}$. Now expanding $h(\psi_1, \psi_2)$ by the Taylor expansion for two variables at ψ_{01}, ψ_{02} we have,

$$\begin{aligned} h(\psi_1,\psi_2) &= h(\psi_{01},\psi_{02}) + \frac{\partial h(\psi_1,\psi_2)}{\partial \psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01}) \\ &+\frac{\partial h(\psi_1,\psi_2)}{\partial \psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_2-\psi_{02}) \\ &+\frac{1}{2!}\{\frac{\partial^2 h(\psi_1,\psi_2)}{\partial^2 \psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})^2 \\ &+2\frac{\partial^2 h(\psi_1,\psi_2)}{\partial \psi_1 \partial \psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) \\ &+\frac{\partial^2 h(\psi_1,\psi_2)}{\partial^2 \psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}(\psi_2-\psi_{02})^2\} \end{aligned}$$

For simplicity denote ,

$$\begin{aligned} \hat{h}_1 &= \frac{\partial h(\psi_1,\psi_2)}{\partial \psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}, \\ \hat{h}_{11} &= \frac{\partial^2 h(\psi_1,\psi_2)}{\partial^2 \psi_1}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}, \\ \hat{h}_2 &= \frac{\partial h(\psi_1,\psi_2)}{\partial \psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}, \\ \hat{h}_{22} &= \frac{\partial^2 h(\psi_1,\psi_2)}{\partial^2 \psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}, \end{aligned}$$

and

$$\hat{h}_{12} = \frac{\partial^2 h(\psi_1,\psi_2)}{\partial \psi_1 \partial \psi_2}|_{\psi_1=\psi_{01},\psi_2=\psi_{02}}.$$

So the above expansion becomes

$$\begin{aligned} h(\psi_1,\psi_2) &= h(\psi_{01},\psi_{02}) + \hat{h}_1(\psi_1-\psi_{01}) + \hat{h}_2(\psi_2-\psi_{02}) \\ &+\frac{1}{2}\hat{h}_{11}(\psi_1-\psi_{01})^2 + \hat{h}_{12}(\psi_1-\psi_{01})(\psi_2-\psi_{02}) + \frac{1}{2}\hat{h}_{22}(\psi_2-\psi_{02})^2. \end{aligned}$$

Now consider the integral (7.51),

$$\begin{aligned}\int\int \exp[Nf(\psi_1,\psi_2)] \times h(\psi_1,\psi_2)d\psi_1 d\psi_2 &= \exp[Nf(\psi_{01},\psi_{02})] \\ &\times C_2\{h(\psi_{01},\psi_{02}) + \hat{h}_1 E(\psi_1-\psi_{01}) \\ &+ \hat{h}_2 E(\psi_2-\psi_{02}) + \frac{1}{2}\hat{h}_{11}E(\psi_1-\psi_{01})^2 \\ &+ \hat{h}_{12}E(\psi_1-\psi_{01})(\psi_2-\psi_{02}) \\ &+ \frac{1}{2}\hat{h}_{22}E(\psi_2-\psi_{02})^2\}.\end{aligned}$$

After suitable re-parametrization the posterior density of ψ_1 and ψ_2 will be independent so we get,

$$\begin{aligned}\int\int \exp[Nf(\psi_1,\psi_2)]h(\psi_1,\psi_2)d\psi_1 d\psi_2 &= \exp[Nf(\psi_{01},\psi_{02})]C_2\{h(\psi_{01},\psi_{02}) \\ &\quad +\frac{1}{2}Var(\psi_1)+\frac{1}{2}Var(\psi_2)\} \\ &= \exp[Nf(\psi_{01},\psi_{02})]C_2\{h(\psi_{01},\psi_{02}) \\ &\quad -\frac{1}{2}\frac{\hat{h}_{11}}{N\hat{g}_{11}}-\frac{1}{2}\frac{\hat{h}_{22}}{N\hat{g_{22}}}\}.\end{aligned}$$

That is $K(N+n(\underline{d}),\sum \underline{t}(x)+\underline{x}(\underline{d}))^{-1}$ is asymptotically,

$$\exp[Nf(\psi_{01},\psi_{02})] \times \frac{2\pi}{N\sqrt{\hat{g}_{11}\hat{g}_{22}}} \times \{h(\psi_{01},\psi_{02}) - \frac{1}{2}\frac{\hat{h}_{11}}{N\hat{g}_{11}} - \frac{1}{2}\frac{\hat{h}_{22}}{N\hat{g_{22}}}\}. \tag{7.54}$$

Since the numerator is a special case of the denominator in (7.38) with $n(d_1,d_2) = x_1(d_1,d_2) = x_1(d_1,d_2) = 0$ gives, $h(\psi_1,\psi_2) = 1$. So, $K(N,\sum t(\underline{x}))^{-1}$ is asymptotically,

$$C_2 \exp[Nf(\psi_{01},\psi_{02})]. \tag{7.55}$$

For large N, applying the result of (7.54) and (7.55) to both numerator and denominator of (7.38) gives $u(d_1,d_2,\psi_1,\psi_2)$ has an approximate value,

$$\left[h(\psi_{01},\psi_{02}) - \frac{1}{2}\frac{\hat{h}_{11}}{N\hat{g}_{11}} - \frac{1}{2}\frac{\hat{h}_{22}}{N\hat{g}_{22}}\right] F(d_1,d_2) \tag{7.56}$$

recognizing $f''(\psi_1, \psi_2) = g''(\psi_1, \psi_2)$. If we ignore the term $O(N^{-1})$ then $u(d_1, d_2)$ is simply

$$h(\psi_{01}, \psi_{02})F(d_1, d_2), \tag{7.57}$$

where,

$$\begin{aligned} h(\psi_{01}, \psi_{02}) &= \exp[\psi_{01}x_1(d_1, d_2) + \psi_{02}x_2(d_1, d_2) + n(d_1, d_2)g(\psi_{01}, \psi_{02})] \\ &= \exp[n(d_1, d_2)\{g(\psi_{01}, \psi_{02}) - \psi_{01}g_1'(\psi_{01}, \psi_{02}) - \psi_{02}g_2'(\psi_{01}, \psi_{02})\}] \end{aligned}$$

and

$$F(d_1, d_2) = \exp[n(d_1, d_2)\{d_1 g_1'(d_1, d_2) + d_2 g_2'(d_1, d_2) - g(d_1, d_2)\}].$$

So the equation (7.57) will be maximum at $\hat{d}_1 = \psi_{01}$ and $\hat{d}_2 = \psi_{02}$. Thus we have proved the following theorem.

Theorem 5 *For large N, under the conditions* (7.42), (7.43) *and* (7.46) *the optimum decision for the two parameter exponential family will be,* ψ_{01} *and* ψ_{02} *where,* ψ_{01} *and* ψ_{02} *are the roots of two simultaneous equations:*

$$\overline{x_1} + g_1'(\psi_1, \psi_2) = 0$$

and

$$\overline{x_2} + g_2'(\psi_1, \psi_2) = 0.$$

Better results can be obtained if we keep $O(N^{-1})$ which we will present in the following theorem.

Theorem 6 *The optimum decisions of* ψ_{01} *and* ψ_{02} *to* $O(N^{-1})$ *will be,* $\psi_{01} - \frac{n'(\psi_{01})}{N[n''(\psi_{01}) + n(\psi_{01})\hat{g}_{11}]}$ *and* $\psi_{02} - \frac{n'(\psi_{02})}{N[n''(\psi_{02}) + n(\psi_{02})\hat{g}_{22})]}$ *respectively.*

Proof

Now expanding d_1 and d_2 of (7.56) near ψ_{01}, ψ_{01} respectively by Taylor series up-to second order term and setting $n(d_1)^{-1} = -k_1\hat{g}_{11}$, $n(d_2)^{-1} = -k_2\hat{g}_{22}$ and $n(d_1, d_2)^{-1} = k_3\hat{g}_{12} = 0$ (as ψ_1 and ψ_2 are assumed independent). We have,

$\exp[-\frac{1}{2}n(\psi_{01}, \psi_{02})\{(d_1 - \psi_{01})^2\hat{g}_{11} + (d_2 - \psi_{02})^2\hat{g}_{22}\}]$

$\times [1 - \frac{1}{N}\{n(\psi_{01}, \psi_{02}) + (d_1 - \psi_{01})n'(\psi_{01}) + (d_2 - \psi_{02})^2 n'(\psi_{02})$

$+ \frac{1}{2}(d_1 - \psi_{01})^2 n''(\psi_{01}) + \frac{1}{2}(d_2 - \psi_{02})^2 n''(\psi_{02})$

$+ \frac{1}{2}n(\psi_{01}, \psi_{02})^2\{(d_1 - \psi_{01})^2\hat{g}_{11} + (d_2 - \psi_{02})^2\hat{g}_{22}\}\}]$.

Now let, $d_1 - \psi_{01} = z_1$ and $d_2 - \psi_{02} = z_2$. So the above expression becomes,

$$u(z) = \exp(A_1 z_1^2 + A_2 z_2^2)(1 + a + b_1 z_1 + b_2 z_2 + c_1 z_1^2 + c_2 z_2^2). \tag{7.58}$$

Where, A_1 , A_2 are O(1) and a , b_1, b_2,c_1, c_2 are $O(N^{-1})$ defined as follows. $a = -n(\psi_{01}, \psi_{02})$, $A_1 = -\frac{1}{2}n(\psi_{01}, \psi_{02})\hat{g}_{11}$, $A_2 = -\frac{1}{2}n(\psi_{01}, \psi_{02})\hat{g}_{22}$, $b_1 = -n'(\psi_{01})$, $b_2 = -n'(\psi_{02})$, $c_1 = -\frac{1}{2}[n''(\psi_{01}) + n^2(\psi_{01}, \psi_{02})\hat{g}_{11}]$, $c_2 = -\frac{1}{2}[n''(\psi_{02}) + n^2(\psi_{01}, \psi_{02})\hat{g}_{22}]$. Now differentiate (7.58) w.r.t. z_1 and z_2 separately and neglecting the powers of z_1 and z_2 with 2 and more then finally setting each equal to zero we have,

$$z_1 = \frac{-b_1}{2(A_1 + c_1 + aA_1)} = -\frac{n'(\psi_{01})}{n''(\psi_{01}) + n(\psi_{01}, \psi_{02})\hat{g}_{11}}$$

and

$$z_2 = \frac{-b_2}{2(A_2 + c_2 + aA_2)} = -\frac{n'(\psi_{02})}{n''(\psi_{02}) + n(\psi_{01}, \psi_{02})\hat{g}_{22}}.$$

So the corrected estimate of ψ_{01} and ψ_{02} will be $\hat{d}_1 = \psi_{01} + z_1$ and $\hat{d}_2 = \psi_{02} + z_2$ respectively.

In the following we will obtain the approximate optimum decisions for the two parameter exponential family when both are unknown using the results of the theorem.

7.3.4 Optimum decisions for a normal mean μ and precision λ when both are unknown

Consider a normal distribution with mean, μ, precision, λ (both parameter unknown) with density function,

$$\begin{aligned} p(x|\mu,\lambda) &= \left[\frac{\lambda}{2\pi}\right]^{\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(x-\mu)^2\right\} \\ &= (2\pi)^{-\frac{1}{2}}\left[\lambda^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\lambda\mu^2\right\}\right]\exp\left(x\mu\lambda-\frac{1}{2}x^2\lambda\right). \end{aligned}$$

Let x_1,x_2,...,x_n be a random sample of size n from the above density, then the likelihood of the sample will be,

$$p(\underline{x}|\mu,\lambda) = (2\pi)^{-\frac{n}{2}}\left[\lambda^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\lambda\mu^2\right\}\right]^n \exp\left(\mu\lambda\sum_{i=1}^n x_i - \frac{1}{2}\lambda\sum_{i=1}^n x_i^2\right) \tag{7.59}$$

Consider a conjugate prior family,

$$p(\mu,\lambda|\tau_0,\tau_1,\tau_2) \propto \left[\lambda^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\lambda\mu^2\right\}\right]^{\tau_0}\exp\left(\tau_1\mu\lambda-\frac{1}{2}\tau_2\lambda\right) \tag{7.60}$$

Now combining (7.59) and (7.60) the posterior density will be,

$$p(\mu,\lambda|\underline{x}) \propto \left[\lambda^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\lambda\mu^2\right\}\right]^{n+\tau_0}\exp\left\{\mu\lambda\left(\sum_{i=1}^n x_i+\tau_1\right)-\frac{1}{2}\left(\sum_{i=1}^n x_i^2+\tau_2\right)\lambda\right\}. \tag{7.61}$$

The canonical form of (7.61) are obtained by setting,

$(\psi_1,\psi_2) = \left(\mu\lambda, -\frac{1}{2}\lambda\right)$, $(x_a,x_b) = (\sum_{i=1}^n x_i+\tau_1, \sum_{i=1}^n x_i^2+\tau_2)$. So, (7.61) becomes,

$$p(\psi_1,\psi_2|\underline{x}) \propto \left[(-2\psi_2)^{\frac{1}{2}}\exp\left\{\frac{\psi_1^2}{4\psi_2}\right\}\right]^{n+\tau_0}\exp[\psi_1 x_a+\psi_2 x_b] \tag{7.62}$$

We can re-write (7.62) as,

$$p(\psi_1,\psi_2|\underline{x}) \propto G(\psi_1,\psi_2)^N \exp[\psi_1 x_a+\psi_2 x_b], \tag{7.63}$$

where, $G(\psi_1, \psi_2) = (-2\psi_2)^{\frac{1}{2}} \exp\left[\frac{\psi_1^2}{4\psi_2}\right]$ and $N = n + \tau_0$. Now find, $g(\psi_1, \psi_2) = \frac{1}{2}\log(-2\psi_2) + \frac{\psi_1^2}{4\psi_2}$. So, $g_1'(\psi_1, \psi_2) = \frac{\psi_1}{2\psi_2}$, $g_2'(\psi_2, \psi_2) = \frac{1}{2\psi_2} - \frac{\psi_1^2}{4\psi_2^2}$. Using (7.48) we have $n(d_1, d_2) = \frac{d_2^2}{d_1}$ assuming $\kappa = \frac{1}{2}$. So from (7.47) we have the conjugate utility function will be,

$$\begin{aligned} u(\psi_1, \psi_2, d_1, d_2) &= \exp[\frac{d_2^2}{d_1}\{\frac{1}{2}\log(-2\psi_2) + \frac{\psi_1^2}{4\psi_2} - \frac{\psi_1}{2\psi_2}(\psi_1 - d_1) \\ &- (\frac{1}{2\psi_2} - \frac{\psi_1^2}{4\psi_2})(\psi_2 - d_2) - \frac{1}{2}\log(-2d_2) - \frac{d_1^2}{4d_2}\}]. \end{aligned} \quad (7.64)$$

Since ψ_1 and ψ_2 are independent, so we will keep $\psi_2 = d_2 = 1$ fixed and in the figure 7.6 we sketched three dimensional shape of the conjugate utility function of (7.64) for different values of decisions d_1 and d_2 when $\psi_1 = 1.5$ and $\psi_2 = -5$ are kept fixed. The reason to consider a large negative value of ψ_2 is because we need to get the positive value of λ with small variance. We observe from the graph that at the point $d_1 = \psi_1 = 1.5$ and $d_2 = \psi_2 = -5$, the shape of the utility function attains its maximum. For large N, $\overline{x_1} = \frac{x_a}{N}$, $\overline{x_2} = \frac{x_b}{N}$. Now by theorem 5, solving equations $\overline{x_1} + g_1'(\psi_1, \psi_2) = 0$ and $\overline{x_2} + g_2'(\psi_1, \psi_2) = 0$ simultaneously, we have the solution for ψ_1 is $\psi_{01} = \frac{\overline{x_1}}{\overline{x_2} - \overline{x_1}^2}$ and the solution for ψ_2 is $\psi_{02} = \frac{1}{2(\overline{x_1}^2 - \overline{x_2})}$. Now going backwards to the data x we have,

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i + \tau_1}{N}$$

and

$$\hat{\lambda} = \frac{N^2}{N(\sum_{i=1}^n x_i^2 + \tau_2) - (\sum_{i=1}^n x_i + \tau_1)^2}.$$

So, $\frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^n x_i^2 + \tau_2}{N} - \hat{\mu}^2$, which is close to sample variance if $\tau_1 = \tau_2 = \tau_0 = 0$. So the approximate value of ψ_1 and ψ_2 will maximize the expected utility defined in (7.64).

7.3.5 Optimum decisions for a trinomial distribution when both parameters are unknown

The probability distribution of the trinomial distribution with parameter θ_1 and θ_2 is,

$$p(r_1, r_2|\theta_1, \theta_2) \propto \theta_1^{r_1}\theta_2^{r_2}[1-(\theta_1+\theta_2)]^{n-(r_1+r_2)}, \tag{7.65}$$

where, $0 < \theta_1 < 1$, $0 < \theta_2 < 1$, $0 < r_1, r_2 < n$, $r_1 + r_2 \leq n$.

The conjugate prior family will be,

$$p(\theta_1, \theta_2) \propto \theta_1^{\alpha_1}\theta_2^{\alpha_2}[1-(\theta_1+\theta_2)]^{\alpha_3}. \tag{7.66}$$

Now combining (7.65) and (7.66) we have the posterior distribution as,

$$\begin{aligned} p(\theta_1, \theta_2|r_1, r_2) &\propto \theta_1^{\alpha_1+r_1}\theta_2^{\alpha_2+r_2}[1-(\theta_1+\theta_2)]^{n+\alpha_3-(r_1+r_2)} \\ &\propto \left[\frac{\theta_1}{1-(\theta_1+\theta_2)}\right]^{\alpha_1+r_1}\left[\frac{\theta_2}{1-(\theta_1+\theta_2)}\right]^{\alpha_2+r_2}[1-(\theta_1+\theta_2)]^{n+\alpha_1+\alpha_2+\alpha_3} \\ &\propto \exp\left[(\alpha_1+r_1)\log\left(\frac{\theta_1}{1-(\theta_1+\theta_2)}\right)+(\alpha_2+r_2)\log\left(\frac{\theta_2}{1-(\theta_1+\theta_2)}\right)\right] \\ &\quad [1-(\theta_1+\theta_2)]^{n+\alpha_1+\alpha_2+\alpha_3} \end{aligned}$$

The canonical form of the above distribution obtained by setting,

$$(\psi_1, \psi_2) = \left(\log\left(\frac{\theta_1}{1-(\theta_1+\theta_2)}\right), \log\left(\frac{\theta_2}{1-(\theta_1+\theta_2)}\right)\right).$$

Now we have,

$$\begin{aligned} p(\theta_1, \theta_2|r_1, r_2) &\propto \exp[(\alpha_1+r_1)\psi_1+(\alpha_2+r_2)\psi_2]\left[\frac{1}{e^{\psi_1}+e^{\psi_2}+1}\right]^{n+\alpha_1+\alpha_2+\alpha_3} \\ &\propto \exp[r_1'\psi_1+r_2'\psi_2]G(\psi_1, \psi_2)^{\alpha_3'}. \end{aligned}$$

Where, $G(\psi_1, \psi_2) = \frac{1}{e^{\psi_1}+e^{\psi_2}+1}$, $r_1' = \alpha_1 + r_1$, $r_2' = \alpha_2 + r_2$ and $N = n + \alpha_1 + \alpha_2 + \alpha_3$.

So, $g(\psi_1, \psi_2) = -\log(e^{\psi_1} + e^{\psi_2} + 1)$, $g_1'(\psi_1, \psi_2) = \frac{\partial g}{\partial \psi_1} = -\frac{e^{\psi_1}}{(e^{\psi_1}+e^{\psi_2}+1)}$, $g_2'(\psi_1, \psi_2) =$

$\frac{\partial g}{\partial \psi_2} = -\frac{e^{\psi_2}}{(e^{\psi_1}+e^{\psi_2}+1)}$. Now using (7.48) we have

$$n(d_1, d_2) = -\frac{(\exp(d_1) + \exp(d_2) + 1)^2}{\exp(d_1 + d_2)}$$

assuming $\kappa = 1$. So from (7.47) we have the conjugate utility function will be,

$$\begin{aligned} u(\psi_1, \psi_2, d_1, d_2) &= \exp[-\frac{(e^{d_1} + e^{d_2} + 1)^2}{e^{d_1+d_2}}\{-\log(e^{\psi_1} + e^{\psi_2} + 1) + \frac{e^{\psi_1}}{e^{\psi_1} + e^{\psi_2} + 1}(\psi_1 - d_1) \\ &+ \frac{e^{\psi_2}}{e^{\psi_1} + e^{\psi_2} + 1}(\psi_2 - d_2) + \log(e^{d_1} + e^{d_2} + 1)\}]. \end{aligned} \tag{7.67}$$

The utility function (7.67) will be maximum at $\psi_1 = d_1$ and $\psi_2 = d_2$. In figure 7.7 we will sketch the conjugate utility function (7.67). We observe from the graph that at the point $d_1 = \psi_1 = 1.5$ and $d_2 = \psi_2 = 1.1$ the utility function (7.67) attains its maximum.

Now let, for large N, $\overline{\psi_1} = \frac{r_1'}{N}$ and $\overline{\psi_2} = \frac{r_2'}{N}$. Now using theorem 5, solving equations $\overline{\psi_1} + g_1'(\psi_1, \psi_2) = 0$, $\overline{\psi_2} + g_2'(\psi_2) = 0$ simultaneously we have,

$$\frac{e^{\psi_1}}{e^{\psi_2}} = \frac{1 + \overline{\psi_1}}{1 + \overline{\psi_1, \psi_2}}.$$

Going backward to the original data we have,

$$\frac{\theta_1}{\theta_2} = \frac{r_1'}{r_2'} = \frac{\alpha_1 + r_1}{\alpha_2 + r_2}.$$

7.3.6 Optimum decisions for an Inverse Gaussian distribution when both parameters are unknown

The probability density function of the inverse Gaussian distribution is as follows,

$$p(x|\mu, \lambda) = \left[\frac{\lambda}{2\pi x^3}\right]^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}; \tag{7.68}$$

where $x > 0$, mean $\mu > 0$, the dispersion parameter $\lambda > 0$. Let, $x_1, ..., x_n$ be a random sample of size n from (7.68), so the likelihood function will be,

$$p(\underline{x}|\mu, \lambda) \propto \lambda^{\frac{n}{2}} \exp\left\{-\frac{n\lambda}{2}V - \frac{n\lambda(\overline{x}-\mu)^2}{2\mu^2\overline{x}}\right\}. \tag{7.69}$$

where, $V = \frac{1}{n}\left(\frac{1}{\sum_{i=1}^{n} x_i} - \frac{1}{\overline{x}}\right)$, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. We can re-write (7.69) as,

$$p(\underline{x}|\mu, \lambda) \propto \lambda^{\frac{n}{2}} \exp\left[\frac{n\lambda}{\mu}\right] \exp\left[-(\frac{nV}{2} + \frac{n}{2\overline{x}})\lambda - \frac{n\lambda\overline{x}}{2\mu^2}\right]. \tag{7.70}$$

Now define a conjugate family as,

$$p(\underline{x}|\mu, \lambda) \propto \lambda^{\frac{n_0}{2}} \exp\left[\frac{n_0\lambda}{\mu}\right] \exp\left[-\frac{\tau_0}{2}\lambda - \frac{\tau_1\lambda}{2\mu^2}\right]. \tag{7.71}$$

Now combining (7.70) and (7.71) we have the posterior distribution will be,

$$p(\mu, \lambda|\underline{x}) \propto \lambda^{\frac{n+n_0}{2}} \exp\left[\frac{(n+n_0)\lambda}{\mu}\right] \exp\left[-\left(\frac{nV}{2} + \frac{n}{2\overline{x}} + \frac{\tau_0}{2}\right)\lambda - \left(\frac{n\overline{x}}{2} + \frac{\tau_1}{2}\right)\frac{\lambda}{\mu^2}\right]. \tag{7.72}$$

Now following Gutierrez-Pena and Smith (1995), the canonical form of the above distribution obtained by setting, $(\psi_1, \psi_2) = \left(-\frac{\lambda}{2}, -\frac{\lambda}{2\mu^2}\right)$ and $(x_a, x_b) = \left(nV + \frac{n}{\overline{x}} + \tau_0, n\overline{x} + \tau_1\right)$. Also let, $N = n + n_0$. So (7.72) becomes,

$$p(\psi_1, \psi_2|\underline{x}) \propto \left[(-2\psi_1)^{\frac{1}{2}} \exp\{-2\sqrt{\psi_1\psi_2}\}\right]^N \exp\left[x_a\psi_1 + x_b\psi_2\right]; \tag{7.73}$$

where, $G(\psi_1, \psi_2) = (-2\psi_1)^{\frac{1}{2}} \exp\{-2\sqrt{\psi_1\psi_2}\}$. So, $g = \frac{1}{2}\log(-2\psi_1) - 2\sqrt{\psi_1\psi_2}$. We have , $g_1'(\psi_1, \psi_2) = \frac{1}{2\psi_1} - \sqrt{\frac{\psi_2}{\psi_1}}$ and $g_2'(\psi_1, \psi_2) = -\sqrt{\frac{\psi_1}{\psi_2}}$. Now using (7.48) we have,

$$n(d_1, d_2) = \sqrt{d_1 d_2}$$

assuming $\kappa = \frac{1}{2}$. So from (7.47) we have the conjugate utility function will be,

$$\begin{aligned} u(\psi_1, \psi_2, d_1, d_2) &= \exp[(\sqrt{d_1 d_2})\{\frac{1}{2}\log(-2\psi_1) - 2\sqrt{\psi_1\psi_2} - (\frac{1}{2\psi_1} - \sqrt{\frac{\psi_2}{\psi_1}})(\psi_1 - d_1) \\ &+ \sqrt{\frac{\psi_1}{\psi_2}}(\psi_2 - d_2) - \frac{1}{2}\log(-2d_1) + 2\sqrt{d_1 d_2}\}]. \end{aligned} \tag{7.74}$$

The utility function (7.74) will be maximum at $\psi_1 = d_1$ and $\psi_2 = d_2$. In figure 7.8 we sketched three dimensional shape of the the conjugate utility function (7.74). We observe from the graph that at the point $d_1 = \psi_1 = -4$ and $d_2 = \psi_2 = -2$ the utility function (7.67) attains its maximum. Because the inverse Gaussian parameters μ and λ are positive, looking at the canonical form of the parameter ψ_1 and ψ_2, we can ensure the positive values of the parameters by taking the negative values of ψ_1 and ψ_2. For large N let, $\overline{x_1} = \frac{x_a}{N}$ and $\overline{x_2} = \frac{x_b}{N}$. Now by theorem 5, solving equations $\overline{x_1} + g_1'(\psi_1) = 0$ and $\overline{x_2} + g_2'(\psi_2) = 0$ simultaneously, we get the solution for ψ_1,

$$\psi_{01} = \left[2\left(\frac{1}{\overline{x_2}} - \overline{x_1}\right)\right]^{-1},$$

and the solution for ψ_2,

$$\psi_{02} = \left[2\overline{x_2}^2\left(\frac{1}{\overline{x_2}} - \overline{x_1}\right)\right]^{-1}.$$

Now going backwards to the original data we have,

$$\hat{\mu} = \frac{1}{N}(n\overline{x} + \tau_1)$$

and

$$\hat{\lambda} = \left[\frac{1}{N}\left(nV + \frac{n}{\overline{x}} + \tau_0\right) - \hat{\mu}^{-1}\right]^{-1}.$$

The approximate value of ψ_1 and ψ_2 will maximize the expected utility defined in (7.74).

7.4 Conclusion

Lindley (1976) clearly explained the concepts of conjugate utility functions and proved theorems to obtain approximate optimum decisions under the utility function for

one parameter exponential family. First of all we have presented his theorem and showed his proof of it. Then using his results we obtained the approximate optimum decisions for the parameter of the exponential, Pareto and Maxwell distributions under the conjugate utility functions. For each distributions we also graphed the shape of the utility function. We have extended the conjugate utility function for the one parameter exponential family of Lindley (1976) to the two parameter exponential family. Here we made the approximation in the expected utility given in (7.36) where we showed the maximum utility will be 1 for $\psi_1 = d_1$ and $\psi_2 = d_2$. We have outlined a method of choosing $\underline{d}$ in equation (7.44). To maximize the expected utility of (7.36) for the two parameter exponential family we have approximated the ratio of two integrals given in (7.38) by Taylor expansion. At the time of doing the expansions, we have considered two roots, given in the equations (7.52) and (7.53), so that the expected utility given in (7.38) is maximized and the approximate decisions for the parameters without $O(N^{-1})$ are summed up in theorem 5. The corrected estimates of the parameters are presented in theorem 6 where we retained $O(N^{-1})$ for the optimum decisions. We did not follow the approximation procedure of Tierney and Kadane (1986) as their intention was to approximate the posterior moments but for us to maximize the expected utility. But the similarity in both methods needs to be a positive g function. For the two parameter exponential family, for a g function, we re-parameterize the parameter $\underline{\theta}$ to $\underline{\psi}$ following the canonical forms of Smith and Gutierrez-Pena (1985), then we applied the results of theorem 5 for an optimum decisions of $\underline{\psi}$. Someone can easily get the optimum decisions for $\underline{\theta}$ going backward to the original data which was shown with examples in section 7.9-7.11. For each case we have written the mathematical form of the utility function

and sketched the shape of the utility function as well. It is to be noted that we have clearly discussed the form of the utility function for the two parameter and developed the approximation procedure of maximizing it under the two parameter exponential family. In the approximations by Lindley (1980), Tierney and Kadane (1986), Tierney, Kass and Kadane (1989) did not consider the form of utility function in their optimization and obtained approximations to the posterior moments. We did not consider the optimum sample size for the exponential family and left it for future research.

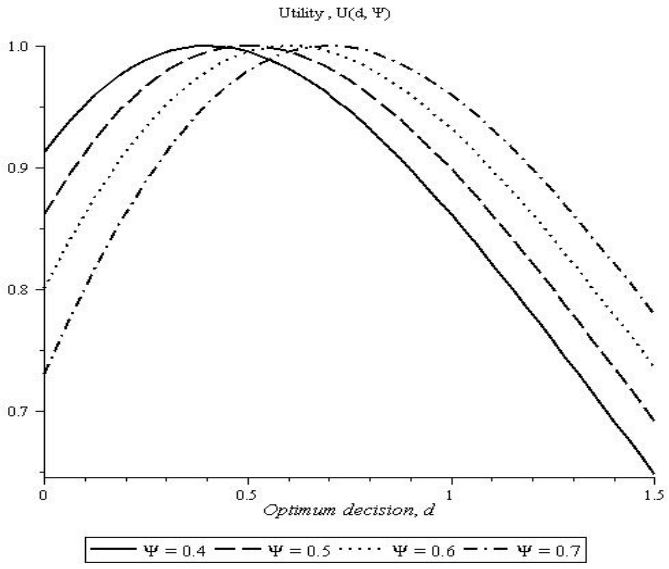

Figure 7.1: Conjugate utility function to estimate a Poisson parameter $\psi = \log\theta$ as a function of d for $k = 1$ and for different values of ψ.

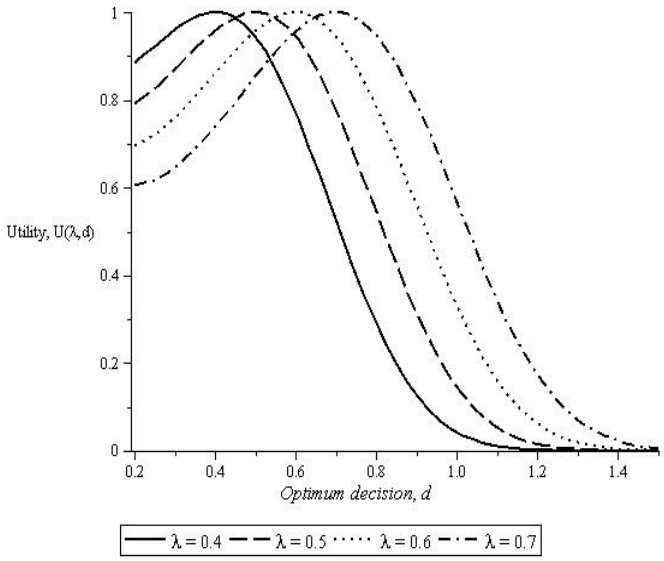

Figure 7.2: Conjugate utility function to estimate an exponential parameter λ as a function of d when $n = 10$, $k = 1$ are fixed.

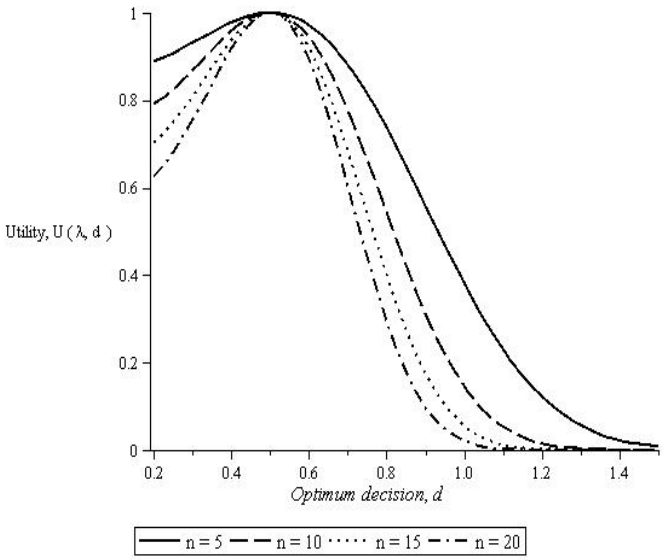

Figure 7.3: Conjugate utility function to estimate an exponential parameter λ as a function of d when $\lambda = 0.5$, $k = 1$ are fixed and for different values of n.

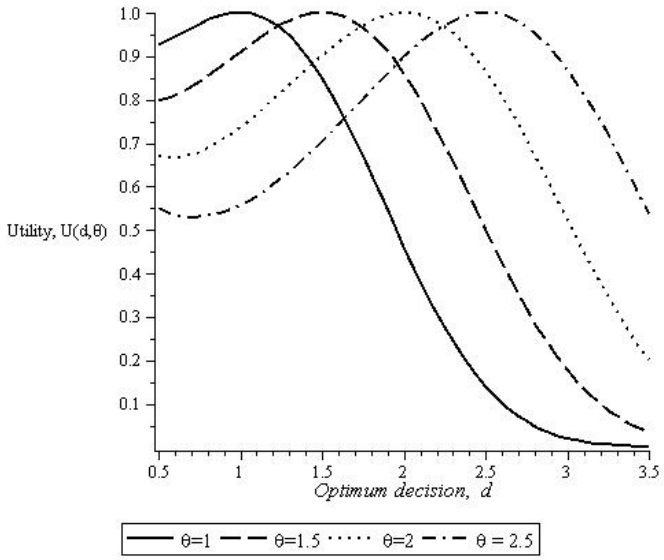

Figure 7.4: Conjugate utility function to estimate a Pareto parameter as a function of d when $k = 1$ are fixed and for different values of the parameter θ.

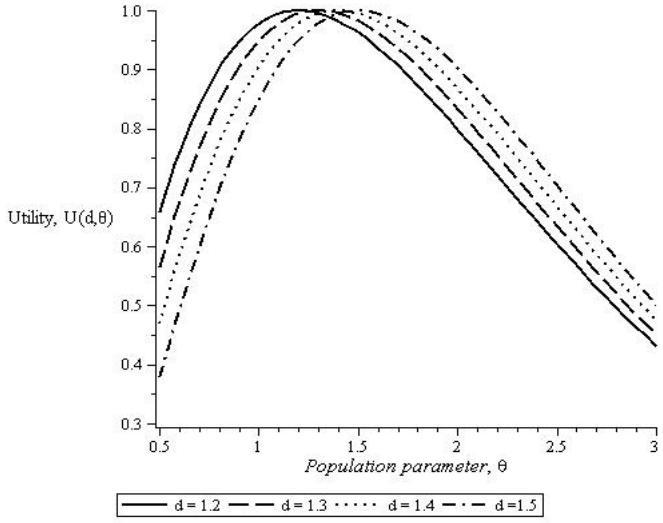

Figure 7.5: Conjugate utility function to estimate a Pareto distribution as a function of the parameter θ when $k = 1$ is fixed and for different values of d.

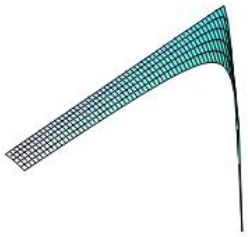

Figure 7.6: Shape of the conjugate utility function for a normal distribution for $\psi_1 = 1.5$ and $\psi_2 = -5$ and in x axis takes the decision d_1 with range -1 to 1 and in y axis takes the decision d_2 with range -1 to 1.

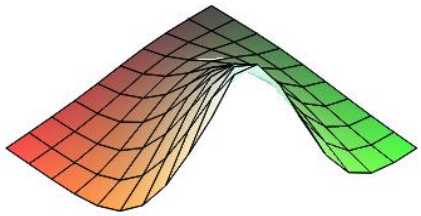

Figure 7.7: Shape of the conjugate utility function for a trinomial distribution for $\psi_1 = 1.5$, $\psi_2 = 1.1$ and in x axis takes the decision d_1 with range 0 to 2.5 and in y axis takes the decision d_2 with range 0 to 2.

Figure 7.8: Shape of the conjugate utility function for an inverse Gaussian distribution for $\psi_1 = -4$, $\psi_2 = -2$ and in x axis takes the decision d_1 with range -5 to 2 and in y axis takes the decision d_2 with range -5 to 2.

Chapter 8

Limitations and future research

8.1 Introduction

In this chapter we shall discuss the applications and limitations of the research done with some indication of possible areas for future research. In the first chapter we reviewed the literature related to loss functions, utility functions and Bayesian sample size determinations. In the chapters 2-6 we have obtained the optimum sample size under various symmetric and asymmetric loss functions. In chapter 7 we studied the Lindley's conjugate utility for the one parameter exponential family and we extended it to the two parameter exponential family. In the next section first we will discuss SSD based on power like calculations and utility (or loss) functions and compare our results with other authors. Then we will discuss Chapters 3-7 indicating the achievements and limitations of the work done in each chapter. We have considered a linear cost function and the posterior risk function for an optimum SSD. In the

following section we will lay out the measurements of costs and risks with a suggestion as to how we can measure these using the same units of measurement. Finally we will give some indications of some possible areas of future research.

8.2 Discussion

We obtained the optimum SSD for a number of models using both symmetric and asymmetric loss functions. Here we will discuss and compare our methods with the methods used by other authors. DeGroot (1970) obtained Bayesian sample size for gamma distribution using a squared error loss function and a linear cost function. He also considered an absolute error loss function for the optimum sample of size n to estimate a normal mean with known precision. In both situations he considered a loss function instead of a utility function and obtained the posterior risk function. Then, adding a linear cost function to the risk function he minimized them together to obtain the optimum sample size which is equivalent to the MEU discussed in Chapter 1. But rather than considering a utility function he considered a loss function. Lindley (1972) obtained the sample size using the squared error loss function for a normal distribution with known variance. In the example (refer to Chapter 2) he minimized a linear cost function and the posterior risk together to obtain the optimum sample of size n. It is of interest for the purposes of this study that DeGroot (1970) considered the absolute error loss function. Muller and Parmigiani (1995) also considered the absolute error utility function to obtain the SSD for a binomial distribution through curve fitting of a Monte Carlo experiment. Adcock (1988) obtained the solution for

an optimum sample size using the ACC method but there has been a ongoing long debate obtaining the sample size between adherence of the ACC method and the MEU method. According to Lindley i) the use of MEU is guaranteed to be coherent : this method is comparable to other methods ii) it is accompanied by a well defined algorithm for its solution iii) a wide range of utility functions can be accommodated. Finally, he suggested that to understand utility and do more research into practical determination of utilities so that the method can be used more effectively. We tried to introduce some new forms of utility function and the SSD under these utility functions as well.

Now we will obtain the sample size using the sampling prior and fitting prior of Wang and Gelfand (2002) and then we will compare our method where we will obtain the sample size using the approach described in Chapter 4. We have discussed Wang and Gelfand's (2002) approach in Chapter 1. For the sampling prior, first we will simulate θ^* from the chosen prior distribution of θ, say, $p(\theta)$. Then we simulate the data $x_i's$ from the distribution $x_i's|\theta^*$. Next we take a non informative prior distribution of θ called the fitting prior and then combine this prior and the data to get the posterior distribution, $p(\theta|\underline{x})$. At this stage we find the posterior expected risk and add this with the cost function to get total cost. Finally, plot sample size against the total cost. We look for the sample size producing minimum cost, which is considered as an optimum sample size. So, this method involves, simulating data from a given prior distribution (for the parameter of interest) then simulating data given those prior samples, finally considering a fitting prior which is non-informative to give the posterior distribution. To do this we have used the R program for the simulation study.

Table 8.1: Optimum sample size for a Poisson parameter under the SE loss.

Pre posterior is a negative binomial Conjugate prior is a Gamma	Sample size	Fitting prior non informative Sampling prior is a Gamma	Sample size
$\lambda = 1$, Gamma(1,1)	35	Gamma(1,1)	33
$Gamma(2, 2)$	30	Gamma(2, 2)	32
$Gamma(3, 3)$	24	Gamma(3,3)	27
$Gamma(4, 4)$	20	Gamma(4,4)	24

In table 8.1 show the optimum sample size to estimate a Poisson parameter under the squared error loss function. Following the approach of Wang and Gelfand (2002), we generated the prior sample from a gamma distribution. Then generated samples from the Poisson distribution given that prior sample. Now giving a fitting prior as non informative and combine it with the data we obtained the posterior distribution. Here we considered a cost function $c_0 + cn = 1 + 0.001n$ and added it with the posterior risk to get the total cost. Finally plot total cost against sample size. Pick the optimum sample size giving minimum cost. It is clear from the table 8.1 that the optimum sample sizes are very close for both under the usual optimization procedure as in Chapter 4 and the approach of Wang and Gelfand (2002) although as the prior becomes more informative the difference increases. Similarly we have considered the optimum sample size for a binomial parameter under the SE loss function considering the cost function $1+0.0008n$. We also found from table 8.2 that the optimum samples sizes are close for both approaches.

In table 8.3 we presented the optimum sample size for a binomial parameter $p = 0.1$ under the absolute error loss function using our approach and the approach of Wang and Gelfand (2002). This time we have taken the sampling cost per unit as 0.0008 as did Muller and Parmigiani (1995). We observe that there is not a huge

Table 8.2: Optimum sample size for a binomial parameter under the SE loss.

Pre posterior is a beta binomial Conjugate prior is a beta	Sample size	Fitting prior non informative Sampling prior is beta	Sample size
Beta(0.3,0.7)	15	Beta(0.3,0.7)	16
Beta(0.2, 0.8)	12	Beta(0.2, 0.8)	14
Beta(0.1,0.5)	8	Beta(0.1,0.5)	10
Beta(0.1,0.9)	5	Beta(0.1,0.9)	7

Table 8.3: Optimum sample size for a binomial parameter under the AE loss.

Pre posterior is a beta binomial Conjugate prior is a beta	Sample size	Fitting prior non informative Sampling prior is a beta	Sample size
Beta(0.3,0.7)	30	Beta(0.3,0.7)	33
Beta(0.2, 0.8)	28	Beta(0.2, 0.8)	29
Beta(0.1,0.5)	26	Beta(0.1,0.5)	27
Beta(0.1,0.9)	23	Beta(0.1,0.9)	25

difference in sample size calculation between our approach and the approach of Wang and Gelfand (2002). To compare both methods, we considered the same loss function and cost functions for both cases. From table 8.3 we observe that, for an absolute error loss function there is some differences found in the optimum sample size in the two methods but not very large. This absolute error loss function is also considered by Muller and Parmigiani (1995) to obtain the optimum sample size. From table 8.3 we can see that, our approach gives very similar results to their results. We obtained the optimum sample size around 30 and they got the optimum sample size around 29 to estimate a binomial parameter $p = 0.1$ with the conjugate prior, Beta(0.3,0.7). On the other hand, when we considered Wang and Gelfand's (2002) approach, then using the usual minimization procedure we found optimum sample size is 33 which is very close to 34, the optimum sample size obtained by Muller and Parmigiani (1995) after

smoothing their fitting curve. So we found some agreement between our approach and others.

But neither Wang and Gelfand (2002) nor Muller and Parmigiani (1995) used the asymmetric loss function or the bounded asymmetric loss function in their studies. So we explored the optimum sample size under these loss functions where we were able to give some closed form optimum sample size for some distributions but for others we used a simulation study for an optimum SSD.

In Chapter 3 we obtained the optimum sample size under the linex loss function for a number of models. In the following practical situations we can use this loss function and in such situations if we need to determine the optimum sample size, we may use the form of SSD discussed in a particular model of Chapter 3 according to the needs of that particular situation.

Example 1. Consider some food which can be displayed on shelves up to 15 days maximum after manufacturing date. If any shop displayed food for only 12-13 days then the shop might lose some money if they hadn't sold it. But if they displayed the food for 16-20 days this may cause a huge danger to customers as it may be poisonous, which in monetary terms could be very serious. So again over estimation of days is more dangerous than under estimation. In that case $b > 0$.

Example-2. Sengupta (2008). Assume a civil engineer is building a dam and is interested in finding the height of the dam that is being built. If, due to some error, the height is estimated to be greater than the optimal value, then the costs the engineer incurs are mainly for materials and labour. On the other hand, if the

estimated height is less than it should be, then the consequences will be dangerous in terms of environmental impact, which in monetary terms can be very high. So in this situation, it is logical to use $b < 0$, such that under estimation is penalized more than over estimation.

There are many situations like this in medicine as well. For example, under estimating (or over estimating) the blood pressure of a patients could have a huge effect on the patient's treatment. In the food industry, when baking biscuits over-heating might be more serious than under heating.

We have obtained the optimum sample size under the blinex loss function in Chapter 4. This overcomes the limitations of the linex loss functions where we are unable to obtain the Bayes estimate if the mgf of the distribution doesn't exist. In practice the loss is bounded for any practical situation. So it is also important to learn about the optimum SSD under the bounded loss function. Highlighting this issue Smith (1980) studied where the Bayes estimates for a parameter θ must lie for a specific posterior distribution of θ under different classes of symmetric bounded loss functions. In such situations we need to find the optimum SSD as well. Looking at the practical importance of bounded loss functions in this chapter we have explored with the optimum SSD under the bounded asymmetric loss functions. But in this case we couldn't manage to get the closed form for an optimum sample size because of the complicated form of the posterior risk function and we used the simulation study using R for the optimum SSD.

Some popular unbounded loss functions like SE loss function may be used in different situations and it is useful to obtain an optimum SSD under those loss functions.

In chapter 5 we have obtained the optimum sample size under the loss function of DeGroot(1970). This is an unbounded loss function and described for the positive values of the parameters. For some distributions we found the optimum SSD in closed form but for others not.

In Chapter 6 we have proposed an asymmetric loss function for a parameter taking positive values which gives the posterior mean as a Bayes estimate. This is the first asymmetric loss function which came in the literature giving posterior mean as a Bayes estimate besides the symmetric SE loss function. This is also a bounded loss function which has practical importance as described by Smith (1980). The limitation of this loss function is it is only for parameters taking positive values. So it doesn't cope with the very commonly used normal distribution.

It is very difficult to obtain the estimates of the distributions with two or more unknown parameters. On the other hand for any unique decision we need to call a loss or utility function which can cope with the distribution under study. Lindley (1976) clearly described the optimum decision procedure for one parameter exponential family proposing one parameter conjugate utility function. First, we reviewed his work in Chapter 7 then we extended it for the two parameter exponential family. We have proposed a conjugate utility function for the bivariate exponential family, then described the optimization procedure for the optimum decisions of both parameters when they are unknown.

8.3 Risks and costs

In Chapter 1, section 1.9, we discussed the cost function, $C(n)$. For an optimum SSD to get the total cost we always added $C(n)$ to the average posterior risk (APR) function. That is

$$E[TC(n)] = C(n) + APR.$$

Here $C(n)$ is measured in pounds but the risk is dimensionless. To choose an appropriate scale for a decision maker I could ask "If I could guarantee to reduce the risk by 1 unit, how much would you pay in pounds?" Let κ be the value in pounds that the decision maker would pay to reduce the the posterior risk by 1. So the total cost is now better expressed as,

$$E[TC(n)] = C(n) + \kappa(APR).$$

Note that Lindley (1976) says that all consequences should be measured in 'utiles' but does not fully explain how this should be done. Note also that scaling the loss function and hence the posterior risk by κ does not change the decision that minimizes the expected loss but it does change the decision for n. It does change the value of the loss or posterior risk in an appropriate way. Similar concepts have been used in health economics to put the costs and the effectiveness of treatments into the same unit of measurements by introducing a suitable scale factor (Jackson and Nixon 2010).

Another way of looking at the costs and the posterior risks in the same units of measurement is to find the optimum value of the total cost $TC(n)$ at $n = n^*$. From Chapter 2, recalling the equation (2.2.1), the expected total cost (to estimate

a normal mean when the precision is known) will be,

$$TC(n)_{n=n^*} = TC(n^*) = c_0 - cn_0 + 2\sigma\sqrt{ca_0},$$

where a_0 is the scale parameter of the squared error loss function which could play the same role as κ described earlier, σ is the sample standard deviation measured in some distance, n_0 is the prior sample size (some numbers), c_0, c are the initial cost and the sampling cost per unit respectively which can both be measured in pounds. Looking

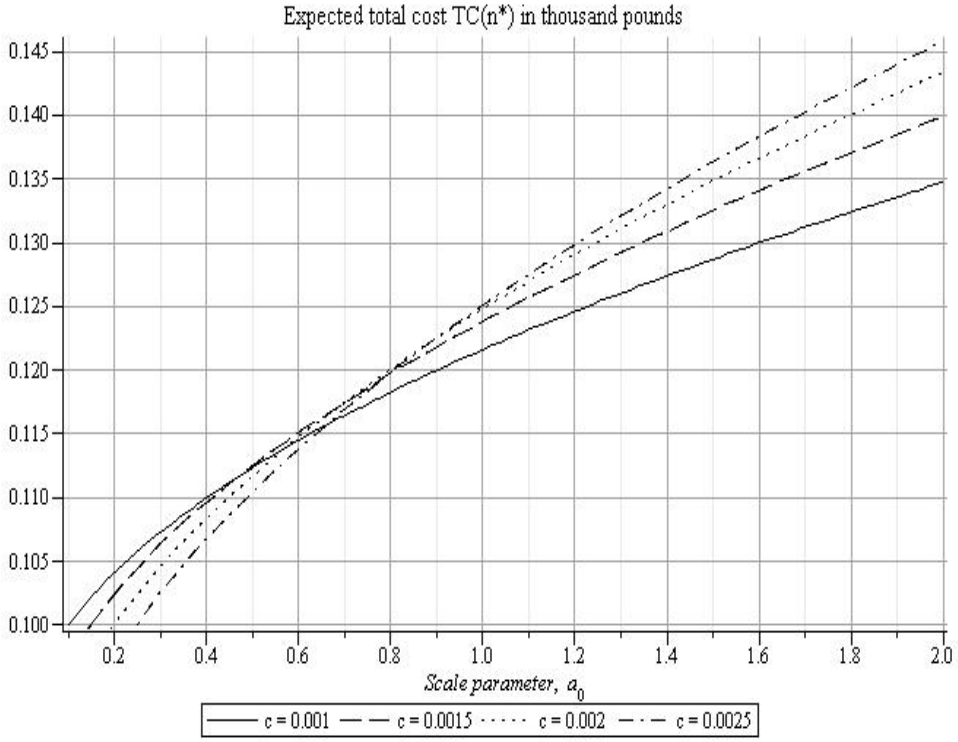

Figure 8.1: The expected total cost for different choices of the scale parameter of the SE loss function a_0 to estimate a normal mean when $\sigma = 0.5$, $n_0 = 10$ and $c_0 = £100$.

at figure 8.1, we can see that the expected total cost is £130 when the sampling set up cost is £100 and the sampling cost per unit is £1. From this we can say that the

experimenter is able to draw up to 30 samples from the budget of £130 by paying up to £1600 (the value of a_0) for reducing the risks by 1 unit. This £1600 is the conversion rate of risk to the equivalent value of costs (same role as κ). We sketched four different scenarios taking different sampling costs per unit. But for a particular experiment the experimenter should have to choose only one strategy out of many possibilities so that he/she has to choose one value a_0, and c looking at his optimal budget for an optimum SSD. It is to be noted that some other parameters (in this example σ, n_0 and c_0) should be known to the experimenter before deciding how many representative samples are needed to perform the study.

8.4 Future research

Because of the limited scope of the present investigation it would be important in the future for further research into the symmetric and asymmetric loss functions and the practical application of these loss functions. Further research would also be done into the optimum decisions and the optimum SSD under these loss functions. There may be some other applications of the linex loss function in different fields, for example, in actuarial science or in clinical trials which require further research.

We have obtained the optimum SSD for different symmetric and asymmetric loss functions. Hopefully, in future, because of the practical importance of the bounded loss functions, we can explore with the optimum sample size under the symmetric bounded loss functions given by Smith (1980).

We have proposed an asymmetric bounded loss function which gives the posterior mean as a Bayes estimate for the positive values of the parameter. There are some areas for improvement in this area, for example, to find the form of the asymmetric bounded loss function or utility function which covers both positive and negative values of the parameter, which can also be used in practical situations.

Throughout the optimum SSD study we have considered a linear cost function but in some other situations a non linear form of the cost function e.g. $c_0 + c\sqrt{n}$ may be used. This requires further research into how it can be used in practical situations.

We have obtained the optimum decisions for the bivariate exponential family when both parameters are unknown under a bivariate conjugate utility function through approximation. This could be extended to more than two parameters. At this stage we are unable to obtain the optimum SSD for the univariate and bivariate exponential family because of the complicated form of the posterior expected utility function and have left this as possible avenue of future research.

Bibliography

[1] Adcock, C. J. (1988) A Bayesian approach to calculating sample sizes. *Statistician,* 37, 433-439.

[2] Adcock, C. J. (1997a) Sample size determination: a review *Statistician,* 46, 2, 262-283.

[3] Adcock, C. J. (1997b) The choice of sample size and the method of maximum expected utility-comments on the paper by Lindley. *Statistician,* 46, 2, 155-162.

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

[5] Bernardo, J .M., and Smith, A. F. M. (1994) *Bayesian Theory,* Chichester, U.K.: John Wiley.

[6] Bolfarine H. (1989) A note on finite population prediction under asymmetric loss functions. *Communications in Statistics-Theory and Methods,* 18(5), 1863-1869.

[7] Britney, R. R. and Winkler, R. L. (1974) Bayesian Point Estimation and Prediction. *Annals Institute of Statistical Mathematics,* Vol.26, 15-34.

[8] Brown, L. (1968) Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters.*Ann Math Statist*; 39 (1): 29-48.

[9] Cain, M. (1991) Minimal quantile measures of predictive cost with a generalized cost error functions. *Economic Research Papers,* No .40, University of Wales, Aberystwyth.

[10] Cain, M and Janssen, C. (1995) Real estate price prediction under asymmetric loss.*Annals of the Institute of Statistical Mathematics,* 47 (3), 401-414.

[11] Calabria, R. and Pulcini, G. (1996) Point estimation under asymmetric loss functions for left-truncated exponential samples. *Communications in Statistics-Theory and Methods,* 25(3), 585-600.

[12] Cain, M and Janssen, C. (1999) A decision framework for nonsymmetric losses. *Information,* 37(4),357-366.

[13] Calabria, R. and Pulcini, G. (1994) An engineering approach to Bayes estimation for the Weibull distribution. *Microelectronic Reliability,* 34 (5),789-802.

[14] Canfield R. V. (1970) A Bayesian approach to reliability estimation using loss function. *IEEE Trans. Reliability,* vol. R-19, 13-16.

[15] Chai, J., Wei, J., Shi, Y.and Li, X. (2007) Emperical Bayes estimators using linex loss under regressively type-II censored samples. *Mathematics in Computer Simulation,* Vol. 73, 320-326.

[16] Chang, Y. and Hung, W. (2007) Linex loss function with application to determining the optimum process parameter. *Quantity and Quantity,* Vol.41, pp.291-301.

[17] Chattopadhyay, S. (2000) Sequential estimation of normal mean using asymmetric loss function. *Communication in Statistics-Theory and Methods,* Vol. 29, pp.738-795.

[18] DeGroot, M. H. (1970) Optimal statistical decision. McGraw-Hill Inc.

[19] Demetrescu, D. (2006) An extension of the Gaussian-Newton algorithm for estimation under asymmetric loss. *Computational Statistics and Data Analysis,* 50, (2006),379-401.

[20] Farsipour, N. S. and Zakerzadeh, H. (2006) Estimation of generalized variance under an asymmetric loss function 'squared log error'. *Communications in Statistics- Theory and Methods* 35, Issue 4, 571-581.

[21] Granger C. W. J. (1969) Prediction with a generalized cost of error function, *Operational Research Quarterly,* 20, 199-207.

[22] Guitierrez-Pena, E. and Smith A. F. M. (1995) Conjugate parameterizations for natural exponential families. *Journal of the American Statistical Associations,* Vol.90, No.432, 1347-1356.

[23] Jackson C. and Nixon R. (2010) Winbugs in health economic evaluations. Notes for MRC Biostatistics unit course 15-16 July, Cambridge.

[24] Joseph L. and Belisle P. (1997) Bayesian sample size determination for normal means and differences between normal means. *Statistician,* 46, 2, 209-226.

[25] Joseph, L., Du Berger, R. and Belisle P. (1997) Bayesian and mixed Bayesian likelihood criteria for sample size determination. *Statist. Med.*, 16, 769-781.

[26] Joseph, L., Wolfson, D.B. and Du Berger, R. (1995) Sample size calculations for binomial proportions via height posterior density intervals. *Statistician,* 44, 143-154.

[27] Lindley, D. V. (1961) the use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkley Symp. Math. Statist. Prob.,* 1 453-468.

[28] Lindley D. V. (1972) Bayesian Statistics, a Review. Philadelphia: Society for Industrial and Applied Mathematics.

[29] Lindley, D. V.(1976) A Class of utility functions. *The Annals of Statistics,* Vol.4, No.1, 1-10.

[30] Lindley, D. V. (1980) Approximate Bayes method. *Trabajos De Estadistica,* 31, 223-237.

[31] Lindley, D. V. (1997a) The choice of sample size, *Statistician,* 46, 2, 129-138.

[32] Lindley, D. V. (1997b) The choice of sample size - A reply to the Discussion. *Statistician,* 46, 2, 163-166.

[33] Maple 13 (1996-2009) Copyright ©Maplesoft, a division of Waterloo Maple Inc.

[34] Marsh, A. C. and Zellner, A. (2004) Bayesian solutions to graduate admissions and related selection problems. *Journal of Econometrics,* 121, 405-426.

[35] Mood, A., Graybill, F. A. and Boes, D. (1974) *Introduction to the Theory of Statistics,* McGraw-Hill Series in Probability and Statistics.

[36] Moore, D. and Papadopoulos, A. S. (2000) The Burr type XII distribution as a failure model under various loss functions. *Microelectronic Reliability,* 40, 2117-2122.

[37] Muller, P. and Parmigiani, G. (1995), Optimal design via curve fitting of Monte Carlo experiments. *American Statistical Associan* Vol. 90, No. 432, Theory and Methods, 1322-1330.

[38] Pandey, B. N. and Rai, O. (1992) Bayesian estimation of mean and square of mean of normal distribution using LINEX loss function.*Communications in Statistics - Theory and Methods,* 21(12), 3369-3391.

[39] Parsian, A. and Farsipour, N. S. (1993) On the admissibility and inadmissibility of estimators of scale parameters using an asymmetric loss function. *Communications in Statistics - Theory and Methods,* 22(10), 2877-2901.

[40] Parsian, A. and Nematollahi, N. (1994) Estimation of scale parameter under entropy loss function. *Journal of Statistical Planning and Inference*. 52, 77-91.

[41] Pham-Gia, T. (1997) Sample size determination in Bayesian analysis. *The Statistician,* 46, 139-144.

[42] Pham-Gia, T. and Turkkan, N. (1992) Sample size determination in Bayesian analysis. *The Statistician,* 41, 389-392.

[43] Pitman, E. J. G. (1939) The estimation of the location and scale parameter of a continuous population of any given form. *Biometrika,* 30, 391-421.

[44] Raiffa, H. and Schalaifer, R. (1961) Applied statistical Decision Theory. Boston, Havard University Graduate School of Business Adminstration.

[45] Rojo, J. (1987) On estimation of of $c\bar{x} + d$ with respect to LINEX loss function. *Communications in Statistics-Theory and Methods,* 16(12), 3754-3748.

[46] Sadooghi-Albandi, S. M. and Nematollahi, N. (1989) A note on admissibility of $c\bar{x} + d$ relative to the LINEX loss function. *Communications in Statistics - Theory and Methods,* 18(5), 1871-1873.

[47] Sadooghi-Albandi S. M. and Parsian, A. (1992) Estimation of binomial parameter n using LINEX loss function. *Communications in Statistics - Theory and Methods,* 21(5), 1427-1439.

[48] Sahu, S. K. and Smith, T. F. M. (2006) A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society A,* 169, Part-2, 235-253.

[49] Santis, F. D. (2006) Sample size determination for robust Bayesian analysis. *Journal of the Royal Statistical Society Series A,* Vol.101, part 1, pp. 278-291.

[50] Santis F. D. (2007) Using historical data for Bayesian sample size determination. *Journal of the American Statistical Associations,* Vol.81, No.473, 446-451.

[51] Schabe, H. (1996) Combining failure rate data from various sources. *IEEE Trans. Reliab.,* Vol.40, No.1, pp.63-67.

[52] Schabe, H. (1991) Bayes estimates under asymmetric loss. *Microelectron. Reliab.,* Vol.36, No.1, 47-54.

[53] Sengupta, R. N. (2008) Use of asymmetric loss functions in sequential estimation problems for multiple linear regression. *Journal of Applied Statistics,* Vol.35, No.3, 245-261.

[54] Shi, Y., Gao, S. and Shi, X. (2005) Convergence rate of emperical Bayes estimation for two dimensional truncation parameters under linex loss function. *Information Science,* 171, 1-11.

[55] Smith, J. Q. (1980) Bayes estimates under bounded loss. *Biometrika,* 67(3), 629-638.

[56] Soliman, A. A. (2001) Linex and quadratic approximate Bayes estimators applied to the Pareto model.*Communications in Statistics - Simulation and Computation,* 30(1), 47-62.

[57] Soliman, A. A. (2002) Reliability estimation in generalized life-model with application to the Burr-XII. *IEEE Transactions on Reliability,* Vol.51, No.3, 337-343.

[58] Stuger, H. P. (2006) Asymmetric loss functions and sample size determination: a Bayesian approach. *Australian Journal of Statistics,* 35 (1), 57-66.

[59] Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* Vol. 81, No. 393, Theory and Methods.

[60] Tierney, L., Kass, R. E. and Kadane, J. B. (1989) Fully exponential Laplace approximations and variances of nonpositive functions. *Journal of the American Statistical Association* Vol. 84, No. 407, Theory and Methods.

[61] Varian, H. R. (1975) A Bayesian approach to real estate assessment. *in Studies in Bayesian Econometrics and Statistics in Honour of Leonard J. Savage, eds. Stephen E. Fienberg and Arnold Zellner, Amsterdam: North Holland,* 195-208.

[62] Wang, F. and Gelfand, E. (2002) A simulation based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science,* Vol.17, No.2, 193-208.

[63] Weiss, R. (1997) Bayesian sample size calculation for hypothesis testing. *The Statistician*, 46, 2, 185-191.

[64] Wen, D. and Levy, S. (2001) BLINEX: A bounded asymmetric loss function.*Communications in Statistics-Theory and Methods,* 30(1), 147-153.

[65] Wen, D. and Levy, S. (2004) Algorithm of fitting BLINEX loss parameters. *Communications in Statistics - Simulation and Computation,* 33(3), 597-608.

[66] Xiao, Y. S., Takada, Y. and Shi, N. Z. (2005) Minimax confidence bound of normal mean under an asymmetric loss function. *Annals of the Institute of Statistical Mathematics,* Vol.57, Issue 1, 167-182.

[67] Zellner, A. (1986) Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Associations,* Vol.81, No.394, 446-451.