# ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

## DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

# A BAYESIAN APPROACH IN DETERMINING THE OPTIMAL SAMPLE SIZE FOR PHASE I DATA

By

**Efstathia N. Giannopoulou**

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece

2006

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΚΑΤΑΛΟΓΟΣ
0 000000 573450

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ
ΒΙΒΛΙΟΘΗΚΗ
εισ. 80013
Αρ.
ταξ.

# ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

## DEPARTMENT OF STATISTICS

## POSTGRADUATE PROGRAM

# A Bayesian approach in determining the optimal sample size for phase I data

By

Efstathia N. Giannopoulou

A THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
January 2006


ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ
ΒΙΒΛΙΟΘΗΚΗ
εισ. 80013
Αρ.
ταξ.

# ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

## Μία Μπεϋζιανή προσέγγιση για τον καθορισμό του βέλτιστου μεγέθους δείγματος για δεδομένα φάσης Ι

Ευσταθία Ν. Γιαννοπούλου

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Ιανουάριος 2006


ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ
ΒΙΒΛΙΟΘΗΚΗ
εισ. 80013
Αρ.
ταξ.

# ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

## DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

# A BAYESIAN APPROACH IN DETERMINING THE OPTIMAL SAMPLE SIZE FOR PHASE I DATA

Efstathia N. Giannopoulou

**Approved by the Graduate Committee**

| | | |
|---|---|---|
| P. Tsiamyrtzis | P. Dellaportas | V. Vasdekis |
| Lecturer | Associate Professor | Assistant Professor |
| Thesis Supervisor | Members of the Committee | |

**Athens, June 2006**

**Michael Zazanis, Professor**
**Director of the Graduate Program**

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΒΙΒΛΙΟΘΗΚΗ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## DEDICATION

To my parents Nikos, Athina and my sister Zaharoula.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Panagiotis Tsiamyrtzis for his invaluable help and advice during the development of this thesis. Furthermore I am grateful to my parents for their moral and economical support during the period of my studies.

Finally, I would like to thank my friends Maria Bakogeorgou and Eleni Papadimitriou for their continuous encouragement.

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΒΙΒΛΙΟΘΗΚΗ



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ



# VITA

I was born in Athens, in January 1981. I graduated from the 12$^{th}$ High School of Peristeri-Athens in 1998. In 1999 I entered the Department of Mathematics in the University of Patras, Greece. I obtained my degree in September 2003 and from October 2003 I am a student of the Master Program in Statistics, offered by the Department of Statistics, Athens University of Economics and Business, Greece. Since October 2004, I worked on my Master Thesis.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# ABSTRACT

Eystathia Giannopoulou

**A Bayesian approach in determining the optimal sample size for phase I data**

January 2006

In this thesis our interest is concentrated on determining the optimal sample size of observations, for phase I data that needs to be taken each time during a process control in order to estimate the unknown parameters of interest, taking into account the fixed cost of every observation along with available prior information. For this we have assumed that the sampling interval and the length of the phase I samples are fixed. The sampling distributions that we referred to are members of a more general family of distributions functions, the regular exponential family and in each separate case an extended analysis has been performed. In defining the optimal sample size for every different sampling distribution we have used basic statistical decision theory properties which are considered to be the major tool of this work.

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΒΙΒΛΙΟΘΗΚΗ



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# ΠΕΡΙΛΗΨΗ

Ευσταθία Γιαννοπούλου

**Μία Μπεϋζιανή προσέγγιση για τον καθορισμό του βέλτιστου μεγέθους δείγματος για δεδομένα σε φάση Ι**

Ιανουάριος 2006

Στην εργασία αυτή το ενδιαφέρον μας επικεντρώνεται στον καθορισμό του βέλτιστου δείγματος μεγέθους παρατηρήσεων για δεδομένα φάσης Ι το οποίο πρέπει να πάρουμε κάθε φορά κατά τη διάρκεια μιας διαδικασίας ποιοτικού ελέγχου με σκοπό να εκτιμήσουμε τις άγνωστες παραμέτρους που μας ενδιαφέρουν, λαμβάνοντας υπ' όψιν το καθορισμένο κόστος για κάθε παρατήρηση μαζί με τη διαθέσιμη πληροφορία που παρέχεται από τις παρατηρήσεις. Για το λόγο αυτό έχουμε υποθέσει ότι το διάστημα δειγματοληψίας και το πλήθος των δειγμάτων μας δεν μεταβάλλονται. Οι κατανομές δειγματοληψίας για τις οποίες γίνεται λόγος είναι μέλη μιας ευρύτερης οικογένειας κατανομών, της κανονικής εκθετικής οικογένειας και σε κάθε ξεχωριστή περίπτωση μία εκτεταμένη ανάλυση έχει πραγματοποιηθεί. Στον καθορισμό του βέλτιστου μεγέθους δείγματος για κάθε ξεχωριστή περίπτωση κατανομής έχουμε χρησιμοποιήσει βασικές αρχές της θεωρίας αποφάσεων κάτι που αποτελεί κύριο εργαλείο αυτής της δουλείας.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## TABLE OF CONTENTS





ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## LIST OF TABLES





ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## LIST OF FIGURES





ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# Chapter 1

# Introduction

In an industrial type setting, products coming out of a process are subject to quality control. In other words each item produced has a measurable quality characteristic (univariate or multivariate) that needs to meet some specification criteria. This characteristic will have an ideal (target) value and some Specification Limits (Lower and/or Higher) that needs to meet or otherwise it will be considered as non-conformable. For example we might consider a process that produces light bulbs. These will have an ideal base-diameter of 1.2cm and will have as Lower Specification Limit (LSL) 1.15 cm and Upper Specification Limit (USL) 1.25cm. If the base-diameter of a light bulb is less than LSL then it will be loose and on the other hand if it is more than USL it will not fit. Ideally (in a dream world) we would produce all items with base-diameter equal to the target value. However, regardless of how well a suitable environment has been designed, it is natural a variability to exist as a result the quality of the raw material and of unavoidable causes during the operation of the process.

In the early stages of the Quality Science the quality characteristic of a product was measured with respect to its specification limits. Thus each product was of acceptable quality if its quality characteristic was in the [LSL, USL] and unacceptable otherwise.

A completely different paradigm arose by Shewhart (1931). He was the first to recognize the role of variability in the quality characteristic and he partitioned the variability in two terms:

a. Variation due to chance causes
b. Variation due to assignable causes



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chance causes are inherent in the process, while assignable causes (if they exist) can be traced to machine, material, worker etc. Operationally a system running under chance causes looks like a realization of a statistical model.

Based on the above, a process is considered to be under the In Control (IC) state if it runs only with chance causes of variation present. On the other hand a process running in the presence of assignable causes of variation is considered to be Out of Control (OOC).

The goal in quality control is to identify when our process moves from the In Control to the Out of Control state. Then corrective actions can be taken and/or quality improvement techniques can be applied. The control charts, originally developed by Shewhart (1931), is the standard approach towards the control of a process characteristic. Constructing a control chart involves a calculation of the Center Line and the (Lower and/or Upper) Control Limits (it is clear that the [LCL, UCL] will be inclusive to the [USL, LSL] indicating that we might have processes within specification but Out of Control). These control limits are chosen so that if the process is In Control, nearly all of the sample points will fall between them. The center line and control limits will be calculated from the early process observations. In other words we split our process in Phase I and Phase II. During Phase I (or so called start-up phase) we are sampling from our process to construct the "historic" data of the process which will be used to derive the center line and control limits, against which samples from phase II of the process will be tested for being In or Out of Control.

During the collection of Phase I data the original approach calls for sampling a fixed number of items with a fixed sampling interval until we have a reliable large number of historic data (samples) to robustly estimate the control chart lines. So in the light bulbs example we might decide to take a sample of 50 light bulbs (for which we will record the average –or some other statistic-) every 10 minutes and the samples drawn the first 3 hours (18 samples of size 50 each) will constitute the Phase I data.

In an attempt to improve the procedures for statistical process control, dynamic programming models have been developed which has led to what is called adaptive or dynamic control charts, introduced by Tagaras and Nikolaidis (2001). Their approach is based on the common characteristic of those charts which allows all three parameters, namely sample size, sampling interval and control limit



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

location, to change during the production, as the sample information becomes available and the state of the process is updated using Bayesian methods. Based on their economic performance, conclusions are being derived relative to the effectiveness of those charts.

A similar work in this area has been done by Tagaras (1994, 1996) where the one-sided control charts for variables is examined first and a theoretical formulation for the two sided adaptive control charts is provided. More recent publications of Calabrese (1995) where we consider a process control procedure with fixed sample sizes and sampling intervals and the fraction defective is the quality variable of interest and an optimal structure is given. Porteus and Angelus (1997) focus on describing how dynamic process control rules can improve static ones by canceling some of the inspections called for by an economically static rule when starting in control. Important early theoretical contributions can be found in Bather (1963), Taylor (1965), Girshik and Rubin (1952) and Carter (1972).

More recent papers are generally more practical. Parkhideh and Case (1989) propose an economic model where the decision parameters of a control chart (sample size, sampling interval and control limits) may change over time but a certain pattern is prespecified. Reynold (1989) studies the statistical properties of one-sided and two-sided Shewhart charts, when the sampling interval after each sample depends on the latest observation, but sample sizes and control limits are kept constant. Daudin (1992) proposes a chart where two samples are taken only from the process at fixed intervals but the second is analyzed only if the first does not suffice to decide if the process is in control. Pradhu, Runger and Keats (1993) and Costa (1994) independently examine the properties of $\bar{X}$ charts with two possible sample sizes, depending on the previous sample statistic. An extended work by considering for a particular cost structure but no prior information of how to choose the sample size which is economically most desirable is developed in V.Barnett (1974).

In this thesis we will assume that the sampling interval and the length of the Phase I is fixed and we will try to determine how many items must be sampled each time, in order to estimate the unknown parameters of interest and at the same time take into account the sampling cost for every observation (i.e. tie this to the economics of the problem). So in this thesis we will work towards determining the



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

optimal sample size for phase I samples, using the current sample information for estimating the unknown parameters along with available prior information in a Bayesian setting. Our approach will be based on properties of Statistical Decision theory.

Decision theory is concerned with the making of decisions in the presence of statistical knowledge which sheds light on some uncertainties involved in the decision problem. We assume that these uncertainties are unknown numerical quantities (parameters) that will commonly be noted by $\theta$ which will either be univariate or multivariate random variables. In our statistical procedures we assign to any parameter a density function (prior distribution) and an attempt is made to combine sample and prior information in order to make the best decision (estimation) about the unknown value of the parameter $\theta$.

In order to specify the optimal sample size to be taken in every different sample of Phase I, we define a loss that would be incurred for each possible decision and for the various possible values of the unknown parameter $\theta$. In all cases studied here we decided to use the square error loss function. We selected this loss function since it is well known to have some nice properties. Important role in the selection of the size of a random sample that needs to be taken has the sampling cost for a single observation as well. In this work the cost per observation is assumed to be constant and as a result the total sampling cost of a single sample will be a function of the size of the sample which is not affected by the value of the parameter $\theta$ or by other magnitudes of the observed values.

The sampling distributions considered in this work are members of a general class, namely the regular exponential family of distributions. More specifically exponential family includes as special cases: Poisson, Bernoulli, Binomial, Gamma, Exponential and Normal sampling distribution functions for the univariate case and Multivariate Normal for the multivariate case analogously. For each one of these models we will derive the optimal (desirable) number of observations that we suggest to be taken, sharing a number of decision theory properties.

This thesis is organized into 6 chapters. The first chapter is the introduction where a basic profile of our work is given. The second chapter provides a theoretical background of the decision theory approach, including definitions,



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

theorems, basic properties and some examples. The third chapter provides a technique for the calculation of the optimal sample size for the case of the general $k$-parameter exponential distribution family where a conjugate prior has been assigned to the unknown parameters. In the fourth chapter optimal sample size is calculated for specific univariate sampling distributions functions that belong to the class of the $k$-parameter exponential family. Furthermore, besides the mathematical part in each case, conclusions about the behavior of the suggested sample quantity accordingly to other specified parameters are derived. In the fifth chapter we derive optimal sample size in the case of Multivariate Normal distribution, with unknown mean vector and a special case for the Bivariate Normal distribution is distinguished. At the end a chapter of some basic conclusions and future research summarizes this work.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# Chapter 2

# Decision theory

## 2.1 A Statistical decision problem

Decision theory as the name implies, is concerned with the problem of making decisions. In statistics, the decision theory framework has a very important role, if we think of the consequences that we might have in the outcome of an experiment from taking a wrong decision. All of the known forms of inference (point estimation, hypothesis testing and interval estimation) can be seen as making of decisions. In this chapter we shall consider problems where decision theory provides a method of analyzing inference problems. The parameter $\theta$ is the true but unknown quantity about which we wish to make an inference. Bayesian analysis is considered to be the most sensible approach to statistical decision analysis where a parameter $\theta$ is a random quantity with a prior distribution function $\pi(\theta)$. In order to draw conclusions about $\theta$ we construct some criteria which are used to compare different decisions and finally lead to the optimal choice. It has been proved that if for a certain problem an optimal decision exists, then this can surly be given through the Bayesian approach.

In general we shall consider decision problems where the statistician has the opportunity to choose a decision after having observed the value of a random variable or vector which is related to the unknown parameter $\theta$. All the elements of a decision problem can be formally defined as follows:

1. The set of all possible values for the unknown parameter $\theta$ for which we wish to obtain information for, is called the parameter space $\Theta$.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

2. The set of **all possible actions under consideration, is the action space** denoted by $\mathcal{A}$.
3. A function of the form $L(\theta,\alpha):\Theta\times\mathcal{A}\to R$ **which expresses the loss that we have** if the true value of the unknown parameter is $\theta$ and an action $\alpha$ is taken, **is called loss function**
4. **The data (outcome of an experiment) are described by a** random vector $\underset{\sim}{x}=(x_1,....,x_n)$. **The set of all possible outcomes is the sample space** $X$.
5. **The sampling model with likelihood function conditioned on** $\theta$: $\{f(\underset{\sim}{x}|\theta):\theta\in\Theta\}$, where $\{f(x|\theta):\theta\in\Theta\}$ **is the probability density function on every observation** $x$.

We have to mention here that parameter space $\Theta$ **and action space** $\mathcal{A}$ **may be** continuous or discrete, finite or infinite. Instead of a loss function $L(\theta,\alpha)$ one may **speak of the utility function (De-Groot (1970), Berger (1985)) which is defined as** the negative of the loss function i.e. $U(\theta,\alpha)=-L(\theta,\alpha)$. Utility function expresses gain **rather than loss and is usually used in theory of economics and in some** branches **of Bayesian analysis.**

## 2.2 Decision rules

**Definition 2.1:** A decision rule also known as strategy is a function of the type $\delta(\underset{\sim}{x})$: $X\to\mathcal{A}$ which specifies what action $\alpha\in\mathcal{A}$ should be taken when $\underset{\sim}{x}\in X$ is observed. The set of all possible decision rules is denoted by $\mathcal{D}$, i.e $\mathcal{D}=\{\delta(\underset{\sim}{x}):X\to\mathcal{A}\}$. **Decision rules of this form are also called nonrandomized decision rules (Ferguson (1967), De Groot (1970), Berger (1985)).**

**Example 2.1:** Consider the two hypotheses testing problem:
$H_0:\theta\in\Theta_0$ and the alternative $H_1:\theta\in\Theta_0^c$. **The action space consists of only two** actions $\mathcal{A}=\{\alpha_0,\alpha_1\}$ where $\alpha_0$ denotes the action to "accept $H_0$" while $\alpha_1$ the



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

action "accept $H_1$" (or equivalently here "reject $H_0$"). Here the decision rule $\delta(\underset{\sim}{x}): X \to \mathcal{A}$

**takes only two possible values:**

- $\delta(\underset{\sim}{x}) = \alpha_0 \ \forall \underset{\sim}{x} \in \{\underset{\sim}{x} : \delta(\underset{\sim}{x}) = \alpha_0\} \equiv$**acceptance region of the test**
- $\delta(\underset{\sim}{x}) = \alpha_1 \ \forall \underset{\sim}{x} \in \{\underset{\sim}{x} : \delta(\underset{\sim}{x}) = \alpha_1\} \equiv$**rejection region of the test**

So the set of all allowable decision rules $\mathcal{D}$ consists of only two possible strategies, one that leads to action $\alpha_0$ and one that leads to action $\alpha_1$.

Usually the class $\mathcal{D}$ of all possible decision rules is quite large (uncountable in most of the cases) and the question of interest becomes which $\delta \in \mathcal{D}$ should we choose as the "best" strategy. Given that the decision rules will be judged through their loss function which is a random quantity (function of $\theta \in \Theta$) we need to define some deterministic ordering of all possible decision rules to be able to find the optimal.

In the following section we will refer to some criteria which will help us to compare different decision rules and will introduce certain ordering criteria.

## 2.3 Risk functions and admissibility

**Definition 2.2:** For every decision rule $\delta(\underset{\sim}{x}) \in \mathcal{D}$ and for a given value of $\theta \in \Theta$ we define the risk function of $\delta(\underset{\sim}{x})$ as the function $R(\theta,\delta): \Theta \to R$

$$R(\theta,\delta) = E_{\underset{\sim}{x}|\theta}\left(L(\theta,\delta)\right) = \int_X L(\theta,\delta)\, dF(\underset{\sim}{x}|\theta) = \int_X L(\theta,\delta)\, f(\underset{\sim}{x}|\theta)\, dx \quad (2.1)$$

(If our problem has no data then $R(\theta,\delta) = L(\theta,\delta)$) which is a function of $\theta$. This definition comes from a frequentist point of view .

Since $\theta$ is unknown we prefer a decision rule with risk function small for all $\theta \in \Theta$, so if two decision rules are to be compared this could be done only through their risk functions.

Let $\delta_1, \delta_2$ two decision rules then if

- $R(\theta,\delta_1) \le R(\theta,\delta_2)$ , $\forall \theta \in \Theta \Rightarrow \delta_1$ is as good as $\delta_2$
- $R(\theta,\delta_1) \le R(\theta,\delta_2)$ , $\forall \theta \in \Theta$ and $R(\theta,\delta_1) < R(\theta,\delta_2)$ for some $\theta \in \Theta \Rightarrow$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$\delta_1$ is better than $\delta_2$ , or $\delta_1$ is preferred to $\delta_2$.

- $R(\theta,\delta_1)=R(\theta,\delta_2)$ , $\forall\theta\in\Theta \Rightarrow \delta_1$ is equivalent to $\delta_2$

**Definition 2.3:** Given a decision rule $\delta(\underset{\sim}{x})\in \mathcal{D}$ and $\pi(\theta), \theta\in\Theta$ to be the prior distribution for the unknown random vector $\theta$ and we can derive the posterior distribution of $\theta$ after $\underset{\sim}{x}=(x_1,.....,x_n)$ has been observed $p(\theta|\underset{\sim}{x})$. We define the posterior risk, or the Bayesian expected loss (Berger (1985)) of the decision rule $\delta(\underset{\sim}{x})$ as a function of the form: $\rho(\pi,\delta(\underset{\sim}{x})): X\rightarrow R$

$$\rho(\pi,\delta(\underset{\sim}{x}))=E_{\theta|\underset{\sim}{x}}\left(L(\theta,\delta(\underset{\sim}{x}))\right)=\int_{\Theta}L(\theta,\delta(\underset{\sim}{x}))\,p(\theta|\underset{\sim}{x})d\theta \qquad (2.2)$$

**If we refereed to a no-data problem**

$$\rho(\pi,\delta)=E_{\theta}\left(L(\theta,\delta)\right)=\int_{\Theta}L(\theta,\delta)\pi(\theta)d\theta \qquad (2.3)$$

Posterior risk is a single number for a given $\underset{\sim}{x}$ regardless of the dimension of $\theta$. **Obviously a desirable decision rule is the one that minimizes the posterior risk for** a given prior $\pi(\theta)$.

Thus if we have to compare two given decision rules $\delta_1(\underset{\sim}{x})$ and $\delta_2(\underset{\sim}{x})$ it is allowable to use one of the two risk functions defined above, conditioning on $\theta\in\Theta$ if we use the frequentist risk function or conditioning on data $\underset{\sim}{x}\in X$ if the posterior risk function is used. Of course both risk functions are random quantities and in order to be able to compare the decision rules we need to have one the two to have uniformly smaller risk function (over all $\theta\in\Theta$ if frequentist risk is used or over all $\underset{\sim}{x}\in X$ if posterior risk is used). Otherwise we can not compare them using risk functions.

**Definition 2.4:** A decision rule $\delta(\underset{\sim}{x})\in \mathcal{D}$ is called admissible if there does not exist any other decision rule $\delta'(\underset{\sim}{x})\in \mathcal{D}$ better than $\delta(\underset{\sim}{x})\in\mathcal{D}$ i.e. $\forall\delta'\in\mathcal{D}-\{\delta\}$

$R(\theta,\delta)\le R(\theta,\delta')$ $\forall\theta\in\Theta$ and $R(\theta,\delta)<R(\theta,\delta')$ for some $\theta\in\Theta$.

A decision rule $\delta(\underset{\sim}{x})\in\mathcal{D}$ is called inadmissible if there exists $\delta'(\underset{\sim}{x})\in\mathcal{D}$ that is better than $\delta(\underset{\sim}{x})$.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

It is obvious that inadmissible decision rules should not be used when a decision rule with smaller risk function exists. On the other hand in many problems there might be a large class of admissible rules and then we are faced with the problem which one to select.

**Example 2.2:** Assume $x|\theta \sim N(\theta,1)$, and that it is desired to estimate $\theta$ under the loss function $L(\theta,\alpha)=(\theta-\alpha)^2$. We will consider decision rules of the form $\delta_c(x)=x+c$ where $c\geq 0$. Clearly we can compute the risk function as

$$\begin{aligned} R(\theta,\delta_c) &= E_{x|\theta}L(\theta-\delta_c) = E_{x|\theta}(\theta-\delta_c)^2 = E_{x|\theta}(\theta-x-c)^2 \\ &= E_{x|\theta}(\theta-c)^2 - 2(\theta-c)E_{x|\theta}(x) + E_{x|\theta}(x^2) \\ &= (\theta-c)^2 - 2(\theta-c)\theta + (1+\theta^2) \\ &= \theta^2 - 2\theta c + c^2 - 2\theta^2 + 2\theta c + 1 + \theta^2 \\ &= c^2 + 1 \end{aligned}$$

If we want to compare two different decision rules that belong in this class $\delta_c$, $\delta_{c'}$ all we have to do is to compare their risk functions. Note that if $c<c'$ then $c^2+1<c'^2+1 \Rightarrow R(\theta,\delta_c)<R(\theta,\delta_{c'})$ for all values of $\theta$. Hence, $\delta_c$ is a better estimator than $\delta_{c'}$. For $c=0$ the minimum value of $c^2+1$ is achieved. So the estimator $\delta_c=x$ is the best estimator in this class or else we can say that in this class of estimators $\delta_c=x$ is the admissible one.

**Definition 2.5:** Let $C$ be a class of decision rules that it is a subclass of the set of all allowable decision rules $\mathcal{D}$. $C$ will be called a complete class if for every decision rule $\delta' \notin C$ there exists $\delta \in C$ which is better than $\delta'$.

**Theorem 2.1:** If $C$ is a complete class then all the admissible decision rules will be contained in $C$.

**Proof:**

Assume that $\exists\delta'$ admissible and $\delta' \notin C$, then because of the fact that $C$ is a complete class $\exists\delta'' \in C$ such that $\delta''$ will be better than $\delta'$. This contradicts the fact that $\delta'$ is admissible so no better rule than $\delta'$ exists. Therefore $\delta' \in C$..



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## 2.4 Bayes rules

Sometimes it is not easy for us to compare decision rules only through their risk functions or equivalently through their posterior risk functions because both of them are random quantities. Thus the only case where two decision rules can be compared is when one is uniformly better than the other (for all $\theta \in \Theta$ in the first and $\forall \underset{\sim}{x} \in X$ in the second). In practice though we are interested in being able "order" all available decision rules and pick the "best" decision rule. In the bibliography one of the ways of ordering the decision rules is done through their expected risk.

More precisely for the frequentist risk function let us consider the prior distribution function $\pi(\theta)$, then the risk function of a decision rule $\delta$ is summarized by the average risk function called Bayes risk with respect to the prior $\pi(\theta)$. A more formal definition is as follows:

Definition 2.6: For every decision rule $\delta(\underset{\sim}{x}) \in \mathcal{D}$ we can define its Bayes risk $r(\pi,\delta)$ as the expected risk function with respect to the prior distribution $\pi(\theta)$

$$r(\pi,\delta) = E_{\pi}\left[R(\theta,\delta)\right] \qquad (2.4)$$

For the discrete case $$r(\pi,\delta) = \sum_{i=1}^{n} R(\theta_i,\delta)\pi(\theta_i) \qquad (2.5)$$

For the continuous case $$r(\pi,\delta) = \int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta \qquad (2.6)$$

Suppose that we are given:

$$r(\pi,\delta) = E_{\pi}\left[R(\theta,\delta)\right] = \int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta = \int_{\Theta}\left[\int_{X} L(\theta,\delta) f(\underset{\sim}{x}\,|\,\theta)dx\right]\pi(\theta)d\theta$$

Under suitable regularity conditions we can apply Fubini's theorem (see for example Chang (1974)) and reverse the order of integration to obtain:

$$r(\pi,\delta) = \int_{\Theta}\left[\int_{X} L(\theta,\delta) f(\underset{\sim}{x}\,|\,\theta)dx\right]\pi(\theta)d\theta = \int_{\Theta}\int_{X} L(\theta,\delta) f(\underset{\sim}{x}\,|\,\theta)\pi(\theta)dxd\theta$$
$$= \int_{\Theta}\int_{X} L(\theta,\delta) p(\theta\,|\,\underset{\sim}{x}) f(\underset{\sim}{x})dxd\theta = \int_{X}\int_{\Theta} L(\theta,\delta) p(\theta\,|\,\underset{\sim}{x}) f(\underset{\sim}{x})d\theta dx$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$= \int_X \left[ \int_\Theta L(\theta,\delta)\, p(\theta \mid \underset{\sim}{x})\, d\theta \right] f(\underset{\sim}{x})\, dx = \int_{x,X} \rho(\pi(\theta),\delta(\underset{\sim}{x}))\, f(\underset{\sim}{x})\, dx$$

$$= E_{\underset{\sim}{x}}\left[\rho(\pi(\theta),\delta(\underset{\sim}{x}))\right] \qquad (2.7)$$

From the above expression it has been proved that Bayes risk can be expressed as the expected posterior risk $\rho(\pi,\delta(\underset{\sim}{x}))$ with respect to the marginal distribution of the data $f(\underset{\sim}{x})$. This is the expected posterior loss before even having observed the data and it may also be called **preposterior** risk (Carlin and T. A. Louis (1996)).

## 2.5 Bayes risk principle

Since Bayes risk is a scalar quantity we can order the decision rules and prefer a decision rule $\delta_1$ instead of a decision rule $\delta_2$ if $r(p,\delta_1) < r(p,\delta_2)$.

**Definition 2.7:** Among all possible decision rules in $\mathcal{D}$, the "best" one according to the Bayes risk principle with respect to the prior $\pi(\theta)$, is the one that minimizes the Bayes risk; is called Bayes rule and is denoted $\delta^*(\underset{\sim}{x})$ i.e. it is a decision rule that satisfies:

$$r(\pi,\delta^*) = \min_{\delta\in D} r(\pi,\delta) \qquad (2.8)$$

**Theorem 2.2:** Under very broad conditions Bayes rule can be simply described as the decision function $\delta^*(\underset{\sim}{x}): X \to A$ which minimizes the corresponding posterior risk $\rho(\pi(\theta),\delta(\underset{\sim}{x}))$.

**Proof:**

We have already seen from the relation (2.7) that Bayes risk is equivalent to the expected posterior risk $\rho(\pi(\theta),\delta(\underset{\sim}{x}))$ with respect to the marginal distribution of the data $f(\underset{\sim}{x})$. From (2.8) we have:

$$r(\pi,\delta^*) = \min_{\delta\in D} r(\pi,\delta) = \min_{\delta\in D} E_{\underset{\sim}{x}} \rho(\pi(\theta),\delta(\underset{\sim}{x}))$$

$$= \min_{\delta\in D} \int_X \rho(\pi(\theta),\delta(\underset{\sim}{x}))\, f(\underset{\sim}{x})\, dx = \int_X f(\underset{\sim}{x}) \min_{\delta\in D} \rho(\pi(\theta),\delta(\underset{\sim}{x}))\, dx$$

$$= \int_X f(\underset{\sim}{x}) \left[ \min_{\delta\in D} \int_\Theta L(\theta,\delta)\, p(\theta \mid \underset{\sim}{x})\, d\theta \right] dx$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Therefore a risk function that minimizes Bayes risk can be found by minimizing the inner integral (posterior risk) for every $\underset{\sim}{x} \in X$.

## 2.6 An Example

**As an illustration of the results that have been developed so far let us consider the** statistical decision problem where $\Theta = \{\theta_1, \theta_2\}$, $A = \{\alpha_1, \alpha_2, \alpha_3\}$ with the loss function **given by the following table:**

| | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ |
|---|---|---|---|
| $\theta_1$ | 0 | 10 | 3 |
| $\theta_2$ | 10 | 0 | 3 |

Table 2.6.1: The loss function for every couple $(\theta_i, \alpha_j)$

Suppose that an observation $x$ is available from the sample space $X = \{0,1\}$ and the **conditional probabilities of the random variable $x$ are given by:**

$$P(x=1 \mid \theta = \theta_1) = \frac{3}{4} \qquad P(x=0 \mid \theta = \theta_1) = \frac{1}{4}$$

$$P(x=1 \mid \theta = \theta_2) = \frac{1}{4} \qquad P(x=0 \mid \theta = \theta_2) = \frac{3}{4}$$

We consider as a prior distribution function for $\theta \in \{\theta_1, \theta_2\}$

$\pi(\theta = \theta_1) = \pi, \pi(\theta = \theta_2) = 1 - \pi$, $0 \le \pi \le 1$

We wish to derive a Bayes decision function against each value of $\pi, (0 \le \pi \le 1)$ and sketch the Bayes risk as a function of $\pi$.

**Case I**

**At first we assume that no observation $x$ is made (no-data problem). We derive the Bayesian expected loss (Bayes risk) for all the possible decisions strategies** $d_1 = \alpha_1$, $d_2 = \alpha_2$, $d_3 = \alpha_3$ $\forall \theta \in \{\theta_1, \theta_2\}$:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\rho(\pi(\theta),d_1) = E_\theta\left(L(\theta,d_1)\right) = \sum_{i=1}^{2} L(\theta_i,d_1)\pi(\theta_i) = 10(1-\pi)$$

$$\rho(\pi(\theta),d_2) = E_\theta\left(L(\theta,d_2)\right) = \sum_{i=1}^{2} L(\theta_i,d_2)\pi(\theta_i) = 10\pi$$

$$\rho(\pi(\theta),d_3) = E_\theta\left(L(\theta,d_3)\right) = \sum_{i=1}^{2} L(\theta_i,d_3)\pi(\theta_i) = 3\pi + 3(1-\pi) = 3$$

- If $0 \le \pi < \frac{3}{10} \Rightarrow d_2$ is the Bayes decision rule
- If $\pi = \frac{3}{10} \Rightarrow d_3$, $d_2$ are both Bayes decision rules
- If $\frac{3}{10} < \pi < \frac{7}{10} \Rightarrow d_3$ is the Bayes decision rule
- If $\pi = \frac{7}{10} \Rightarrow d_3, d_1$ are both Bayes decision rules
- If $\frac{7}{10} < \pi \le 1 \Rightarrow d_1$ is the Bayes decision rule

**Hence, all the above results can be summarized into the following diagram**

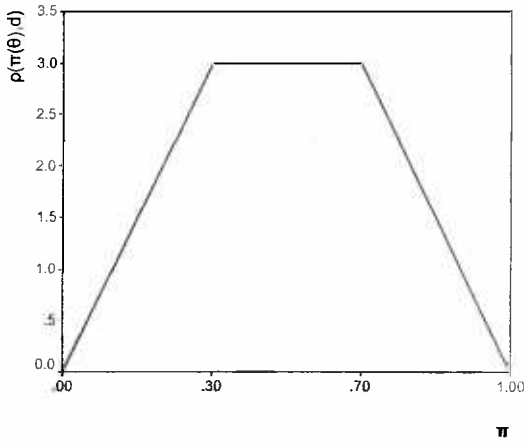


**Figure 2.6.1: Bayes risk function against the prior in the** no-data **case**

Figure 2.6.1 reveals the Bayes risk function $\rho(\pi(\theta),d)$ as it has already been defined against the prior probability $\pi$. It can be easily seen that $\rho(\pi(\theta),d)$ increases and decreases linearly for $0 \le \pi < \frac{3}{10}$ and for $\frac{7}{10} < \pi \le 1$ equivalently, except for $\frac{3}{10} \le \pi \le \frac{7}{10}$ where $\rho(\pi(\theta),d)$ remains stable.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**Case II**

We proceed now to the case where an observation $x$ is made before choosing the optimal decision. Let $p(\theta|x)$ denote the posterior probability (an updated version of the prior $\pi$) that $\theta=\theta_1$ if the value $x\in\{0,1\}$ has been observed i.e.

$P(\theta=\theta_1|X=x)$ and $1-P(\theta=\theta_1|X=x)=P(\theta=\theta_2|X=x)$

**We derive the posterior probabilities for every possible value of $x$:**

$$P(\theta=\theta_1|x=0)=\frac{P(x=0|\theta=\theta_1)P(\theta_1)}{\sum_{i=1}^{2}P(x=0|\theta=\theta_i)P(\theta_i)}=\frac{\frac{\pi}{4}}{\frac{\pi}{4}+\frac{3}{4}(1-\pi)}=\frac{\pi}{-2\pi+3}$$

$$P(\theta=\theta_1|x=1)=\frac{P(x=1|\theta=\theta_1)P(\theta_1)}{\sum_{i=1}^{2}P(x=1|\theta=\theta_i)P(\theta_i)}=\frac{\frac{3\pi}{4}}{\frac{3\pi}{4}+\frac{(1-\pi)}{4}}=\frac{3\pi}{2\pi+1}$$

$$P(\theta=\theta_2|x=0)=1-P(\theta=\theta_1|x=0)=1-\frac{\pi}{-2\pi+3}=\frac{3(1-\pi)}{-2\pi+3}$$

$$P(\theta=\theta_2|x=1)=1-P(\theta=\theta_1|x=1)=1-\frac{3\pi}{2\pi+1}=\frac{1-\pi}{2\pi+1}$$

Calculation of the posterior risks for each possible value of $x\in\{0,1\}$ yields

$$\rho(\pi(\theta),d_1(0))=\sum_{i=1}^{2}L(\theta_i,d_1)P(\theta=\theta_i|X=0)=0P(\theta=\theta_1|X=0)+10(1-P(\theta=\theta_1|X=0))$$
$$=10(1-P(\theta=\theta_1|X=0))$$

$$\rho(\pi(\theta),d_1(1))=\sum_{i=1}^{2}L(\theta_i,d_1)P(\theta=\theta_i|X=1)=0P(\theta=\theta_1|X=1)+10(1-P(\theta=\theta_1|X=1))$$
$$=10(1-P(\theta=\theta_1|X=1))$$

$$\rho(\pi(\theta),d_2(0))=\sum_{i=1}^{2}L(\theta_i,d_2)P(\theta=\theta_i|X=0)=10P(\theta=\theta_1|X=0)+0(1-P(\theta=\theta_1|X=0))$$
$$=10P(\theta=\theta_1|X=0)$$

$$\rho(\pi(\theta),d_2(1))=\sum_{i=1}^{2}L(\theta_i,d_2)P(\theta=\theta_i|X=1)=10P(\theta=\theta_1|X=1)+0(1-P(\theta=\theta_1|X=1))$$
$$=10P(\theta=\theta_1|X=1)$$

$$\rho(\pi(\theta),d_3(0))=\sum_{i=1}^{2}L(\theta_i,d_3)P(\theta=\theta_i|X=0)=3P(\theta=\theta_1|X=0)+3(1-P(\theta=\theta_1|X=0))=3$$

$$\rho(\pi(\theta),d_3(1))=\sum_{i=1}^{2}L(\theta_i,d_3)P(\theta=\theta_i|X=1)=3P(\theta=\theta_1|X=1)+3(1-P(\theta=\theta_1|X=1))=3$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

So in the end **we summarize regardless of the value of the observation** $x$ **the following:**

$$\begin{cases} \rho(\pi, d_1(x)) = 10(1 - P(\theta = \theta_1 \mid X = x)) = 10P(\theta = \theta_2 \mid X = x) \\ \rho(\pi, d_2(x)) = 10P(\theta = \theta_1 \mid X = x) \\ \rho(\pi, d_3(x)) = 3 \end{cases}$$

**It has been proved that Bayes risk minimizes the posterior risk function for every** $x \in \{0,1\}$. According to this we will have:

- If $0 \le P(\theta = \theta_1 \mid X = x) < \frac{3}{10} \Rightarrow 0 \le \pi < \frac{9}{16}$ when $x = 0$ is observed and $0 \le \pi < \frac{1}{8}$ when $x = 1$. Then $d_2$ is the Bayes decision function.
- If $P(\theta = \theta_1 \mid X = x) = \frac{3}{10} \Rightarrow \pi = \frac{9}{16}$ when $x = 0$ is observed and $\pi = \frac{1}{8}$ when $x = 1$ so both $d_3, d_2$ are Bayes decision rules.
- If $\frac{3}{10} < P(\theta = \theta_1 \mid X = x) < \frac{7}{10} \Rightarrow \frac{9}{16} < \pi < \frac{7}{8}$ when value $x = 0$ is observed and $\frac{1}{8} < \pi < \frac{7}{16}$ when $x = 1$. Then $d_3$ is the Bayes rule.
- If $P(\theta = \theta_1 \mid X = x) = \frac{7}{10} \Rightarrow \pi = \frac{7}{8}$ when value $x = 0$ is observed and $\pi = \frac{7}{16}$ when $x = 1$. Then both $d_1, d_3$ are Bayes decision rules.
- If $\frac{7}{10} < P(\theta = \theta_1 \mid X = x) \le 1 \Rightarrow \frac{7}{8} < \pi \le 1$ when value $x = 0$ is observed and $\frac{7}{16} < \pi \le 1$ when $x = 1 \Rightarrow d_1$ **is the Bayes decision function.**

We shall know compute Bayes risk for any given prior probability $\pi$:

- For $0 \le \pi \le \frac{1}{8}$ $d_2$ is the Bayes rule regardless of $x$ and $r(\pi, d_2) = \rho(\pi(\theta), d_2) = 10\pi$ is the Bayes risk.
- For $\frac{7}{8} \le \pi \le 1$ $d_1$ is the Bayes rule regardless of $x$ and $r(\pi, d_1) = \rho(\pi(\theta), d_1) = 10(1 - \pi)$ is the Bayes **risk**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- For $\frac{1}{8}<\pi<\frac{7}{16}$ $d_2$ is the Bayes rule when $x=0$ and $d_3$ is the Bayes rule when $x=1$, $r(\pi,d)=\pi(10\frac{1}{4}+3\frac{3}{4})+(1-\pi)(0\frac{3}{4}+3\frac{1}{4})=4\pi+\frac{3}{4}$ is the Bayes risk .
- For $\frac{7}{16}<\pi<\frac{9}{16}$ $d_2$ is the Bayes rule when $x=0$ and $d_1$ is the Bayes rule when $x=1$

  $r(\pi,d)=\pi(10\frac{1}{4}+0\frac{3}{4})+(1-\pi)(0\frac{3}{4}+10\frac{1}{4})=\frac{5}{2}$ is the Bayes risk.
- For $\frac{9}{16}<\pi<\frac{7}{8}$ $d_3$ is the Bayes rule when $x=0$ and $d_1$ is the Bayes rule when when $x=1$

  $r(\pi,d)=\pi(3\frac{1}{4}+0\frac{3}{4})+(1-\pi)(3\frac{3}{4}+10\frac{1}{4})=-4\pi+\frac{19}{4}$ is the Bayes risk.

**As it has already been done for the** no-data case we present all the above **conclusions into the following diagram.**

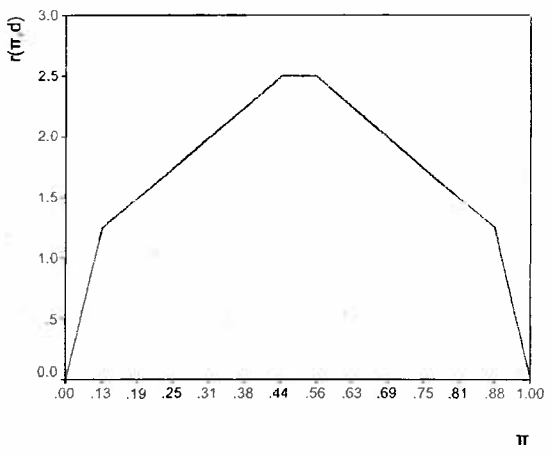


**Figure 2.6.2: The Bayes risk function against the prior when some data have been collected**

**In Figure 2.6.2 Bayes risk function is sketched against the prior probability** $\pi$ **in the case where we have the opportunity to observe a random variable** $x$. Again we can clearly see that the Bayes risk function $r(\pi,d)$ has a linear behavior against $\pi$. From Figures 2.6.1, 2.6.2 it can be seen that for any prior probability



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$\pi$ in either of these intervals $0 \le \pi \le \frac{1}{8}$ or $\frac{7}{8} \le \pi \le 1$ one can do just as well without any observation as he can by observing $x$ ( $\rho(\pi(\theta), d) = r(\pi, d)$ ). If $\frac{7}{16} \le \pi \le \frac{9}{16}$ then $r(\pi, d) < \rho(\pi(\theta), d)$ $(2.5 < 3)$.

**The method which has been described here for the construction of a Bayes decision function is called the extensive form of analysis ( by Raiffa and Schlaifer 1961) .**

## 2.7 Admissibility of Bayes rules

**In the following theorem we are going to prove that under specific assumptions on the prior distribution, the Bayes rule is an admissible rule.**

**Theorem 2.3:** Suppose that we have a statistical decision problem with parameter space $\Theta \subseteq \mathbb{R}$ and sample space $X$. Given any decision rule $\delta \in \mathcal{D}$, **assume that the** frequentist risk $R(\theta, \delta)$ **is a continuous function of** $\theta$ **with respect to the likelihood** $f(\underset{\sim}{x} | \theta)$. Furthermore, if $\pi(\theta)$ reflects our prior belief for $\theta$ and assume that for every $\theta \in \Theta$ and for every $\varepsilon > 0$ the interval $(\theta - \varepsilon, \theta + \varepsilon)$ **has a positive** probability under the prior distribution $\pi(\theta)$. If $\delta^*$ is a Bayes rule (or Bayes estimator) and $-\infty < r(\pi, \delta^*) < \infty$ with respect to the prior $\pi(\theta)$, then $\delta^*$ **admissible.**

Proof:

Let us assume that $\delta^*$ is inadmissible, then according to the earlier definitions, there exists a decision rule $\delta' \in \mathcal{D}$ which is better than $\delta^* \in \mathcal{D}$ i.e. $R(\theta, \delta') \le R(\theta, \delta^*)$ $\forall \theta \in \Theta$ and $R(\theta_1, \delta') < R(\theta_1, \delta^*)$ for $\theta_1 \in \Theta$. Let $R(\theta_1, \delta^*) - R(\theta_1, \delta') = \mu > 0$ , since $R(\theta, \delta)$ is a continuous function of $\theta$ for every $\delta \in \mathcal{D} \Rightarrow R(\theta, \delta^*) - R(\theta, \delta')$ is also continuous, then $\exists\, \varepsilon > 0$ such that



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$R(\theta,\delta^*)-R(\theta,\delta')>\frac{\mu}{2},\ \forall\theta\in(\theta_1-\varepsilon,\theta_1+\varepsilon)$$

$$\Rightarrow r(\pi,\delta^*)-r(\pi,\delta')=\int_{-\infty}^{\infty}R(\theta,\delta')\pi(\theta)d\theta-\int_{-\infty}^{\infty}R(\theta,\delta')\pi(\theta)d\theta=\int_{-\infty}^{\infty}\left[R(\theta,\delta^*)-R(\theta,\delta')\right]\pi(\theta)d\theta$$

**Then:**

$$\int_{\theta_1-\varepsilon}^{\theta_1+\varepsilon}\left[R(\theta,\delta^*)-R(\theta,\delta')\right]\pi(\theta)d\theta\geq\int_{\theta_1-\varepsilon}^{\theta_1+\varepsilon}\frac{\mu}{2}\pi(\theta)d\theta\geq\frac{\mu}{2}\int_{\theta_1-\varepsilon}^{\theta_1+\varepsilon}p(\theta)d\theta\geq\frac{\mu}{2}>0$$

$$\Rightarrow r(\pi,\delta^*)-r(\pi,\delta')>0\Leftrightarrow r(\pi,\delta')<r(\pi,\delta^*)$$

**This last inequality contradicts the definition that Bayes rule minimizes Bayes risk** and so $\delta^*$ is admissible.

Wald (1950) proved the converse result, where every admissible decision rule is **also a Bayes rule with respect to some prior distribution under certain conditions. In general Bayes rules can not be admissible if their Bayes risks** $r(\pi,\delta)=E_\theta\left[R(\theta,\delta)\right]$ are infinite. If we choose a proper prior $\pi(\theta)$ then $r(\pi,\delta)<\infty$ and $r(\pi,\delta)$ is surly the Bayes risk. If prior $\pi(\theta)$ is improper then $r(\pi,\delta)=\infty$ and **Bayes rules are inadmissible. As it has already been defined all admissible decision rules are contained in a complete class. Since under general conditions every Bayes rule is also admissible, then Bayes rules are usually contained in a complete class. It would be desirable the set of all Bayes rules to form a complete class.**

## 2.8 Point estimation problem

A point estimation problem is a statistical decision problem where the decision to be made is the estimator of the unknown parameter $\theta\in\Theta$. So all the possible decisions $\delta\in\mathcal{D}$ are possible values of $\theta\in\Theta$ and decision space $\mathcal{D}$ coincides with $\Theta$ and for reasons of simplicity we shall assume that $\Theta=\mathcal{D}$. In case that the unknown parameter $\theta$ is univariate we have $\Theta=\mathcal{D}=\mathbb{R}$.

**The most common form of the loss functions used in the univariate case are of the** form $L(\theta,\delta)=\alpha|\theta-\delta|^\beta$ where $\alpha>0,\ \beta>0$

For $\alpha=1,\beta=2\Rightarrow L(\theta,\delta)=(\theta-\delta)^2$, which is known as a square error loss function



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

For $\alpha=1,\beta=1 \Rightarrow L(\theta,\delta)=|\theta-\delta|$, which is known as an absolute error loss function

**Square error loss and absolute error loss are the most popular loss functions in a point estimation problem. In both, the loss occurred, increases** (quadratic and linearly) as the discrepancy of $\delta$ from $\theta$ **increases. Furthermore both are symmetric, penalizing under and over estimation similarly. The absolute error loss** penalizes more than the square error loss in $(-1,1)$ while the opposite holds in $(-\infty,-1)\cup(1,+\infty)$. The risk function $R(\theta,\delta)$ if we will use square error loss then takes the form:

$$R(\theta,\delta)=E_{\underset{\sim}{x}|\theta}\left[L(\theta,\delta)\right]=E_{\underset{\sim}{x}|\theta}(\theta-\delta)^2=\text{var}_{\underset{\sim}{x}|\theta}(\theta-\delta)+\left[E_{\underset{\sim}{x}|\theta}(\theta-\delta)\right]^2$$
$$=\text{var}_{\underset{\sim}{x}|\theta}\,\delta+(\theta-E_{\underset{\sim}{x}|\theta}(\delta))^2=\text{var}_{\underset{\sim}{x}|\theta}\,\delta+\left(Bias_{\cdot|\theta}(\delta)\right)^2=MSE(\theta)$$

So the risk function under square error loss is simply the known mean squared error. From this last expression of the risk function a good estimator $\delta$ of $\theta$ must have small variance combined with a small (usually nonzero) bias. If we were restricted to the class of all unbiased estimators the optimal estimator for this class would be found by minimizing only the variance. Because of the fact that variance and bias are contained in this expression of risk function an ideal estimator has to minimize simultaneously this quantities.

**Theorem 2.4:** In a point estimation problem where $\theta\in\mathbb{R}$ is the unknown real valued parameter and square error loss function is used $L(\theta,\delta)=(\theta-\delta)^2$, then Bayes rule is proved to be the posterior mean i.e. $\delta^*=E(\theta|\underset{\sim}{x})$ and

Bayes risk for $\delta^*\in\mathbb{R}$ will be given by the expected posterior variance with respect to the marginal distribution of the data i.e. $r(\pi,\delta^*)=E_{\underset{\sim}{x}}\left[\text{var}(\theta|\underset{\sim}{x})\right]$

**Proof:**

Bayes rule is found by minimizing the posterior risk

$$\frac{\partial}{\partial\delta}\rho(\pi,\delta)=\frac{\partial}{\partial\delta}\left(\int_\Theta L(\theta,\delta)\,p(\theta|\underset{\sim}{x})d\theta\right)=\int_\Theta\frac{\partial}{\partial\delta}(\theta-\delta)^2\,p(\theta|\underset{\sim}{x})d\theta=\int_\Theta -2(\theta-\delta)p(\theta/\underset{\sim}{x})d\theta$$
$$\Rightarrow\int_\Theta -2(\theta-\delta)p(\theta|\underset{\sim}{x})d\theta=0\Rightarrow\delta\int_\Theta p(\theta|\underset{\sim}{x})d\theta=\int_\Theta\theta\,p(\theta|\underset{\sim}{x})d\theta$$
$$\Rightarrow\delta^*=E(\theta|\underset{\sim}{x})$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

In order to show that in $\delta^* = E(\theta|\underset{\sim}{x})$ we have minimum we need to take the second derivative in $\delta^*$ as well and show that it is positive. For every possible decision rule $\delta \in \mathcal{D}$ we will have

$$\frac{\partial^2}{\partial\delta^2}\rho(\pi,\delta) = \frac{\partial}{\partial\delta}\left(\frac{\partial}{\partial\delta}\int_{\Theta} L(\theta,\delta)\,p(\theta|\underset{\sim}{x})d\theta\right) = \frac{\partial}{\partial\delta}\left(\int_{\Theta}\frac{\partial}{\partial\delta}(\theta-\delta)^2\,p(\theta|\underset{\sim}{x})d\theta\right)$$

$$= \frac{\partial}{\partial\delta}\left(\int_{\Theta}-2(\theta-\delta)p(\theta|\underset{\sim}{x})d\theta\right) = \int_{\Theta}\frac{\partial}{\partial\delta}\left(-2(\theta-\delta)\,p(\theta|\underset{\sim}{x})\right)d\theta$$

$$= \int_{\Theta} 2\,p(\theta|\underset{\sim}{x})d\theta = 2\int_{\Theta} p(\theta|\underset{\sim}{x})d\theta = 2 > 0$$

i.e. $\delta^* = E(\theta|\underset{\sim}{x})$ is the Bayes rule for the square error loss and the corresponding **Bayes risk is given by**

$$r(\pi,\delta^*) = E_{\underset{\sim}{x}}\left[\rho(\pi,\delta^*)\right] = E_{\underset{\sim}{x}}\left[E_{\theta|\underset{\sim}{x}}\left(\theta-\delta^*\right)^2\right] = E_{\underset{\sim}{x}}\left[E_{\theta|\underset{\sim}{x}}\left(\theta-E(\theta|\underset{\sim}{x})\right)^2\right]$$

$$= E_{\underset{\sim}{x}}\left[\mathrm{var}(\theta|\underset{\sim}{x})\right]$$

Theorem 2.5: In a point estimation problem where $\theta \in \mathbb{R}$ **is the unknown real** valued parameter and the absolute error loss function is used $L(\theta,\delta) = |\theta-\delta|$, then **Bayes rule is the median of the posterior distribution**

**Proof:**

Let $m$ denotes the median of the posterior $p(\theta|\underset{\sim}{x})$ and $\delta > m$ is another decision **rule. Then**

$$L(\theta,m) - L(\theta,\delta) = \begin{cases} m-\delta, & \theta \le m \\ 2\theta-(m+\delta), & m < \theta < \delta \\ \delta-m, & \theta \ge \delta \end{cases}$$ **from which it follows that**

$L(\theta,m) - L(\theta,\delta) \le (m-\delta)I_{(-\infty,m]}(\theta) + (\delta-m)I_{(m,\infty)}(\theta)$. Since $m$ is the median of the posterior we have $p(\theta \le m/\underset{\sim}{x}) \ge \frac{1}{2}$, so that $p(\theta > m/\underset{\sim}{x}) \le \frac{1}{2}$. Then,

$$E_{\theta|\underset{\sim}{x}}\left(L(\theta,m)-L(\theta,\delta)\right) \le (m-\delta)p(\theta \le m|\underset{\sim}{x}) + (\delta-m)p(\theta > m|\underset{\sim}{x}) \Rightarrow$$

$$\Rightarrow E_{\theta|\underset{\sim}{x}}\left(L(\theta,m)-L(\theta,\delta)\right) \le (m-\delta)\frac{1}{2} + (\delta-m)\frac{1}{2} \Rightarrow$$

$$\Rightarrow E_{\theta|\underset{\sim}{x}}\left(L(\theta,m)-L(\theta,\delta)\right) \le 0 \Rightarrow E_{\theta|\underset{\sim}{x}}L(\theta,m) \le E_{\theta|\underset{\sim}{x}}L(\theta,\delta) \Rightarrow \rho(\pi,m) \le \rho(\pi,\delta)$$

we end up to the conclusion that $m$ has posterior risk at least as small as any decision rule $\delta > m$ . A similar proof holds if $\delta < m$. So in general for every



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

decision rule $\delta$, median $m$ of $p(\theta|\underset{\sim}{x})$ has the smaller posterior risk if absolute error loss function is used **therefore the posterior median** $m$ **is the Bayes rule in** this case.

We described before the case where the unknown parameter $\theta \in \Theta$ is real -valued. If $\underset{\sim}{\theta}$ is a random vector i.e. $\underset{\sim}{\theta} = (\theta_1, ....., \theta_n)'$ then $\Theta$ will be a subset of $\mathbb{R}^n$ and the estimator of $\underset{\sim}{\theta}$ in a statistical decision problem is a decision $\underset{\sim}{\delta} \in \mathcal{D}$ of the form $\underset{\sim}{\delta} = (\delta_1, ....., \delta_n)'$ where $\Theta \subseteq \mathcal{D}$. Again for reasons of simplicity we shall assume that $\Theta = \mathcal{D} = \mathbb{R}^n$. **In this case the loss function is often assumed** to have the following form:

$$L(\underset{\sim}{\theta}, \underset{\sim}{\delta}) = \beta(\underset{\sim}{\theta}) \mathrm{A}(\underset{\sim}{\theta} - \underset{\sim}{\delta})$$

Where A is a nonnegative function of the vector $\underset{\sim}{\theta} - \underset{\sim}{\delta}$ such that $\mathrm{A}(0) = 0$ and $\beta(\underset{\sim}{\theta})$ is a nonnegative weighting function of $\underset{\sim}{\theta}$.

There may be problems where not all components of the random vector $\underset{\sim}{\theta}$ need to **be estimated. Let us assume that the first** $k$ **components** from $\underset{\sim}{\theta} = (\theta_1, ....., \theta_n)$ are to be estimated but it is not required the remaining $n-k$ components of $\underset{\sim}{\theta}$ to be estimated. In this situation the last $n-k$ components are called nuisance parameters and $L(\underset{\sim}{\theta}, \underset{\sim}{\delta}) = \beta(\underset{\sim}{\theta}) \mathrm{A}(\underset{\sim}{\theta} - \underset{\sim}{\delta})$ can still be used with the difference that the function A will take into account only the first $k$ components of the vector $\underset{\sim}{\theta} - \underset{\sim}{\delta}$, while the weighting function $\beta$ may involve any number of the components of $\underset{\sim}{\theta}$. If we still retain the assumption that $\mathcal{D} = \mathbb{R}^n$, then it is required that all $n$ components of $\underset{\sim}{\theta}$ are to be estimated but the estimates $\delta_{k+1}, ....\delta_n$ of the nuisance parameters are irrelevant.

We return to the arbitrary estimation problem where we are interested for all coordinates of $\underset{\sim}{\theta}$ so $\Theta = \mathcal{D} = \mathbb{R}^n$. **The Bayes risk is given by:**

$$r(\pi, \underset{\sim}{\theta}) = \int_{\Theta} \int_{X} L(\underset{\sim}{\theta}, \underset{\sim}{\delta}) f(\underset{\sim}{x} | \underset{\sim}{\theta}) \pi(\underset{\sim}{\theta}) d\underset{\sim}{x} d\underset{\sim}{\theta}$$

**and according to a previous theorem Bayes rule minimizes the posterior risk**

**The most popular used loss function for this statistical problem is the quadratic** loss function; $L(\underset{\sim}{\theta}, \underset{\sim}{\delta}) = (\underset{\sim}{\theta} - \underset{\sim}{\delta})' Q (\underset{\sim}{\theta} - \underset{\sim}{\delta})$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

where $Q$ is a $n\times n$ symmetric nonnegative definite matrix. If $Q$ is diagonal then

$$L(\theta,\delta)=\sum_{i=1}^{n} q_i(\theta_i-\delta_i)^2$$

**Theorem 2.6:** For the quadratic loss function $L(\underset{\sim}{\theta},\underset{\sim}{\delta})=(\underset{\sim}{\theta}-\underset{\sim}{\delta})'Q(\underset{\sim}{\theta}-\underset{\sim}{\delta})$ we have:

**a)** The Bayes estimator of the unknown random vector $\underset{\sim}{\theta}\in\mathbb{R}^n$ **is the posterior mean:**

$\underset{\sim}{\delta}^*=E(\underset{\sim}{\theta}\mid\underset{\sim}{x})\in\mathbb{R}^n$

**b)** Bayes risk for $\underset{\sim}{\delta}^*\in\mathbb{R}^n$ **will be given by the relation:**

$r(\pi,\underset{\sim}{\delta}^*)=tr\left\{QE_{\underset{\sim}{x}}\left[Cov(\underset{\sim}{\theta}\mid\underset{\sim}{x})\right]\right\}$

**Proof:**

Assume that no data are being observed, then Bayes rule $\underset{\sim}{\delta}^*\in\mathcal{D}=\mathbb{R}^n$ **minimizes** the expected loss with respect to the prior distribution of $\underset{\sim}{\theta}$, $\pi(\underset{\sim}{\theta})$ with $E(\underset{\sim}{\theta})=\underset{\sim}{\mu}$, $Cov(\underset{\sim}{\theta})=\Sigma$, where $\underset{\sim}{\mu}\in\mathbb{R}^n$ is the prior mean vector and $\Sigma$ is the $n\times n$ covariance matrix of the prior distribution

$$E_{\underset{\sim}{\theta}}\left[L(\underset{\sim}{\theta},\underset{\sim}{\delta})\right]=E_{\underset{\sim}{\theta}}\left[(\underset{\sim}{\theta}-\underset{\sim}{\delta})'Q(\underset{\sim}{\theta}-\underset{\sim}{\delta})\right]=E_{\underset{\sim}{\theta}}\left\{\left[\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}\right)'+\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)'\right]Q\left[\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}\right)+\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)\right]\right\}$$

$$=E_{\underset{\sim}{\theta}}\left[\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}\right)'Q\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}\right)\right]+\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)'Q\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)$$

$E_{\underset{\sim}{\theta}}\left[L(\underset{\sim}{\theta},\underset{\sim}{\delta})\right]$ is analyzed into two terms. In the first term no decision rule $\delta$ is contained so a Bayes rule must minimize the second term $(\mu-\delta)'Q(\mu-\delta)$. Since $Q$ is a nonnegative definite matrix $\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)'Q\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)$ is a nonnegative quantity for all decision rules $\delta\in\mathcal{D}$. The minimum value of $\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)'Q\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)$ is zero i.e. $\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)'Q\left(\underset{\sim}{\mu}-\underset{\sim}{\delta}\right)=0\Rightarrow\underset{\sim}{\delta}^*=\underset{\sim}{\mu}\Rightarrow\underset{\sim}{\delta}^*=E(\underset{\sim}{\theta})\in\mathbb{R}^n$ **is the Bayes** estimator for $\theta\in\mathbb{R}^n$. If $Q$ is a positive definite matrix then $\underset{\sim}{\delta}^*=E(\underset{\sim}{\theta})=\underset{\sim}{\mu}$ is the only Bayes estimator for $\underset{\sim}{\theta}$ and

$$r(\pi,\underset{\sim}{\delta}^*)=E_{\underset{\sim}{\theta}}\left[L(\underset{\sim}{\theta},\underset{\sim}{\delta}^*)\right]=E_{\underset{\sim}{\theta}}\left[\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}\right)'Q\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}\right)\right]=tr(Q\Sigma).$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

If $\underset{\sim}{x}=(x_1,....x_k)$ is the observation vector with likelihood density function conditioned on $\underset{\sim}{\theta}$: $f(\underset{\sim}{x}|\underset{\sim}{\theta})$ then all the conclusions about $\underset{\sim}{\theta}$ are based on the posterior information $p(\theta|\underset{\sim}{x})$ with corresponding prior $\pi(\theta)$, then the Bayes estimator for $\underset{\sim}{\theta}$ after $\underset{\sim}{x}=(x_1,....x_k)$ is observed will be given from the posterior mean $\underset{\sim}{\delta}^* = E(\underset{\sim}{\theta}|\underset{\sim}{x})$.

and $r(p,\delta^*)=tr\left\{QE_{\underset{\sim}{x}}\left[\mathrm{cov}\left(\theta|\underset{\sim}{x}\right)\right]\right\}$.

## 2.9 Optimal sample size

**In many statistical decision problems there is a sampling cost of receiving an observation before taking a decision. This cost reflects on our decision of how many random observations should be taken, or if it is better to draw a decision without even seeing any observations at all. More specifically, in many problems the statistician needs to decide on the size of the sample which is to be taken. The sampling cost depends on the size of the sample and we will denote it by $c(n)$ which will be a non-decreasing function of the sample size $n$. This cost must be definitely considered when we are interested in evaluating the risk of any decision function which makes use of a number of random observations.**

Let us consider the sampling model of the form $\{f(x|\theta):\theta\in\Theta\}$ where $x$ is a random observation with probability density function conditioned on $\theta$. $\theta$ is the unknown parameter that we want to estimate. Suppose now that for a given prior distribution $\pi(\theta)$ of the parameter $\theta$ and for a specified loss function $L(\theta,\alpha)$ it is desired to draw a decision about $\theta$ using the set $\mathcal{A}$. We have the opportunity either to choose a decision function without any observation, or to observe a random vector $\underset{\sim}{x}=(x_1,....,x_n)$ that is related to $\theta$. In case that no observations are made, a Bayes decision function against the prior $\pi(\theta)$ would be optimal with Bayes risk

$$r(\pi,\delta_{Bayes})=\min_{\delta\in D} r(\pi,\delta)=E_\theta\left[L(\theta,\delta_{Bayes})\right] \qquad (2.9)$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

If $\underset{\sim}{x}=(x_1,....,x_n)$ is to be observed before a decision is chosen the decision problem is basically the same as it was in the first case. The only difference is that the distribution $\pi(\theta)$ has been updated to the posterior $p(\theta|\underset{\sim}{x})$. Hence, a Bayes decision against the posterior $p(\theta|\underset{\sim}{x})$ of $\theta$ would know be the optimal one with Bayes risk

$$r(\pi,\delta_{Bayes})=\min_{\delta\in D} r(\pi,\delta)=E_{\underset{\sim}{x}}\left[\rho(\pi(\theta),\delta_{Bayes}(\underset{\sim}{x}))\right] \tag{2.10}$$

Because of the fact that in general $r(\pi,\delta)$ is a decreasing function of $n$, we can derive the conclusion from (2.9) that since $n=0$, the quantity $E_\theta\left[L(\theta,\delta_{Bayes})\right]$ will take larger values than the quantity $E_{\underset{\sim}{x}}\left[\rho(\pi(\theta),\delta_{Bayes}(\underset{\sim}{x}))\right]$ which depends on $n$. However if $\underset{\sim}{x}=(x_1,....,x_n)$ is made we have to take up in mind the total sampling cost $c(n)$ in the calculation of the risk function when a Bayes decision rule is chosen.

**Definition 2.9:** The total risk of observing $x=(x_1,....,x_n)$ using a Bayes decision function $\delta_{Bayes}(\underset{\sim}{x})$ can be expressed as the sum of the Bayes risk $r(\pi,\delta_{Bayes}(\underset{\sim}{x}))$ and the sampling cost of the given sample $c(n)$.

$$r_{total}(\pi(\theta),\delta_{Bayes}(\underset{\sim}{x}))=r(\pi(\theta),\delta_{Bayes}(\underset{\sim}{x}))+c(n) \tag{2.11}$$

Clearly in a statistical decision problem we desire to choose the best decision while we attain small sampling cost. The question is how many observations should be taken in order to accomplish the optimal procedure and pay as less as possible?

As it has already been mentioned $r(\pi,\delta_{Bayes}(\underset{\sim}{x}))$and $c(n)$ are a decreasing and a non-decreasing function of $n$ correspondently. In determining the optimal number of observations, noted by $n_{optimal}$, it is rather useful for us to use a sample size that balances the Bayes risk function and the sampling cost. So the optimal sample size ($n_{optimal}$) is clearly that $n$ which minimizes the total risk function $r_{total}(p(\theta),\delta_{Bayes}(\underset{\sim}{x}))$ and can be easily found by minimizing (2.11) with respect to $n$.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# Chapter 3

## Optimal sample size on general exponential family parametric models

### 3.1 A general Form of the exponential family

Definition 3.1: A probability density function $f(x|\theta)$, $x \in \mathcal{X}$ which is labeled by $\theta \in \Theta \subseteq \mathbb{R}^k$, is said to belong to the $k$-parameter exponential family if it is of the form:

$$f(x|\theta) = h(x)c(\theta)\exp\left\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\right\} \tag{3.1}$$

where $t(x) = (t_1(x),.....,t_k(x))$ is a sufficient statistic, $w(\theta) = (w_1(\theta),.....,w_k(\theta))$ and given the functions $h(x)$ , $w(\theta)$ , $t(x)$

$$[c(\theta)]^{-1} = \int_{\mathcal{X}} h(x)\exp\left\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\right\}dx < \infty$$

is the normalizing constant

Let $\underset{\sim}{x} = (x_1,...,x_n)$ to be a random sample from the $k$-parameter exponential family then the likelihood function has the form:

$$f(\underset{\sim}{x}|\theta) = \prod_{j=1}^{n} f(x_j|\theta) = \left[\prod_{j=1}^{n} h(x_j)\right][c(\theta)]^n \exp\left\{\sum_{i=1}^{k} w_i(\theta)\left[\sum_{j=1}^{n} t_i(x_j)\right]\right\} \tag{3.2}$$

where $T = (T_1,.....,T_k) = \left(\sum_{j=1}^{n} t_1(x_j),.....,\sum_{j=1}^{n} t_k(x_j)\right)$ is a sufficient statistic.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

### 3.1.1 Conjugate families for exponential families

**Proposition 3.1:** Assume that we have a random sample $\underset{\sim}{x}=(x_1,...,x_n)$ from the $k$-parameter exponential family with likelihood density function $f(\underset{\sim}{x}|\theta)$ given from (3.1.3). Then there exists a conjugate prior density function for the unknown parameter $\theta\in\Theta$ (Bernardo and Smith (1994)) which has the form

$$\pi(\theta|\tau)=[k(\tau)]^{-1}c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\tau_i\right\} \qquad (3.3)$$

where $\tau=(\tau_0,\tau_1,.....\tau_k)$ is the $k+1$-vector of the specified parameters of the prior (hyperparameters) and

$$k(\tau)=\int_{\theta\in\Theta}c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\tau_i\right\}d\theta<\infty \qquad (3.4)$$

is the normalizing constant .

We can also recognize (3.3) as a member of a $(k+1)$-parameter exponential family for $\theta$ (O'Hagan and Forster (2003)). A family $\mathcal{F}$ that is closed under sampling is easily identified with members density functions defined by (3.3) which lead to proper posterior distributions as it can be seen in the following proposition.

**Proposition 3.2:** For the $k$-parameter exponential family likelihood model (defined by (3.2)) and the conjugate prior density function for $\theta\in\Theta$ (defined in proposition 3.1) the posterior density function for $\theta\in\Theta$ after a random sample $\underset{\sim}{x}=(x_1,...,x_n)$ has been observed is

$$p(\theta|\underset{\sim}{x},\tau)=[k(\tau')]^{-1}c(\theta)^{\tau_0'}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\tau_i'\right\}=p(\theta|\tau') \qquad (3.5)$$

where $\tau'=(\tau_0',\tau_1',....,\tau_k')=\left(n+\tau_0,\tau_1+\sum_{j=1}^{n}t_1(x_j),....., \tau_k+\sum_{j=1}^{n}t_k(x_j)\right)$, and

$[k(\tau')]=\int_{\Theta}c(\theta)^{\tau_0'}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\tau_i'\right\}d\theta<\infty$ is the normalizing constant



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**Proof:**

From Bayes' theorem,

$$\begin{aligned} p(\theta|\underset{\sim}{x},\tau) &\propto \pi(\theta|\tau) f(\underset{\sim}{x}|\theta) \\ &\propto [c(\theta)]^n \exp\left\{\sum_{i=1}^{k} w_i(\theta)\left[\sum_{j=1}^{n} t_i(x_j)\right]\right\} c(\theta)^{\tau_0} \exp\left\{\sum_{i=1}^{k} w_i(\theta)\tau_i\right\} \\ &\propto [c(\theta)]^{n+\tau_0} \exp\left\{\sum_{i=1}^{k} w_i(\theta)\left[\tau_i + \sum_{j=1}^{n} t_i(x_j)\right]\right\} \\ &\propto \pi(\theta|\tau') \end{aligned}$$

With $\tau_i' = \tau_i + \sum_{j=1}^{n} t_i(x_j)$ for $i = 1,\ldots,k$ and dividing by $f(\underset{\sim}{x}) = \int_{\Theta} \pi(\theta|\tau) f(\underset{\sim}{x}|\theta) d\theta$ we obtain (3.5).

## 3.1.2 Moments of the conjugate prior

As we will see in the next chapter all the parameters under estimation are real valued. Hence the parameter space $\Theta$ is considered to be a subset of $\mathbb{R}$ i.e. $\Theta \subseteq \mathbb{R}$. For this case we prove the following Lemma:

**Lemma 3.1:** Consider the $(k+1)$- parameter exponential family for $\theta \in \Theta \subseteq \mathbb{R}$ which has already been defined from (3.3). The first and second moments of $\theta$ are given by:

$$E[\theta] = \frac{k^*(\tau,1)}{k(\tau)}, \quad E[\theta^2] = \frac{k^*(\tau,2)}{k(\tau)}$$

**Proof:**

$$\begin{aligned} E[\theta] &= \int_{\Theta} \theta \pi(\theta|\tau) d\theta \\ &= \int_{0} \theta [k(\tau)]^{-1} c(\theta)^{\tau_0} \exp\left\{\sum_{i=1}^{k} w_i(\theta)\tau_i\right\} d\theta \\ &= \int_{\Theta} [k(\tau)]^{-1} c(\theta)^{\tau_0} \exp\left\{\sum_{i=1}^{k} w_i(\theta)\tau_i + \ln\theta\right\} d\theta \\ &= [k(\tau)]^{-1} \int_{\Theta} c(\theta)^{\tau_0} \exp\left\{\sum_{i=1}^{k+1} w_i(\theta)\tau_i\right\} d\theta \end{aligned}$$

where: $w_{k+1}(\theta) = \ln\theta$ and $\tau_{k+1} = 1$ , so $E[\theta]$ becomes



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$E[\theta]=[k(\tau)]^{-1}\frac{1}{[k^*(\tau,1)]^{-1}}\int_{\theta\in\Theta}[k^*(\tau,1)]^{-1}c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k+1}w_i(\theta)\tau_i\right\}d\theta$$

$$=\frac{k^*(\tau,1)}{k(\tau)}$$

$$E[\theta]=\frac{k^*(\tau,1)}{k(\tau)} \qquad (3.6)$$

**Following the same argument we obtain that**

$$E[\theta^2]=\int_\Theta\theta^2\pi(\theta|\tau)d\theta$$

$$=\int_{\theta\in\Theta}\theta^2[k(\tau)]^{-1}c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\tau_i\right\}d\theta$$

$$=\int_\Theta[k(\tau)]^{-1}c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\tau_i+2\ln\theta\right\}d\theta$$

$$=[k(\tau)]^{-1}\int_\Theta c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k+1}w_i(\theta)\tau_i\right\}d\theta$$

Where, $w_{k+1}(\theta)=\ln\theta$ and $\tau_{k+1}=2$. Then

$$E[\theta^2]=[k(\tau)]^{-1}\frac{1}{[k^*(\tau,2)]^{-1}}\int_{\theta\in\Theta}[k^*(\tau,2)]^{-1}c(\theta)^{\tau_0}\exp\left\{\sum_{i=1}^{k+1}w_i(\theta)\tau_i\right\}d\theta$$

$$=\frac{k^*(\tau,2)}{k(\tau)}$$

$$E[\theta^2]=\frac{k^*(\tau,2)}{k(\tau)} \qquad (3.7)$$

**From (3.6) and (3.7) we calculate** the prior variance

$$\text{var}(\theta)=E[\theta^2]-E^2[\theta]$$

$$=\frac{k^*(\tau,2)}{k(\tau)}-\left[\frac{k^*(\tau,1)}{k(\tau)}\right]^2$$

$$=\frac{k^*(\tau,2)}{k(\tau)}-\frac{k^*(\tau,1)^2}{k(\tau)^2}$$

$$=\frac{k(\tau)k^*(\tau,2)-k^*(\tau,1)^2}{k(\tau)^2} \qquad (3.8)$$

We have to mention here that all the results summarized in the above relations can take different forms if there exists an $i$, $i=1,\ldots,k$ such that $w_i(\theta)=w_{k+1}(\theta)$. Then $k^*(\tau,1)=k(\tau_0,\tau_1^1,\ldots\tau_k^1)$, with $\tau_i^1=\tau_i+1$ for the same $i$ and $\tau_i^1=\tau_j\ \forall j\neq i$. Similarly under the same circumstances $k^*(\tau,2)=k(\tau_0,\tau_1^1,\ldots\tau_k^1)$ where $\tau_i^1=\tau_i+2$ and $\tau_j^1=\tau_j$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$\forall j \neq i$ , $j = 1,....,k$. If we are interested in the posterior mean and variance of $\theta$ the formulas will be given by (3.6) and (3.8) equivalently, but with different parameters.

More precisely:

$$E[\theta | \underset{\sim}{x}] = \frac{k^*(\tau', 1)}{k(\tau')} \qquad (3.9)$$

$$\text{var}(\theta | \underset{\sim}{x}) = \frac{k(\tau')k^*(\tau', 2) - k^*(\tau', 1)^2}{k(\tau')^2} \qquad (3.10)$$

Where, again for an $i$, $i = 1,....,k$ such that $w_i(\theta) = w_{k+1}(\theta)$, $k^*(\tau', 1) = k(\tau_0', \tau_1'^{+1}, ...\tau_k'^{+1})$, with $\tau_i'^{+1} = \tau_i' + 1$ for the same $i$ and $\tau_j'^{+1} = \tau_j'$ $\forall j \neq i$. For the same $i, i = 1,....,k$, $k^*(\tau', 2) = k(\tau_0', \tau_1'^{+1}, ...\tau_k'^{+1})$ and $\tau_i'^{+1} = \tau_i' + 2$ , $\tau_j'^{+1} = \tau_j'$ $\forall j \neq i$.

### 3.1.3 Derivation of the optimal sample size

**Proposition 3.3:** Let us consider a random sample $\underset{\sim}{x} = (x_1, ..., x_n)$ from the $k$-parameter exponential family for which the value of the parameter $\theta$ is unknown. The conjugate prior distribution of $\theta$ is defined from (3.1.1.1) . We are interested in estimating the value of $\theta$, under squared error loss function i.e. $L(\theta, d) = (\theta - \alpha)^2$. If the sampling cost per observation is $c, (c > 0)$. Then the optimal number of observations that must be drawn is specified by the following:

**Proof:**

The likelihood function of the data is specified by (3.2) and the posterior distribution of $\theta$ by (3.5) . Bayes estimator for $\theta$ will be given by the posterior mean, from (3.9) where

$$\delta^*(\underset{\sim}{x}) = \frac{k^*(\tau', 1)}{k(\tau')} \qquad (3.11)$$

The Bayes risk will be

$$r(\pi(\theta), \delta^*(\underset{\sim}{x})) = E_{\underset{\sim}{x}}[\text{var}(\theta | \underset{\sim}{x})] = E\left[\frac{k(\tau')k^*(\tau', 2) - k^*(\tau', 1)^2}{k(\tau')^2}\right].$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

In order to obtain the optimal sample we have to minimize with respect to $n$ the total risk function:

$$\begin{aligned} r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x})) &= nc + E_{\underset{\sim}{x}}\left[\mathrm{var}(\theta\,|\,\underset{\sim}{x})\right] \\ &= nc + \mathrm{var}(\theta) - \mathrm{var}_{\underset{\sim}{x}}\left[E(\theta\,|\,\underset{\sim}{x})\right] \\ &= nc + \mathrm{var}(\theta) - E_{\underset{\sim}{x}}\left[E^2(\theta\,|\,\underset{\sim}{x})\right] + E_{\underset{\sim}{x}}^2\left[E(\theta\,|\,\underset{\sim}{x})\right] \\ &= nc + \mathrm{var}(\theta) + E^2(\theta) - E_{\underset{\sim}{x}}\left[E^2(\theta\,|\,\underset{\sim}{x})\right] \\ &= nc + \frac{k(\tau)k^*(\tau,2) - k^*(\tau,1)^2}{k(\tau)^2} + \left(\frac{k^*(\tau,1)}{k(\tau)}\right)^2 - \int_{\underset{\sim}{x}\in\underset{\sim}{X}} \left(\frac{k^*(\tau',1)}{k(\tau')}\right)^2 f(\underset{\sim}{x})d\underset{\sim}{x} \end{aligned} \qquad (3.12)$$

For the marginal distribution of $\underset{\sim}{x}$ we have:

$$\begin{aligned} f(\underset{\sim}{x}) &= \int_{\Theta} f(\underset{\sim}{x}\,|\,\theta)\pi(\theta\,|\,\tau)d\theta \\ &= \left[k(\tau)\right]^{-1}\prod_{j=1}^{n}h(x_j)\int_{\theta\in\Theta} c(\theta)^{n+\tau_0}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\left(\tau_i+\sum_{j=1}^{n}h_i(x_j)\right)\right\}d\theta \\ &= \left[k(\tau)\right]^{-1}\prod_{j=1}^{n}h(x_j)\frac{1}{\left[k(\tau')\right]^{-1}}\int_{\theta\in\Theta}\left[k(\tau')\right]^{-1}c(\theta)^{n+\tau_0}\exp\left\{\sum_{i=1}^{k}w_i(\theta)\left(\tau_i+\sum_{j=1}^{n}h_i(x_j)\right)\right\}d\theta \Rightarrow \end{aligned}$$

$$f(\underset{\sim}{x}) = \left[k(\tau)\right]^{-1}\prod_{j=1}^{n}h(x_j)\frac{1}{\left[k(\tau')\right]^{-1}} = \frac{k(\tau')}{k(\tau)}\prod_{j=1}^{n}h(x_j) \qquad (3.13)$$

**Minimizing (3.12) we take the following result**

$$\frac{\partial}{\partial n}\left[nc + \frac{k(\tau)k^*(\tau,2) - k^*(\tau,1)^2}{k(\tau)^2} + \left(\frac{k^*(\tau,1)}{k(\tau)}\right)^2 - \int_{\underset{\sim}{X}}\left(\frac{k^*(\tau',1)}{k(\tau')}\right)^2\frac{k(\tau')}{k(\tau)}\prod_{j=1}^{n}h(x_j)d\underset{\sim}{x}\right] = 0 \Leftrightarrow$$

$$c - \frac{\partial}{\partial n}\int_{\underset{\sim}{X}}\frac{k^*(\tau',1)^2}{k(\tau')k(\tau)}\prod_{j=1}^{n}h(x_j)d\underset{\sim}{x} = 0 \qquad (3.14)$$

**In order to obtain (3.14) we have considered that**

$\frac{\partial}{\partial n}\left[\frac{k(\tau)k^*(\tau,2) - k^*(\tau,1)^2}{k(\tau)^2} + \left(\frac{k^*(\tau,1)}{k(\tau)}\right)^2\right] = 0$ because this term contains only parameters of the prior so it does not depend on $n$ at all. Furthermore we will have to check that $\frac{\partial^2 r_{total}}{\partial n^2} > 0$ for the estimated $n$ which will be the solution of the equation (3.14) in order to verify that achieves the minimum total risk.

From (3.14) we need to have the exact form of the functions $k(\cdot)$, $k^*(\cdot)$ and $h(\cdot)$ in order to derive the optimal sample size. Therefore we are not able to have a closed



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

form solution for the general case. However in the next chapter we will explore and derive the exact formulas for several well known members of the exponential family.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# Chapter 4

## Determination of the optimal sample size for specified sample distributions of the exponential family

### 4.1 Introduction

In this chapter we will restrict our focus to specific univariate sampling distributions from the general exponential family $f(\underset{\sim}{x}|\theta)$. Our goal is to estimate the parameter $\theta$ of the exponential family under square error loss function. We will determine the optimal sample size needed, as a function of the rest of the parameters of the sampling distribution and the prior. The specific members of the exponential family that will be examined are:
Normal, Poisson, Binomial (Bernoulli), Gamma (Exponential).

### 4.2 Optimal sample for Gamma distribution

**Proposition 4.1** Suppose $x_1,....,x_n$ is a random sample from a Gamma distribution with $k$ known and $\theta$ unknown $(k,\theta>0)$ i.e. $f(x_i|\theta)\sim Ga(k,\theta)$. The conjugate prior for $\theta$ is
again a gamma distribution with specified values of the parameters $\alpha,\beta$ where $\alpha>0,\beta>0$ i.e. $\pi(\theta)\sim Ga(\alpha,\beta)$. If we are interested in estimating $\theta$ under squared error loss function i.e. $L(\theta,d)=(\theta-d)^2$ and the sampling cost per observation is $c,(c>0)$, then the optimal sample size is defined by the equation:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$n_{optimal} = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k} \qquad (4.1)$$

**Proof:**

For **every random observation** $x$ **the conditional density function will be**

$$f(x\,|\,\theta) = \frac{\theta^k}{\Gamma(k)} x^{k-1} \exp\{-\theta x\}$$

**This relation can also be written as**

$f(x\,|\,\theta) = \frac{\theta^k}{\Gamma(k)} x^{k-1} \exp\{-\theta x\} = \frac{x^{k-1}}{\Gamma(k)} \theta^k \exp\{-\theta x\}$ and from (3.1) it is amember of the

**exponential family distribution with:**

$h(x) = \frac{x^{k-1}}{\Gamma(k)}$, $c(\theta) = \theta^k$, $t(x) = x$ and $w(\theta) = -\theta$

The calculation of the likelihood from **(3.2) will give:**

$$f(\underset{\sim}{x}\,|\,\theta) = \left[\prod_{i=1}^{n} \frac{x_i^{k-1}}{\Gamma(k)}\right] \theta^{nk} \exp\left\{-\theta \sum_{i=1}^{n} x_i\right\} = \frac{\left(\prod_{i=1}^{n} x_i\right)^{k-1}}{\Gamma(k)^n} \theta^{nk} \exp\left\{-\theta \sum_{i=1}^{n} x_i\right\}$$

From (3.3) the conjugate prior density for $\theta$ :

$$\pi(\theta\,|\,\tau_0, \tau_1) = \left[k(\tau_0, \tau_1)\right]^{-1} \theta^{k\tau_0} \exp\{-\theta\tau_1\} \qquad (4.2)$$

$$\pi(\theta\,|\,\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp\{-\theta\beta\} \qquad (4.3)$$

Because of the equality $\pi(\theta\,|\,\tau_0, \tau_1) = Ga(\alpha, \beta) = \pi(\theta\,|\,\alpha, \beta)$ we may derive from (4.2), (4.3) the following conditions

$\left[k(\tau_0, \tau_1)\right]^{-1} = \frac{\beta^\alpha}{\Gamma(\alpha)}$, $k\tau_0 = \alpha - 1$, $\tau_1 = \beta$

The posterior distribution of $\theta$ can be determined from proposition (3.5)as

$$p(\theta\,|\,\underset{\sim}{x}, \tau_0, \tau_1) = \pi(\theta\,|\,\tau_0', \tau_1') = \pi(\theta\,|\,\tau_0 + n, \tau_1 + \sum_{i=1}^{n} x_i)$$

$$= \left[k(\tau_0 + n, \tau_1 + \sum_{i=1}^{n} x_i)\right]^{-1} \theta^{k(n+\tau_0)} \exp\left\{-\theta\left(\tau_1 + \sum_{i=1}^{n} x_i\right)\right\}$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Where,

$k(n+\tau_0)=\alpha^*-1 \Leftrightarrow \alpha^*=k(n+\tau_0)+1 \Leftrightarrow \alpha^*=kn+k\tau_0+1 \Leftrightarrow \alpha^*=kn+\alpha-1+1 \Leftrightarrow \alpha^*=kn+\alpha$ and

$\tau_1+\sum_{i=1}^{n}x_i=\beta^* \Leftrightarrow \beta^*=\beta+\sum_{i=1}^{n}x_i$

So the posterior distribution of $\theta$ is derived to be gamma with parameters

$\alpha^*=nk+\alpha$ and $\beta^*=\beta+\sum_{i=1}^{n}x_i$

Then the **Bayes rule under** square error loss is given by theorem 2.4 as:

$$\delta^*(\underset{\sim}{x})=E(\theta\,|\,\underset{\sim}{x})=\frac{\alpha^*}{\beta^*}=\frac{\alpha+nk}{\beta+\sum_{i=1}^{n}x_i}$$

**and the Bayes risk defined also from 2.4:**

$$r(\pi(\theta),\delta^*(\underset{\sim}{x}))=E\left[\text{var}(\theta\,|\,\underset{\sim}{x})\right]=\text{var}(\theta)-\text{var}\left[E(\theta\,|\,\underset{\sim}{x})\right]$$

$$=\frac{\alpha}{\beta^2}-\text{var}\left(\frac{\alpha+nk}{\beta+\sum_{i=1}^{n}x_i}\right)$$

$$=\frac{\alpha}{\beta^2}-(\alpha+nk)^2\,\text{var}\left(\frac{1}{\beta+\sum_{i=1}^{n}x_i}\right) \qquad (4.4)$$

We must now define the probability density function of the random variable $Z=\frac{1}{Y}$

where

$Y=\beta+\sum_{i=1}^{n}x_i$. We have to mention here that because of the fact that

$x_i\,|\,\theta\sim gamma(k,\theta) \Rightarrow x_i>0$ for all $i=1,..,n$. So $\sum_{i=1}^{n}x_i>0 \Rightarrow \beta+\sum_{i=1}^{n}x_i>\beta \Rightarrow Y>\beta$

First we derive the marginal probability density function for $\sum_{i=1}^{n}x_i$:

$x_i\,|\,\theta\sim gamma(k,\theta) \Rightarrow \sum_{i=1}^{n}x_i\,|\,\theta\sim gamma(nk,\theta)$ so for the marginal distribution

we will have:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$f\left(\sum_{i=1}^{n} x_i\right) = \int_{\Theta} f\left(\sum_{i=1}^{n} x_i \mid \theta\right)\pi(\theta)d\theta$$

$$= \int_{\Theta} \frac{\theta^{nk}}{\Gamma(nk)}\left(\sum_{i=1}^{n} x_i\right)^{nk-1} \exp\left\{-\theta\left(\sum_{i=1}^{n} x_i\right)\right\}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp\{-\beta\theta\}d\theta$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\left(\sum_{i=1}^{n} x_i\right)^{nk-1}}{\Gamma(nk)}\int_{\Theta} \theta^{\alpha+nk-1}\exp\left\{-\theta\left(\beta+\sum_{i=1}^{n} x_i\right)\right\}d\theta$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\left(\sum_{i=1}^{n} x_i\right)^{nk-1}}{\Gamma(nk)}\frac{\Gamma(\alpha+nk)}{\left(\beta+\sum_{i=1}^{n} x_i\right)^{\alpha+nk}}$$

Next we shall derive the density function for the random variable $Y$

$$Y = \beta + \sum_{i=1}^{n} x_i \Leftrightarrow \sum_{i=1}^{n} x_i = Y - \beta = h^{-1}(Y) \Rightarrow$$

$$f_Y(y) = f_{\sum_{i=1}^{n} x_i}(y)\left|\frac{d}{dy}h^{-1}(y)\right|$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{(y-\beta)^{nk-1}}{\Gamma(nk)}\frac{\Gamma(\alpha+nk)}{(\beta+y-\beta)^{\alpha+nk}}\left|\frac{d}{dy}(y-\beta)\right|$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{(y-\beta)^{nk-1}}{y^{\alpha+nk}} \qquad (4.5)$$

Following the same strategy for the random variable $Z$ we have: $Z = \frac{1}{Y} \Leftrightarrow Y = \frac{1}{Z} = h^{-1}(Z)$ where $Z$ takes values in the interval $\left(0, \frac{1}{\beta}\right)$

$$f_Z(z) = f_Y(h^{-1}(z))\left|\frac{d}{dz}h^{-1}(z)\right|$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{\left(\frac{1}{z}-\beta\right)^{nk-1}}{\left(\frac{1}{z}\right)^{\alpha+nk}}\left|-\frac{1}{z^2}\right|$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{z^{\alpha+nk}(1-\beta z)^{nk-1}}{z^{nk-1}}\frac{1}{z^2}$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}(1-\beta z)^{nk-1}z^{\alpha-1}$$

So we have derived the density function of the random variable $Z = \dfrac{1}{\beta + \sum_{i=1}^{n} x_i}$ to have the following formula:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$f_Z(z)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}(1-\beta z)^{nk-1}z^{\alpha-1},\quad 0<Z<\frac{1}{\beta} \tag{4.6}$$

**First and second moment of the random variable** $Z$

$$\begin{aligned}
E(z)&=\int_{\angle}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}(1-\beta z)^{nk-1}z^{\alpha-1}z\,dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\int_{\angle}(1-\beta z)^{nk-1}z^{\alpha}dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\int_{\angle}z^{(\alpha+1)-1}(1-\beta z)^{nk-1}dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{\Gamma(\alpha+1)\Gamma(nk)}{\beta^{\alpha+1}\Gamma(\alpha+1+nk)}\int_{\angle}\frac{\beta^{\alpha+1}\Gamma(\alpha+1+nk)}{\Gamma(\alpha+1)\Gamma(nk)}(1-\beta z)^{nk-1}z^{(\alpha+1)-1}dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{\Gamma(\alpha+1)\Gamma(nk)}{\beta^{\alpha+1}\Gamma(\alpha+1+nk)}\\
&=\frac{\alpha}{\beta(\alpha+nk)}
\end{aligned}$$

$$\begin{aligned}
E(z^2)&=\int_{\angle}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}(1-\beta z)^{nk-1}z^{\alpha-1}z^2dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\int_{\angle}(1-\beta z)^{nk-1}z^{\alpha+1}dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\int_{\angle}z^{(\alpha+2)-1}(1-\beta z)^{nk-1}dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{\Gamma(\alpha+2)\Gamma(nk)}{\beta^{\alpha+2}\Gamma(\alpha+2+nk)}\int_{\angle}\frac{\beta^{\alpha+2}\Gamma(\alpha+2+nk)}{\Gamma(\alpha+2)\Gamma(nk)}(1-\beta z)^{nk-1}z^{(\alpha+2)-1}dz\\
&=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+nk)}{\Gamma(nk)}\frac{\Gamma(\alpha+2)\Gamma(nk)}{\beta^{\alpha+2}\Gamma(\alpha+2+nk)}\\
&=\frac{\alpha(\alpha+1)}{\beta^2(\alpha+nk)(\alpha+nk+1)}
\end{aligned}$$

**So the relation (4.4)** takes the following form




ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$r(\pi(\theta),\delta^*(\underset{\sim}{x})) = \frac{\alpha}{\beta^2} - (\alpha+nk)^2 \operatorname{var}(z)$$

$$= \frac{\alpha}{\beta^2} - (\alpha+nk)^2\left[E(z^2) - E(z)^2\right]$$

$$= \frac{\alpha}{\beta^2} - (\alpha+nk)^2\left[\frac{\alpha(\alpha+1)}{\beta^2(\alpha+nk)(\alpha+nk+1)} - \frac{\alpha^2}{\beta^2(\alpha+nk)^2}\right]$$

$$= \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} - \frac{\alpha(\alpha+1)}{\beta^2}\frac{(\alpha+nk)}{(\alpha+nk+1)} \qquad (4.7)$$

**The total risk function from (2.11) is specified from relation**

$$r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x})) = \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} - \frac{\alpha(\alpha+1)}{\beta^2}\frac{(\alpha+nk)}{(\alpha+nk+1)} + cn \qquad (4.8)$$

**Minimization of (4.8) with respect to $n$ gives**

$$\frac{\partial}{\partial n}\left[\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} - \frac{\alpha(\alpha+1)}{\beta^2}\frac{(\alpha+nk)}{(\alpha+nk+1)} + cn\right] = 0$$

$$-\frac{\alpha(\alpha+1)}{\beta^2}\left[\frac{k(\alpha+nk+1)-(\alpha+nk)k}{(\alpha+nk+1)^2}\right] + c = 0$$

$$c\beta^2(\alpha+nk+1)^2 = \alpha k(\alpha+1)$$

$$(\alpha+nk+1)^2 = \frac{\alpha k(\alpha+1)}{c\beta^2} \Leftrightarrow (\alpha+nk+1) = \frac{\sqrt{\alpha k(\alpha+1)}}{\beta\sqrt{c}}$$

$$n = \frac{\sqrt{\alpha k(\alpha+1)}}{k\beta\sqrt{c}} - \frac{\alpha+1}{k} = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{kc}} - \frac{\alpha+1}{k}$$

**Taking the second derivative we get:**

$$\frac{\partial^2 r_{total}}{\partial n^2} = \frac{\partial}{\partial n}\left\{-\frac{\alpha(\alpha+1)}{\beta^2}\left[\frac{k(\alpha+nk+1)-(\alpha+nk)k}{(\alpha+nk+1)^2}\right] + c\right\}$$

$$= \frac{\partial}{\partial n}\left\{-\frac{\alpha(\alpha+1)}{\beta^2}\frac{k}{(\alpha+nk+1)^2} + c\right\}$$

$$= \frac{\alpha(\alpha+1)}{\beta^2}\frac{2k^2}{(\alpha+nk+1)^3} > 0 \quad \forall n > 0$$

**Hence, we end up to the conclusion that** $n_{optimal} = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{kc}} - \frac{\alpha+1}{k}$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**A discussion on the behaviour of the optimal sample as a function of its parameters**

$n_{optimal}$ at the first step is considered to be a positive integer. As it has already been done for other distributions we consider $n_{optimal}$ as a function $g$ of the hyper parameters i.e.

$n_{optimal} = g(\alpha,\beta,k,c) = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k}$. In order to study the dependence of $n_{optimal}$ on the parameters $\alpha$, $\beta$, $k$, $c$ we calculate the partial derivatives of $g(\alpha,\beta,k,c)$:

$$\frac{\partial g(\alpha,\beta,c,k)}{\partial\beta} = \frac{\partial}{\partial\beta}\left\{\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k}\right\} = -\frac{1}{\beta^2}\frac{\sqrt{\alpha(\alpha+1)}}{\sqrt{ck}}$$

$$= -\sqrt{\frac{\alpha(\alpha+1)}{\beta^4 ck}} \quad (<0)$$

$$\frac{\partial g(\alpha,\beta,c,k)}{\partial c} = \frac{\partial}{\partial c}\left\{\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k}\right\} = -\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{k}}\frac{1}{2c\sqrt{c}}$$

$$= -\sqrt{\frac{\alpha(\alpha+1)}{4k\beta^2 c^3}} \quad (<0)$$

$$\frac{\partial g(\alpha,\beta,c,k)}{\partial\alpha} = \frac{\partial}{\partial\alpha}\left\{\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k}\right\} = \frac{\partial}{\partial\alpha}\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{1}{k}$$

$$= \frac{2\alpha+1}{2\beta\sqrt{kc\alpha(\alpha+1)}} - \frac{1}{k} = \frac{(2\alpha+1)\sqrt{k} - 2\beta\sqrt{c\alpha(\alpha+1)}}{2\beta k\sqrt{c\alpha(\alpha+1)}}$$

$$\frac{\partial g(\alpha,\beta,c,k)}{\partial k} = \frac{\partial}{\partial k}\left\{\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k}\right\} = -\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{c}}\frac{1}{2\sqrt{k}k} + \frac{\alpha+1}{k^2}$$

$$= \frac{-\sqrt{k}\sqrt{\alpha(\alpha+1)} + 2\beta\sqrt{c}(\alpha+1)}{2\beta k^2\sqrt{c}}$$

**Summarizing the above we can have the following:**

$n_{optimal}$ is a decreasing function of $\beta$. For $\beta\to+\infty$ the variance of the prior distribution of $\theta$, $\frac{\alpha}{\beta^2}\to 0$ and $n_{optimal} \xrightarrow{\beta\to+\infty} -\frac{\alpha+1}{k} < 0$ so if our prior distribution is very informative and no observations needs to be taken i.e. $n_{optimal} = 0$. If $\beta\to 0$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

then $\frac{\alpha}{\beta^2} \to +\infty$ and $n_{optimal} \xrightarrow{\beta \to 0} +\infty$. So if our prior distribution is non-informative a large number of observations must be taken to estimate the unknown value of the parameter $\theta$.

$\forall c > 0$ $n_{optimal}$ is a decreasing function of $c$. As $c$ increases the optimal sample size $n_{optimal}$ decreases. In other words if the cost for every observation is very large then we cannot afford to take many observations because the total sampling cost becomes too high. If $c \to +\infty$ then $n_{optimal} \xrightarrow{c \to \infty} -\frac{\alpha+1}{k} < 0$ i.e. $n_{optimal} = 0$. For $c \to 0$, $n_{optimal} \xrightarrow{c \to 0} +\infty$

$\forall \alpha > 0$ we derive the critical points of the expression

$$\frac{(2\alpha+1)\sqrt{k} - 2\beta\sqrt{c\alpha(\alpha+1)}}{2\beta k\sqrt{c\alpha(\alpha+1)}} = 0 \Leftrightarrow (2\alpha+1)\sqrt{k} - 2\beta\sqrt{c\alpha(\alpha+1)} = 0$$

$$(2\alpha+1)\sqrt{k} = 2\beta\sqrt{c\alpha(\alpha+1)} \Leftrightarrow (2\alpha+1)^2 k = 4\beta^2 c\alpha(\alpha+1)$$

$$4k\alpha^2 + 4k\alpha + k - 4c\beta^2\alpha^2 - 4c\beta^2\alpha = 0$$

$$(4k - 4c\beta^2)\alpha^2 + (4k - 4c\beta^2)\alpha + k = 0$$

$$\Delta = 16(k - c\beta^2)^2 - 16k(k - c\beta^2) = 16(k - c\beta^2)(k - c\beta^2 - k) = -16c\beta^2(k - c\beta^2)$$

We have three different cases for $\Delta$

1. For $\Delta > 0 \Rightarrow k - c\beta^2 < 0 \Leftrightarrow c > \frac{k}{\beta^2}$ we will have to possible critical points $\alpha_1$, $\alpha_2$ with $\alpha_1 \neq \alpha_2$ i.e. $\alpha_1 = \frac{-4(k - c\beta^2) + \sqrt{-16c\beta^2(k - c\beta^2)}}{8(k - c\beta^2)} < 0$ but as we have already mentioned $\alpha > 0$ so $\alpha_1$ is rejected and

$$\alpha_2 = \frac{-4(k - c\beta^2) - \sqrt{-16c\beta^2(k - c\beta^2)}}{8(k - c\beta^2)} = \frac{4(k - c\beta^2) + \sqrt{-16c\beta^2(k - c\beta^2)}}{8(c\beta^2 - k)} > 0 \qquad (4.9)$$

Indeed the numerator is always a positive quantity because if we assume that:

$$4(k - c\beta^2) + \sqrt{-16c\beta^2(k - c\beta^2)} > 0 \Leftrightarrow \sqrt{16c\beta^2(c\beta^2 - k)} > 4(c\beta^2 - k)$$

$$16c\beta^2(c\beta^2 - k) > 16(c\beta^2 - k)^2 \Leftrightarrow 16(c\beta^2 - k)(c\beta^2 - c\beta^2 + k) > 0$$

$$\Rightarrow 16k(c\beta^2 - k) > 0$$

This last expression is always true so the initial quantity $4(k - c\beta^2) + \sqrt{-16c\beta^2(k - c\beta^2)}$ is positive hence, $\alpha_2 > 0$. In the interval $(0, \alpha_2)$ $g(\alpha, \beta, k, c)$ is an increasing function of $\alpha$ and in the interval $(\alpha_2, +\infty)$, $g(\alpha, \beta, k, c)$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

is a decreasing function of $\alpha$. For $\alpha=\alpha_2$ $n_{optimal}$ takes its maximum value(due to the complexity of the type **(4.9) we are** not going to proceed **to further replacements ).**

2. For $\Delta=0\Rightarrow k-c\beta^2=0\Rightarrow c=\frac{k}{\beta^2}$, then from the expression

$4(k-c\beta^2)\alpha^2+4(k-c\beta^2)\alpha+k=0\Rightarrow k=0$ which can not be true.

3. For $\Delta<0\Rightarrow k-c\beta^2>0\Rightarrow c<\frac{k}{\beta^2}$, then for every $\alpha>0$ the function

$g(\alpha,\beta,c,k)$

is an increasing function of $\alpha$.

$\forall k>0$ **we derive the critical points** also for the expression

$$\frac{-\sqrt{k}\sqrt{\alpha(\alpha+1)}+2\beta\sqrt{c}(\alpha+1)}{2\beta k^2\sqrt{c}}=0\Leftrightarrow -\sqrt{k}\sqrt{\alpha(\alpha+1)}+2\beta\sqrt{c}(\alpha+1)=0$$

$$2\beta\sqrt{c}(\alpha+1)=\sqrt{k}\sqrt{\alpha(\alpha+1)}\Leftrightarrow 4\beta^2c(\alpha+1)^2=k\alpha(\alpha+1)$$

$$(\alpha+1)\left[4\beta^2c\alpha+4\beta^2c-k\alpha\right]=0$$

$(\alpha+1)$ is always a positive quantity so $4\beta^2c\alpha+4\beta^2c-k\alpha=0\Rightarrow k=\frac{4\beta^2c(\alpha+1)}{\alpha}>0$.

In the interval $\left(0,\frac{4\beta^2c(\alpha+1)}{\alpha}\right)$ $g(\alpha,\beta,k,c)$ is an increasing function of $k$ and in the interval $\left(\frac{4\beta^2c(\alpha+1)}{\alpha},+\infty\right)$ $g(\alpha,\beta,k,c)$ **is a decreasing function of** $k$. For $k=\frac{4\beta^2c(\alpha+1)}{\alpha}$ function $g(\alpha,\beta,k,c)$ takes its maximum value i.e.

$$n_{optimal\max}(\alpha,\beta,k,c)=\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{c}\frac{\sqrt{\alpha+1}}{\sqrt{\alpha}}\sqrt{4\beta^2c}}-\frac{\alpha+1}{\left(\frac{\alpha+1}{\alpha}\right)4\beta^2c}$$

$$=\frac{\alpha}{2\beta^2c}-\frac{\alpha}{4\beta^2c}=\frac{\alpha}{4\beta^2c}$$

In case that $n_{optimal}=\frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}}-\frac{\alpha+1}{k}$ **is a non-positive quantity then,**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$n_{optimal} \le 0 \Leftrightarrow \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k} \le 0 \Leftrightarrow \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} \le \frac{\alpha+1}{k}$$

$$\Leftrightarrow \frac{\alpha(\alpha+1)}{\beta^2 ck} \le \frac{(\alpha+1)^2}{k^2} \Leftrightarrow \frac{\alpha}{\beta^2 c} \le \frac{\alpha+1}{k} \Leftrightarrow c \ge \frac{\alpha k}{\beta^2(\alpha+1)}$$

When the cost per observation $c$ becomes greater or equal to the ratio $\frac{\alpha k}{\beta^2(\alpha+1)}$ then $n_{optimal}$ becomes negative. Ratio $\frac{\alpha k}{\beta^2(\alpha+1)}$ increases as $\alpha$, $k$ increases and $\beta$ decreases. In that case the prior variance $\frac{\alpha}{\beta^2}$ becomes very large, so the prior distribution of $\theta$ is not informative at all and it is preferable that no observations should be taken because the sampling cost $c$ is very high for us to pay.

**Corollary 4.1** Suppose $x_1,....,x_n$ is a random sample from an exponential distribution i.e. $f(\underset{\sim}{x}|\theta) \sim Exp(\theta)$ with unknown the value of the parameter $\theta$, $(\theta > 0)$. If $\theta$ is desired to be estimated under the same prior and error loss function as they defined in **Proposition 4.1** then, when the sampling cost per observation is $c, (c > 0)$ :

$$n_{optimal} = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{c}} - (\alpha+1) \qquad (4.10)$$

The case of the exponential distribution may be considered as a special case of the Gamma distribution i.e. $f(\underset{\sim}{x}|\theta) \sim Ga(1,\theta)$. So from (4.1) and for $k=1$ we can obtain relation (4.10).

## 4.3 Optimal sample size from normal distribution

**Proposition 4.2** Suppose that $x_1,.....,x_n$ is a random sample from a normal distribution with an unknown value of the mean $\theta$ and a specified value of the



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

precision $r, (r>0)$, $f(x_i|\theta) \sim N\left(\theta, \frac{1}{r}\right)$. The prior distribution for $\theta$ is selected from the conjugate family, i.e. $\pi(\theta) \sim N\left(\mu, \frac{1}{\tau}\right)$ with $\mu, \tau$ both known. If the value of $\theta$ is to be estimated under the squared error loss, i.e. $L(\theta,d)=(\theta-d)^2$ and the sampling cost per observation is $c$, $(c>0)$ then the optimal number of observations $n$ is specified by the equation:

$$n_{optimal} = \left(\frac{1}{rc}\right)^{\frac{1}{2}} - \frac{\tau}{r} \qquad (4.11)$$

**Proof:**

For any random observation $x$ **the density function is**

$f(x|\theta) = (2\pi)^{-\frac{1}{2}} r^{\frac{1}{2}} \exp\left\{-\frac{r}{2}(x-\theta)^2\right\}$ from proposition (3.1) we can write it in the **form of an exponential family:**

$$f(x|\theta) = (2\pi)^{-\frac{1}{2}} r^{\frac{1}{2}} \exp\left\{-\frac{r}{2}\left(x^2-2\theta x+\theta^2\right)\right\} = \left[(2\pi)^{-1} r\right]^{\frac{1}{2}} \exp\left\{-\frac{r}{2}x^2\right\} \exp\left\{-\frac{r}{2}\theta^2\right\} \exp\left\{x(r\theta)\right\}$$

where: $h(x) = \left[(2\pi)^{-1} r\right]^{\frac{1}{2}} \exp\left\{-\frac{r}{2}x^2\right\}$, $c(\theta) = \exp\left\{-\frac{r}{2}\theta^2\right\}$, $w(\theta) = r\theta$, $t(x) = x$

**The likelihood function from (3.2) becomes**

$$f(\underline{x}|\theta) = \left[(2\pi)^{-1} r\right]^{\frac{n}{2}} \exp\left\{-\frac{r}{2}\sum_{i=1}^{n} x_i^2\right\} \exp\left\{-\frac{nr}{2}\theta^2\right\} \exp\left\{(r\theta)\sum_{i=1}^{n} x_i\right\}$$

From (3.3) the conjugate prior density for $\theta$ :

$$\pi(\theta|\tau_0,\tau_1) = \left[k(\tau_0,\tau_1)\right]^{-1} \exp\left\{-\frac{r\tau_0}{2}\theta^2\right\} \exp\left\{(r\theta)\tau_1\right\} \qquad (4.12)$$

$$\pi(\theta|\mu,\tau) = \left[(2\pi)^{-1}\tau\right]^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\mu^2\right\} \exp\left\{-\frac{\tau}{2}\theta^2\right\} \exp\left\{\tau\mu\theta\right\} \qquad (4.13)$$

**From these two last expressions (4.2), (4.3) which are equivalent we have:**

$$\left[k(\tau_0,\tau_1)\right]^{-1} = \left[(2\pi)^{-1}\tau\right]^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\mu^2\right\}, \quad -\frac{r\tau_0}{2} = -\frac{\tau}{2} \Leftrightarrow \tau_0 = \frac{\tau}{r}, \quad r\tau_1 = \tau\mu \Leftrightarrow \tau_1 = \frac{\tau\mu}{r} = \tau_0\mu$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

The posterior distribution of the unknown value of the mean $\theta$ after having observed $\underset{\sim}{x}=(x_1,...,x_n)$ is available from proposition (3.5)

$$p(\theta\,|\,\underset{\sim}{x},\tau_0,\tau_1)=\pi(\theta\,|\,\tau_0',\tau_1')=\pi(\theta\,|\,n+\tau_0,\tau_1+\sum_{i=1}^{n}x_i)$$

$$=\left[k(n+\tau_0,\tau_1+\sum_{i=1}^{n}x_i)\right]^{-1}\exp\left\{-\frac{r(\tau_0+n)}{2}\theta^2\right\}\exp\left\{(r\theta)\left(\tau_1+\sum_{i=1}^{n}x_i\right)\right\}$$

Accordingly to the previous relations that hold for the prior the following conditions must be satisfied for the posterior as well:

$$n+\tau_0=\frac{\tau^*}{r}\Leftrightarrow\tau^*=nr+r\tau_0\Leftrightarrow\tau^*=nr+r\frac{\tau}{r}\Leftrightarrow\tau^*=nr+\tau,$$

$$\tau_1+\sum_{i=1}^{n}x_i=\mu^*(n+\tau_0)\Leftrightarrow\mu^*=\frac{\tau_1+\sum_{i=1}^{n}x_i}{n+\tau_0}\Leftrightarrow\mu^*=\frac{\tau_0\mu+\sum_{i=1}^{n}x_i}{n+\tau_0}\Leftrightarrow\mu^*=\frac{\mu\frac{\tau}{r}+\sum_{i=1}^{n}x_i}{n+\frac{\tau}{r}}$$

$$\Leftrightarrow\mu^*=\frac{\tau\mu+nr\bar{x}}{nr+\tau}$$

So we have derived that the posterior distribution of $\theta$ will be normal with mean $\mu^*=\frac{\tau\mu+nr\bar{x}}{\tau+nr}$, and precision $\tau^*=\tau+rn$. We can easily observe that the posterior mean can be expressed as the weighted average of the sample mean $\bar{x}$ and the prior mean $\mu$.

$$\mu^*=\frac{\tau\mu+nr\bar{x}}{\tau+nr}=\frac{\tau}{\tau+nr}\mu+\frac{nr}{\tau+nr}\bar{x}=(1-k)\mu+k\bar{x},\text{ where }k=\frac{nr}{\tau+nr}.$$

Then the Bayes rule under square error loss will be:

$$\delta^*(\underset{\sim}{x})=E(\theta\,|\,\underset{\sim}{x})=\mu^*=(1-k)\mu+k\bar{x}$$

For the above estimator of $\theta$, Bayes and total risk are specified by the equations:

$$r(\pi(\theta),\delta^*(\underset{\sim}{x}))=E(\text{var}(\theta\,|\,\underset{\sim}{x}))=E(\tau+rn)^{-1}=E\left(\frac{1}{\tau+rn}\right)=\frac{1}{\tau+rn}\qquad(4.14)$$

$$r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x}))=r(\pi(\theta),\delta^*(\underset{\sim}{x}))+c(n)=\frac{1}{\tau+rn}+nc\qquad(4.15)$$

We will derive the optimal sample size by minimizing (4.15) with respect to $n$ :



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\frac{\partial r_{total}(\pi(\theta),\delta^*(\underline{x}))}{\partial n}=0 \Leftrightarrow \frac{\partial}{\partial n}\left(r(\pi(\theta),\delta^*(\underline{x}))+c(n)\right)=0$$

$$-\frac{r}{(\tau+rn)^2}+c=0 \Leftrightarrow (\tau+rn)^2=\frac{r}{c} \Leftrightarrow (\tau+rn)=\left(\frac{r}{c}\right)^{\frac{1}{2}}$$

$$rn=\left(\frac{r}{c}\right)^{\frac{1}{2}}-\tau \Leftrightarrow n=\left(\frac{1}{rc}\right)^{\frac{1}{2}}-\frac{\tau}{r}$$

$$\frac{\partial^2 r_{total}(\pi(\theta),\delta^*(\underline{x}))}{\partial n^2}=\frac{\partial}{\partial n}\left(-\frac{r}{(\tau+rn)^2}+c\right)=\frac{2r^2}{(\tau+rn)^3}>0 \quad \forall\, n>0$$

**So in order to have the minimum total risk the optimal sample size choice** will be **given by:**

$$n_{optimal}=\left(\frac{1}{rc}\right)^{\frac{1}{2}}-\frac{\tau}{r}$$

**A discussion on the behaviour of the optimal sample size as a function of its parameters**

We have proved that the optimal sample size in the case of normal distribution with unknown mean will be given by the equation $n_{optimal}=\left(\frac{1}{rc}\right)^{\frac{1}{2}}-\frac{\tau}{r}$. This quantity needs to be a positive integer i.e. $n_{optimal}>0$. Next, we consider $n_{optimal}$ as a function of $r,\tau,c$ i.e. $n_{optimal}=f(r,\tau,c)$. We are interested in studying the dependence of the optimal sample size on each parameter separately considering all the others constant. For this reason we calculate the partial derivatives of $f(r,\tau,c)$

$$\frac{\partial f(r,\tau,c)}{\partial r}=\frac{\partial}{\partial r}\left[\left(\frac{1}{rc}\right)^{\frac{1}{2}}-\frac{\tau}{r}\right]=\frac{1}{\sqrt{c}}\frac{\partial}{\partial r}\frac{1}{\sqrt{r}}-\frac{\partial}{\partial r}\frac{\tau}{r}$$

$$=\frac{1}{\sqrt{c}}\left(-\frac{1}{2\sqrt{r}r}\right)+\frac{\tau}{r^2}=-\frac{1}{2}\frac{1}{\sqrt{rc}r}+\frac{\tau}{r^2}$$

$$=-\frac{1}{2}\frac{1}{\sqrt{r^3c}}+\frac{\tau}{r^2}$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\frac{\partial f(r,\tau,c)}{\partial \tau} = \frac{\partial}{\partial \tau}\left[\left(\frac{1}{rc}\right)^{\frac{1}{2}} - \frac{\tau}{r}\right]$$

$$= \frac{\partial}{\partial \tau}\left[\left(\frac{1}{rc}\right)^{\frac{1}{2}}\right] - \frac{\partial}{\partial t}\frac{\tau}{r} = 0 - \frac{1}{r}$$

$$= -\frac{1}{r} \ (<0)$$

$$\frac{\partial f(r,\tau,c)}{\partial c} = \frac{\partial}{\partial c}\left[\left(\frac{1}{rc}\right)^{\frac{1}{2}} - \frac{\tau}{r}\right] = \frac{\partial}{\partial c}\left[\left(\frac{1}{rc}\right)^{\frac{1}{2}}\right] - \frac{\partial}{\partial c}\frac{\tau}{r}$$

$$= \frac{1}{\sqrt{r}}\frac{\partial}{\partial c}\frac{1}{\sqrt{c}} - 0 = -\frac{1}{\sqrt{r}}\frac{1}{2c\sqrt{c}} = -\frac{1}{\sqrt{r}}\frac{1}{2c\sqrt{c}}$$

$$= -\frac{1}{2}\frac{1}{\sqrt{c^3 r}} \ (<0)$$

**Based on the partial derivatives we conclude the following:**

$n_{optimal}$ will be a decreasing function of $\tau$. As $\tau$ **increases the** variance of the prior of $w$, $\frac{1}{\tau}$ gets small and $n_{optimal}$ decreases too. Hence, as $\tau \to +\infty$ $\frac{1}{\tau} \to 0$, which means that we have a complete apriori knowledge about the value of the parameter $\theta$, (point mass) and thus no observation should be taken in order to estimate $\theta$ i.e. $n_{optimal} = 0$. If conversely $\tau \to 0$ $\frac{1}{\tau} \to +\infty$. So when the prior distribution of $\theta$ becomes flat **then** $n_{optimal}$ increases to take the asymptotic value

$$n_{optimal} \xrightarrow{\tau \to 0} \left(\frac{1}{rc}\right)^{\frac{1}{2}}$$

$n_{optimal}$ is a decreasing function of $c, \forall c > 0$. If $c \to +\infty$ then, $n_{optimal} \xrightarrow{c \to \infty} -\frac{\tau}{r} < 0$.

**In other words if the cost for every observation** is very large **then we** cannot afford **to take many observations because the total** sampling cost **becomes** too high.

**For the derivative with respect** to $r$ **we had:**

$\frac{\partial f(r,\tau,c)}{\partial r} = -\frac{1}{2}\frac{1}{\sqrt{r^3 c}} + \frac{\tau}{r^2}$. We derive the critical points of this expression



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$-\frac{1}{2}\frac{1}{\sqrt{r^3c}}+\frac{\tau}{r^2}=0 \Leftrightarrow \frac{1}{2}\frac{1}{\sqrt{r^3c}}=\frac{\tau}{r^2} \Leftrightarrow$$
$$\frac{1}{4}\frac{1}{r^3c}=\frac{\tau^2}{r^4} \Leftrightarrow \frac{1}{4}\frac{1}{c}=\frac{\tau^2}{r} \Leftrightarrow$$
$$r=4c\tau^2=r^*$$

For the values of $r$ which satisfy the relation $r<r^* \Rightarrow r<4c\tau^2$, $\frac{\partial f(r,\tau,c)}{\partial r}>0$ and in the interval $\left(0,4c\tau^2\right)$ $n_{optimal}$ is an increasing function of $r$.

For $r>r^* \Rightarrow r>4c\tau^2$, $\frac{\partial f(r,\tau,c)}{\partial r}<0$. So in the interval $\left(4c\tau^2,+\infty\right)$ $n_{optimal}$ is a decreasing function of $r$. For $r=r^*=4c\tau^2$ the function $f(r,\tau,c)=n_{optimal}$ takes its **maximum value:**

$$n_{optimal\max}=\left(\frac{1}{r^*c}\right)^{\frac{1}{2}}-\frac{\tau}{r^*}=\frac{1}{\sqrt{4c\tau^2c}}-\frac{\tau}{4c\tau^2}$$
$$=\frac{1}{2c\tau}-\frac{1}{4c\tau}=\frac{2}{4c\tau}-\frac{1}{4c\tau}$$
$$=\frac{1}{4c\tau}$$

All the above analysis **was done in the case that $n_{optimal}$ is a positive integer. But** under what scenario we have $n_{optimal}$ **to be a negative quantity?**

$$n_{optimal}\le 0 \Rightarrow \left(\frac{1}{rc}\right)^{\frac{1}{2}}-\frac{\tau}{r}\le 0 \Leftrightarrow \left(\frac{1}{rc}\right)^{\frac{1}{2}}\le\frac{\tau}{r}$$
$$\Leftrightarrow \frac{1}{rc}\le\frac{\tau^2}{r^2} \Leftrightarrow \frac{1}{c}\le\frac{\tau^2}{r} \Leftrightarrow c\tau^2\ge r$$
$$\Leftrightarrow c\ge\frac{r}{\tau^2}$$

If the cost per observation $c$ is greater or equal to the ratio $\frac{r}{\tau^2}$, then $n_{optimal}$ is a **negative number and so we may consider that it is better** for us not to pay the price of any observation i.e. $n_{optimal}=0$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

As $r\uparrow \Rightarrow \frac{1}{r}\downarrow$ and $\tau\downarrow \Rightarrow \frac{1}{\tau}\uparrow$ the ratio $\frac{r}{\tau^2}$ increases. In other words, when the sample distribution is very informative and the prior distribution is non-informative the sampling cost of an observation becomes too high and we choose the strategy of not taking any observations at all.

**Proposition 4.3** Suppose that $x_1,.....,x_n$ is a random sample from a normal distribution for which both the mean $\theta$ and the precision $r$ are unknown. The likelihood will be given by: $f(x_i|\theta,r) \sim N\left(\theta, \frac{1}{r}\right)$ and we are interested in estimating $\theta$ using squared error loss i.e. $L(\theta,d)=(\theta-d)^2$. The unknown precision $r$ will be a nuisance parameter. So the conjugate choice for the joint prior distribution of $(\theta,r)$ is the Normal-gamma distribution where: The conditional distribution of $\theta$ given $r$ is a normal distribution with mean $\mu$ and precision $tr(t>0)$, i.e. $\pi(\theta|r) \sim N\left(\mu, \frac{1}{tr}\right)$ and the marginal distribution of $r$ is a gamma distribution with parameters $\alpha$ and $\beta$, $\alpha>0, \beta>0$ i.e. $\pi(r) \sim Gamma(\alpha, \beta)$. So the joint prior distribution of $(\theta,r)$ will be $\pi(\theta,r)=\pi(\theta|r)\pi(r)=N\left(\mu,\frac{1}{tr}\right)G(\alpha,\beta)$. If the sampling cost per observation is $c,(c>0)$ then the optimal number observations $n$ to be sampled is:

$$n_{optimal}=\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t \qquad (4.16)$$

**Proof:**

If $x$ is a random observation from the sample then,

$f(x|\theta,r)=(2\pi)^{-\frac{1}{2}}r^{\frac{1}{2}}\exp\left\{-\frac{r}{2}(x-\theta)^2\right\}$ from proposition (3.1.1) this relation can be formed as follows:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$f(x\,|\,\theta,r)=(2\pi)^{-\frac{1}{2}}r^{\frac{1}{2}}\exp\left\{-\frac{r}{2}x^2+x\theta r-\frac{r}{2}\theta^2\right\}$$

$$=(2\pi)^{-\frac{1}{2}}\left[r^{\frac{1}{2}}\exp\left\{-\frac{r}{2}\theta^2\right\}\right]\exp\left\{x\theta r-\frac{r}{2}x^2\right\}$$

**Then:**

$h(x)=(2\pi)^{-\frac{1}{2}}$, $c(\theta,r)=r^{\frac{1}{2}}\exp\left\{-\frac{r}{2}\theta^2\right\}$, $w_1(\theta,r)=\theta r$, $t_1(x)=x$, $w_2(\theta,r)=-\frac{r}{2}$,

$t_2(x)=x^2$

**and the likelihood function from (3.1.3) is written as**

$$f(\underset{\sim}{x}\,|\,\theta,r)=(2\pi)^{-\frac{n}{2}}\left[r^{\frac{n}{2}}\exp\left\{-\frac{rn}{2}\theta^2\right\}\right]\exp\left\{\theta r\sum_{i=1}^{n}x_i-\frac{r}{2}\sum_{i=1}^{n}x_i^2\right\}$$

From (3.1.1.1) the conjugate prior density for $\theta$:

$$\pi(\theta,r\,|\,\tau_0,\tau_1,\tau_2)=\left[k(\tau_0,\tau_1,\tau_2)\right]^{-1}\left[r^{\frac{\tau_0}{2}}\exp\left\{-\frac{r\tau_0}{2}\theta^2\right\}\right]\exp\left\{\theta r\tau_1-\frac{1}{2}r\tau_2\right\} \qquad (4.17)$$

$$\pi(\theta,r\,|\,\mu,t,\alpha,\beta)=(2\pi)^{-\frac{1}{2}}(tr)^{\frac{1}{2}}\exp\left\{-\frac{tr}{2}(\theta-\mu)^2\right\}\times\frac{\beta^\alpha}{\Gamma(\alpha)}r^{\alpha-1}\exp\{-\beta r\}$$

$$=(2\pi)^{-\frac{1}{2}}\frac{\beta^\alpha}{\Gamma(\alpha)}t^{\frac{1}{2}}r^{\frac{1}{2}+\alpha-1}\exp\left\{-\frac{tr}{2}(\theta-\mu)^2-\beta r\right\}$$

$$=(2\pi)^{-\frac{1}{2}}\frac{\beta^\alpha}{\Gamma(\alpha)}t^{\frac{1}{2}}\left[r^{\frac{1}{2}+\alpha}\exp\left\{-\frac{rt}{2}\theta^2\right\}\right]\exp\left\{\theta rt\mu-\frac{r}{2}(t\mu^2+2\beta)\right\} \qquad (4.18)$$

**Because of the fact that:**

$$\pi(\theta,r\,|\,\tau_0,\tau_1,\tau_2)=\pi(\theta,r\,|\,\mu,t,\alpha,\beta)=\pi(\theta\,|\,r,\mu,t)\times\pi(r\,|\,\alpha,\beta)=Ng(\mu,t,\alpha,\beta)$$

From (4.17) and (4.18) we can obtain the following:

$$\left[k(\tau_0,\tau_1,\tau_2)\right]^{-1}=(2\pi)^{-\frac{1}{2}}\frac{\beta^\alpha}{\Gamma(\alpha)}t^{\frac{1}{2}},\ \frac{\tau_0}{2}=\alpha-\frac{1}{2},\ t=\tau_0,\ \tau_1=t\mu,$$

$$\tau_2=t\mu^2+2\beta=\tau_0\frac{\tau_1^2}{\tau_0^2}+2\beta\Rightarrow\tau_2=\frac{\tau_1^2}{\tau_0}+2\beta$$

The posterior distribution of the unknown value of the mean $\theta$ and the precision $r$ is defined from proposition (3.5)



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$p(\theta,r\mid \underset{\sim}{x},\tau_0,\tau_1,\tau_2)=\pi(\theta,r\mid \tau_0',\tau_1',\tau_2')=\pi(\theta,r\mid n+\tau_0,\tau_1+\sum_{i=1}^{n}x_i,\tau_2+\sum_{i=1}^{n}x_i^2)$$

$$=\left[k\left(n+\tau_0,\tau_1+\sum_{i=1}^{n}x_i,\tau_2+\sum_{i=1}^{n}x_i^2\right)\right]^{-1}\left[r^{\frac{n+\tau_0}{2}}\exp\left\{-\frac{r(n+\tau_0)}{2}\theta^2\right\}\right]$$

$$\times\exp\left\{\theta r\left(\tau_1+\sum_{i=1}^{n}x_i\right)-\frac{1}{2}r\left(\tau_2+\sum_{i=1}^{n}x_i^2\right)\right\}$$

**Where, based on the relations that hold for the prior we derive equivalent conditions for the posterior. Thus,**

$$\frac{n+\tau_0}{2}=\alpha^*-\frac{1}{2}\Leftrightarrow\alpha^*=\frac{n+\tau_0+1}{2}\Leftrightarrow\alpha^*=\frac{n+2\alpha-1+1}{2}\Leftrightarrow\alpha^*=\alpha+\frac{n}{2}$$

$$t^*=n+\tau_0\Leftrightarrow t^*=n+t,\ \tau_1+\sum_{i=1}^{n}x_i=(n+t)\mu^*\Leftrightarrow\mu^*=\frac{t\mu+\sum_{i=1}^{n}x_i}{n+t}\Leftrightarrow\mu^*=\frac{t\mu+n\bar{X}}{n+t}$$

$$\tau_2+\sum_{i=1}^{n}x_i^2=\frac{(\tau_1+\sum_{i=1}^{n}x_i)^2}{n+\tau_0}+2\beta^*\Leftrightarrow\frac{\tau_1^2}{\tau_0}+2\beta+\sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2=\frac{(\tau_1+n\bar{x})^2}{n+\tau_0}+2\beta^*$$

$$\Leftrightarrow\frac{(t\mu)^2}{t}+2\beta+\sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2=\frac{(t\mu+n\bar{x})^2}{n+t}+2\beta^*$$

$$\Leftrightarrow\beta^*=\beta+\frac{1}{2}\left\{n\bar{x}^2+t\mu^2+\sum_{i=1}^{n}(x_i-\bar{x})^2-\frac{(t\mu+n\bar{x})^2}{(n+t)}\right\}$$

$$=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{n\bar{x}^2+t\mu^2}{2}-\frac{(t\mu+n\bar{x})^2}{(n+t)}$$

$$=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{n^2\bar{x}^2+t^2\mu^2+nt\bar{x}^2+nt\mu^2-n^2\bar{x}^2-t^2\mu^2-2tn\mu\bar{x}}{2(n+t)}$$

$$=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{nt(\bar{x}-\mu)^2}{2(n+t)}$$

**So the posterior joint distribution** will be normal-gamma i.e.

$p(\theta,r\mid \underset{\sim}{x})=Ng(\theta,r\mid\mu^*,t^*,\alpha^*,\beta^*)$.

We derive from this last expression that the posterior conditional distribution of $\theta$ given $r$ will be normal as well with mean $\mu^*=\frac{n\bar{x}+t\mu}{t+n}$ and precision $k^*=rt^*=r(t+n)$. The posterior marginal distribution of $r$ will be gamma with parameters $\alpha^*=\alpha+\frac{n}{2}$, $\beta^*=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{X})^2+\frac{tn}{2}\frac{(\bar{X}-\mu)^2}{t+n}$

Posterior marginal distribution of $\theta$ will be given by:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$p_{\Theta}(\theta \mid \underline{x}) = \int_{R} p(\theta, r \mid \underline{x}) dr = \int_{R} (t+n)^{\frac{1}{2}} r^{\frac{1}{2}} \exp\left\{-\frac{(t+n)r}{2}(\theta-\mu^{*})^{2}\right\} r^{\alpha^{*}-1} \exp\left\{-r\beta^{*}\right\} dr$$

$$= \int_{R} (t+n)^{\frac{1}{2}} r^{\alpha^{*}+\frac{1}{2}-1} \exp\left\{-r\left[\frac{t+n}{2}(\theta-\mu^{*})^{2}+\beta^{*}\right]\right\} dr$$

$$= (2\pi)^{-\frac{1}{2}} (t+n)^{\frac{1}{2}} \frac{\beta^{*\alpha^{*}}}{\Gamma(\alpha^{*})} \int_{R} r^{\alpha^{*}+\frac{1}{2}-1} \exp\left\{-r\left[\frac{t+n}{2}(\theta-\mu^{*})^{2}+\beta^{*}\right]\right\} dr$$

$$= (2\pi)^{-\frac{1}{2}} (t+n)^{\frac{1}{2}} \frac{\beta^{*\alpha^{*}}}{\Gamma(\alpha^{*})} \frac{\Gamma\left(\alpha^{*}+\frac{1}{2}\right)}{\left[\beta^{*}+\frac{t+n}{2}(\theta-\mu^{*})^{2}\right]^{\alpha^{*}+\frac{1}{2}}}$$

If we use the proportionality symbol and drop all the factors that they do not involve $\theta$ we take can take as a result

$$p_{\Theta}(\theta \mid \underline{x}) \propto \left[\beta^{*}+\frac{t+n}{2}(\theta-\mu^{*})^{2}\right]^{-\alpha^{*}-\frac{1}{2}} \propto \left[\beta^{*}+\frac{t+n}{2}(\theta-\mu^{*})^{2}\right]^{-\frac{2\alpha^{*}+1}{2}}$$

$$\propto \left[1+\frac{t+n}{2\beta^{*}}(\theta-\mu^{*})^{2}\right]^{-\frac{2\alpha^{*}+1}{2}} \propto \left[1+\frac{\alpha^{*}(t+n)}{2\alpha^{*}\beta^{*}}(\theta-\mu^{*})^{2}\right]^{-\frac{2\alpha^{*}+1}{2}}$$

$$\propto \left[1+\frac{\alpha^{*}\left(\frac{t+n}{\beta^{*}}\right)}{2\alpha^{*}}(\theta-\mu^{*})^{2}\right]^{-\frac{2\alpha^{*}+1}{2}}$$

Thus the posterior marginal distribution of $\theta$ is a $t$ distribution with $2\alpha^{*}$ degrees of freedom location parameter $\mu^{*}=\frac{n\bar{x}+t\mu}{t+n}$ and scale parameter $k^{**}=\frac{\alpha^{*}(t+n)}{\beta^{*}}$

Similarly we can derive for the prior marginal distribution of $\theta$:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\pi(\theta,r) \propto r^{\frac{1}{2}} r^{\alpha-1} \exp\left\{-\frac{tr}{2}(\theta-\mu)^2 - \beta r\right\}$$

$$\Rightarrow \pi_{\Theta}(\theta) = \int_R \pi(\theta,r)dr \propto \int_R r^{\frac{1}{2}} r^{\alpha-1} \exp\left\{-\frac{tr}{2}(\theta-\mu)^2 - \beta r\right\} dr$$

$$\propto \int_R r^{\alpha+\frac{1}{2}-1} \exp\left\{-r\left[\frac{t}{2}(\theta-\mu)^2 + \beta\right]\right\} dr \propto \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\left[\frac{t}{2}(\theta-\mu)^2 + \beta\right]^{\alpha+\frac{1}{2}}}$$

$$\propto \left[\frac{t}{2}(\theta-\mu)^2 + \beta\right]^{-\frac{2\alpha+1}{2}} \propto \left[\frac{t\alpha}{2\alpha\beta}(\theta-\mu)^2 + 1\right]^{-\frac{2\alpha+1}{2}}$$

$$\propto \left[\frac{\left(\frac{t\alpha}{\beta}\right)}{2\alpha}(\theta-\mu)^2 + 1\right]^{-\frac{2\alpha+1}{2}}$$

So the prior marginal distribution of $\theta$ is a $t$ distribution with $2\alpha$ degrees of freedom location parameter $\mu$ and scale parameter $\frac{t\alpha}{\beta}$.

**First and second moment of $t$ distribution:**

Let a random variable $y$ follows student distribution with $\alpha$ degrees of freedom, mean parameter $\mu$ and scale parameter $t$. Then $z = t^{\frac{1}{2}}(y-\mu)$ follows a standardized $t$ distribution with $\alpha$ degrees of freedom. For $\alpha > 2$ we have $E(z) = 0$, $\text{var}(z) = \frac{\alpha}{\alpha-2}$.

Then for the posterior of $\theta|\underset{\sim}{x}$ we have:

$\left[\frac{\alpha^*(t+n)}{\beta^*}\right]^{\frac{1}{2}}(\theta-\mu^*)|\underset{\sim}{x} \sim t$ with $2\alpha^*$ degrees of freedom so:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$E\left(\left[\frac{\alpha^*(t+n)}{\beta^*}\right]^{\frac{1}{2}}(\theta-\mu^*)\,|\,\underset{\sim}{x}\right)=0\Rightarrow\left[\frac{\alpha^*(t+n)}{\beta^*}\right]^{\frac{1}{2}}E\left((\theta-\mu^*)\,|\,\underset{\sim}{x}\right)=0$$

$$\Rightarrow E(\theta\,|\,\underset{\sim}{x})=\mu^*=\frac{n\bar{x}+t\mu}{t+n}$$

$$\text{var}\left(\left[\frac{\alpha^*(t+n)}{\beta^*}\right]^{\frac{1}{2}}(\theta-\mu^*)\,|\,\underset{\sim}{x}\right)=\frac{2\alpha^*}{2\alpha^*-2}\Rightarrow\frac{\alpha^*(t+n)}{\beta^*}\text{var}(\theta\,|\,\underset{\sim}{x})=\frac{\alpha^*}{\alpha^*-1}$$

$$\Rightarrow\text{var}(\theta\,|\,\underset{\sim}{x})=\frac{\beta^*}{(\alpha^*-1)(t+n)}$$

equivalently for the prior distribution of $\theta$ we have that $\left(\frac{\alpha t}{\beta}\right)^{\frac{1}{2}}(\theta-\mu)\sim t$ with

$2\alpha$ degrees of freedom so:

$$E\left[\left(\frac{\alpha t}{\beta}\right)^{\frac{1}{2}}(\theta-\mu)\right]=0\Rightarrow E(\theta)=\mu$$

$$\text{var}\left[\left(\frac{\alpha t}{\beta}\right)^{\frac{1}{2}}(\theta-\mu)\right]=\frac{2\alpha}{2\alpha-2}\Rightarrow\frac{\alpha t}{\beta}\text{var}(\theta)=\frac{\alpha}{\alpha-1}\Rightarrow\text{var}(\theta)=\frac{\beta}{t(\alpha-1)}$$

Under squared error loss function we end up to the conclusion from theorem 2.7.1 that the Bayes rule will be given by:

$$\delta^*(\underset{\sim}{x})=E(m\,|\,\underset{\sim}{x})=\mu^*=\frac{n\bar{X}+t\mu}{t+n}$$

From 2.7.1 Bayes risk is defined by the equation

$$\begin{aligned}r(\pi(\theta),\delta^*(\underset{\sim}{x}))&=E\left[\text{var}(\theta\,|\,\underset{\sim}{x})\right]=\text{var}(\theta)-\text{var}\left[E(\theta\,|\,\underset{\sim}{x})\right]\\&=\frac{\beta}{t(\alpha-1)}-\text{var}\left(\frac{n\bar{x}+t\mu}{t+n}\right)\\&=\frac{\beta}{t(\alpha-1)}-\frac{n^2}{(t+n)^2}\text{var}(\bar{x})\\&=\frac{\beta}{t(\alpha-1)}-\frac{n^2}{(t+n)^2}\left\{\text{var}\left[E(\bar{x}\,|\,\theta,r)\right]+E\left[\text{var}(\bar{x}\,|\,\theta,r)\right]\right\}\quad(4.19)\end{aligned}$$

Given that $\bar{x}\,|\,\theta,r\sim N\left(\theta,\frac{1}{rn}\right)$ we obtain



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$E\left[\text{var}(\bar{x}\mid\theta,r)\right]=E\left(\frac{1}{rn}\right)=\frac{1}{n}E\left(\frac{1}{r}\right)=\frac{1}{n}\int_R\frac{1}{r}p(r)dr$$

$$=\frac{1}{n}\int_R\frac{1}{r}\frac{\beta^\alpha}{\Gamma(\alpha)}r^{\alpha-1}\exp\{-r\beta\}dr=\frac{1}{n}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_R r^{(\alpha-1)-1}\exp\{-r\beta\}$$

$$=\frac{1}{n}\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}}=\frac{1}{n}\frac{\beta\Gamma(\alpha-1)}{\Gamma(\alpha-1)(\alpha-1)}$$

$$=\frac{1}{n}\frac{\beta}{(\alpha-1)}$$

$$\text{var}\left[E(\bar{x}\mid\theta,r)\right]=\text{var}(\theta)=\frac{\beta}{t(\alpha-1)}$$

Replacing in (4.19) we take

$$r(\pi(\theta),\delta^*(\underset{\sim}{x}))=\frac{\beta}{t(\alpha-1)}-\frac{n^2}{(t+n)^2}\frac{\beta(t+n)}{nt(\alpha-1)}=\frac{\beta}{t(\alpha-1)}\left(1-\frac{n}{t+n}\right)$$

$$=\frac{\beta t}{t(\alpha-1)(t+n)}=\frac{\beta}{(\alpha-1)(t+n)}$$

The total risk function from (2.8.2) will be:

$$r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x}))=\frac{\beta}{(\alpha-1)(t+n)}+nc \qquad (4.20)$$

For the calculation of the optimal sample size we minimize (4.20) with respect to $n$

$$\frac{\partial r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x}))}{\partial n}=0\Leftrightarrow\frac{\partial}{\partial n}\left\{\frac{\beta}{(\alpha-1)(t+n)}+cn\right\}=0$$

$$c-\frac{\beta}{\alpha-1}\frac{1}{(t+n)^2}=0\Leftrightarrow c(\alpha-1)(t+n)^2-\beta=0$$

$$(t+n)^2=\frac{\beta}{c(\alpha-1)}\Leftrightarrow(t+n)=\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}$$

$$n=\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t$$

$$\frac{\partial^2 r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x}))}{\partial n^2}=\frac{\partial}{\partial n}\left[c-\frac{\beta}{\alpha-1}\frac{1}{(t+n)^2}\right]$$

$$=\frac{\beta}{\alpha-1}\frac{2}{(t+n)^3}>0\quad\forall n>0$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

The optimal solution for the minimum total risk is given by equation

$$n_{optimal}=\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t$$

**A study of the function of the optimal sample size**

$n_{optimal}=\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t$ is expected to be a positive integer and it can be considered as a function of $\alpha$, $\beta, c, t$ i.e. $n_{optimal}=g(\alpha,\beta,c,t)$. In order to investigate how these

**parameters affect** separately $n_{optimal}$, we calculate the partial derivatives of $g(\alpha,\beta,t,c)$

$$\frac{\partial g(\alpha,\beta,t,c)}{\partial\alpha}=\frac{\partial}{\partial\alpha}\left\{\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t\right\}=\left(\frac{\beta}{c}\right)^{\frac{1}{2}}\frac{\partial}{\partial\alpha}\frac{1}{\sqrt{\alpha-1}}=\left(\frac{\beta}{c}\right)^{\frac{1}{2}}\left[-\frac{1}{2(\alpha-1)^{\frac{3}{2}}}\right]$$

$$=-\left[\frac{\beta}{4c(\alpha-1)^3}\right]^{\frac{1}{2}}(<0)$$

$$\frac{\partial g(\alpha,\beta,t,c)}{\partial\beta}=\frac{\partial}{\partial\beta}\left\{\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t\right\}=\frac{1}{\left[c(\alpha-1)\right]^{\frac{1}{2}}}\frac{\partial}{\partial\beta}\sqrt{\beta}=\frac{1}{\left[c(\alpha-1)\right]^{\frac{1}{2}}}\frac{1}{2\sqrt{\beta}}$$

$$=\left[\frac{1}{4\beta c(\alpha-1)}\right]^{\frac{1}{2}}(>0)$$

$$\frac{\partial g(\alpha,\beta,t,c)}{\partial t}=\frac{\partial}{\partial t}\left\{\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t\right\}=0-\frac{\partial}{\partial t}t=-1\ (<0)$$

$$\frac{\partial g(\alpha,\beta,t,c)}{\partial c}=\frac{\partial}{\partial c}\left\{\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}-t\right\}=\frac{\partial}{\partial c}\left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}}=\left[\frac{\beta}{(\alpha-1)}\right]^{\frac{1}{2}}\frac{\partial}{\partial c}\frac{1}{\sqrt{c}}$$

$$=\left[\frac{\beta}{(\alpha-1)}\right]^{\frac{1}{2}}\left(-\frac{1}{2c\sqrt{c}}\right)=-\left[\frac{\beta}{4(\alpha-1)c^3}\right]^{\frac{1}{2}}(<0)$$

**Summarizing all the above results we conclude:**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$\forall \alpha > 1$ $n_{optimal}$ is a decreasing function of $\alpha$. As $\alpha$ increases it can be easily seen that $\text{var}(\theta) = \frac{\beta}{t(\alpha - 1)}$ which is the prior variance of $\theta$, decreases so our prior becomes very informative and eventually we do not need many observations to be taken. So if $\alpha \to +\infty$, $\frac{\beta}{t(\alpha-1)} \to 0$ and $n_{optimal} \xrightarrow{\alpha \to +\infty} -t < 0$ i.e. $n_{optimal} = 0$. In this case $\theta$ is directly estimated from the prior.

$n_{optimal}$ is an increasing function of $\beta, \forall \beta > 0$. More specifically if $\beta \to +\infty$ then, $\frac{\beta}{t(\alpha-1)} \to +\infty$ and $n_{optimal} \xrightarrow{\beta \to \infty} +\infty$ which means that if $\beta$ becomes very big then variance of the prior of $\theta$ becomes non-informative and a very big number of observations is need to be taken in order to have a good decision for $\theta$. If $\beta \to 0$ $\frac{\beta}{t(\alpha-1)} \to 0$ and $n_{optimal} \xrightarrow{\beta \to 0} -t < 0$. So as the value of $\beta$ gets very small $n_{optimal}$ decreases also to take the minimum value $-t \Rightarrow n_{optimal} = 0$ so $\theta$ is estimated from the prior i.e. $\delta^*(\underline{x}) = E(m) = \mu$ is now the Bayes rule for $\theta$.

$\forall t > 0$ $n_{optimal}$ is a decreasing function of $t$. If $t \to +\infty$ then the variance of the prior $\text{var}(\theta) = \frac{\beta}{t(\alpha-1)} \to 0$, and $n_{optimal} \xrightarrow{t \to +\infty} -\infty$. So in this case we have a fully prior knowledge for the mean $\theta$ and eventually we need no observations i.e. $n_{optimal} = 0$. Otherwise as $t \to 0$ $\text{var}(\theta) \to +\infty$ and the optimal sample size will increase to take the maximum value

$$n_{optimal} \xrightarrow{t \to 0} \left[ \frac{\beta}{c(\alpha - 1)} \right]^{\frac{1}{2}}$$

As we should expect $n_{optimal}$ is a decreasing function of $c$. More precisely:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

If $c \to +\infty$ then, $n_{optimal} \xrightarrow{c \to \infty} -t < 0$. In other words if an observation is very **expensive then we can not afford to take too many observations because the total** sampling cost becomes too high and it is better to make a decision about $\theta$ without taking any observations.

**If the optimal sample size turns** out to be a non-positive number hence, $n_{optimal} \leq 0$ **then it can be proved that**

$$n_{optimal} \leq 0 \Rightarrow \left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}} - t \leq 0 \Leftrightarrow \left[\frac{\beta}{c(\alpha-1)}\right]^{\frac{1}{2}} \leq t$$

$$\Leftrightarrow \frac{\beta}{c(\alpha-1)} \leq t^2 \Leftrightarrow \beta \leq c(\alpha-1)t^2$$

$$\Leftrightarrow c \geq \frac{\beta}{(\alpha-1)t^2}$$

Thus, if the cost per observation $c$ **is greater or equal** to the ratio $\frac{\beta}{t^2(\alpha-1)}$ the optimal sample size is a negative **quantity which means that** we should not proceed into a sampling strategy. The ratio $\frac{\beta}{t^2(\alpha-1)}$ increases as $\beta\uparrow$, $t\downarrow$ and $\alpha\downarrow$. **In that** case the prior distribution of $m$ becomes vague: $\text{var}(m) = \frac{\beta}{t(\alpha-1)}\uparrow$ and the **sampling cost** $c$ **increases very much that force us not** to pay for further **observations in order to gain some information about** $m$.

**Proposition 4.4** Suppose that $x_1, \ldots, x_n$ is a random sample from a normal distribution with unknown precision $\theta$, $\theta > 0$ and specified the value of the mean $\mu$, i.e. $f(x_i|\theta) \sim N\left(\mu, \frac{1}{\theta}\right)$. The prior distribution for $\theta$ is selected from the conjugate family, i.e. $\pi(\theta) \sim Ga(\alpha, \beta)$ with $\alpha, \beta > 0$ both known. If the value of $\theta$ is to be estimated under the squared error loss, i.e. $L(\theta, d) = (\theta - d)^2$ and the sampling cost per observation is $c$, $(c > 0)$ then the optimal number of observations $n$ is specified by the equation:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$n_{required} = \frac{\sqrt{2}}{\beta\sqrt{c}}\sqrt{\alpha(\alpha+1)} - 2(\alpha+1) \qquad (4.21)$$

**Proof**

**For every observation $x$ the density function is**

$f(x|\theta) = (2\pi)^{-\frac{1}{2}}\theta^{\frac{1}{2}}\exp\left\{-\frac{\theta}{2}(x-\mu)^2\right\}$ from proposition (3.1) we can write it in the form of an **exponential family:**

$f(x|\theta) = (2\pi)^{-\frac{1}{2}}\theta^{\frac{1}{2}}\exp\left\{-\frac{\theta}{2}(x-\mu)^2\right\}$ where: $h(x) = (2\pi)^{-\frac{1}{2}}$, $c(\theta) = \theta^{\frac{1}{2}}$, $w(\theta) = -\theta$,

$t(x) = \frac{1}{2}(x-\mu)^2$

The likelihood function from (3.2) becomes

$$f(\underset{\sim}{x}|\theta) = (2\pi)^{-\frac{n}{2}}\theta^{\frac{n}{2}}\exp\left\{-\frac{\theta}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

From (3.3) the conjugate prior density for $\theta$ :

$$\pi(\theta|\tau_0,\tau_1) = [k(\tau_0,\tau_1)]^{-1}\theta^{\frac{\tau_0}{2}}\exp\{-\theta\tau_1\} \qquad (4.22)$$

$$\pi(\theta|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp\{-\beta\theta\} \qquad (4.23)$$

**From these two last expressions** (4.22), (4.23) we obtain:

$$[k(\tau_0,\tau_1)]^{-1} = \frac{\beta^{\alpha}}{\Gamma(\alpha)},\ \frac{\tau_0}{2} = \alpha - 1 \Leftrightarrow \alpha = \frac{\tau_0}{2} + 1,\ \tau_1 = \beta$$

**The posterior distribution of the unknown value of the precision $\theta$ after having** observed $\underset{\sim}{x} = (x_1,\ldots,x_n)$ **is calculated directly from proposition (3.5)**

$$p(\theta|\underset{\sim}{x},\tau_0,\tau_1) = \pi(\theta|\tau_0',\tau_1') = \pi\left(\theta|n+\tau_0,\tau_1+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$= \left[k\left(n+\tau_0,\tau_1+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right]^{-1}\theta^{\frac{n+\tau_0}{2}}\exp\left\{-\theta\left(\tau_1+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right\}$$

where: $\frac{n+\tau_0}{2} = \alpha^* - 1 \Leftrightarrow \alpha^* = \frac{n+\tau_0}{2} + 1 = \frac{n+2\alpha-2+2}{2} \Leftrightarrow \alpha^* = \alpha + \frac{n}{2}$,



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\tau_1+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2=\beta^*\Leftrightarrow\beta^*=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

So it is proved that the posterior distribution of $\theta$ will be gamma with parameters $\alpha^*=\alpha+\frac{n}{2}$ and $\beta^*=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$. **Then the Bayes rule under square error loss is given by theorem 2.4 as:**

$$\delta^*(\underline{x})=E(\theta\,|\,\underline{x})=\frac{\alpha^*}{\beta^*}=\frac{\alpha+\frac{n}{2}}{\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

**and the Bayes risk also from 2.4 is defined to be**

$$r(\pi(\theta),\delta^*(\underline{x}))=E(\mathrm{var}(\theta\,|\,\underline{x}))=\mathrm{var}(\theta)-\mathrm{var}\left[E(\theta\,|\,\underline{x})\right]$$

$$=\frac{\alpha}{\beta^2}-\mathrm{var}\left[\frac{\alpha+\frac{n}{2}}{\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}\right]$$

$$=\frac{\alpha}{\beta^2}-\left(\alpha+\frac{n}{2}\right)^2\mathrm{var}\left[\frac{1}{\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}\right]\qquad(4.24)$$

Our next step is to define the density function of the random variable $Z=\frac{1}{Y}$ where $Y=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$. It is well known that:

$\sqrt{\theta}(x_i-\mu)\,|\,\theta\sim N(0,1)\Rightarrow\theta(x_i-\mu)^2\,|\,\theta\sim X_1^2\Rightarrow\theta\sum_{i=1}^{n}(x_i-\mu)^2\,|\,\theta\sim X_n^2=Ga\left(\frac{n}{2},\frac{1}{2}\right)$. We set know $\Lambda\,|\,\theta=\sum_{i=1}^{n}(x_i-\mu)^2\,|\,\theta$ and we are going to obtain the density function of $\Lambda\,|\,\theta$ when $\Lambda=\frac{1}{\theta}\mathrm{K}$ with $\mathrm{K}\,|\,\theta\sim X_n^2$ and $\Lambda>0$ since $\mathrm{K}>0$.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\Lambda=\frac{1}{\theta}\mathrm{K}\Leftrightarrow \mathrm{K}=\theta\Lambda=h^{-1}(\Lambda)\Rightarrow$$

$$f_{\Lambda|\theta}(\Lambda)=f_{\mathrm{K}|\theta}\left(h^{-1}(\Lambda)\right)\left|\frac{d}{d\Lambda}h^{-1}(\Lambda)\right|$$

$$=\frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}(\theta\Lambda)^{\frac{n}{2}-1}\exp\left\{-\frac{1}{2}\theta\Lambda\right\}\theta$$

$$=\frac{\left(\frac{\theta}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\Lambda^{\frac{n}{2}-1}\exp\left\{-\frac{\theta}{2}\Lambda\right\}\Rightarrow\Lambda\,|\,\theta\sim Ga\left(\frac{n}{2},\frac{\theta}{2}\right)$$

We know define the marginal probability density function of $\Lambda=\sum_{i=1}^{n}(x_i-\mu)^2$

$$f(\Lambda)=\int_{\Theta}f(\Lambda\,|\,\theta)\pi(\theta)d\theta$$

$$=\int_{\Theta}\frac{\left(\frac{\theta}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\Lambda^{\frac{n}{2}-1}\exp\left\{-\frac{\theta}{2}\Lambda\right\}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp\{-\beta\theta\}d\theta$$

$$=\frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\Lambda^{\frac{n}{2}-1}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_{\Theta}\theta^{\left(\alpha+\frac{n}{2}\right)-1}\exp\left\{-\theta\left(\beta+\frac{\Lambda}{2}\right)\right\}d\theta$$

$$=\frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\Lambda^{\frac{n}{2}-1}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma\left(\alpha+\frac{n}{2}\right)}{\left(\beta+\frac{\Lambda}{2}\right)^{\alpha+\frac{n}{2}}}$$

. **Next we proceed to obtain the distribution of the random variable**

$Y=\beta+\frac{1}{2}\Lambda\Leftrightarrow\Lambda=2Y-2\beta=h^{-1}(Y)$ **where** $Y>\beta$ **and following the same well known strategy we have:**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$f_Y(y) = f_\Lambda\left(h^{-1}(y)\right)\left|\frac{d}{d\Lambda}h^{-1}(y)\right|$$

$$= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\frac{\beta^\alpha}{\Gamma(\alpha)}\Gamma\left(\alpha+\frac{n}{2}\right)\frac{(2y-2\beta)^{\frac{n}{2}-1}2}{\left(\beta+\frac{2y-2\beta}{2}\right)^{\alpha+\frac{n}{2}}}$$

$$= \frac{\beta^\alpha\Gamma\left(\alpha+\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma(\alpha)}\frac{(y-\beta)^{\frac{n}{2}-1}}{y^{\alpha+\frac{n}{2}}}$$

As it has been shown in the proof **of proposition 4.1, if a random variable** $Y$ **has** density function given by the form (4.5), then the random variable $Z=\frac{1}{Y}$ **with** $Z\in\left(0,\frac{1}{\beta}\right)$ **has density function defined by** (4.6). The Bayes risk function also when $\theta\sim Ga(\alpha,\beta)$ **is given by** (4.7). If we conceder our case as a special case of $Z=\frac{1}{Y}$ when $k=\frac{1}{2}$, **then the Bayes risk defined in (4.24) is transformed using** (4.7) **into the form**

$$r_{Bayes}(\pi(\theta),\delta^*(\underset{\sim}{x})) = \frac{\alpha}{\beta^2}+\frac{\alpha^2}{\beta^2}-\frac{\alpha(\alpha+1)}{\beta^2}\frac{\left(\alpha+\frac{n}{2}\right)}{\left(\alpha+\frac{n}{2}+1\right)} \qquad (4.25)$$

**and the corresponding total risk**

$$r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x})) = \frac{\alpha}{\beta^2}+\frac{\alpha^2}{\beta^2}-\frac{\alpha(\alpha+1)}{\beta^2}\frac{\left(\alpha+\frac{n}{2}\right)}{\left(\alpha+\frac{n}{2}+1\right)}+cn \qquad (4.26)$$

**Equivalently by minimizing (4.26) with respect** to $n$ and taking the second **derivative we obtain that the optimal sample size is:**

$$n_{optimal} = \frac{\sqrt{2}}{\beta\sqrt{c}}\sqrt{\alpha(\alpha+1)}-2(\alpha+1)$$

This type is derived as a special case of (4.1) for $k=\frac{1}{2}$.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**Comments on the behavior of the optimal sample size as a function of its parameters**

Based on the discussion on $n_{optimal}$ defined in proposition 4.1 we have the following conclusions about $n_{optimal} = \frac{\sqrt{2}}{\beta\sqrt{c}}\sqrt{\alpha(\alpha+1)} - 2(\alpha+1) = f(\alpha,\beta,c)$

$n_{optimal}$ will be a decreasing function of $\beta$. For $\beta \to +\infty$ the variance of the prior distribution of $\theta$, $\frac{\alpha}{\beta^2} \to 0$ and $n_{optimal} \xrightarrow{\beta\to+\infty} -2(\alpha+1) < 0$ so if our prior distribution is very informative and no observations needs to be taken i.e. $n_{optimal} = 0$. If $\beta \to 0$ then $\frac{\alpha}{\beta^2} \to +\infty$ and $n_{optimal} \xrightarrow{\beta\to 0} +\infty$. So if our prior distribution is non-informative a large number of observations must be taken to estimate the unknown value of the parameter $\theta$.

$\forall c > 0$ $n_{optimal}$ is a decreasing function of $c$. As $c$ increases the optimal sample size $n_{optimal}$ decreases. In other words if the cost for every observation is very large then we cannot afford to take many observations because the total sampling cost becomes too high. If $c \to +\infty$ then $n_{optimal} \xrightarrow{c\to\infty} -2(\alpha+1) < 0$ i.e. $n_{optimal} = 0$. For $c \to 0$, $n_{optimal} \xrightarrow{c\to 0} +\infty$

$\forall \alpha > 0$ we can distinguish two different cases

1. For $c > \frac{1}{2\beta^2}$, $f(\alpha,\beta,c)$ in the interval $(0,\alpha_2)$ is an increasing function of $\alpha$ and in the interval $(\alpha_2,+\infty)$, $f(\alpha,\beta,c)$ is a decreasing function of $\alpha$. For $\alpha = \alpha_2$ $n_{optimal}$ takes its maximum value. We have to mention here that $\alpha_2$ is defined by the quantity $\alpha_2 = \frac{4\left(\frac{1}{2} - c\beta^2\right) + \sqrt{-16c\beta^2\left(\frac{1}{2} - c\beta^2\right)}}{8\left(c\beta^2 - \frac{1}{2}\right)} > 0$. Which may be found by (4.9) if we replace the value of $k$ with $\frac{1}{2}$

2. For $c < \frac{1}{2\beta^2}$, then for every $\alpha > 0$ $f(\alpha,\beta,c)$ is an increasing function of $\alpha$.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

In case that $n_{optimal} = \frac{\sqrt{2}}{\beta\sqrt{c}}\sqrt{\alpha(\alpha+1)} - 2(\alpha+1) \le 0$ then, $c \ge \frac{\alpha}{2\beta^2(\alpha+1)}$

When the cost per observation $c$ becomes greater or equal to the ratio $\frac{\alpha}{2\beta^2(\alpha+1)}$ then $n_{optimal}$ becomes negative. Ratio $\frac{\alpha}{2\beta^2(\alpha+1)}$ increases as $\alpha$ increases and $\beta$ decreases. In that case the prior variance $\frac{\alpha}{\beta^2}$ becomes very large, so the prior distribution of $\theta$ is not informative at all and it is preferable that no observations should be taken because the sampling cost $c$ is very high for us to pay.

## 4.4 Optimal sample for Poisson sample distribution

**Proposition 4.5:** Suppose that $x_1,.....,x_n$ is a random sample from a Poisson distribution for which the value of the mean $\theta$ is unknown$(\theta > 0)$, where $f(x_i \mid \theta) = \exp\{-\theta\}\theta^{x_i}\frac{1}{x_i!}$. The prior distribution of the mean $\theta$ is selected from the conjugate family and is a gamma distribution with parameters $\alpha, \beta > 0$, $\pi(\theta) \sim Gamma(\alpha, \beta)$. The value of $\theta$ is desired to be estimated under the squared error loss, i.e. $L(\theta, d) = (\theta - d)^2$. If the sampling cost per observation is $c, (c > 0)$ then the optimal number observations $n$ is specified by the equation:

$$n_{optimal} = \left[\frac{\alpha}{c\beta}\right]^{\frac{1}{2}} - \beta \qquad (4.27)$$

**Proof:**

For every observation $x$ the density function is defined by



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$f(x|\theta)=\frac{1}{x!}\theta^x\exp\{-\theta\}$, from proposition (3.1) this could be written in the form

$$f(x|\theta)=\frac{1}{x!}\exp\{-\theta\}\exp\{x\ln\theta\}$$

where: $h(x)=\frac{1}{x!}$, $c(\theta)=\exp\{-\theta\}$, $w(\theta)=\ln\theta$, $t(x)=x$

**The corresponding likelihood from (3.2) is formed as**

$$f(\underset{\sim}{x}|\theta)=\frac{1}{\prod_{i=1}^{n}x_i!}\exp\{-n\theta\}\exp\left\{(\ln\theta)\sum_{i=1}^{n}x_i\right\}$$

From (3.3) the conjugate prior density for $\theta$ :

$$\pi(\theta|\tau_0,\tau_1)=[k(\tau_0,\tau_1)]^{-1}\exp\{-\tau_0\theta\}\exp\{(\ln\theta)\tau_1\} \qquad (4.28)$$

$$\pi(\theta|\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\exp\{-\beta\theta\}=\frac{\beta^\alpha}{\Gamma(\alpha)}\exp\{(\alpha-1)\ln\theta\}\exp\{-\beta\theta\} \qquad (4.29)$$

**under the following condition**

$\pi(\theta|\tau_0,\tau_1)=Ga(\alpha,\beta)=\pi(\theta|\alpha,\beta)$

We derive from expressions **(4.28)** and (4.29)

$$[k(\tau_0,\tau_1)]^{-1}=\frac{\beta^\alpha}{\Gamma(\alpha)},\ \tau_0=\beta,\ \tau_1=\alpha-1$$

**The posterior distribution of $\theta$ is specified from proposition (3.5)**

$$p(\theta|\underset{\sim}{x},\tau)=\pi(\theta|\tau_0',\tau_1')=\pi(\theta|\tau_0+n,\tau_1+\sum_{i=1}^{n}x_i)$$

$$=\left[k(\tau_0+n,\tau_1+\sum_{i=1}^{n}x_i)\right]^{-1}\left[\exp\{-\theta\}\right]^{\tau_0+n}\exp\left\{(\ln\theta)\left(\tau_1+\sum_{i=1}^{n}x_i\right)\right\}$$

where: $\tau_0'=\beta^*\Leftrightarrow\beta^*=\beta+n$ **and**

$$\tau_1'=\alpha^*-1\Leftrightarrow\alpha^*=\tau_1'+1\Leftrightarrow\alpha^*=\alpha+\sum_{i=1}^{n}x_i-1+1\Leftrightarrow\alpha^*=\alpha+\sum_{i=1}^{n}x_i$$

It is then proved that the posterior distribution of $\theta$ after having observed $x_1,.....,x_n$ is a gamma distribution with parameters $\alpha^*=\sum_{i=1}^{n}x_i+\alpha$, $\beta^*=n+\beta$

**Accordingly to the theorem 2.4 we derive that the Bayes estimator is given by the expression**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\delta^*(\underset{\sim}{x}) = E(\theta \mid \underset{\sim}{x}) = \frac{\alpha^*}{\beta^*} = \frac{\sum_{i=1}^{n} x_i + \alpha}{n+\beta}$$

and the corresponding Bayes risk function:

$$r(\pi(\theta), \delta^*(\underset{\sim}{x})) = E\left[\text{var}(\theta \mid \underset{\sim}{x})\right] = E\left(\frac{\alpha^*}{\beta^{*2}}\right) = E\left[\frac{\alpha + \sum_{i=1}^{n} x_i}{(n+\beta)^2}\right]$$

$$= \frac{1}{(n+\beta)^2} E\left(\alpha + \sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{(n+\beta)^2}\left[\alpha + nE(x_i)\right] \qquad (4.30)$$

We know that $E(x_i) = E\left[E(x_i \mid \theta)\right] = E(\theta) = \frac{\alpha}{\beta}$, so (4.30) becomes

$$r(\pi(\theta), \delta^*(\underset{\sim}{x})) = \frac{1}{(n+\beta)^2}\left[\alpha + n\frac{\alpha}{\beta}\right]$$

$$= \frac{1}{(n+\beta)^2}\left[\frac{\alpha\beta + n\alpha}{\beta}\right]$$

$$= \frac{\alpha(\beta+n)}{\beta(n+\beta)^2}$$

$$= \frac{\alpha}{\beta(n+\beta)}$$

From (2.11) the total risk function will be:

$$r_{total}(\pi(\theta), \delta^*(\underset{\sim}{x})) = \frac{\alpha}{\beta(n+\beta)} + nc \qquad (4.31)$$

In order to obtain the optimal sample size we minimize (4.31) with respect to $n$

$$\frac{\partial r_{total}(\pi(\theta), \delta_{Bayes}(\underset{\sim}{x}))}{\partial n} = 0 \Leftrightarrow \frac{\partial}{\partial n}\left[\frac{\alpha}{\beta(n+\beta)}\right] + c = 0$$

$$-\frac{\alpha}{\beta(n+\beta)^2} + c = 0 \Leftrightarrow -\alpha + c\beta(n+\beta)^2 = 0$$

$$c\beta(n+\beta)^2 = \alpha \Leftrightarrow (n+\beta)^2 = \frac{\alpha}{c\beta}$$

$$(n+\beta) = \left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}} \Leftrightarrow n = \left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}} - \beta$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

For the second derivative with respect to $n$ we will have:

$$\frac{\partial^2 r_{total}(\pi(\theta),\delta_{Bayes}(\underset{\sim}{x}))}{\partial n^2}=\frac{\partial}{\partial n}\left[-\frac{\alpha}{\beta(n+\beta)^2}+c\right]$$

$$=\frac{2\alpha}{\beta(n+\beta)^3}>0 \quad \forall n>0$$

Therefore the value of $n$ found above will correspond to a minimum and the optimal sample size in this case will be given by the formula:

$$n_{optimal}=\left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}}-\beta$$

We will refer first to the case where $n_{optimal}$ is a positive integer i.e. $n_{optimal}>0$. For this we consider a function $f$ such that $n_{optimal}=f(\alpha,\beta,c)$. The behaviour of this function with respect to $\alpha$, $\beta$, $c$ can be studied through the partial derivatives of $f$

$$\frac{\partial f(\alpha,\beta,c)}{\partial \alpha}=\frac{\partial}{\partial \alpha}\left[\left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}}-\beta\right]=\left(\frac{1}{c\beta}\right)^{\frac{1}{2}}\frac{\partial}{\partial \alpha}\sqrt{\alpha}$$

$$=\left(\frac{1}{c\beta}\right)^{\frac{1}{2}}\frac{1}{2\sqrt{\alpha}}=\left(\frac{1}{4c\alpha\beta}\right)^{\frac{1}{2}} \quad (>0)$$

$$\frac{\partial f(\alpha,\beta,c)}{\partial \beta}=\frac{\partial}{\partial \beta}\left[\left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}}-\beta\right]=\frac{\partial}{\partial \beta}\left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}}-\frac{\partial}{\partial \beta}\beta$$

$$=\left(\frac{\alpha}{c}\right)^{\frac{1}{2}}\frac{\partial}{\partial \beta}\frac{1}{\sqrt{\beta}}-1=-\left(\frac{\alpha}{c}\right)^{\frac{1}{2}}\frac{1}{2\beta\sqrt{\beta}}-1$$

$$=-\left(\frac{\alpha}{c}\right)^{\frac{1}{2}}\frac{1}{2\beta^{\frac{3}{2}}}-1=-\left(\frac{\alpha}{4c\beta^3}\right)^{\frac{1}{2}}-1 \quad (<0)$$

$$\frac{\partial f(\alpha,\beta,c)}{\partial c}=\frac{\partial}{\partial c}\left[\left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}}-\beta\right]=\frac{\partial}{\partial c}\left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}}$$

$$=\left(\frac{\alpha}{\beta}\right)^{\frac{1}{2}}\frac{\partial}{\partial c}\frac{1}{\sqrt{c}}=\left(\frac{\alpha}{\beta}\right)^{\frac{1}{2}}\left(-\frac{1}{2c\sqrt{c}}\right)=-\left(\frac{\alpha}{4\beta c^3}\right)^{\frac{1}{2}} \quad (<0)$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

We end up to the following conclusions

$n_{optimal}$ is an increasing function of $a$. Hence, as $\alpha \to +\infty$, $\text{var}(\theta) = \frac{\alpha}{\beta^2} \to +\infty$ and $n_{optimal} \xrightarrow{\alpha \to +\infty} +\infty$. So when the prior distribution of $\theta$ is vague then the optimal sample **size becomes extremely** large and as a **result we need more observations** for **valid inference**

On the other hand as $\alpha \to 0$, $\frac{\alpha}{\beta^2} \to 0$ and $n_{optimal} \xrightarrow{\alpha \to 0} -\beta < 0$. We see that when the prior distribution of $\theta$ is very informative $n_{optimal}$ is negative so no observations are taken i.e. $n_{optimal} = 0$ and the value of the parameter $\theta$ is estimated directly from **the prior.**

$n_{optimal}$ is a decreasing function of $\beta$. For $\beta \to 0$, $\frac{\alpha}{\beta^2} \to +\infty$ and $n_{optimal} \xrightarrow{\beta \to 0} +\infty$.

When the prior distribution of $\theta$ is vague then the optimal sample size becomes very large. For $\beta \to +\infty$, $\frac{\alpha}{\beta^2} \to 0$ and $n_{optimal} \xrightarrow{\beta \to +\infty} -\infty < 0$. I.e. if the prior distribution of $\theta$ is very informative no observations are taken $\left(n_{optimal} < 0 \Rightarrow\right)$ $n_{optimal} = 0$ and the value of the parameter $\theta$ is estimated from the prior.

$n_{optimal}$ is a decreasing function of $c$. As $c \to 0$ then $n_{optimal} \to +\infty$, so if the cost per observation $c$ is very small, then the optimal sample size increases very much. Otherwise as $c \to +\infty$ then $n_{optimal} \to -\beta < 0$ which means that it is better not to take any observations in order to derive further information for $\theta$ from the posterior distribution because the total sampling cost will be too high.

If the optimal sample is negative then we end up to the following assumption:

$$n_{optimal} \le 0 \Rightarrow \left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}} - \beta \le 0 \Leftrightarrow \left(\frac{\alpha}{c\beta}\right)^{\frac{1}{2}} \le \beta$$

$$\Leftrightarrow \frac{\alpha}{c\beta} \le \beta^2 \Leftrightarrow c \ge \frac{\alpha}{\beta^3}$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

When the cost per observation $c$ becomes greater or equal to the ratio $\frac{\alpha}{\beta^3}$, then $n_{optimal}$ becomes negative. Ratio $\frac{\alpha}{\beta^3}$ increases as $\alpha(\uparrow)$ and $\beta(\downarrow)$. In this situation prior variance $\frac{\alpha}{\beta^2}$ becomes very large, so the prior distribution of $\theta$ is not informative at all and it is preferable no observations to be taken because sampling cost $c$ is way too expensive.

## 4.5 Optimal sample for Binomial distribution

**Proposition 4.6:** Suppose that $x_1,.....,x_n$ is a random sample from Binomial distribution with parameters: $k>0$ specified and $\theta$ unknown, $0<\theta<1$, i.e. $f(x_i|\theta) \sim B(k,\theta)$.

For the prior distribution of the parameter $\theta$ we choose from the conjugate family the beta distribution with parameters $\alpha, \beta>0$ $\pi(\theta) \sim Be(\alpha,\beta)$ . We are interested in estimating the value of $\theta$ under squared error loss i.e. $L(\theta,d)=(\theta-d)^2$. If the sampling cost per observation is $c,(c>0)$ then the optimal number observations $n$ is specified by the equation:

$$n_{optimal}=\left[\frac{\alpha\beta}{ck(\alpha+\beta)(\alpha+\beta+1)}\right]^{\frac{1}{2}}-\frac{\alpha+\beta}{k} \qquad (4.32)$$

**Proof:**

For every random observation $x$ the density function is defined

$$f(x|\theta)=\binom{k}{x}\theta^x(1-\theta)^{k-x}$$

The above expression can also be written according to the proposition (3.1) as:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$f(x|\theta)=\binom{k}{x}\theta^{x}(1-\theta)^{k-x}$$
$$=\binom{k}{x}(1-\theta)^{k}\left(\frac{\theta}{1-\theta}\right)^{x}$$
$$=\binom{k}{x}(1-\theta)^{k}\exp\left\{x\ln\left(\frac{\theta}{1-\theta}\right)\right\}$$

Where: $h(x)=\binom{k}{x}$, $c(\theta)=(1-\theta)^{k}$, $t(x)=x$ and $w(\theta)=\ln\left(\frac{\theta}{1-\theta}\right)$

**Calculating the likelihood from (3.2) provides:**

$$f(\underset{\sim}{x}|\theta)=\left[\prod_{i=1}^{n}\binom{k}{x_i}\right](1-\theta)^{nk}\exp\left\{\left(\ln\left(\frac{\theta}{1-\theta}\right)\right)\sum_{i=1}^{n}x_i\right\}$$

From (3.3) the conjugate prior density for $\theta$:

$$\pi(\theta|\tau_0,\tau_1)=\left[k(\tau_0,\tau_1)\right]^{-1}(1-\theta)^{\tau_0}\exp\left\{\left(\ln\left(\frac{\theta}{1-\theta}\right)\right)\tau_1\right\} \qquad (4.33)$$
$$=\left[k(\tau_0,\tau_1)\right]^{-1}\theta^{\tau_1}(1-\theta)^{\tau_0-\tau_1}$$

$$\pi(\theta|\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad (4.34)$$

Where $\pi(\theta|\tau_0,\tau_1)=Be(\alpha,\beta)=\pi(\theta|\alpha,\beta)$ and so from expressions (4.33), (4.34) **which are equivalent we will have**

$$\left[k(\tau_0,\tau_1)\right]^{-1}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)},\ \tau_1=\alpha-1,\ \beta-1=\tau_0-\tau_1$$

**The posterior distribution of $\theta$ from proposition (3.5) will be:**

$$p(\theta|\underset{\sim}{x},\tau_0,\tau_1)=\pi(\theta|\tau_0+n,\tau_1+\sum_{i=1}^{n}x_i)$$
$$=\left[k(\tau_0+n,\tau_1+\sum_{i=1}^{n}x_i)\right]^{-1}(1-\theta)^{nk+\tau_0}\exp\left\{\left(\ln\left(\frac{\theta}{1-\theta}\right)\right)\left(\tau_1+\sum_{i=1}^{n}x_i\right)\right\}$$
$$=\left[k(\tau_0+n,\tau_1+\sum_{i=1}^{n}x_i)\right]^{-1}(1-\theta)^{nk+\tau_0-\left(\tau_1+\sum_{i=1}^{n}x_i\right)}\theta^{\tau_1+\sum_{i=1}^{n}x_i}$$

where, $\tau_1+\sum_{i=1}^{n}x_i=\alpha^*-1\Leftrightarrow\alpha-1+\sum_{i=1}^{n}x_i=\alpha^*-1\Leftrightarrow\alpha^*=\alpha+\sum_{i=1}^{n}x_i$ and

$$\beta^*-1=nk+\tau_0-\tau_1-\sum_{i=1}^{n}x_i\Leftrightarrow\beta^*-1=nk+\beta-1-\sum_{i=1}^{n}x_i\Leftrightarrow\beta^*=nk+\beta-\sum_{i=1}^{n}x_i$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

so the posterior distribution of $\theta$ after having observed $x_1,.....,x_n$ is also a Beta distribution with parameters $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$ and $\beta^* = nk + \beta - \sum_{i=1}^{n} x_i$.

The Bayes rule **under square error loss function is defined from theorem 2.4 to be**

$$\delta^*(\underset{\sim}{x}) = E(\theta \mid \underset{\sim}{x}) = \frac{\alpha^*}{\alpha^* + \beta^*} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \sum_{i=1}^{n} x_i + kn + \beta - \sum_{i=1}^{n} x_i} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + kn}$$

**and the Bayes risk function also from 2.4**

$$\begin{aligned} r(\pi(\theta), \delta^*(\underset{\sim}{x})) &= E\left[\operatorname{var}(\theta \mid \underset{\sim}{x})\right] = \operatorname{var}(\theta) - \operatorname{var}\left[E(\theta \mid \underset{\sim}{x})\right] \\ &= \operatorname{var}(\theta) - \operatorname{var}\left(\frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + kn}\right) \\ &= \operatorname{var}(\theta) - \frac{1}{(\alpha + \beta + kn)^2} \operatorname{var}\left(\sum_{i=1}^{n} x_i\right) \quad (4.35) \end{aligned}$$

The conditional distribution of $\sum_{i=1}^{n} x_i \mid \theta$ **will be also Binomial i.e.** $\sum_{i=1}^{n} x_i \mid \theta \sim B(nk, \theta)$

with $E\left(\sum_{i=1}^{n} x_i \mid \theta\right) = nk\theta$ and $\operatorname{var}\left(\sum_{i=1}^{n} x_i \mid \theta\right) = nk\theta(1-\theta)$. Hence,

$$\begin{aligned} \operatorname{var}\left(\sum_{i=1}^{n} x_i\right) &= E\left[\operatorname{var}\left(\sum_{i=1}^{n} x_i \mid \theta\right)\right] + \operatorname{var}\left[E\left(\sum_{i=1}^{n} x_i \mid \theta\right)\right] \\ &= E\left[nk\theta(1-\theta)\right] + \operatorname{var}(nk\theta) = nk\left[E(\theta) - E(\theta^2)\right] + n^2k^2 \operatorname{var}(\theta) \\ &= nk\left[E(\theta) - \operatorname{var}(\theta) - \left(E(\theta)\right)^2\right] + n^2k^2 \operatorname{var}(\theta) \\ &= -nk \operatorname{var}(\theta) + n^2k^2 \operatorname{var}(\theta) + nkE(\theta) - nk\left(E(\theta)\right)^2 \\ &= nk \operatorname{var}(\theta)(nk-1) + nkE(\theta)\left[1 - E(\theta)\right] \end{aligned}$$

Replacing in (4.35) we will get:



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$r(\pi(\theta),\delta^*(\underset{\sim}{x})) = \text{var}(\theta) - \frac{1}{(\alpha+\beta+nk)^2}\left\{nk\,\text{var}(\theta)(nk-1) + nkE(\theta)\left[1-E(\theta)\right]\right\}$$

$$= \text{var}(\theta) - \frac{nk(nk-1)}{(\alpha+\beta+nk)^2}\text{var}(\theta) - \frac{nk}{(\alpha+\beta+nk)^2}E(\theta)\left[1-E(\theta)\right]$$

$$= \text{var}(\theta)\left[1-\frac{nk(nk-1)}{(\alpha+\beta+nk)^2}\right] - \frac{nk}{(\alpha+\beta+nk)^2}E(\theta)\left[1-E(\theta)\right]$$

$$= \text{var}(\theta)\left[1-\frac{n^2k^2-nk}{(\alpha+\beta+nk)^2}\right] - \frac{nk}{(\alpha+\beta+nk)^2}E(\theta)\left[1-E(\theta)\right] \qquad (4.36)$$

**From (2.11) the total risk is specified as**

$$r_{total}(\pi(\theta),\delta^*(\underset{\sim}{x})) = \text{var}(\theta)\left[1-\frac{n^2k^2-nk}{(\alpha+\beta+nk)^2}\right] - \frac{nk}{(\alpha+\beta+nk)^2}E(\theta)\left[1-E(\theta)\right] + cn \qquad (4.37)$$

The mean **and the variance of the prior distribution are given** by the formulas

$$\text{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$E(\theta) = \frac{\alpha}{\alpha+\beta}$$

Minimizing (4.26) with respect to $n$ we have:

$$\frac{\partial}{\partial n}\left\{\text{var}(\theta)\left[1-\frac{n^2k^2-nk}{(\alpha+\beta+nk)^2}\right] - \frac{nk}{(\alpha+\beta+nk)^2}E(\theta)\left[1-E(\theta)\right]\right\} + c = 0 \qquad (4.38)$$

$$\frac{\partial}{\partial n}\left\{\text{var}(\theta)\left[1-\frac{n^2k^2-nk}{(\alpha+\beta+nk)^2}\right] - \frac{nk}{(\alpha+\beta+nk)^2}E(\theta)\left[1-E(\theta)\right]\right\}$$

$$= \text{var}(\theta)\left[-\frac{(2nk^2-k)(\alpha+\beta+nk)^2 - 2k(n^2k^2-nk)(\alpha+\beta+nk)}{(\alpha+\beta+nk)^4}\right]$$

$$-E(\theta)\left[1-E(\theta)\right]\frac{k(\alpha+\beta+nk)^2 - 2nk(\alpha+\beta+nk)k}{(\alpha+\beta+nk)^4}$$

$$= -\text{var}(\theta)\frac{\left[(2nk^2-k)(\alpha+\beta+nk) - 2k(n^2k^2-nk)\right]}{(\alpha+\beta+nk)^3}$$

$$-E(\theta)\left[1-E(\theta)\right]\frac{(k\alpha+k\beta+nk^2-2nk^2)}{(\alpha+\beta+nk)^3}$$

$$= -\text{var}(\theta)\frac{2\alpha nk^2+2\beta nk^2+2n^2k^3-k\alpha-k\beta-nk^2-2n^2k^3+2nk^2}{(\alpha+\beta+nk)^3}$$

$$-E(\theta)\left[1-E(\theta)\right]\frac{k\alpha+k\beta-nk^2}{(\alpha+\beta+nk)^3}$$

$$= -\text{var}(\theta)\frac{2\alpha nk^2+2\beta nk^2-k\alpha-k\beta+nk^2}{(\alpha+\beta+nk)^3} - E(\theta)\left[1-E(\theta)\right]\frac{k\alpha+k\beta-nk^2}{(\alpha+\beta+nk)^3}$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

We replace now the mean and the variance of the prior and so we obtain

$$=-\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\frac{2\alpha nk^2+2\beta nk^2-k\alpha-k\beta+nk^2}{(\alpha+\beta+nk)^3}-\frac{\alpha}{\alpha+\beta}\left(1-\frac{\alpha}{\alpha+\beta}\right)\frac{k\alpha+k\beta-nk^2}{(\alpha+\beta+nk)^3}$$

$$=-\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\frac{2\alpha nk^2+2\beta nk^2-k\alpha-k\beta+nk^2}{(\alpha+\beta+nk)^3}-\frac{\alpha\beta}{(\alpha+\beta)^2}\frac{k\alpha+k\beta-nk^2}{(\alpha+\beta+nk)^3}$$

$$=-\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+nk)^3}\left[\frac{2\alpha nk^2+2\beta nk^2-k\alpha-k\beta+nk^2+(\alpha+\beta+1)(k\alpha+k\beta-nk^2)}{\alpha+\beta+1}\right]$$

$$=-\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+nk)^3}\left(\frac{2\alpha nk^2+2\beta nk^2-k\alpha-k\beta+nk^2+k\alpha^2+k\alpha\beta-\alpha nk^2}{\alpha+\beta+1}\right.$$

$$\left.+\frac{k\alpha\beta+k\beta^2-\beta nk^2+k\alpha+k\beta-nk^2}{\alpha+\beta+1}\right)$$

$$=-\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+nk)^3}\frac{\alpha nk^2+\beta nk^2+k\alpha^2+2k\alpha\beta+k\beta^2}{\alpha+\beta+1}$$

$$=-\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+nk)^3}\frac{(\alpha+\beta)(nk^2+k\alpha+k\beta)}{\alpha+\beta+1}$$

$$=-\frac{k\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+nk)^2}$$

Replacing in relation (4.38) we prove that $n_{optimal}$ will be the solution of the following equation

$$-\frac{k\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+nk)^2}+c=0$$

$$c(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+nk)^2-k\alpha\beta=0$$

$$(\alpha+\beta+nk)^2=\frac{k\alpha\beta}{c(\alpha+\beta+1)(\alpha+\beta)}\Leftrightarrow(\alpha+\beta+nk)=\left[\frac{k\alpha\beta}{c(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}$$

$$nk=\left[\frac{k\alpha\beta}{c(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}-(\alpha+\beta)\Leftrightarrow n=\left[\frac{\alpha\beta}{ck(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}-\frac{\alpha+\beta}{k}$$

Taking the second derivative of the total risk function we obtain:

$$\frac{\partial^2 r_{total}}{\partial n^2}=\frac{\partial}{\partial n}\left[-\frac{k\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+nk)^2}+c\right]$$

$$=\frac{2k^2\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+nk)^3}>0\quad\forall n>0$$

Thus, the optimal sample size is defined by:

$$n_{optimal}=\left[\frac{\alpha\beta}{ck(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}-\frac{\alpha+\beta}{k}$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

If we consider $n_{optimal}$ to be a positive integer, then:

$$\frac{\partial n_{optimal}}{\partial c}=\frac{\partial}{\partial c}\left\{\left[\frac{\alpha\beta}{ck(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}-\frac{\alpha+\beta}{k}\right\}=\left[\frac{\alpha\beta}{k(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}\frac{\partial}{\partial c}\frac{1}{\sqrt{c}}$$

$$=-\frac{1}{2}\left[\frac{\alpha\beta}{kc^{3}(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}<0$$

It is proved then, that $\forall c>0$ $n_{optimal}$ **is a decreasing function of the sampling cost** and more specifically $n_{optimal}\xrightarrow{c\to 0}+\infty$ and $n_{optimal}\xrightarrow{c\to\infty}-\frac{\alpha+\beta}{k}<0$ i.e. $n_{optimal}=0$.

**Due to the complexity of the partials first derivatives with respect to the** parameters $\alpha,\beta,k$ it is rather difficult for us to derive some exact conclusions about the behavior of $n_{optimal}$ function corresponding to these parameters.

In the case that $n_{optimal}\le 0$ we derive the following:

$$\left[\frac{\alpha\beta}{ck(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}-\frac{\alpha+\beta}{k}\le 0\Leftrightarrow\left[\frac{\alpha\beta}{ck(\alpha+\beta+1)(\alpha+\beta)}\right]^{\frac{1}{2}}\le\frac{\alpha+\beta}{k}$$

$$\Leftrightarrow\frac{\alpha\beta}{ck(\alpha+\beta+1)(\alpha+\beta)}\le\left(\frac{\alpha+\beta}{k}\right)^{2}\Leftrightarrow c\ge\frac{\alpha\beta k}{(\alpha+\beta+1)(\alpha+\beta)^{3}}$$

**We observe from this last inequality that as the cost per sampling unit becomes** greater or equal than the ratio $\frac{\alpha\beta k}{(\alpha+\beta+1)(\alpha+\beta)^{3}}$, $n_{optimal}$ becomes negative and we **decide that the observations are way too expensive, so we** do not draw a sample.

**Corollary 4.2** Suppose that our random observations $x_1,.....,x_n$ come from a Bernoulli distribution i.e. $f(\underset{\sim}{x}|\theta)\sim B(1,\theta)$. If we are interested to make an inference about $\theta$ **using as a prior the Beta distribution and the square error loss function as they have been defined in Proposition 4.6 then it is proved that if the** sampling cost per observation is $c,(c>0)$ the optimal number observations $n$ is **specified by the equation:**

$$n_{optimal}=\left[\frac{\alpha\beta}{c(\alpha+\beta)(\alpha+\beta+1)}\right]^{\frac{1}{2}}-(\alpha+\beta)\qquad(4.39)$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

The formula (4.39) can be derived if we replace in relation (4.32) $k=1$, since we may consider the Bernoulli distribution as a special case of the Binomial when the value of $k$ is specified to be 1.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ΒΙΒΛΙΟΘΗΚΗ

# Summary of Basic Formulae

In this section the following table is provided for reference. This records the sampling distribution for each one of the statistical models which are studied in **Chapter 4**, the conjugate prior and the corresponding optimal sample size which is suggested to be taken through the Bayesian approach, at each separate case, when square error loss function is used.

## Table of optimal sample sizes

### Discrete sampling distributions

***Bernoulli model***

$\underset{\sim}{x} = (x_1, \ldots, x_n)$, $\quad x_i \in \{0,1\}$

$p(x_i \mid \theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $\quad 0 < \theta < 1$

***Conjugate prior*** : $\quad p(\theta) = Be(\theta \mid \alpha, \beta)$, $\quad \alpha > 0, \beta > 0$

***Optimal sample*** : $\quad n_{optimal} = \left[\frac{\alpha\beta}{c(\alpha+\beta)(\alpha+\beta+1)}\right]^{\frac{1}{2}} - (\alpha+\beta)$

***Binomial model***

$\underset{\sim}{x} = (x_1, \ldots, x_n)$, $\quad x_i \in \{0,1,\ldots,k\}$

$p(x_i \mid \theta) = \binom{k}{x_i}\theta^{x_i}(1-\theta)^{k-x_i}$, $\quad 0 < \theta < 1$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

*Conjugate prior* : $p(\theta) = Be(\theta \mid \alpha, \beta),\ \alpha > 0, \beta > 0$

*Optimal sample* : $n_{optimal} = \left[\frac{\alpha\beta}{ck(\alpha+\beta)(\alpha+\beta+1)}\right]^{\frac{1}{2}} - \frac{\alpha+\beta}{k}$

---

***Poisson model***

$\underset{\sim}{x} = (x_1, \ldots, x_n),\qquad x_i \in \{0, 1, \ldots\}$

$p(x_i \mid \theta) = \exp\{-\theta\}\frac{\theta^{x_i}}{x_i!},\quad \theta \geq 0$

*Conjugate prior* : $p(\theta) = Ga(\theta \mid \alpha, \beta),\ \alpha > 0, \beta > 0$

*Optimal sample* : $n_{optimal} = \left[\frac{\alpha}{c\beta}\right]^{\frac{1}{2}} - \beta$

## Continuous sampling distributions

---

*Normal model* (Specified precision $r$)

$\underset{\sim}{x} = (x_1, \ldots, x_n),\qquad x_i \in (-\infty, +\infty)$

$p(x_i \mid \theta) = (2\pi)^{-\frac{1}{2}} r^{\frac{1}{2}} \exp\left\{-\frac{r}{2}(x_i - \theta)^2\right\},\quad \theta \in (-\infty, +\infty)$

*Conjugate prior* : $p(\theta) = N\left(\theta \mid \mu, \frac{1}{\tau}\right),\quad \tau > 0,\ \mu \in (-\infty, +\infty)$

*Optimal sample* : $n_{optimal} = \left(\frac{1}{rc}\right)^{\frac{1}{2}} - \frac{\tau}{r}$

---

*Normal model* (specified mean $\mu$)

$\underset{\sim}{x} = (x_1, \ldots, x_n),\qquad x_i \in (-\infty, +\infty)$

$p(x_i \mid \theta) = (2\pi)^{-\frac{1}{2}} \theta^{\frac{1}{2}} \exp\left\{-\frac{\theta}{2}(x_i - \mu)^2\right\},\quad \mu \in (-\infty, +\infty),\ \theta > 0$

*Conjugate prior* : $p(\theta) = Ga(\alpha, \beta)\quad \alpha > 0, \beta > 0$

*Optimal sample* : $n_{optimal} = \frac{\sqrt{2}}{\beta\sqrt{c}}\sqrt{\alpha(\alpha+1)} - 2(\alpha+1)$

---



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

*Normal model* (Both parameters unknown)

$\underset{\sim}{x} = (x_1, \ldots, x_n)$, $\quad x_i \in (-\infty, +\infty)$

$$p(x_i \mid m, r) = (2\pi)^{-\frac{1}{2}} r^{\frac{1}{2}} \exp\left\{-\frac{r}{2}(x_i - m)^2\right\}, \quad m \in (-\infty, +\infty),\ r > 0$$

***Conjugate prior*** : $p(m, r) = Ng(m, r \mid \mu, t, \alpha, \beta) = N\left(m \mid \mu, \frac{1}{tr}\right) Ga(r \mid \alpha, \beta)$

$t > 0$, $\mu \in (-\infty, +\infty)$ , $\alpha > 1, \beta > 0$

***Optimal sample*** : $n_{optimal} = \left[\frac{\beta}{c(\alpha - 1)}\right]^{\frac{1}{2}} - t$

---

*Gamma model*

$\underset{\sim}{x} = (x_1, \ldots, x_n)$, $\quad x_i \in [0, +\infty)$

$$p(x_i \mid \theta) = \frac{\theta^k}{\Gamma(k)} x_i^{k-1} \exp\{-\theta x_i\} \quad k > 0, \theta > 0$$

***Conjugate prior*** : $\quad p(\theta) = Ga(\theta \mid \alpha, \beta), \quad \alpha > 0, \beta > 0$

***Optimal sample*** : $\quad n_{optimal} = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{ck}} - \frac{\alpha+1}{k}$

---

***Exponential model***

$\underset{\sim}{x} = (x_1, \ldots, x_n)$, $\quad x_i \in [0, +\infty)$

$p(x_i \mid \theta) = \theta \exp\{-\theta x_i\}$, $\quad \theta > 0$

***Conjugate prior*** : $\quad p(\theta) = Ga(\theta \mid \alpha, \beta), \quad \alpha > 0, \beta > 0$

***Optimal sample*** : $\quad n_{optimal} = \frac{\sqrt{\alpha(\alpha+1)}}{\beta\sqrt{c}} - (\alpha + 1)$

---



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# Chapter 5

# Sampling from a multivariate normal distribution

We shall know consider the problem where samples are taken from a non-singular, $k$ $(k \geq 1)$ dimensional multivariate normal distribution. The mean vector of the distribution is a $k$ - dimensional vector and the precision matrix of the distribution must be a symmetric positive definite $k \times k$ matrix. Any observation $x$ will have the form of a random vector in $\mathbb{R}^k$ : $x = (x_1, ...., x_k)$.

## 5.1 Determination of the optimal sample size

Let $\underset{\sim}{x} = (x_1, ...., x_n)$ to be a random sample from a multivariate normal distribution with unknown mean vector $\theta \in \mathbb{R}^k$ and a specified precision matrix $r \in \mathbf{M}_k(\mathbb{R})$ with properties as they were defined above. The prior distribution of $\theta$ is a multivariate normal distribution too with known mean vector $\mu \in R^k$ and precision matrix $t \in \mathbf{M}_k(\mathbb{R})$

Symmetric and positive definite in order to find the Bayes estimator for $\theta$ we use the quadratic error loss function $L(\theta, d) = (\theta - d)' \mathrm{A} (\theta - d)$. We consider here for



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

reasons of simplicity that $A = I \in M_k(\mathbb{R}) = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}$. Also the cost per observation $x = (x_1, \dots, x_k)$ is $c$, $c > 0$.

**Proof:**

For every random variable $x \in \mathbb{R}^k$

$$f(x|\theta) = (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\theta)' r (x-\theta)\right\}$$

$$= (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}x'rx + \frac{1}{2}x'r\theta + \frac{1}{2}\theta'rx - \frac{1}{2}\theta'r\theta\right\}$$

$$= (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}x'rx\right\} \exp\left\{-\frac{1}{2}\theta'r\theta\right\} \exp\left\{\frac{1}{2}(\theta'rx + x'r\theta)\right\}$$

$$= (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}x'rx\right\} \exp\left\{-\frac{1}{2}\theta'r\theta\right\} \exp\left\{\frac{1}{2}\left(\theta'rx + (\theta'rx)'\right)\right\}$$

$$= (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}x'rx\right\} \exp\left\{-\frac{1}{2}\theta'r\theta\right\} \exp\{\theta'rx\}$$

**So the density function of $x$ conditioned on $M$, from relation** (3.1.1) will have

$h(x) = (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}x'rx\right\}$, $c(\theta) = \exp\left\{-\frac{1}{2}\theta'r\theta\right\}$, $w(\theta) = \theta'r$ and $t(x) = x$

From (3.1.3) **the data likelihood** is proved to be:

$$f(\underline{x}|\theta) = \left[\prod_{i=1}^{n} (2\pi)^{-\frac{k}{2}} |r|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}x_i'rx_i\right\}\right]\left(\exp\left\{-\frac{1}{2}\theta'r\theta\right\}\right)^n \exp\left\{\theta'r\sum_{i=1}^{n} x_i\right\}$$

$$= (2\pi)^{-\frac{nk}{2}} |r|^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i'rx_i\right\} \exp\left\{-\frac{n}{2}\theta'r\theta\right\} \exp\left\{\theta'r\sum_{i=1}^{n} x_i\right\}$$

From (3.1.1.1) the conjugate prior density for $\theta$:

$$\pi(\theta|\tau_0, \tau_1) = MN(\mu, t^{-1}) = \pi(\theta|\mu, t)$$

$$\pi(\theta|\tau_0, \tau_1) = [k(\tau_0, \tau_1)]^{-1} \left(\exp\left\{-\frac{1}{2}\theta'r\theta\right\}\right)^{\tau_0} \exp\{\theta'r\tau_1\}$$

$$= [k(\tau_0, \tau_1)]^{-1} \exp\left\{-\frac{\tau_0}{2}\theta'r\theta\right\} \exp\{\theta'r\tau_1\}$$

$$= [k(\tau_0, \tau_1)]^{-1} \exp\left\{-\frac{1}{2}\theta'(\tau_0 r)\theta\right\} \exp\{\theta'r\tau_1\}$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$\pi(\theta \mid \mu, t) = (2\pi)^{\frac{k}{2}} |t|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\theta-\mu)' t(\theta-\mu)\right\}$$

$$= (2\pi)^{\frac{k}{2}} |t|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\theta' t\theta + \frac{1}{2}\theta' t\mu + \frac{1}{2}\mu' t\theta - \frac{1}{2}\mu' t\mu\right\}$$

$$= (2\pi)^{\frac{k}{2}} |t|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\mu' t\mu\right\} \exp\left\{-\frac{1}{2}\theta' t\theta\right\} \exp\{\theta' t\mu\}$$

Where: $\left[k(\tau_0, \tau_1)\right]^{-1} = (2\pi)^{\frac{k}{2}} |t|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\mu' t\mu\right\}$, $\tau_0 r = t$ and $r\tau_1 = t\mu$ (5.1)

Posterior distribution of $\theta$ from proposition (3.1.1.2) is defined to be

$$p(\theta \mid \underset{\sim}{x}, \tau_0, \tau_1) = \pi\left(\theta \mid n+\tau_0, \tau_1 + \sum_{i=1}^{n} x_i\right)$$

$$= k\left(n+\tau_0, \tau_1 + \sum_{i=1}^{n} x_i\right)^{-1} \left(\exp\left\{-\frac{1}{2}\theta' r\theta\right\}\right)^{n+\tau_0} \exp\left\{\theta' r\left(\tau_1 + \sum_{i=1}^{n} x_i\right)\right\}$$

$$= k\left(n+\tau_0, \tau_1 + \sum_{i=1}^{n} x_i\right)^{-1} \exp\left\{-\frac{1}{2}\theta'\left[(n+\tau_0)r\right]\theta\right\} \exp\left\{\theta' r\left(\tau_1 + \sum_{i=1}^{n} x_i\right)\right\}$$

Where according to the relations on (5.1) **we obtain the following**

$(n+\tau_0)r = t^* \Leftrightarrow t^* = nr + \tau_0 r \Leftrightarrow t^* = nr + t$ and

$$r\left(\tau_1 + \sum_{i=1}^{n} x_i\right) = t^*\mu^* \Leftrightarrow r\left(r^{-1}t\mu + n\bar{X}\right) = (t+nr)\mu^* \Leftrightarrow t\mu + nr\bar{X} = (t+nr)\mu^*$$

$$\Leftrightarrow \mu^* = (t+nr)^{-1}\left(t\mu + nr\bar{X}\right)$$

Thus, the posterior distribution of $\theta$ is a $k$- dimensional multivariate normal with mean vector $\mu^* = (t+nr)^{-1}\left(t\mu + nr\bar{X}\right)$ and precision matrix $t^* = (t+nr)$, where $t^*$ is a symmetric positive definite $k \times k$ matrix. Using the quadratic error loss function $L(\theta, d) = (\theta - d)'\mathrm{I}(\theta - d)$ it has been proved from theorem 2.7.3 that the Bayes estimator will be given by the posterior mean

$\delta^*(\underset{\sim}{x}) = E(\theta \mid \underset{\sim}{x}) = \mu^* = (t+nr)^{-1}\left(t\mu + nr\bar{X}\right)$

**The corresponding Bayes risk function and total risk equivalently will be**

$$r(\pi(\theta), \delta^*(\underset{\sim}{x})) = tr\left[\mathrm{IE}\left(\mathrm{cov}(\theta \mid \underset{\sim}{x})\right)\right] = tr\left[(t+nr)^{-1}\right]$$

$$r_{total}(\pi(\theta), \delta^*(\underset{\sim}{x})) = tr\left[(t+nr)^{-1}\right] + c(n) = tr\left[(t+nr)^{-1}\right] + cn \qquad (5.2)$$



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

We minimize (5.1.2) with respect to $n$ in order to derive the optimal sample size

$$\frac{\partial}{\partial n}\left\{tr\left[\left(t+nr\right)^{-1}\right]+cn\right\}=0 \Leftrightarrow \frac{\partial}{\partial n}tr\left[\left(t+nr\right)^{-1}\right]+c=0$$

$$\frac{\partial}{\partial n}tr\left[\frac{1}{|t+nr|}adj\left(t+nr\right)\right]+c=0 \Leftrightarrow \frac{\partial}{\partial n}\left[\frac{1}{|t+nr|}tr\left[adj\left(t+nr\right)\right]\right]+c=0$$

$$\left[\frac{\partial}{\partial n}\frac{1}{|t+nr|}\right]tr\left[adj\left(t+nr\right)\right]+\frac{1}{|t+nr|}\frac{\partial}{\partial n}tr\left[adj\left(t+nr\right)\right]+c=0$$

$$\frac{1}{|t+nr|}\frac{\partial}{\partial n}tr\left[adj\left(t+nr\right)\right]-\frac{|t+nr|tr\left[\left(t+nr\right)^{-1}r\right]}{|t+nr|^{2}}tr\left[adj\left(t+nr\right)\right]+c=0$$

$$\frac{1}{|t+nr|}\frac{\partial}{\partial n}tr\left[adj\left(t+nr\right)\right]-\frac{tr\left[\left(t+nr\right)^{-1}r\right]}{|t+nr|}tr\left[adj\left(t+nr\right)\right]+c=0$$

$$\frac{\partial}{\partial n}tr\left[adj\left(t+nr\right)\right]-tr\left[\left(t+nr\right)^{-1}r\right]tr\left[adj\left(t+nr\right)\right]+c|t+nr|=0 \quad (5.3)$$

From this last expression (5.3) we can derive $n_{optimal}$ as the positive integer solution of this equation for a specific decision problem.

Due to the complexity we shall concentrate to the case of the bivariate normal distribution ($k=2$) where the symmetric precision matrices for the data and the prior have general forms:

$r=\begin{pmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{pmatrix}$, $t=\begin{pmatrix} t_{11} & t_{12} \\ t_{12} & t_{22} \end{pmatrix} \Rightarrow t^{*}=t+nr=\begin{pmatrix} t_{11}+nr_{11} & t_{12}+nr_{12} \\ t_{12}+nr_{12} & t_{22}+nr_{22} \end{pmatrix}$ is also a symmetric precision matrix of the posterior density function of $\theta \in \mathbb{R}^{2}$.

Under these conditions equation (5.3) becomes

$$\frac{\partial}{\partial n}tr\left[\left(t+nr\right)\right]-tr\left[\left(t+nr\right)^{-1}r\right]tr\left[\left(t+nr\right)\right]+c|t+nr|=0$$

$$tr\left[\frac{\partial}{\partial n}\left(t+nr\right)\right]-tr\left[\left(t+nr\right)^{-1}r\right]\left(tr(t)+ntr(r)\right)+c|t+nr|=0$$

$$tr(r)-tr\left[\left(t+nr\right)^{-1}r\right]\left(tr(t)+ntr(r)\right)+c|t+nr|=0 \quad (5.4)$$

The analytical form of this equation when all the appropriate replacements have been done on (5.4) leads to a polynomial of forth degree i.e.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

$$n^4(cr_{12}^4-2cr_{11}r_{12}^2r_{22}+cr_{11}^2r_{22}^2)+n^3(-2cr_{12}^2r_{22}t_{11}+2cr_{11}r_{22}^2t_{11}+4cr_{12}^3t_{12}-4cr_{11}r_{22}r_{12}t_{12}-2cr_{11}r_{12}^2t_{22}$$
$$+2cr_{11}^2r_{22}t_{22})+n^2(r_{11}r_{12}^2-r_{11}^2r_{22}+r_{12}^2r_{22}-r_{11}r_{22}^2+cr_{22}^2t_{11}^2-4cr_{12}r_{22}t_{11}t_{12}+6cr_{12}^2t_{12}^2-2cr_{11}r_{22}t_{12}^2$$
$$-2cr_{12}^2t_{11}t_{22}+4cr_{11}r_{12}t_{12}t_{22}+cr_{11}^2t_{11}^2)+n(2r_{12}^2t_{11}-2r_{11}r_{22}t_{11}-2cr_{22}t_{11}t_{12}^2+4cr_{12}t_{12}^3+2r_{12}^2t_{22}$$
$$-2r_{11}r_{22}t_{22}+2cr_{22}t_{11}^2t_{22}-4cr_{12}t_{11}t_{12}t_{22}-2cr_{11}t_{12}^2t_{22}+2cr_{11}t_{11}t_{22}^2)-r_{22}t_{11}^2+2r_{12}t_{11}t_{12}-r_{11}t_{12}^2$$
$$-r_{22}t_{12}^2+ct_{12}^4+2r_{12}t_{12}t_{22}-2ct_{11}t_{12}^2t_{22}-r_{11}t_{22}^2+ct_{11}^2t_{22}^2=0 \qquad (5.5)$$

**It can be clearly seen that the optimal sample size in this case cannot be given in a closed form but it can always be found as the positive integer solution of (5.4) or more analytically of (5.5) for a specified sampling model and a specified prior** distribution each time.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# Chapter 6

# Discussion

In this thesis we worked towards determining what is the optimal number of observations to be drawn for each sample during Phase I data subject to constant sampling interval and constant cost of observations. The sampling distributions considered here were from the regular exponential family, where a conjugate prior always exist. More precisely in the univariate case we derived the optimal sample size when the sampling distribution was Gamma, Normal, Poisson and Binomial while in the multivariate setting of a Multivariate Normal we showed that the optimal sample size is not given by a closed form but as the positive integer solution of a rather complicated equation.

The theoretic development was on a conjugate-based Bayesian approach of decision theory techniques with squared error loss. Using basic decision theory properties we have derived the corresponding Bayes rules, Bayes risk and total risk functions for every given sample distribution. The optimal sample choice has been given then as the solution that minimizes the total risk function which balances the Bayes risk and the total sampling cost.

This method of defining optimal sample size for phase I samples in a production process can also be extended to other distribution functions using various error loss functions and/or sampling cost functions. It may also be expanded to several other multivariate cases. All these can develop a subject of future research.



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

# *References*


**Barnett, V. (1974).** Economic Choice of Sample Size for Sampling Inspection Plans, *Applied Statistics*, 23, 149-157

**Bather, J.A. (1963).** Control charts and minimization of costs, *Journal of the Royal Statistical Society,* 25, 49-80

**Berger, J.O. (1985).** *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York

**Bernardo, J.M. and Smith, A.F.M. (1995).** *Bayesian Theory.* Wiley, Chichester

**Calabrese, J.M. (1995).** Bayesian process control for attributes, *Management Science*, 41, 637-645

**Carlin, B.P. and Louis, T.A. (1996).** *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, London

**Carter, P.L. (1972).** A Bayesian approach to quality control, *Management Science,* 18, 647-655

**Casella, G. and Berger, R.L. (1990).** *Statistical Inference.* Duxbury, Belmot, CA

**Costa, A.F.B. (1994).** $\bar{X}$ charts with variable sample size, *Journal of Quality Technology,* 26, 155-163

**Daudin, J.J. (1992).** Double sampling $\bar{X}$ charts, *Journal of Quality Technology,* 24, 78-87

**DeGroot, M.H. (1970).** *Optimal Statistical Decisions.* McGraw-Hill, New York

**Diaconis, P. and Ylvisaker, D. (1979).** Conjugate priors for exponential families, *The Annals of Statistics*, 7, 269-281

**Ferguson, T.S. (1967).** *Mathematical Statistics: a decision theoretic approach.* Academic, New York

**Girshik, M.A. and Rubin, H. (1952).** A Bayes' approach to a quality control model, *Ann.Math.Statistics*, 23, 114-125

**Gutierrez-Pena, E. (1997).** Moments for the canonical parameter of an exponential family under a conjugate distribution, *Biometrika*, 84, 727-732




ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**Montgomery, D.C. (2001).** *Introduction to Statistical Quality Control.* Lohn Wiley& Sons, Inc., New York

**Morris, C.N. (1983).** Natural exponential families with quadratic variance functions: Statistical theory, *The Annals of Statistics*, 11,515-529

**O'Hagan, A. and Forster, J. (1994).** *Kendall's advanced theory of statistics.* Arnold, London

**Parkhideh, B. and Case, K.E. (1989).** The economic design of a dynamic $\bar{X}$ control chart, *IIE Transactions,* 21, 313-323

**Pericchi, L.R., Sanso, B. and A. Smith, A. F. M. (1993).** Posterior cumulant relationships in Bayesian inference involving the exponential family, *Journal of the American Statistical Association,* 88, 1419-1426

**Porteus, E.L. and Angelus, A. (1997).** Opportunities for Improved Statistical Process Control, *Management Science,* 43, 1214-1228

**Prabhu, S.S., Runger, G.C. and Keats, J.B. (1993).** An adaptive sample size $\bar{X}$ chart, *International Journal of Production Research,* 31, 2895-2909

**Reynolds, Jr., M.R. (1989).** Optimal variable sampling interval control charts, *Sequential Analysis,* 8, 361-379

**Shewhart, W.A. (1931).** ***Economic Control of Quality*** *Manufactured Product.* D.Van Nostrand, New York

**Tagaras, G. (1994).** A dynamic programming approach to the economic design of $\bar{X}$ charts, *IIE Transactions*, 26(3), 48-56

**Tagaras, G. (1996). Dynamic control charts for finite production runs,** *European Journal of Operational Research,* **91, 38-55**

**Tagaras, G. and Nikolaidis, Y. (2002).** Comparing the effectiveness of various Bayesian $\bar{X}$ control charts, *Operations Research*, 50, 878-888

**Taylor, H.M. (1965). Markovian Sequential Replacement Processes,** *Ann.Math.Statistics*, **36, 1677-1694**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Δωρεά

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ