THE UNIVERSITY OF MANITOBA

ON THE STUDY OF THE POSTERIORS OF PARAMETERS OF SOME WELL KNOWN DISTRIBUTIONS UNDER PROPER AND IMPROPER PRIORS

by

HATEM ALI HOWLADER

A THESIS

SUBMITTED TO THE FACULTY OF

GRADUATE STUDIES IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Department of Statistics

WINNIPEG, MANITOBA

FEBRUARY, 1982

ON THE STUDY OF THE POSTERIORS OF PARAMETERS
OF SOME WELL KNOWN DISTRIBUTIONS
UNDER PROPER AND IMPROPER PRIORS

BY

HATEM ALI HOWLADER

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

TABLE OF CONTENTS

ABSTRACT

In this dissertation we have studied the posteriors of parameters of some well-known distributions under proper and improper priors. Based on Monte Carlo studies we have deduced the appropriate prior and appropriate loss function on the basis of minimum MSE criterion from among a family of priors and loss functions.

In Chapter One, we have given a brief intorduction of Bayesian analysis including different types of priors developed so far.

In Chapter Two, Bayesian estimators of the Pareto-parameter, β are obtained under the Natural Conjugate Prior (NCP), Minimal Information Prior (MIP), and Jeffreys' Invariant Prior (JIP) using the squared error loss (SEL), the logarithmic squared error loss (LSEL), and a general loss function (GLF). The properties of these estimators are discussed. Bayes estimators of β under MIP and LSEL appeared intractable; as a result we used Lindley's (1980) approximate method for the evaluation of the ratio of integrals for the one parameter case. Monte Carlo studies showed that the LSEL function with JIP and NCP give smaller MSE than (i) SEL function with JIP and NCP or (ii) GLF with JIP and NCP. MIP with all the three loss functions considered gives larger

MSE. We have also obtained the sampling distributions of the Bayes estimators under JIP and MIP. Both distributions were found to fit Pearson's type II curve.

In Chapter Three, we considered the scale parameter a and shape parameter m of Pearson's type VII distribution under JIP when both the parameters are unknown, and a family of improper priors proportional to $1/a^t$ and $1/m^t$ when one of the parameters is assumed known. The type VII distribution does not belong to the general exponential family. As a result no NCP can be generated and hence no Bayes estimators of the parameters can be obtained under NCP. The Bayes estimators of the parameters under JIP did not take simple form. Lindley's methods for one and two parameters were used. The estimators based on the maximum likelihood (ML) and the method of moments (MM) were also obtained. For joint estimation of a and m, Monte Carlo studies showed that the Bayesian method led to smaller MSE than the other two. The goodness of fit for Pearson's type I curve, for both the sampling distributions of a and m (for joint estimation), were tested by Chi-square, Cramér von-Mises and Kolmogorov-Smirnov statistics.

In Chapter Four, we considered the Rayleigh distribution and obtained Bayes estimators of σ, σ^2 and reliability function R_t under JIP and MIP with SEL. Monte

Carlo studies showed that the MLE of σ, σ^2 and R_t have slight edge over their Bayesian counterparts. We have also obtained the equal-tail and Highest Posterior Density (HPD) credible intervals for σ, R_t, and a future observation .

In Chapter Five, we considered the Bayesian estimators of the parameters μ and σ^2 of a normal distribution under NCP and using SEL, LSEL and GLF. For estimation of σ^2, Monte Carlo studies showed that the LSEL results in a smaller MSE than the GLF and SEL for all combinations of the prior parameters. Bayes estimates of μ were found to be quite robust for the choice of various loss functions and prior parameters while those of σ^2 were less so.

In Chapter Six, we considered the simple regression model in orthogonal form and obtained the Bayes estimators of (α,β) and σ^2 when the prior distributions of α and β are uniform and that of σ^2 is inverted gamma. Bayes estimators of (α,β) in this situation are the same as the MLE of (α,β). Bayes estimators of (α,β) were also obtained under a bivariate normal prior for (α,β); σ^2 was assumed known. Under the bivariate normal prior for (α,β), the Bayes estimators of α and β did not yield simple form. Lindley's (1980) method was again used. Monte Carlo studies showed that the Bayes estimators of

(α, β) based on the bivariate normal prior have considerably lower MSE than their ML counterparts.

ACKNOWLEDGEMENTS

I would like to express my indebtedness to my supervisor, Professor S.K. Sinha, for suggesting the problem and for his advice, encouragement and guidance throughout the preparation of this dissertation.

I would also like to thank the entire staff of the Department of Statistics, University of Manitoba, for help in one form or another, especially for the financial assistance, without which it would not have been possible to complete this project.

I am grateful to Mrs. Eva Loewen for her excellent job of typing this dissertation.

CHAPTER ONE

INTRODUCTION AND SUMMARY

1.1 BAYES' THEOREM

The basic ingredient of Bayesian analysis is Bayes' Theorem (1763), the first attempt known to us to rationalize the process of inductive reasoning. The process of learning from experience had been in practice from time immemorial. Bayes' theorem can be used to bring the class of inductive inferences within the domain of the theory of probability. A philosophical extension to Bayes' theorem can be made where probability represents a degree of belief rather than a relative frequency of 'success'. The belief that we have in a proposition depends on the state of our current information, and is therefore, in general, always a conditional probability, conditional on the state of information. As the information changes, we revise the degree of belief in the proposition. This process of revising probabilities after obtaining new information is the essence of learning from experience and can be made by using a simple rule of probability theory, viz., Bayes' theorem. Bayes' problem was: given a certain data, what is the probability that a particular type of population might have given rise to it? Thus, Bayes' theorem is often referred to as the theorem of inverse probability.

Consider a set of k mutually exclusive and exhaustive events $E_1, E_2, \ldots, E_k$, of which none has zero probability. Let D be another event which can occur only if one of the events $E_1, E_2, \ldots, E_k$ occurs; then by the definition of conditional probability,

$$P(E_i|D) = \frac{P(E_i)P(D|E_i)}{\sum_1^k P(E_i)P(D|E_i)}, \qquad (1.1)$$

for i = 1,2,... or k.

This formula, known as Bayes' theorem (Mosteller, Rourke and Thomas (1970)) was proposed by Reverend Thomas Bayes (1702-1761) and was posthumously published in 1763. Although (1.1) represents an application of conditional probability which gives the change in the probability of E_i with the additional information that D has occurred, it has far-reaching implications in the development of a new school of thought in Statistical Inference, widely known as the school of BAYESIAN INFERENCE.

In (1.1), $P(E_i)$ is called the prior probability and $P(E_i|D)$ is called the posterior probability of the event E_i conditional on the event D.

The equation (1.1) simply represents a conditional probability — a probability that the event E_i occurs on the basis of the prior information about E_i as well as information contained in or provided by the event (or data) D.

1.2 BAYES' THEOREM IN CONJUNCTION WITH LIKELIHOOD FUNCTION

Let $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ be a random sample of n observations whose probability density $f(x|\theta)$ is completely known except it depends on the parameter θ, where θ may be a vector or real valued parameter. Let $f(x,\theta)$ denote the joint densoty of the random variable x and θ. Then

$$f(x,\theta) = g(\theta)f(x|\theta)$$
$$= h(x)\Pi(\theta|x).$$

Thus, given the observed data $\underset{\sim}{x}$, the conditional density of θ is

$$\Pi(\theta|\underset{\sim}{x}) = \frac{g(\theta)f(\underset{\sim}{x}|\theta)}{h(\underset{\sim}{x})}.$$

Also, $f(\underset{\sim}{x}|\theta)$ is a function of θ for given $\underset{\sim}{x}$ and is called the likelihood of θ. Denoting $f(\underset{\sim}{x}|\theta)$ by $\ell(\underset{\sim}{x}|\theta)$, the posterior density of θ can be written as

$$\Pi(\theta|\underset{\sim}{x}) \propto g(\theta)\ell(\underset{\sim}{x}|\theta),$$

that is,

$$\Pi(\theta|\underset{\sim}{x}) = k\cdot g(\theta)\ell(\underset{\sim}{x}|\theta),$$

where k is a normalizing factor and is given by

$$k^{-1} = \int_\Omega \Pi(\theta|\underset{\sim}{x})d\theta = \int_\Omega g(\theta)\ell(\underset{\sim}{x}|\theta)d\theta,$$

where Ω is the parameter space.

Hence,

$$\Pi(\theta|\underset{\sim}{x}) = \frac{g(\theta)\ell(\underset{\sim}{x}|\theta)}{\int_\Omega g(\theta)\ell(\underset{\sim}{x}|\theta)d\theta} \qquad (1.2)$$

where $g(\theta)$ is the prior density of θ representing our knowledge about parameter θ before the data $\underset{\sim}{x}$ are drawn and the likelihood $\ell(\underset{\sim}{x}|\theta)$ representing the information about θ contained in the sample $\underset{\sim}{x}$.

$\Pi(\theta|\underset{\sim}{x})$ is known as the posterior density of θ given the data $\underset{\sim}{x}$, which summarizes all the information we have about θ.

1.3 INFERENCES ON THE BASIS OF BAYES' THEOREM

The inferences in classical approach are based on the likelihood function in which a real or vector-valued parameter θ is treated strictly as an unknown constant. On the other hand, in Bayesian approach θ is treated as a random variable having a probability distribution representing a formalization of the information on prior to any experimentation. For example, a quality control department of an industry may want to know the process average at the present point in time. The department knows that the process average is affected by random factors that cause it to change over time. Also the past relative frequency distribution over time of the process average is available to the department. Thus, the process average at any point in time may be treated as a random variable and the relative frequency of the past data may be used as the prior distribution of the processed average.

The prior information on θ may be objective (data based), or subjective (non-data based), or a mixture of both. The concept of objective probability requires ideas of repeated trials and stabilizing of relative frequencies. If we have enough information available from past records, then we may set up a data based prior distribution for θ. However, the most controversial form of the prior is the subjective prior which takes into account one's degree of belief about θ based on personal judgement and experience. For example, consider the hypothesis that the percentage of mercury content in pickerel in Lake Winnipeg is p_o. This is an uncertainty and as a result people will have different opinions regarding this hypothesis. Someone might believe one person's estimate to be too high or too low. Through the concept of subjective probability, it is possible to assign to a hypothesis a probability, P(H), which will numerically represent a person's degree of belief about uncertainty. Then an experiment is undertaken to combine each person's prior probability with the experimental evidence. Bayes' theorem allows us to calculate the conditional probability $P(H|DATA)$ for each person which is simply the revised belief about the hypothesis after observing the experimental evidence.

Why Bayesian?

Bayes' theorem provides a formal mechanism by which prior information about a parameter θ may be combined with information contained in the data to give the posterior distribution of θ, which effectively summarizes all the information we have about θ.

The Bayesian approach to inference incorporating prior and sample information is illustrated in Figure 1 (Evans, 1976).

Pre-sample information on θ $\rightarrow$ Prior distribution, $g(\theta)$

Sample information, $\underset{\sim}{x}$ $\rightarrow$ Likelihood function, $\ell(\underset{\sim}{x}|\theta)$

$\downarrow$

Apply Bayes' Theorem

$\downarrow$

Posterior distribution, $\Pi(\theta|\underset{\sim}{x}) = g(\theta)\ell(\underset{\sim}{x}|\theta)$

$\downarrow$

Inference

FIGURE 1

Further, given the prior and the likelihood, the posterior distribution is unique and there is no such problem as the choice of which statistic to be used, as occurs in the sampling theory framework.

In the case of a scale parameter θ $(0 < \theta < \infty)$, the Bayesian estimate will always be positive, while in the sampling theory framework, the estimate for a positive parameter can be negative. For example, a uniformly minimum variance unbiased estimator for θ^2 in $N(\theta,1)$ is $\bar{x}^2 - 1/n$ which may be negative for $\bar{x}^2 < 1/n$, a result which is totally unacceptable.

Bayesian framework is specially suited for dealing with nuisance parameters. Considerable difficulties arise in dealing with nuisance parameters using sampling theory approach. In Bayesian framework one may obtain the posterior distribution of parameter(s) of interest by integrating out the nuisance parameters.

It can be applied to a very wide range of problems not necessarily restricted to normal family.

It allows explicit use of priors and offers a well defined and straightforward procedure for analyzing a problem.

1.4 ATTEMPTS MADE TO OVERCOME 'NOT KNOWING' THE PRIOR DISTRIBUTIONS

Bayes' theorem has been subjected to extreme criticism since the time of Bayes. In recent years Bayesian analysis has gained a much greater applicability in Statistical Inference. As the use of Bayes' theorem increased, so did the criticism. The main thrust of the criticism by the classical theorists, is that it is difficult to accurately formulate and assess prior densities. Much work has been done on the problem of 'not knowing' priors, notably by de Finetti (1964), Jeffreys (1961), Raiffa and Schlaifer (1961), Edwards, Lindman and Savage (1963), and Zellner (1971). We consider below a few approaches:

(i) The Principle of Stable Estimation or the Principle of Precise Measurements:

This principle (Edwards, Lindman and Savage (1963)), states that the posterior distributions derived from uniform prior densities are adequate approximations to the actual posterior distributions. That is, we can approximate the posterior distribution by the likelihood function in situations in which the prior $g(\theta)$ is gently changing in the region where the likelihood is large and if $g(\theta)$ at no other point is of sufficiently great magnitude as to become appreciable when multiplied by the likelihood (Box and Tiao (1973)). As a result we may assume the prior distribution to be uniform if this condition is attained. The principle of stable estimation is useful in practice since in many experimental situations, the likelihood is expected to exert a much stronger influence on the final result than the initial opinions, for otherwise there would be little justification for carrying out the experiment.

(ii) Jeffreys' Rule of Non-Informative Prior (NIP)

The adoption of improper priors when we know little about the parameters were justified by Jeffreys (1961) based on the invariance of the densities under parametric transformations. The development of invariance property follows.

If we have no prior knowledge or vague prior knowledge about the location parameter θ, then we also have vague prior knowledge about the linear transformation

$$\eta(\theta) = \alpha\theta + \beta$$

where α, β are constants and $\alpha \neq 0$. If we take a uniform density for θ, then we should also take a uniform prior for $\eta(\theta)$ since from the distribution theory if θ has a uniform distribution, then $\eta(\theta)$ has also a uniform distribution if it is a linear function of θ.

Again, if θ is a scalar parameter, obviously it has a range 0 to ∞, and if we have vague prior knowledge about θ, then we have also the same vague prior knowledge about any power trnasformation of the form

$$\psi(\theta) = \theta^{\nu}, \quad \nu \neq 0.$$

Also, if $\log \theta$ is uniformly distributed, the prior $g(\theta) \propto 1/\theta$, then the distribution of $\psi(\theta)$ is proportional to $1/\psi(\theta)$, for $\psi(\theta) > 0$. We show this fact with the following example:

Suppose

$$\psi(\theta) = \theta^{\nu}, \quad \psi'(\theta) = \nu\theta^{\nu-1},$$

$$f[\psi(\theta)] = f(\theta)\left|\psi'(\theta)\right|^{-1} = \frac{1}{\theta}\cdot\frac{1}{\nu\theta^{\nu-1}} = \frac{1}{\nu\theta^{\nu}} = \frac{1}{\nu\psi(\theta)} .$$

Therefore,

$$f[\psi(\theta)] \propto \frac{1}{\psi(\theta)} .$$

Jeffreys' Invariant Prior (JIP) is defined by (Jeffreys, 1961)

$$g(\theta) \propto \sqrt{|I(\theta)|},$$

where $I(\theta)$ is Fisher's information about θ contained in a single observation. This prior is 'improper' in the sense

$$\int_{\Omega} g(\theta)d\theta \neq 1.$$

When the range of a random variable does not depend upon the parameter under consideration, Jeffreys' Invariant Priors (JIP) follow the following working rules:

(i) if the parameter space is the entire real line, $g(\theta)$ may be taken to be constant, i.e., to be uniformly distributed;

(ii) if the parameter space is the non-negative real line, $g(\theta)$ may be taken proportional to $1/\theta$, i.e., $\log\theta$ to be uniformly distributed.

We will illustrate how to find JIP when X has the well-known normal probability density function

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},$$

$$-\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0;$$

$$\log f(x|\mu,\sigma) = -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}(x-\mu)^2,$$

and

$$\left.\begin{aligned}
&\frac{\partial \log f}{\partial \mu} = \frac{1}{\sigma^2}(x-\mu), \qquad \frac{\partial^2 \log f}{\partial \mu^2} = -\frac{1}{\sigma^2} \\
&\frac{\partial \log f}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2 \\
&\frac{\partial^2 \log f}{\partial \sigma^2} = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x-\mu)^2 \\
&\frac{\partial^2 \log f}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3}(x-\mu)
\end{aligned}\right\} \qquad (1.3)$$

$$-E\left(\frac{\partial^2 \log f}{\partial \sigma^2}\right) = \frac{2}{\sigma^2}, \quad -E\left(\frac{\partial^2 \log f}{\partial \sigma \partial \mu}\right) = 0, \quad -E\left(\frac{\partial^2 \log f}{\partial \mu^2}\right) = \frac{1}{\sigma^2}$$

Hence, the Fisher's information matrix (Kendall and Stuart, 1973) is given by

$$I(\mu,\sigma) = \begin{pmatrix} 2/\sigma^2 & 0 \\ 0 & 1/\sigma^2 \end{pmatrix}$$

Thus, the joint JIP for (μ,σ) is given by

$$g(\mu,\sigma) \propto |I(\mu,\sigma)|^{\frac{1}{2}} \propto \frac{1}{\sigma^2}.$$

If σ is known,

$$I(\mu) = -E\left(\frac{\partial^2 \log f}{\partial \mu^2}\right) = \frac{1}{\sigma^2},$$

and

$$g(\mu) \propto |I(\mu)|^{\frac{1}{2}} = \frac{1}{\sigma} = \text{constant}.$$

Similarly, if μ is known,

$$g(\sigma) \propto |I(\sigma)|^{\frac{1}{2}} \propto \frac{1}{\sigma}.$$

Bayes and Laplace themselves used uniform priors in related problems. The weight of the authority of these 'master minds' has led to the general idea that 'uniform prior is the answer'. Jeffreys (1961) makes an interesting comment that there is no more need for this kind of belief than to say that an oven which successfully cooked roast beef once cannot be used for cooking anything but roast beef.

Why improper priors?

These priors are very convenient for dealing with complicated situations, particularly with nuisance parameters. They are powerful techniques but indiscriminate use of them, without a proper understanding of the problem and information available, may lead to inconsistent and dubious results (Dawid, Stone and Zidek, 1973).

(iii) Minimal Information Prior

The main objectives of generating improper or non-informative priors are to express ignorance and to obtain priors which possess certain invariance property. Sometimes, the second objective is in conflict with the first; i.e., by insisting on invariance we may face problems in achieving the objective of expressing ignorance adequately. A solution has been proposed (Zellner, 1971) to this problem in information theoretic approach — what is termed as the 'minimal information prior'. These priors are dependent on a particular parameterization used. Following Zellner (1971), we develop the technique of obtaining a minimal information prior (MIP).

Let $f(x|\theta)$ be the probability density function of X given θ; then define

$$I_x(\theta) = \int f(x|\theta)\log\{f(x|\theta)\}dx$$

as measuring the information in $f(x|\theta)$. Define

$$\bar{I}_x = \int I_x(\theta)g(\theta)d\theta$$

as measuring the prior average information in the data, where $g(\theta)$ is a proper prior density function.

Let

$$\int g(\theta)\log\{g(\theta)\}d\theta$$

measure the information in the prior $g(\theta)$. Then

$$G = \int I_x(\theta)g(\theta)d\theta - \int g(\theta)\log\, g(\theta)d\theta$$

represents the gain in information associated with an observation x over the information in the prior $g(\theta)$. The minimal information prior density is defined as one which maximizes G by varying $g(\theta)$ subject to

$$\int g(\theta)d\theta = 1,$$

that is, maximizes

$$U = \left[\int I_x(\theta)g(\theta)d\theta - \int g(\theta)\log\, g(\theta)d\theta\right] + \lambda\left[\int g(\theta)d\theta - 1\right]$$

where λ is the Lagrange multiplier.

To maximize U,

$$\frac{\partial U}{\partial g(\theta)} = 0$$

gives

$$I_x(\theta) - 1 - \log\, g(\theta) + \lambda = 0,$$

that is,

$$g(\theta) = k\exp\{I_x(\theta)\}, \qquad (1.4\ )$$

k is a normalizing constant.

Consider the density

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\}, \quad -\infty < x,\ \mu < \infty,$$

then

$$I_x(\mu) = \int_{-\infty}^{\infty} \left\{-\frac{1}{2}\log 2\pi - \frac{1}{2}(x-\mu)^2\right\} f(x|\mu)\,dx$$

$$= -\frac{1}{2}(\log 2\pi + 1)$$

which is independent of μ. For proper $g(\mu)$,

$$G = -\frac{1}{2}(\log 2\pi + 1) - \int g(\mu)\log g(\mu)\,d\mu$$

is maximized if

$$\int g(\mu)\log g(\mu)\,d\mu$$

is minimized subject to

$$\int g(\mu)\,d\mu = 1.$$

Thus,

$$1 + \log g(\mu) + \lambda = 0, \quad \lambda \text{ is Lagrange multiplier,}$$

giving,

$$g(\mu) = \exp\{-(1+\lambda)\} = \text{constant},$$

which is the same as Jeffreys' NIP.

We again consider the density

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

when both μ and σ are unknown. Working the same line, we obtain

$$g(\mu,\sigma) \propto \frac{1}{\sigma} ,$$

i.e., the minimal information prior for μ and σ shows that they are independent and μ and $\log \sigma$ are uniformly distributed. While Jeffreys' prior,

$$g(\mu,\sigma) = |I(\mu,\sigma)|^{\frac{1}{2}} \propto \frac{1}{\sigma^2} . \tag{1.5}$$

Jeffreys justified this departure keeping the fact in mind that if μ and σ are known to be independently distributed then the rule should be applied separately to obtain the joint prior, $g(\mu,\sigma)$ as the product of their marginal priors, i.e.,

$$g(\mu,\sigma) \doteq g(\mu)g(\sigma) \propto \frac{1}{\sigma} . \tag{1.6}$$

(iv) The Natural Conjugate Prior (NCP) Densities

A class of priors was developed by Raiffa and Schlaifer (1961) for the densities belonging to the general exponential family possessing sufficient statistics of fixed dimensionality. This family is mathematically tractable in the sense that it is reasonably easy to obtain a posterior distribution. It possesses a very interesting property that the posteriors obtained by using a member of this class also belongs to the same class. This property is termed as "closure under sampling" by Wetherill (1961).

We now generate an NCP and demonstrate the above property. Let the distribution of the random variable

X be binomial with n trials and the probability, θ of a "success", and having the likelihood

$$\ell(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0,1,2,\ldots,n. \qquad (1.7)$$

The binomial distribution belongs to the general exponential family and hence to generate NCP we replace all the quantities in the kernel of (1.7) dependent on the sample by the parameters α and β, known as prior parameters (Evans, 1976), to obtain the NCP

$$g(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 < \theta < 1;\ \alpha,\ \beta > 0, \qquad (1.8)$$

that is, θ has apriori a beta distribution with parameters α and β. The posterior density of θ given $X = x$ is obtained by combining (1.8) with (1.7) and using (1.2) as

$$\Pi(\theta|x) = \frac{\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}}{\int_0^1 \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}d\theta}$$

$$= \frac{1}{B(\alpha+x,\beta+n-x)}\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}, \quad 0 < \theta < 1$$

which is also a beta distribution with parameters $\alpha+x$ and $\beta+n-x$. Thus, the posterior has the same functional form as the prior, and hence the beta priors are closed under sampling.

(v) Exchangeable Prior Distributions

The idea of exchangeability is due to de Finetti (1964). It states that the order of occurrence of a sequence of events does not affect the probabilities of happening of the events. The classical theory terms "random" and "random sample" may be replaced by the terms "exchangeable" and "exchangeable items". It is easier to attach the concept of exchangeability to a sequence of events than the concept of randomness which involves a condition of independence and such a condition is sometimes untenable. To clarify the preceding, we consider the following:

Suppose a box contains an unknown number of cold tablets and an unknown number of identical looking aspirin tablets. If we draw the tablets from the box one after another and replace each tablet after it is drawn, a classical theorist will say that the probability that the second tablet is aspirin is independent of the fact that the first tablet is aspirin. That is, for him the conditional probability that the second tablet is aspirin given that the first tablet is aspirin is exactly the same as the unconditional probability that the second tablet is aspirin. The events in a sequence of drawings are not independent to an exchangeablist, in the sense that events which have already occurred give evidence about occurrence of future events. Thus, the degree of belief is changed. In this sense, the conditional probability of the k-th tablet being aspirin given that the type of the first (k-1) tablets is known, is dependent on the proportion of aspirins observed in the first (k-1) drawings. And that the probability of obtaining an aspirin in the k-th draw is not dependent on the particular order in which the aspirins appeared on the first (k-1) drawings, but how many of them were aspirins.

For any k distinct events, the probability that a specified r of them occur and (k-r) of them do not occur, depends only on k and r and these events follow the binomial likelihood

$$P(r,k|\theta) = \theta^r(1-\theta)^{k-r}$$

for some θ, with θ having a probaiblity distribution, $g(\theta)$ on $[0,1]$ (Lindley, 1972).

Definition: The quantities $X_1, X_2, \ldots, X_n$ are exchangeable if the probabilities are invariant under permutations of the suffixes. Lindley and Smith (1972) used this notion in Bayesian estimation for the parameters in linear model in which they described that one way of obtaining an exchangeable prior, $p(\theta)$ is to suppose

$$p(\theta) = \int \prod_{i=1}^{n} p(\theta_i|\mu)dQ(\mu)$$

where $p(\theta_i|\mu)$ for each μ, and $Q(\mu)$ are arbitrary probability distributions. Hewitt and Savage (1955) showed that if exchangeability is assumed for every n, then a mixture is the only way to generate an exchangeable distribution. Also that exchangeability implies that the θ's have the probability structure of a random sample from a distribution.

As an example, we consider the linear model

$$E(\underset{\sim}{Y}) = \underset{\sim}{A}\underset{\sim}{\theta} \tag{1.9}$$

considered by Lindley and Smith (1972), where,

$\underset{\sim}{Y}$ = vector of observations

$\underset{\sim}{A}$ = known design matrix

$\underset{\sim}{\theta}$ = vector of parameters.

For simplicity let us suppose that A is a unit matrix, then

$$E(y_i) = \theta_i, \quad i = 1,2,\ldots,n \tag{1.10}$$

and

$$y_i \sim N(\theta_i, \sigma^2)$$

This model may arise in agricultural experimentation when y_i is observed on the i-th variety of average yield θ_i. It often seems reasonable to say that there is the same amount of information about each of the means θ_i, i.e., the joint distribution of the means remains invariant under permutation of the suffixes. In particular, the prior opinion about θ_3 is the same as θ_9, or any other θ_i; similarly for pairs, triplets, etc. That is, θ_i are exchangeable.

We note that the exchangeability assumption will not hold when one or more varieties are controls and the others are experimental. In such a situation separate exchangeability assumptions are to be made — one for within control varieties and another for within experimental varieties.

In the model (1.9), we suppose

$$E(\theta_i) = \mu$$

and

$$\theta_i \sim N(\mu, \tau^2). \tag{1.11}$$

That is, apriori the linear structure for θ_i is analogous to the linear structure assumed for y_i. Lindley and Smith (1972) following I.J. Good, termed μ as a hyper-parameter. Equations (1.10) and (1.11) above constitute a two-stage model. We can go further by supposing the hyper-parameters to have a linear structure that will lead to a three-stage model, and so on. The specification will be complete when a prior is attached to the ultimate hyper-parameter. In this example if we consider a two-stage model, then a prior has to be attached to μ.

Assuming vague prior knowledge for μ, Lindley (1971) has obtained the posterior distribution of θ_i with posterior mean,

$$E(\theta_i \mid \underset{\sim}{y}) = \frac{y_i/\sigma^2 + y_o/\tau^2}{1/\sigma^2 + 1/\tau^2},$$

where

$$y_o = \Sigma y_i/n.$$

Recently, Lindley and Novick (1981) have used the idea of exchangeability in situations like analysis of variance and covariance, contingency tables, and calibration.

In this dissertation, we have not used exchangeable prior density but there is enough scope for research with this prior which we hope to pursue in future.

A Final Note:

The reviews of Bayesian Statistics by Lindley (1972) and (1978) give excellent summaries of recent advancement in this viewpoint.

Lindley (1975) suggested a moratorium on research for two years so that all of us in the field of statistics can read de Finetti (1974, 1975).

1.5 OUTLINE OF THE DISSERTATION

This dissertation is concerned with the study of the posteriors of parameters of some well-known distributions under proper and improper priors. Based on Monte Carlo studies we have deduced, on the basis of minimum MSE criterion, the appropriate prior and the appropriate loss function from among a family of priors and loss functions.

In Chapter Two we considered the Pareto distribution with density

$$f(x|\theta_o,\beta) = \frac{\beta\theta_o^{\beta}}{x^{\beta+1}}, \quad 0 < \beta < \infty, \quad 0 < \theta_o < x < \infty.$$

This distribution is useful in economics to represent the distribution of incomes of individuals whose income exceeds a certain limit θ_o. It has also application in modelling economic problems involving distributions of incomes. Rodolfo Bonini (Allais, 1968) extended and modified Pareto distribution in the analysis of the distribution of property. This distribution is also used in the study of size distributions of firms in economics (Allais, 1968). It will be interesting to study the behaviour of the Bayes estimator of β as against the well-known classical estimators. In this study, we have assumed the threshold parameter θ_o known and obtain the Bayes estimator of β under JIP, MIP, and NCP and a variety of loss functions. Properties of these estimators are also discussed. Based on a Monte Carlo study, the sampling distributions of the Bayes estimators of β under JIP and MIP are obtained. Appropriate Pearsonian curves are fitted to these sampling distributions and Chi-square 'goodness of fit' test performed.

In Chapter Three, Bayes estimators of the scale parameter a and shape parameter m in Pearson's type VII distribution with density

$$f(x|a,m) = \frac{\Gamma(m)}{a\sqrt{\pi}\ \Gamma(m-\frac{1}{2})}\left(1+\frac{x^2}{a^2}\right)^{-m}, \quad -\infty < x < \infty,\ a > 0,\ m > \frac{1}{2}$$

are obtained. The families of improper priors used have densities

$$g_1(a) \propto \frac{1}{a^t}, \quad t > 0, \quad \text{for estimation of } a \text{ when } m \text{ is known}$$

$$g_2(m) \propto \frac{1}{m^t}, \quad t > 0, \quad \text{for estimation of } m \text{ when } a \text{ is known}$$

$$g_3(a,m) \propto \frac{1}{am}, \quad \text{for estimation of } a \text{ and } m \text{ when both are unknown.}$$

A special case of the family of improper priors is Jeffreys' Invariant Prior (JIP) when $t = 1$.

The estimators based on the classical methods, namely, the method of maximum likelihood and the method of moments are obtained. Based on a Monte Carlo study these estimators are compared with their Bayesian counterparts. Appropriate Pearsonian curves are fitted to the sampling distributions of the Bayesian estimators under $g_3(a,m)$. The Chi-square, Kolmogorov-Smirnov and Cramer von Mises goodness of fit tests are performed.

In Chapter 4, we considered the Rayleigh probability density function

$$f(x|\sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad x, \sigma > 0$$

and obtained the Bayes' estimators of σ and reliability function, R_t, under MIP and JIP. Based on a Monte Carlo study, these estimators were compared with their ml and umvu estimators. HPD and equal-tail credible intervals for σ and R_t have been obtained. We also considered the prediction of a future observation.

In Chapter 5, we considered NCP for σ for the two parameter normal density

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \qquad -\infty < x, \mu < \infty, \quad \sigma > 0,$$

and obtained Bayes' estimators using NCP as well as a family of improper priors for σ and μ under a variety of loss functions. Based on a Monte Carlo study, we compared the estimators and searched for the optimum loss function which yields the minimum mean square estimators for all combinations of prior parameters involved.

In Chapter 6, we considered the model

$$E(Y|x) = \alpha + \beta(x-\bar{x}).$$

Bayes estimators of (α, β) are obtained when the prior distributions of α and β are uniform, and that of the error variance σ^2 is inverted gamma, and the distributions are independent. The case when (α, β) has a bivariate normal distribution and σ^2 is known is also considered. A Monte Carlo study is used to compare the Bayes estimators with their maximum likelihood counterparts.

The programmes used for Monte Carlo studies were written in Fortran WATFIV. The University of Manitoba computer facilities were used. These programmes are available on request.

For a variety of reasons not all the priors and loss functions mentioned were exhausted in this dissertation. For example, NCP was not used in the type VII distribution, since there are no sufficient statistics for the parameters of this distribution and hence no NCP can be generated. Again for the estimation of α and β in the regression model, the logarithmic squared error loss function cannot be used for obvious reason of $-\infty < \alpha,\ \beta < \infty$.

In the limited space of this dissertation, it was not possible to use all the possible combinations of loss functions and priors; as a result, we had to make some compromise. This also applies to the goodness of fit tests to the Pearsonian curves, fitted to the sampling distributions of the Bayesian estimators in Chapter Two.

One should be careful that a global conclusion can hardly be made from Monte Carlo studies no matter how extensive they are and as such, the conclusions drawn may not apply in general.

The priors and loss functions used in this dissertation for various distributions are shown in Table 1.1.

TABLE 1.1

Priors and Loss Functions Used in This Dissertation

	JIP	MIP	NCP	Loss Functions
Pareto	√	√	√	SEL, LSEL, GLF
Type VII	√	*	No suff. stat.	SEL (LSEL and GLF appeared intractable)
Rayleigh	√	**	Similar to Pareto	SEL (LSEL and GLF are similar to Pareto)
Normal	√	**	√	SEL, LSEL, GLF
Regression	√	*	√	SEL (LSEL not possible since $-\infty < \alpha, \beta < \infty$; GLF is intractable)

* Requires more extensive simulation which will be continued.

** Under independence JIP and MIP are the same.

CHAPTER TWO

PARETO DISTRIBUTION (*)

2.0 INTRODUCTION

We consider the Pareto probability density function

$$f(x|\theta_o,\beta) = \frac{\beta\theta_o^{\beta}}{x^{\beta+1}}, \quad 0 < \beta < \infty, \ 0 < \theta_o < x < \infty \qquad (2.1)$$

The Pareto pdf is often used to represent the distribution of income X above a known income θ_o. β measures the degree of inequality. Zellner (1971) considered the estimation of β under Jeffreys' invariant prior. In this chapter we will examine the behaviour of the Bayesian estimator of the shape parameter β under the Natural Conjugate Prior (NCP), Minimal Information Prior (MIP) (Zellner, 1971), and a family of Non-Informative Prior (NIP) distributions of β subject to a vareity of loss functions. We will further work out a Monte Carlo study on the sampling distributions of β^*, fit appropriate Pearsonian system of curves to these distributions and test their goodness of fit.

2.1 DERIVATION OF NCP, MIP, AND JIP FOR THE PARETO-PARAMETER

Let $\underset{\sim}{x} = (x_1,x_2,\ldots,x_n)$ be a random sample of size n from the pdf (2.1). The likelihood function is given by

$$\ell(\underset{\sim}{x}|\theta_o,\beta) = \beta^n\theta_o^{n\beta}/\prod_{i=1}^{n} x_i^{\beta+1} = (\frac{\beta}{G})^n\exp(-n\beta c_o) \qquad (2.2)$$

where $G = (x_1\cdot x_2\cdot\ldots\cdot x_n)^{1/n}$ and $c_o = \log(\frac{G}{\theta_o})$.

(*) Part of the results in this chapter have been published in the Journal of the Society of Management Sciences and Applied Cybernetics (SCIMA), Vol. 9, No. 1 (Sinha and Howlader (1980)).

In order to generate the NCP for β, we have to show that there exists a sufficient statistic. We may express the likelihood (2.2) as belonging to the general exponential family by putting

$$\ell(\underset{\sim}{x}|\theta_o,\beta) \equiv A(\beta)\cdot B(x)\cdot \exp\{c(\beta)\cdot d(x)\}$$

where

$$A(\beta) = \exp\{n\log\beta + (n\beta)\log\theta_o\},$$

$$B(x) = \exp\left\{-\sum_{i=1}^{n}\log x_i\right\},$$

$$c(\beta) = \beta$$

and

$$d(x) = -\sum_{i=1}^{n}\log x_i$$

which implies that $\sum_{i=1}^{n}\log x_i$ is sufficient for β (Hogg and Craig, 1978)

We may re-write (2.2) in the form

$$\ell(\underset{\sim}{x}|\theta_o,\beta) = \beta^n\left\{\frac{\exp(-\beta c_o)}{G}\right\}^n \tag{2.3}$$

and obtain the NCP for β as

$$g_1(\beta) \propto \beta^{c-1}\exp(-p\beta), \quad p > 0,\ c \geq 1 \tag{2.4}$$

where c and p are the prior parameters.

The normalizing constant of (2.4) is $\frac{p^c}{\Gamma(c)}$. Thus, the NCP for the Pareto distribution is a gamma distribution, G(p,c).

The MIP (Zellner, 1971) is given by

$$g_2(\beta) \propto \exp[I_x(\beta)]$$

where

$$I_x(\beta) = \int_{\theta_o}^{\infty} f(x|\theta_o,\beta)\log f(x|\theta_o,\beta)dx$$

$$= \int_{\theta_o}^{\infty} \{\log\beta + \beta\log\theta_o - (\beta+1)\log x\} f(x|\theta_o,\beta)dx$$

$$= \log\beta + \beta\log\theta_o - (\beta+1)E(\log x)$$

$$= \log\beta + \beta\log\theta_o - (\beta+1)(\log\theta_o + \frac{1}{\beta})$$

$$= \log\beta - \frac{1}{\beta} - 1 - \log\theta_o.$$

Thus

$$g_2(\beta) \propto \beta\exp(-\frac{1}{\beta}), \quad 0 < \beta < \infty. \tag{2.5}$$

Jeffreys' Invariant Prior (JIP) distribution of β is given by

$$g_3(\beta) \propto |I(\beta)|^{\frac{1}{2}}$$

where $I(\beta)$ is Fisher's Information matrix, and is obtained as

$$g_3(\beta) \propto \frac{1}{\beta}, \quad 0 < \beta < \infty.$$

With $g_3(\beta) = g_1(\beta)$ in (2.4) when $c = 0$, $p = 0$.

2.2 POSTERIOR DENSITIES OF β UNDER NCP, MIP, AND JIP

Combining (2.3) and (2.4), the Posterior density of β is given by

$$\Pi_1(\beta|\underset{\sim}{x},\theta_o) = k\beta^{c+n-1}\exp\{-\beta(nc_o+p)\}$$

where k is a normalizing constant and is given by

$$k^{-1} = \int_0^{\infty} \Pi_1(\beta|\underset{\sim}{x},\theta_o)d\beta = \frac{\Gamma(c+n)}{(nc_o+p)^{c+n}}.$$

Thus

$$\Pi_1(\beta|\underset{\sim}{x},\theta_o) = \frac{(nc_o+p)^{c+n}}{\Gamma(c+n)}\,\beta^{c+n-1}\exp\{-\beta(nc_o+p)\} \tag{2.6}$$

The posterior (2.6) may be seen to be a $G(nc_o+p,\ c+n)$.

From (2.3) and (2.5), we have the Posterior density of β under MIP,

$$\Pi_2(\beta|\underset{\sim}{x},\theta_o) = \frac{\beta^{n+1}\exp\{-(n\beta c_o + \frac{1}{\beta})\}}{\int_0^\infty \beta^{n+1}\exp\{-(n\beta c_o + \frac{1}{\beta})\}d\beta},\quad 0 < \beta < \infty. \tag{2.7}$$

We evaluate the denominator by using the modified Bessel function of the third kind of order ν (Erdelyi, et al, 1953) given by

$$\frac{2}{a^\nu}\,k_\nu(az) = \int_0^\infty t^{-\nu-1}\exp\{-\frac{z}{2}(t+\frac{a^2}{t})\}dt \tag{2.8}$$

where

$$k_\nu(z) = k_{-\nu}(z).$$

Replacing z, a, ν in (2.8) respectively by $2nc_o$, $(nc_o)^{-\frac{1}{2}}$ and $-(n+2)$ we have from (2.7),

$$\Pi_2(\beta|\theta_o,\underset{\sim}{x}) = \frac{(nc_o)^{\frac{n+2}{2}}}{2k_{n+2}\{2(nc_o)^{\frac{1}{2}}\}}\,\beta^{n+1}\exp\{-(n\beta c_o+\frac{1}{\beta})\},\quad 0 < \beta < \infty. \tag{2.9}$$

We now consider a family of improper priors

$$g(\beta) \propto \frac{1}{\beta^t},\quad t > 0. \tag{2.10}$$

of which JIP is a particular case with $t = 1$.

Under the prior (2.10) we have the corresponding posterior density

$$\Pi_3(\beta|\underset{\sim}{x}) = \frac{(nc_o)^{n-t+1}}{\Gamma(n-t+1)}\,\beta^{n-t}\exp(-n\beta c_o),\quad 0 < \beta < \infty,\quad t < (n+1). \tag{2.11}$$

We generated a sample of size $n = 100$ from the Pareto-pdf (2.1) with $\theta_o = 5000$ and $\beta = .7$ and plotted the posterior distribution (2.11) for $t = 1,2,3$ in Figure 1. It appears that the farther one deviates from the JIP, the posterior tends to be less and less robust since the curves corresponding to $t = 2$ and $t = 3$ slip away from $\beta = .7$. It is also evident from the fact that as t increases from unity the prior proportional to β^{-t} also increases for $\beta < 1$.

2.3 LOSS FUNCTIONS

In making a decision, the consequence of estimating the true parameter θ by a quantity T is measured by the loss, $L(T,\theta)$. Since θ is unknown, the actual loss in any situation will not be known but T can be determined for the data $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ for which the expected Posterior loss $E[L(T,\theta)|\underset{\sim}{x}]$ is minimized. We consider the following loss functions:

(i) squared error loss (SEL):

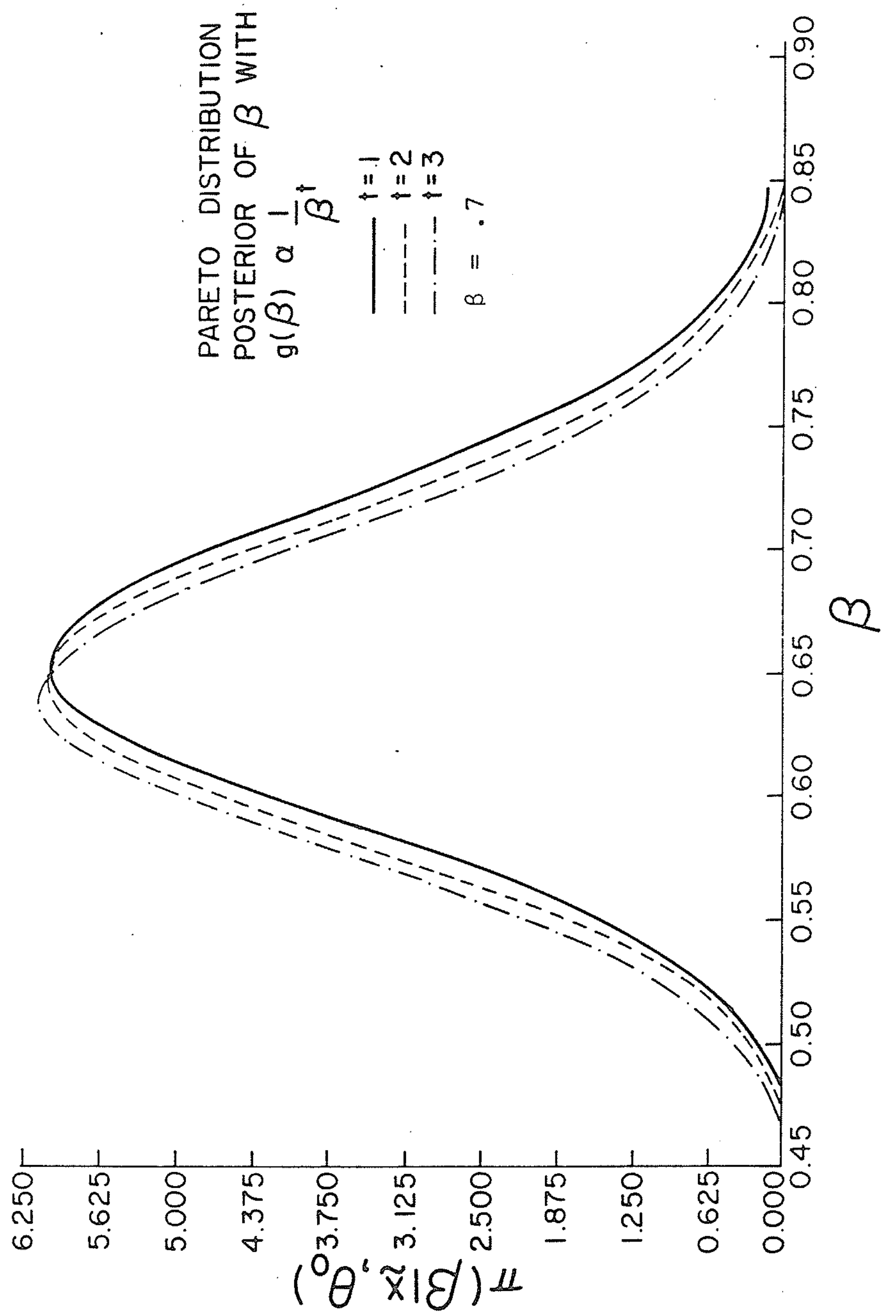
$$L(T,\theta) = (T-\theta)^2 \tag{2.12}$$

which may be used when decisions taken become gradually more damaging for a larger error in estimating θ by T, i.e., one pays more for a larger error. The expected posterior loss for this loss function is

$$E[L(T,\theta)|\underset{\sim}{x}] = \int (T-\theta)^2 \Pi(\theta|\underset{\sim}{x})\,d\theta$$

which is minimized at

$$T = E(\theta|\underset{\sim}{x}).$$



PARETO DISTRIBUTION
POSTERIOR OF β WITH
g(β) α 1/β^t
t = 1
t = 2
t = 3
β = .7
6.250
5.625
5.000
4.375
3.750
3.125
2.500
1.875
1.250
0.625
0.000
π(β|x̃, θ_o)
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
β

Fig. 1.

Thus, under SEL, the posterior mean is the point estimator for θ.

(ii) Logarithmic squared error loss (LSEL):

$$L(T,\theta) = (\log T - \log\theta)^2, \quad \theta > 0 \tag{2.13}$$

$$\frac{\partial L(T,\theta)}{\partial T} = \frac{2}{T}(\log T - \log\theta).$$

Thus, $\frac{\partial}{\partial T} \int L(T,\theta)\Pi(\theta|\underset{\sim}{x})d\theta = 0$, gives

$$\log T - E(\log\theta|\underset{\sim}{x}) = 0$$

or

$$\log T = E(\log\theta|\underset{\sim}{x}).$$

In a Bayesian analysis, a loss function represents a realistic monetary penalty. The loss functions discussed above may not be appropriate in certain situations and hence other types of loss functions such as those proposed by Goodman (1960) and used in El-Sayyad (1967) need to be investigated.

(iii) A general loss function (GLF) of the form

$$L(T,\theta) = \lambda(\theta)(T^m - \theta^m)^2 \quad \text{(El-Sayyad, 1967)} \tag{2.14}$$

where $\lambda(\theta)$ is a weight function and m may be positive or negative. In this dissertation we will use $\lambda(\theta) = \theta^{\ell}$. The Bayesian estimator of θ is the solution of

$$\frac{\partial}{\partial T} \int L(T,\theta)\Pi(\theta|\underset{\sim}{x})d\theta = 0$$

or

$$\int \lambda(\theta)2(T^m-\theta^m)mT^{m-1}\Pi(\theta|\underset{\sim}{x})d\theta = 0$$

or

$$T^m E[\lambda(\theta)|\underset{\sim}{x}] - E[\lambda(\theta)\theta^m|\underset{\sim}{x}] = 0$$

or

$$T = \left\{\frac{E[\lambda(\theta)\theta^m|\underset{\sim}{x}]}{E[\lambda(\theta)|\underset{\sim}{x}]}\right\}^{1/m} . \tag{2.15}$$

For $\lambda(\theta) = 1$, $m = 1$, we have the SEL estimator.

It is easy to show that for $\lambda(\theta) = 1$, $m \to 0$, GLF tends to LSEL. To prove this, we divide both sides of (2.14) by m^2 which will not change the loss since m is a constant, giving

$$L^*(T,\theta) = \lambda(\theta)\left(\frac{T^m-1}{m} - \frac{\theta^m-1}{m}\right)^2 .$$

Using the result $\lim_{m\to 0}\left(\frac{p^m-1}{m}\right) = \log p$, $L^*(T,\theta)$ takes the form

$$L^*(T,\theta) = \lambda(\theta)(\log T - \log\theta)^2 \quad \text{(El-Sayyad, 1967).}$$

2.4 LOSS FUNCTIONS AND CORRESPONDING ESTIMATORS

It is shown in Section 2.3 that under the Squared-Error-Loss (SEL) function, the Bayes estimator for β,

$$\beta^* = E(\beta|\underset{\sim}{x}).$$

Considering the posterior (2.6) under NCP, we have

$$\beta_1^* = \int_0^\infty \beta \Pi_1(\beta|\underset{\sim}{x})d\beta$$

$$= \frac{c+n}{nc_o+p} \to \frac{1}{c_o} \qquad (2.16)$$

max-likelihood estimator (MLE) of β as $n \to \infty$. The corresponding posterior variance

$$V(\beta|\underset{\sim}{x}) = \frac{c + n}{(nc_o+p)^2}.$$

Similarly, from (2.9), the Bayes estimator of β under MIP and SEL is

$$\beta_2^* = \frac{(nc_o)^{\frac{n+2}{2}}}{2k_{n+2}\{2(nc_o)^{\frac{1}{2}}\}} \int_0^\infty \beta^{n+2} \exp\{-(nc_o + \frac{1}{\beta})\}d\theta$$

$$= (nc_o)^{-\frac{1}{2}} \left[\frac{k_{n+3}\{2(nc_o)^{\frac{1}{2}}\}}{k_{n+2}\{2(nc_o)^{\frac{1}{2}}\}}\right]. \qquad (2.17)$$

Using an asymptotic expansion (Copson, 1971)

$$k_\nu(x) \to \frac{2^{\nu-\frac{1}{2}}\nu^{\nu-\frac{1}{2}}\sqrt{\pi}\ \exp(-\nu)}{x^\nu} \qquad (2.18)$$

$$\beta_2^* \approx \frac{(n+3)^{n+\frac{5}{2}}\exp(-1)}{(nc_o)(n+2)^{n+\frac{3}{2}}} \to \frac{1}{c_o}$$

the MLE as $n \to \infty$ and

$$\mathrm{Var}(\beta|\underset{\sim}{x}) = (nc_o)^{-1}\left[\frac{k_{n+4}(a)}{k_{n+2}(a)} - \frac{k_{n+3}^2(a)}{k_{n+2}^2(a)}\right]$$

where $a = 2(nc_o)^{\frac{1}{2}}$.

The posterior (2.11) yields

$$\beta_3^* = \frac{n-t+1}{nc_o} \tag{2.19}$$

with posterior $\mathrm{var}(\beta|\underset{\sim}{x}) = \dfrac{n-t+1}{(nc_o)^2}$.

We note that under JIP (t = 1), β_3^* equals $\dfrac{1}{c_o}$, the mle of β, and for $t \neq 1$, $\beta_3^* \to \dfrac{1}{c_o}$ as $n \to \infty$.

We now consider the LSEL function and GLF given in (2.13) and (2.14) respectively. Under the LSEL, the Bayes estimator of β is $T_1 = \exp[E \log(\beta|\underset{\sim}{x},\theta_o)]$.
To evaluate $E(\log\beta|\underset{\sim}{x},\theta_o)$ we will use the digamma function, $\psi(M) = \dfrac{\partial}{\partial M} \log \Gamma(M)$ which is extensively tabulated (Abramowitz and Stegum, 1964). We have

$$\psi(M)\Gamma(M) = \frac{\partial}{\partial M} \Gamma(M) = \int_0^\infty \exp(-u)u^{M-1}(\log u)du$$

where

$$\begin{aligned}\psi(M) &= \frac{1}{M-1} + \psi(M-1)\\ &= \sum_{\gamma=1}^{M-1} \frac{1}{\gamma} + \psi(1)\end{aligned} \tag{2.20}$$

and $-\psi(1) = -\displaystyle\int_0^\infty \exp(-u)\log u\, du$ is known as Euler's constant and approximately equals 0.5772.

From the posterior of β under NCP in (2.6), we have

$$E(\log\beta|\underset{\sim}{x},\theta_o) = \frac{(nc_o+p)^{c+n}}{\Gamma(c+n)} \int_0^\infty (\log\beta)\beta^{c+n-1} \cdot\exp\{-\beta(nc_o+p)\}d\beta$$

$$= \frac{1}{\Gamma(c+n)} \int_0^\infty (\log w) \exp(-w) w^{c+n-1} dw - \log(cn_o+p)$$

$$= \frac{1}{\Gamma(M)} \int_0^\infty (\log w) \exp(-w) w^{M-1} dw - \log(nc_o+p)$$

$$= \psi(M) - \log(nc_o+p)$$

where

$$\beta(nc_o+p) = w$$

and

$$c+n = M.$$

Thus, the Bayes estimator of β under NCP and LSEL is

$$T_1^N = \frac{\exp\{\psi(M)\}}{(nc_o+p)} \tag{2.21}$$

where $\psi(M)$ is given by (2.20).

For positive integral M, using Stirling's approximation $\Gamma(M) \approx \sqrt{2\pi}\, M^{M-\frac{1}{2}} \exp(-M)$, one can easily obtain an asymptotic expression for T_1^N

$$\log \Gamma(M) = \text{constant} + (M-\tfrac{1}{2})\log M - M$$

$$\psi(M) = \log M - \frac{1}{2M} \tag{2.22}$$

$$\exp\{\psi(M)\} = M \exp(-\frac{1}{2M}) \approx M-\tfrac{1}{2},$$

ignoring terms of order $\frac{1}{M}$ and lower which leads to

$$T_1^N \approx \frac{c+n-\frac{1}{2}}{nc_o+p} \to \frac{1}{c_o}$$

$$\approx \text{MLE of } \beta \text{ as } n \to \infty.$$

Putting $c = p = 0$ in (2.21), we obtain the Bayes estimator of β under JIP and LSEL as

$$T_1^J = \frac{\exp\left\{\sum\limits_{\gamma=1}^{n-1} \frac{1}{\gamma} + \psi(1)\right\}}{nc_o}$$

$$\approx \frac{n-\frac{1}{2}}{nc_o} \to \frac{1}{c_o} = \text{MLE of } \beta \text{ as } n \to \infty.$$

Under MIP for β, using the posterior (2.7), we have

$$E(\log\beta|\underset{\sim}{x},\theta_o) = \frac{\int_0^\infty (\log\beta)\beta^{n+1}\exp\left[-nc_o\left\{\beta+\frac{(nc_o)^{-1}}{\beta}\right\}\right]d\beta}{\int_0^\infty \beta^{n+1}\exp\left[-nc_o\left\{\beta+\frac{(nc_o)^{-1}}{\beta}\right\}\right]d\beta}. \quad (2.23)$$

We evaluate the ratio of integrals (2.23) by using Lindley's technique discussed in Appendix A.1. It follows that

$$E(\log\beta|\underset{\sim}{x},\theta_o) = \log\hat{\beta} + \frac{\sigma^2}{2}\left[\frac{1}{\hat{\beta}^2} + \frac{2}{\hat{\beta}^3}\right] + \frac{\sigma^4}{2\hat{\beta}} L_3(\hat{\beta}). \quad (2.24)$$

where $\sigma^2 = -L_2^{-1}(\hat{\beta})$ defined in (3a) of Appendix A.1.
We generate a sample of size 100 from the Pareto pdf (2.1) with $\theta_o = 5000$ and $\beta = 0.7$ which yields $G = 22824.23926$, $c_o = 1.518385$, and $\hat{\beta} = 0.658594$. The log-likelihood and the differences are shown in Table 2.1.

By interpolation from Table 2.1, $L_2(\hat{\beta}) = -230.549652$ so that $\sigma^2 = 0.00433746$, $\sigma^4 = 0.000018814$ and $L_3(\hat{\beta}) = 700.125494$. Substituting in (2.24), we obtain

$$\log T_1^M = E(\log\beta|\underset{\sim}{x},\theta_o) = -0.387463.$$

Hence, the Bayes estimate of β under MIP and the LSEL function (2.13) is

$$T_1^M = 0.678777.$$

TABLE 2.1

$L(\beta) = 100(\log \beta - \log G) - 151.8385184\beta$

β	$L(\beta)$	L_1	L_2	L_3
0.656	-1145.323354650			
		0.484435631		
0.657	.322870215		-231.669981360	
		0.252765650		703.629552845
0.658	.322617449		-230.966351808	
		0.021799298		700.429268363
0.659	.322595650		-230.265922539	
		-0.208466624		
0.660	.322804116			

TABLE 2.2

Bayes Estimators T_1 Derived Under Different Priors and the LSEL Function (2.13)

$g(\beta)$	Estimator T_1
NCP	$\dfrac{\exp\{\psi(M)\}}{nc_o+p}$ where $\psi(M)$ is given by (2.20)
MIP	$\exp E(\log\beta \mid \underset{\sim}{x},\theta_o)$ where $E(\log\beta \mid \underset{\sim}{x},\theta_o)$ is given by (2.24)
JIP	$\dfrac{\exp\left\{\sum_{\gamma=1}^{n-1} \frac{1}{\gamma} + \psi(1)\right\}}{nc_o}$

It follows from (2.15) that under the GLF (2.14), Bayes estimator T_2 is given by

$$T_2 = \left[\frac{E(\beta^{\ell+m} \mid \underset{\sim}{x},\theta_o)}{E(\beta^{\ell} \mid \underset{\sim}{x},\theta_o)}\right]^{1/m} . \tag{2.25}$$

Using the natural conjugate prior for β from (2.6), we have

$$E(\beta^{\ell+m} \mid \underset{\sim}{x},\theta_o) = \frac{\Gamma(c+n+\ell+m)}{\Gamma(c+n)(nc_o+p)^{\ell+m}}$$

and

$$E(\beta^{\ell} \mid \underset{\sim}{x},\theta_o) = \frac{\Gamma(c+n+\ell)}{\Gamma(c+n)(nc_o+p)^{\ell}} .$$

Thus, the Bayes estimator of β under NCP and GLF, T_2^N, is

$$T_2^N = \frac{1}{nc_o+p}\left[\frac{\Gamma(c+n+\ell+m)}{\Gamma(c+n+\ell)}\right]^{1/m}. \tag{2.26}$$

For positive integral c, ℓ, m, using Stirling's approximation to $n! \approx \exp(-n)n^{n+\frac{1}{2}}\sqrt{2\pi}$,

$$\lim_{n\to\infty} T_2^N = \lim_{n\to\infty} \frac{1}{nc_o+p} \Big[\exp(-c-n-\ell-m+1)(c+n+\ell+m-1)^{c+n+\ell+m-\frac{1}{2}} /$$

$$\exp(-c-n-\ell+1)(c+n+\ell-1)^{c+n+\ell-\frac{1}{2}}\Big]^{1/m}$$

$$= \lim_{n\to\infty} \frac{1}{nc_o+p} \Big[\exp(-m)n^m\left(1 + \frac{c+\ell+m-1}{n}\right)^n$$

$$\cdot\left(1 + \frac{c+\ell+m-1}{n}\right)^{c+\ell+m-\frac{1}{2}} \Big/ \Big\{\left(1 + \frac{c+\ell-1}{n}\right)^n$$

$$\cdot\left(1 + \frac{c+\ell-1}{n}\right)^{c+\ell-\frac{1}{2}}\Big\}\Big]^{1/m}$$

$$= \lim_{n\to\infty} \frac{1}{c_o + \frac{p}{n}} \Big[\exp(c+\ell-1)/\exp(c+\ell-1)\Big]^{1/m} \to \frac{1}{c_o}$$

$= \text{MLE of } \beta, \text{ independently of } m.$

Under MIP for β, from (2.9) and using (2.8),

$$E(\beta^{\ell+m}|\underset{\sim}{x},\theta_o) = \frac{(nc_o)^{\frac{n+1}{2}}}{2k_{n+2}(x)}\int_0^\infty \beta^{\ell+m+n+1}\exp\{-(n\beta c_o + \frac{1}{\beta})\}d\beta$$

$$= \frac{1}{(nc_o)^{\frac{\ell+m}{2}}} \frac{k_{n+\ell+m+2}(a)}{k_{n+2}(a)}$$

and

$$E(\beta^{\ell}|\underset{\sim}{x},\theta_o) = \frac{1}{(nc_o)^{\frac{1}{2}}} \frac{k_{n+\ell+2}(a)}{k_{n+2}(a)} .$$

Thus from (2.25) the Bayes estimator of β under MIP and GLF, T_2^M, is

$$T_2^M = \frac{1}{(nc_o)^{\frac{1}{2}}}\left[\frac{k_{n+\ell+m+2}(a)}{k_{n+\ell+2}(a)}\right]^{1/m} \tag{2.27}$$

where $a = 2(nc_o)^{\frac{1}{2}}$.

Using the asymptotic expansion of $k_\nu(a)$ in (2.18),

$$T_2^M \simeq \frac{\exp(-1)}{nc_o}\left[\frac{(n+\ell+m+2)^{n+\ell+m+\frac{3}{2}}}{(n+\ell+2)^{n+\ell+\frac{3}{2}}}\right]^{1/m}$$

and

$$\lim_{n\to\infty} T_2^M = \exp(-1)\lim_{n\to\infty}\frac{1}{nc_o}\left[\frac{n^m\exp(\ell+m+2)}{\exp(\ell+2)}\right]^{1/m} \to \frac{1}{c_o}$$

$$= \text{MLE of } \beta, \text{ independently of } m.$$

Putting $c = p = 0$ in (2.26), we obtain the Bayes estimator of β under JIP and GLF, T_2^J, as

$$T_2^J = \frac{1}{nc_o}\left[\frac{\Gamma(n+\ell+m)}{\Gamma(\ell+n)}\right]^{1/m}. \tag{2.28}$$

For positive integral ℓ, m, again by Stirling's approximation of n!,

$$\lim_{n\to\infty} T_2^J \to \frac{1}{c_o} = \text{MLE of } \beta, \text{ independent of } m.$$

Further, putting $\ell = 0$, $m = 1$ we obtain our earlier squared-error-loss function estimators β_1^*, β_2^*, $[\beta_3^*]_{t=1}$ in (2.16), (2.17) and (2.19).

Bayes estimators of β with NCP, MIP and NIP and under SEL are presented in Table 2.3. Similarly, Bayes estimators of β with NCP, MIP and JIP and under GLF are presented in Table 2.4.

TABLE 2.3

Bayes Estimators β^* Derived Under Different Priors and the SEL Function $\ell(\beta^*,\beta)$

$g(\beta)\propto$	$\Pi(\beta \mid \underset{\sim}{x})$	β^*	Posterior Variance
$\beta^{c-1}\exp(-p\beta)$	$\dfrac{(nc_o+p)^{n+c}}{\Gamma(c+n)}\beta^{c+n-1} \cdot \exp\{-\beta(nc_o+p)\}$	$\dfrac{c+n}{nc_o+p}$	$\dfrac{c+n}{(nc_o+p)^2}$
$\beta\exp(-\frac{1}{\beta})$	$\dfrac{(nc_o)^{\frac{n+2}{2}}\beta^{n+1}}{2k_{n+2}\{2(nc_o)^{\frac{1}{2}}\}} \cdot \exp\{-(n\beta c_o+\frac{1}{\beta})\}$	$(nc_o)^{-\frac{1}{2}}\left[\dfrac{k_{n+3}(a)}{k_{n+2}(a)}\right]$ where $a = 2(nc_o)^{\frac{1}{2}}$	$(nc_o)^{-1}\left[\dfrac{k_{n+4}(a)}{k_{n+2}(a)} - \dfrac{k_{n+3}^2(a)}{k_{n+2}^2(a)}\right]$, where $a = 2(nc_o)^{\frac{1}{2}}$
$\dfrac{1}{\beta^t}$	$\dfrac{(nc_o)^{n-t+1}}{(n-t+1)}\beta^{n-t} \cdot \exp(-n\beta c_o)$	$\dfrac{n-t+1}{nc_o}$	$\dfrac{n-t+1}{(nc_o)^2}$

TABLE 2.4

Bayes Estimators T_2 Derived Under Different Priors and the GLF (2.14)

$g(\beta)$	Estimator T_2
NCP	$\dfrac{1}{(nc_o+p)}\left[\dfrac{\Gamma(c+n+\ell+m)}{\Gamma(c+n+\ell)}\right]^{\frac{1}{m}}$
MIP	$\dfrac{1}{(nc_o)^{\frac{1}{2}}}\left[\dfrac{k_{n+\ell+m+2}(a)}{k_{n+\ell+2}(a)}\right]^{\frac{1}{m}}$, where $a = 2(nc_o)^{\frac{1}{2}}$
JIP	$\dfrac{1}{nc_o}\left[\dfrac{\Gamma(n+\ell+m)}{\Gamma(n+\ell)}\right]^{\frac{1}{m}}$

2.5 MONTE CARLO STUDY

In order to compare the mean-squared-errors (mse) of the various estimators β^*, T_1 and T_2, 500 (= N) samples of size 100 (= n) were generated from the Pareto pdf (2.1) with $\theta_o = 5000$, $\beta = 0.7$. The mean of the N estimators and the corresponding

$$\text{mse} = \frac{\text{(sum of the squares of the N deviations of the estimators from the true } \beta)}{N}$$

were computed and tabulated. We report the results in the following tables. Entries in the parentheses indicate the corresponding mse.

TABLE 2.5

β_1^*: Bayes Estimates of β Under NCP and SEL Function (N = 500, n = 100)

c \ p	0	1	2	3	4
0	0.711954 (0.005234)	0.706872 (0.004996)	0.701862 (0.004814)	0.696923 (0.004687)	0.692053 (0.004612)
1	0.719074 (0.005558)	0.713940 (0.005242)	0.708880 (0.004986)	0.703892 (0.004786)	0.698974 (0.004641)
2	0.726193 (0.005983)	0.721009 (0.005590)	0.715899 (0.005258)	0.710861 (0.004984)	0.705895 (0.004767)
3	0.733313 (0.006511)	0.728078 (0.006038)	0.722917 (0.005629)	0.717830 (0.005280)	0.712815 (0.004990)
4	0.740433 (0.007142)	0.735146 (0.006588)	0.729936 (0.006099)	0.724800 (0.005674)	0.719736 (0.005309)

An extension of Table 2.5 is given in Appendix A.3.

Note that for a given c, the mse monotonically decreases as p increases which suggests that for minimum mse-estimator of β the prior $g(\beta) \propto \frac{1}{\beta} \exp(-p\beta)$, $\beta > 0$ with $p > 0$ and large should be recommended.

TABLE 2.6

β_2^*, β_3^*: Bayes Estimates of β Under SEL Function with MIP and JIP ($N = 500$, $n = 100$)

MIP β_2^*	JIP β_3^*	MLE $\hat{\beta}$
0.7262	0.7120	0.7120
(0.0060)	(0.0052)	(0.0052)

Note that β_2^* in (2.17) was computed by using the asymptotic expansion of $k_\nu(a)$ in (2.18) and β_3^* is obtained by putting $c = p = 0$ in Table 2.5. Comparing the corresponding mean-squared-errors in Table 2.6 it follows that the estimator based on JIP is more desirable than the one based on MIP for β.

Tables 2.7 and 2.8 clearly indicate that under the GLF (2.14), the mse of Bayes estimators based on JIP in this study consistently have smaller mse than those derived with MIP as the prior for β. If MIP is to be used for β, one should prefer the SEL function to the class of GLF (2.14).

TABLE 2.7

T_2^M: Bayes Estimates of β Under MIP and GLF (2.14) for Specific Combinations of ℓ, m (N = 500, n = 100)

ℓ \ m	1	2	3
1	0.733319 (0.006512)	0.736870 (0.006814)	0.740415 (0.007140)
2	0.740438 (0.007142)	0.743989 (0.007495)	0.747535 (0.007873)
3	0.747558 (0.007875)	0.751109 (0.008279)	0.754655 (0.008708)

TABLE 2.8

T_2^J: Bayes Estimates of β Under JIP and GLF (2.14) for Specific Combinations of ℓ, m (N = 500, n = 100)

ℓ \ m	1	2	3
1	0.719074 (0.005558)	0.722625 (0.005757)	0.726170 (0.005982)
2	0.726193 (0.005983)	0.729745 (0.006234)	0.733290 (0.006509)
3	0.733313 (0.006511)	0.736865 (0.006813)	0.740410 (0.007140)

We find the same trend in Table 2.9 as in Tables 2.5 and 2.7 to 2.8, viz., for a given c, the mse monotonically decreases as p increases; thus, for minimum mse with NCP for β, the recommended prior would be $g(\beta) \propto \frac{1}{\beta} \exp(-p\beta)$, $\beta > 0$ and $p > 0$ and large.

TABLE 2.9

T_1^N: Bayes Estimates of β under NCP and LSEL Function (2.13) (N = 500, n = 100)

c \ p	0	1	2	3	4
0	0.708398 (0.005111)	0.703340 (0.004910)	0.698355 (0.004765)	0.693441 (0.004674)	0.688596 (0.004600)
1	0.715517 (0.005384)	0.710409 (0.005106)	0.705374 (0.004888)	0.700410 (0.004724)	0.695517 (0.004614)
2	0.722637 (0.005758)	0.717478 (0.005404)	0.712392 (0.005109)	0.707379 (0.004873)	0.702437 (0.004692)
3	0.729756 (0.006235)	0.724546 (0.005802)	0.719411 (0.005431)	0.714349 (0.005120)	0.709358 (0.004866)
4	0.736876 (0.006814)	0.731615 (0.006300)	0.726430 (0.005852)	0.721318 (0.005465)	0.716278 (0.005138)

Table 2.10 shows that T_1^J has a smaller mse compared to T_1^M. Comparing Tables 2.5, 2.6, 2.9, and 2.10, we find that LSEL yields a smaller mse compared to the SEL function as well as GLF under JIP and NCP for all combinations of c, p, ℓ, m.

TABLE 2.10

T_1^J, T_1^M: Bayes Estimates of β Under LSEL Function (2.13) with JIP and MIP (N = 500, n = 100)

T_1^J	T_1^M
0.708398	0.734159
(0.005111)	(0.006825)

For the family of loss functions $\ell(T,\beta) = \beta^{\ell}(T^m-\beta^m)^2$ and the prior $g(\beta) \propto \frac{1}{\beta}\exp(-p\beta)$, $\beta > 0$ with $p > 0$ and large, the combination $\ell = 1 = m$ yields the minimum-mse-estimator of β as shown in Tables 2.11 - 2.19. It further follows from Tables 2.5 and 2.11 - 2.19 that if it is a question of choice between the SEL function $\ell(\beta^*,\beta) = (\beta^*-\beta)^2$ and the class of GLF (2.14), one should use the SEL function which yields uniformly smaller mse for all combinations of c, p, ℓ and m.

The posteriors (2.7) and (2.11) plotted in Figure 2 support the findings in the tables that compared to JIP, MIP consistently overestimates β.

TABLES 2.11 - 2.19

T_2^C: Bayes Estimate of β Under NCP and GLF (2.14) for Specific Combinations of ℓ, m (N = 500, n = 100)

TABLE 2.11: (ℓ = 1, m = 1)

c \ p	0	1	2	3	4
0	0.719074 (0.005558)	0.713940 (0.005242)	0.708880 (0.004986)	0.703892 (0.004786)	0.698974 (0.004641)
1	0.726193 (0.005983)	0.721009 (0.005590)	0.715899 (0.005258)	0.710861 (0.004984)	0.705895 (0.004767)
2	0.733313 (0.006511)	0.728078 (0.006038)	0.722917 (0.005629)	0.717830 (0.005280)	0.712815 (0.004990)
3	0.740433 (0.007142)	0.735146 (0.006588)	0.729936 (0.006099)	0.724800 (0.005674)	0.719736 (0.005309)
4	0.747552 (0.007875)	0.742215 (0.007238)	0.736955 (0.006669)	0.731769 (0.006166)	0.726656 (0.005725)

TABLE 2.12: (ℓ = 1, m = 2)

c \ p	0	1	2	3	4
0	0.722625 (0.005757)	0.717466 (0.005403)	0.712381 (0.005109)	0.707368 (0.004873)	0.702426 (0.004692)
1	0.729745 (0.006234)	0.724535 (0.005801)	0.719400 (0.005430)	0.714337 (0.005119)	0.709346 (0.004866)
2	0.736864 (0.006813)	0.731604 (0.006300)	0.726418 (0.005851)	0.721307 (0.005464)	0.716267 (0.005137)
3	0.743984 (0.007495)	0.738672 (0.006899)	0.733437 (0.006371)	0.728276 (0.005907)	0.723188 (0.005505)
4	0.751103 (0.008278)	0.745741 (0.007600)	0.740456 (0.006991)	0.735245 (0.006448)	0.730108 (0.005969)

TABLE 2.13: (ℓ = 1, m = 3)

c \ p	0	1	2	3	4
0	0.726170 (0.005982)	0.720986 (0.005588)	0.715876 (0.005257)	0.710838 (0.004983)	0.705872 (0.004767)
1	0.733290 (0.006509)	0.728055 (0.006037)	0.722895 (0.005627)	0.717808 (0.005279)	0.712793 (0.004989)
2	0.740410 (0.007140)	0.735124 (0.006586)	0.729913 (0.006098)	0.724777 (0.005673)	0.719713 (0.005308)
3	0.747529 (0.007872)	0.742193 (0.007236)	0.736932 (0.006667)	0.731747 (0.006164)	0.726634 (0.005724)
4	0.754649 (0.008707)	0.749262 (0.007987)	0.743951 (0.007336)	0.738716 (0.006754)	0.733555 (0.006237)

TABLE 2.14: (ℓ = 2, m = 1)

0	0.726193 (0.005983)	0.721009 (0.005590)	0.715899 (0.005258)	0.710861 (0.004984)	0.705895 (0.004767)
1	0.733313 (0.006511)	0.728078 (0.006038)	0.722917 (0.005629)	0.717830 (0.005280)	0.712815 (0.004990)
2	0.740433 (0.007142)	00.735146 (0.006588)	0.729936 (0.006099)	0.724800 (0.005674)	0.719736 (0.005309)
3	0.747552 (0.007875)	0.742215 (0.007238)	0.736955 (0.006669)	0.731769 (0.006166)	0.726656 (0.005725)
4	0.754672 (0.008710)	0.749284 (0.007989)	0.743973 (0.007339)	0.738738 (0.006756)	0.733577 (0.006238)

TABLE 2.15: (ℓ = 2, m = 2)

0	0.729745 (0.006511)	0.728078 (0.006038)	0.722917 (0.005629)	0.717830 (0.005280)	0.712815 (0.004990)
1	0.740433 (0.007142)	0.735146 (0.006588)	0.729936 (0.006099)	0.724800 (0.005674)	0.719736 (0.005309)
2	0.747552 (0.007875)	0.742215 (0.007238)	0.736955 (0.006669)	0.731769 (0.006166)	0.726656 (0.005725)
3	0.754672 (0.008710)	0.749284 (0.007989)	0.743973 (0.007339)	0.738738 (0.006756)	0.733577 (0.006238)
4	0.761791 (0.009647)	0.756353 (0.008841)	0.750992 (0.008108)	0.745707 (0.007444)	0.740497 (0.006848)

TABLE 2.18: (ℓ = 3, m = 2)

c \ p	0	1	2	3	4
0	0.736864 (0.006813)	0.731604 (0.006300)	0.726418 (0.005851)	0.721307 (0.005464)	0.716267 (0.005137)
1	0.743984 (0.007495)	0.738672 (0.006899)	0.733437 (0.006371)	0.728276 (0.005907)	0.723188 (0.005505)
2	0.751103 (0.008278)	0.745741 (0.007600)	0.740456 (0.006991)	0.735245 (0.006448)	0.730108 (0.005969)
3	0.758223 (0.009165)	0.752810 (0.008401)	0.747474 (0.007710)	0.742214 (0.007087)	0.737029 (0.006530)
4	0.765343 (0.010153)	0.759879 (0.009304)	0.754493 (0.008528)	0.749184 (0.007824)	0.743949 (0.007188)

TABLE 2.19: (ℓ = 3, m = 3)

0	0.740410 (0.007140)	0.735124 (0.006586)	0.729913 (0.006098)	0.724777 (0.005673)	0.719713 (0.005308)
1	0.747529 (0.007872)	0.742193 (0.007236)	0.736932 (0.006667)	0.731747 (0.006164)	0.726634 (0.005724)
2	0.754649 (0.008707)	0.749262 (0.007987)	0.743951 (0.007336)	0.738716 (0.006754)	0.733555 (0.006237)
3	0.761769 (0.009644)	0.756331 (0.008838)	0.750970 (0.008105)	0.745686 (0.007442)	0.740476 (0.006846)
4	0.768889 (0.010684)	0.763399 (0.009791)	0.757989 (0.008973)	0.752655 (0.008228)	0.747396 (0.007552)

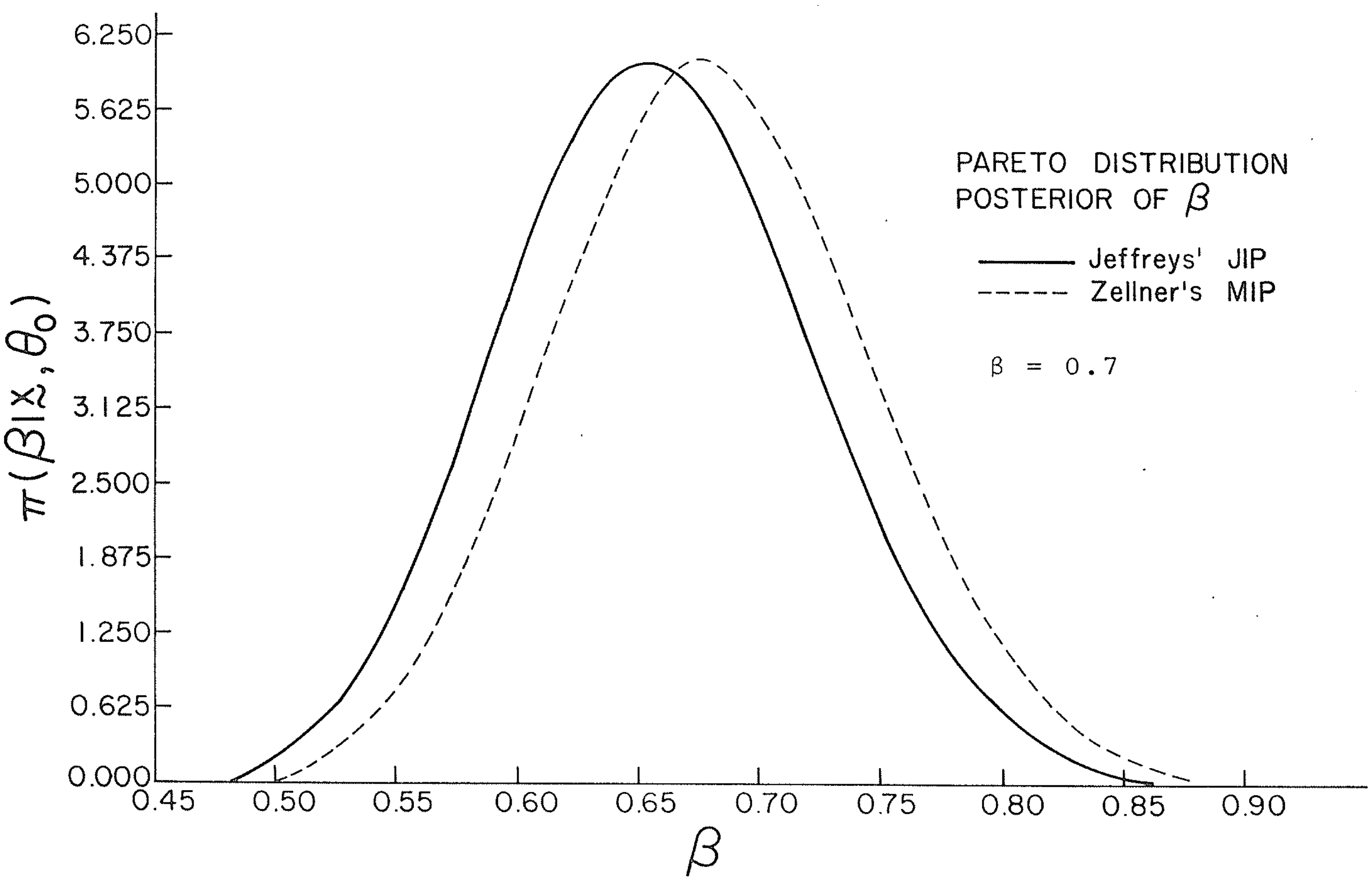

PARETO DISTRIBUTION
POSTERIOR OF β
Jeffreys' JIP
Zellner's MIP
β = 0.7
$\pi(\beta | \underset{\sim}{x}, \theta_0)$
6.250
5.625
5.000
4.375
3.750
3.125
2.500
1.875
1.250
0.625
0.000
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
β

Fig. 2.

The results of the simulation study as reported in Tables 2.5 - 2.19 lead to the following recommendations about the loss functions and priors of β:

TABLE 2.20

Minimum-Mean Square-Estimator

Prior / Loss Function	$g(\beta)$: With Increasing Order of MSE	
SEL	NCP	$g_1(\beta) \propto \frac{1}{\beta} \exp(-p\beta)$,
		$\beta > 0$, $p > 0$ and large
	JIP	$g_3(\beta) \propto \frac{1}{\beta}$
	MIP	$g_2(\beta) \propto \beta \exp(-\frac{1}{\beta})$
$[GLF]_{\ell=m=1}$	NCP	$g_1(\beta) \propto \frac{1}{\beta} \exp(-p\beta)$,
		$\beta > 0$, $p > 0$ and large
All (ℓ, m)	JIP	$g_3(\beta) \propto \frac{1}{\beta}$
	MIP	$g_2(\beta) \propto \beta \exp(-\frac{1}{\beta})$
LSEL	NCP and JIP (optimal choice) each leads to smaller mse estimator compared to SEL or GLF.	

Except for $p > 0$ and large in NCP, Jeffreys' invariant prior $g(\beta) \propto \frac{1}{\beta}$ yields smaller mse-estimators for each of the loss functions and prior distributions under consideration. JIP is specially suited for the LSEL function (2.13). If a choice is to be made between the GLF (2.14) and LSEL function (2.13), one should choose the latter with NCP or JIP for β. If one would prefer NCP/MIP, the study recommends that one should use LSEL/SEL.

Our observations in the Monte Carlo study may be summarized as follows:

(i) The loss function $\ell(T,\beta) = (\log T - \log \beta)^2$ leads to uniformly minimum mse estimators with JIP and NCP for β for all combinations of (c,p,ℓ,m) under consideration as against the corresponding priors with SEL and GLF.

(ii) The JIP, which is often objected to on the grounds that it is 'improper', still leads to sound results.

(iii) If one would prefer MIP, the SEL (2.12) and GLF (2.14) with $\ell = m = 1$ are recommended in that order.

2.6 SAMPLING DISTRIBUTIONS OF β^* WITH JIP AND MIP

The sampling distributions of β^* derived under the SEL function with JIP and MIP for β ($N = 500$, $n = 100$) are represented in the histograms in Figure 3. The momental constants and $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$ were computed using Sheppard's corrections and the point (β_1, β_2) was plotted on the chart relating the type of Pearson frequency curve to the value of β_1, β_2 (Pearson and Hartley, 1966, Vol. 1).

For the sampling distribution of β^* with JIP for β, using Sheppard's corrections we obtain

$$\mu_2 = 0.00510923, \quad \mu_3 = 0.00003366, \quad \mu_4 = 0.00006626$$

$$\beta_1 = 0.008495, \quad \beta_2 = 2.538173.$$

The point (β_1, β_2) suggests that Pearson's Type II curve

$$f(x) = \frac{\Gamma(m+\frac{3}{2})N}{a\sqrt{\pi}\,\Gamma(m+1)}\left(1 - \frac{x^2}{a^2}\right)^m, \quad -a < x < a \tag{2.29}$$

should fit the data where

$$x = X - \mu_1' = X - 0.7122, \quad X = \text{class mark},$$

$$m = \frac{5\beta_2 - 9}{2(3-\beta_2)} = 0.3995939$$

and

$$a = \sqrt{\frac{2\mu_2\beta_2}{3-\beta_2}} = 0.236981,$$

$$N = 500.$$

HISTOGRAMS OF THE SAMPLING DISTRIBUTIONS OF β^* UNDER JEFFREYS' NIP AND ZELLNER'S MIP

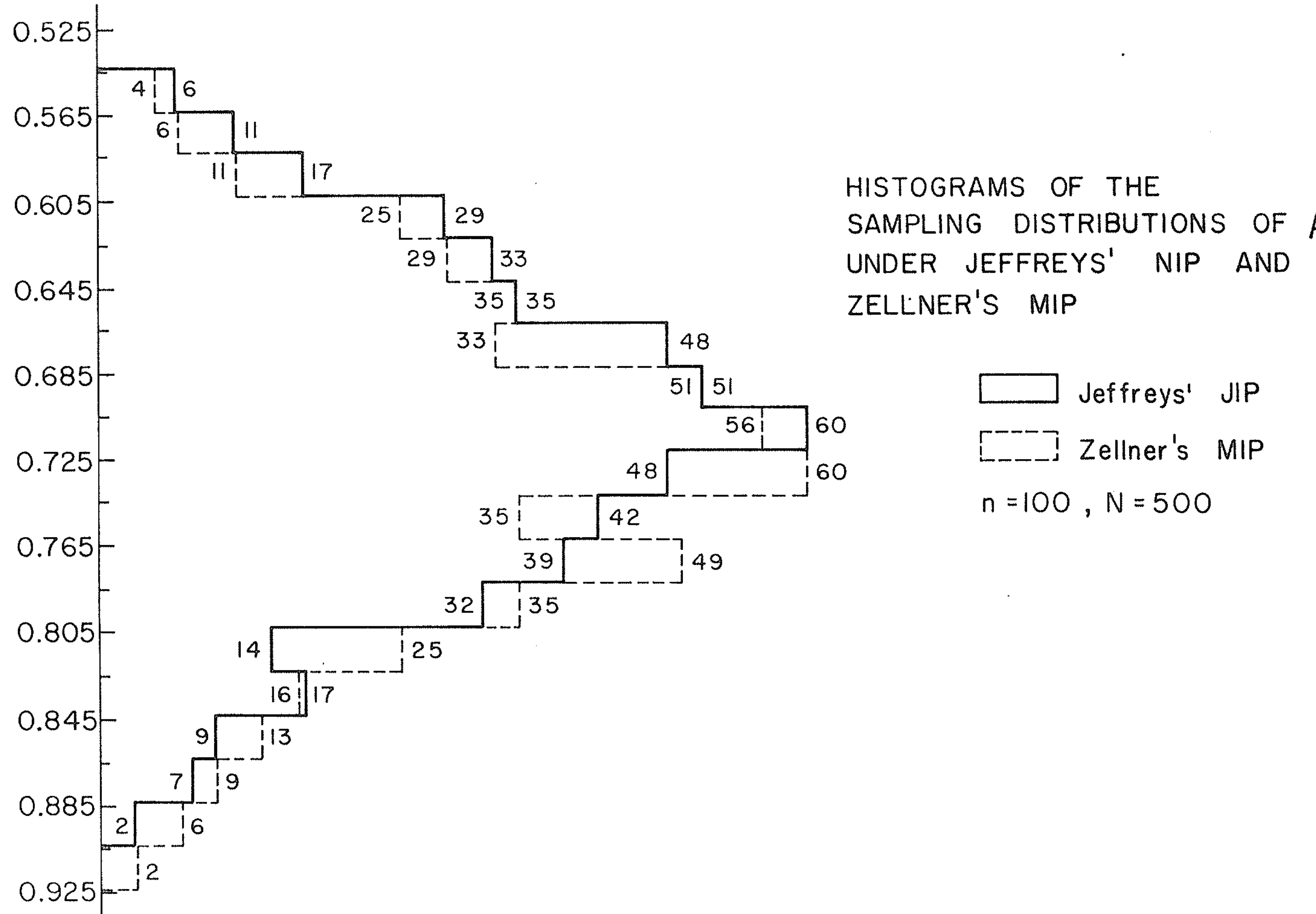


Fig. 3.

Substituting the estimated parameters we obtain

$$f(x) = 2595.0236\left(1 - \frac{x^2}{0.05616}\right)^{3.995939}, \qquad (2.30)$$

$$-0.2370 < x < 0.2370.$$

TABLE 2.21

Ordinates of Pearson's Type II Distribution for the Sampling Distribution of β* with JIP for β

X	x	f(x)	X	x	f(x)
.485	-.2272	0.1120	.735	.0228	2500.3615
.495	-.2172	1.7123	.745	.0328	2402.0053
.505	-.2072	8.0349	.755	.0428	2272.9579
.515	-.1972	23.3292	.765	.0528	2117.2972
.525	-.1872	52.0736	.775	.0628	1939.9076
.535	-.1772	98.3764	.785	.0728	1746.2991
.545	-.1672	165.5390	.795	.0828	1542.3992
.555	-.1572	255.7623	.805	.0928	1334.3242
.565	-.1472	369.9766	.815	.1028	1128.1355
.575	-.1372	507.7752	.825	.1128	929.5906
.585	-.1272	667.4371	.835	.1228	743.8971
.595	-.1172	846.0200	.845	.1328	575.4802
.605	-.1072	1039.5107	.855	.1428	427.7746
.615	-.0972	1243.0184	.865	.1528	303.0540
.625	-.0872	1450.9981	.875	.1628	202.3106
.635	-.0772	1657.4910	.885	.1728	125.2000
.645	-.0672	1856.3747	.895	.1828	70.0659
.655	-.0572	2041.6092	.905	.1928	34.0625
.665	-.0472	2207.4738	.915	.2028	13.3919
.675	-.0372	2348.7849	.925	.2128	3.6748
.685	-.0272	2461.0895	.935	.2228	0.4756
.695	-.0172	2540.8283	.945	.2328	0.0039
.705	-.0072	2585.4649			
.715	.0028	2593.5763			
.725	.0128	2564.9036			

TABLE 2.22

Fitting of Pearson's Type II Distribution to the Sampling Distribution of β^* with JIP for β

Class Boundaries	Observed Frequencies f_o		Expected Frequencies f_e	
< 0.545	0	6	2.64	7.84
0.545 - 0.565	6		5.20	
0.565 - 0.585	11		10.23	
0.585 - 0.605	17		16.97	
0.605 - 0.625	29		24.88	
0.625 - 0.645	33		33.12	
0.645 - 0.665	35		40.78	
0.665 - 0.685	48		46.89	
0.685 - 0.705	51		50.70	
0.705 - 0.725	60		51.75	
0.725 - 0.745	48		49.89	
0.745 - 0.765	42		45.37	
0.765 - 0.785	39		38.74	
0.785 - 0.805	32		30.83	
0.805 - 0.825	14		22.59	
0.825 - 0.845	17		14.94	
0.845 - 0.865	9		8.63	
0.865 - 0.885	7	9	4.12	5.87
> 0.885	2		1.75	
TOTAL	500		500.02	

The expected frequency for each class was computed by using Simpson's $\frac{1}{3}$ rule. Consider the class 0.545-0.565. From Table 2.22:

X	x	f(x)
.545	-0.1672	165.5390
.555	-0.1572	255.7623
.656	-0.1472	369.9766

The expected frequency for the class 0 545 - 0.565 is given by

$$\frac{0.01}{3}\,[f(-0.1672) + 4f(-0.1572) + f(-0.1472)] = 5.20$$

Similarly for the class 0.865 - 0.885:

X	x	f(x)
.865	0.1528	303.0540
.875	0.1628	202.3106
.885	0.1728	125.2000

$$\frac{0.01}{3}\,[f(-.1528) + 4f(0.1628) + f(0.1728)] = 4.12.$$

From Table 2.22 we have

$$\chi^2 = 8.895 \quad \text{with} \quad \nu = 14 \text{ degrees of freedom}$$

$$\chi^2(5\%,14) = 5\% \text{ point on a } \chi^2\text{-distribution with 14 degrees of freedom}$$

$$= 23.685$$

which suggests that at 5% level, Pearson's Type II distribution (2.26) is a 'good' fit to the sampling distribution of β^* in Table 2.22. We now consider the sampling distribution of β^* with MIP for β. Using Sheppard's corrections we obtain

$\mu_2 = 0.00531550, \quad \mu_3 = 0.00003532, \quad \mu_4 = 0.00007305$

$\beta_1 = 0.008306, \quad \beta_2 = 2.58525.$

The point (β_1, β_2) suggests Pearson's Type II distribution (2.29) where m = 4.73327, a = 0.257422 and

$$f(x) = 2567.409914\left(1 - \frac{x^2}{0.066266}\right)^{4.73327} \tag{2.31}$$

$$-0.2574 < x < 0.2574.$$

The ordinates f(x) and the expected frequencies have been computed in the following tables.

TABLE 2.23

Ordinates of Pearson's Type II Distribution for the

Sampling Distribution of β^* with MIP for β

X	x = X - 0.72588	f(x)
.485	-0.24088	0.133312
.495	-0.23088	1.13543
.505	-0.22088	4.67613
.515	-0.21088	13.29529
.525	-0.20088	30.15873
.535	-0.19088	58.72503
.545	-0.18088	102.38881
.555	-0.17088	164.14265
.565	-0.16088	246.28486
.575	-0.15088	350.19043
.585	-0.14088	476.15432
.595	-0.13088	623.30993
.605	-0.12088	789.62140
.615	-0.11088	971.94415
.625	-0.10088	1166.14599
.635	-0.09088	1367.27866
.645	-0.08088	1569.78873
.655	-0.07088	1767.75602
.665	-0.06088	1955.14731
.675	-0.05088	2126.07360

X	x = X - 0.72588	f(x)
.685	-0.04088	2275.03942
.695	-0.03088	2397.17395
.705	-0.02088	2488.43440
.715	-0.01088	2545.77400
.725	-0.00088	2567.26790
.735	0.00912	2552.19261
.745	0.01912	2501.05567
.755	0.02912	2415.57487
.765	0.03912	2298.60739
.775	0.04912	2154.03176
.785	0.05912	1986.58692
.795	0.06912	1801.67448
.805	0.07912	1605.13186
.815	0.08912	1402.98548
.825	0.09912	1200.19407
.835	0.10912	1005.39347
.845	0.11912	820.65469
.855	0.12912	651.26727
.865	0.13912	500.55994
.875	0.14912	370.76993
.885	0.15912	262.97109
.895	0.16912	177.06915
.905	0.17912	111.86994
.915	0.18912	65.22303
.925	0.19912	34.23849
.934	0.20912	15.56892
.945	0.21912	5.74123
.955	0.22912	1.51259
.965	0.23912	0.21156
.975	0.24912	0.00551

TABLE 2.24

Fitting of Pearson's Type II Distribution to the

Sampling Distribution of β* with MIP for β

Class Boundaries	f_o	f_e
< 0.585	10	11.89
0.585 - 0.605	11	12.53
0.605 - 0.625	25	19.48
0.625 - 0.645	29	27.35
0.645 - 0.665	35	35.32
0.665 - 0.685	33	42.45
0.685 - 0.705	51	47.84
0.705 - 0.725	56	50.80
0.725 - 0.745	60	50.92
0.745 - 0.765	35	48.21
0.765 - 0.785	49	43.00
0.785 - 0.805	35	35.99
0.805 - 0.825	25	28.06
0.825 - 0.845	16	20.14
0.845 - 0.865	13	13.09
0.865 - 0.885	9	7.49
> 0.885	8	5.44
TOTAL	500	500

$\chi^2 = 13.80, \quad \nu = 14$

$\chi^2(5\%,14) = 23.685$ which suggests a 'good' fit.

CHAPTER THREE

PEARSON'S TYPE VII DISTRIBUTION(*)

3.0 INTRODUCTION

Estimation of the parameters in the Pearsonian system of distributions by the classical approach is not very satisfactory. The conventional method of moments, which equates the sample moments to the population moments, sometimes gives unrealistic results. The method of maximum likelihood has rarely been used to estimate the parameters in this system, possibly because the likelihood equations are not easily solvable unless numerical methods are employed.

Pearson's type VII is one of the important distributions in the system. The density of this distribution is given by

$$f(x|a,m) = \frac{\Gamma(m)}{a\sqrt{\pi}\ \Gamma(m-\frac{1}{2})}\left(1+\frac{x^2}{a^2}\right)^{-m},$$

$$-\infty < x < \infty,\ a > 0,\ m > \tfrac{1}{2} \qquad (3.1)$$

which is related to a number of well known distributions. It is symmetrical at x = 0, bell-shaped with $\beta_1 = 0$ and $\beta_2 > 3$. It resembles a normal distribution;

(*) Some of the results in this chapter have been presented at the Canadian Conference in Applied Statistics - STATISTICS '81 Canada (Montreal, April 29 - May 1, 1981).

in the normal distribution the expectation of any power of x is finite but in Type VII the expectation of any even power $\geq$ 2m-1 is infinite, where m may or may not be integral. This is a useful property which makes type VII a more general model for errors of measurements. A special case of (3.1) with m = 1 is the Cauchy density

$$\frac{1}{a\pi}\left(1+\frac{x^2}{a^2}\right)^{-1}, \quad a > 0, \; -\infty < x < \infty.$$

With $a^2 = \nu$ and $m = \frac{1}{2}(\nu+1)$ in (3.1) we have Student's t-distribution with ν degrees of freedom with density

$$\frac{\nu^{-\frac{1}{2}}}{B(\frac{1}{2}, \frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}, \quad \nu > 0, \; -\infty < x < \infty.$$

Fisher (1922) has shown that the estimation of the parameters of a type VII distribution by the method of moments is inefficient except in the region near normality.

The method of moments leads to unacceptable estimators for any type VII distribution with shape parameter $m \leq \frac{5}{2}$. Jeffreys (1961) pointed out: "when $m \leq \frac{5}{2}$, the expectation of the fourth moment is infinite. The actual fourth moment of any set of observations is finite, and therefore any set of observations derived from such a law would be interpreted as implying $m \geq \frac{5}{2}$".

Thus, this is a field in which Bayesian method of estimation of the parameters is of practical importance.

We will obtain Bayes estimators of the scale parameter a and shape parameter m in Pearson's type VII distribution with density given by (3.1). The families of improper priors we propose to use have densities

$g_1(a) \propto \frac{1}{a^t}$, $t > 0$, for estimation of a when m is known

$g_2(m) \propto \frac{1}{m^t}$, $t > 0$, for estimation of m when a is known

$g_3(a,m) \propto \frac{1}{am}$, for estimation of a and m when both are unknown.

A special case of the improper family is Jeffreys' Invariant Prior (JIP) when $t = 1$.

The estimators based on the classical methods, namely the method of maximum likelihood and the method of moments wlll also be obtained. Based on a Monte Carlo study we will compare these estimators with their Bayesian counterparts. We will also fit appropriate Pearsonian curves to the sampling distributions of the Bayesian estimators under $g_3(a,m)$, and test the goodness of fit by chi-square, Kolmogorov-Smirnov and Cramér-von Mises tests.

3.1 ESTIMATION OF a AND m BY THE CLASSICAL METHODS

Let $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ be a random sample of size n from (3.1). The likelihood is given by

$$\ell(\underline{x}|a,m) = \left\{\frac{\Gamma(m)}{\sqrt{\pi}\ \Gamma(m-\frac{1}{2})}\right\}^n a^{-n} \sum_{i=1}^{n} \left(1 + \frac{x_i^2}{a^2}\right)^{-m} \tag{3.2}$$

$$L = \log \ell(\underline{x}|a,m) = -\frac{n}{2}\log\pi + n[\log\Gamma(m)$$

$$- \log\Gamma(m-\tfrac{1}{2})] - n\log a$$

$$- m\sum_{1}^{n} \log\left(1 + \frac{x_i^2}{a^2}\right). \tag{3.3}$$

Differentiating (3.3) with respect to a, we obtain

$$\frac{\partial L}{\partial a} = -\frac{n}{a} + \frac{2m}{a}\sum_{1}^{n}\frac{x_i^2}{a^2+x_i^2}$$

$$\frac{\partial L}{\partial a} = 0$$

yields

$$\sum_{1}^{n}\frac{x_i^2}{\hat{a}^2+x_i^2} = \frac{n}{2m}. \tag{3.4}$$

The mle of a is the solution of the equation (3.4) for known m.

Similarly, differentiating (3.3) w.r.t. m and setting $\frac{\partial L}{\partial m} = 0$, we obtain

$$n[\psi(\hat{m}) - \psi(\hat{m}-\tfrac{1}{2})] = \sum_{1}^{n}\log\left(1 + \frac{x_i^2}{a^2}\right) \tag{3.5}$$

where

$$\psi(P) = \frac{\partial \log\Gamma(P)}{\partial p}$$

is the digamma function, which can be evaluated by using the asymptotic formula

$$\psi(P) = \log p - \frac{1}{2p} - \frac{1}{12p^2} + \frac{1}{120p^4} - \frac{1}{252p^6} + \ldots \qquad (3.6)$$

(Abramowitz and Stegum, 1964).

Thus, for known a, mle of m is the solution of the equation (3.5).

When a and m are both unknown, mle of a and m are the solutions of the simultaneous equations

$$\left.\begin{aligned} &\sum_1^n \frac{x_i^2}{\hat{a}^2+x_i^2} = \frac{n}{2\hat{m}} \\ &\text{and} \\ &n[\psi(\hat{m}) - \psi(\hat{m}-\tfrac{1}{2})] = \sum_1^n \log\left(1 + \frac{x_i^2}{a^2}\right). \end{aligned}\right\} \qquad (3.7)$$

To obtain the solution of (3.7) for a and m, we use the iterative method.

We wish to find a and m which satisfy

$$g(a,m) = 0 \qquad (3.8)$$

$$h(a,m) = 0 \qquad (3.9)$$

where

$$g(a,m) = \frac{\partial L}{\partial a} = -\frac{n}{a} + \frac{2mS_1}{a}$$

$$h(a,m) = \frac{\partial L}{\partial m} = nD_1 - S_2$$

$$S_1 = \sum_1^n \frac{x_i^2}{a^2+x_i^2}$$

$$S_2 = \sum_1^n \log_e\left(1 + \frac{x_i^2}{a^2}\right)$$

$$D_1 = \psi(m) - \psi(m-\tfrac{1}{2}).$$

Suppose we have initial estimates for a and m, namely, a_o and m_o; then assuming continuity of g(a,m) in the neighbourhood of (a_o, m_o) and the existence and continuity of all derivatives, by Taylor series expansion of (3.8), we have

$$g(a,m) = g(a_o,m_o) + \left.\frac{\partial g}{\partial a}\right|_{\substack{a=a_o\\ m=m_o}} (a-a_o) + \left.\frac{\partial g}{\partial m}\right|_{\substack{a=a_o\\ m=m_o}} (m-m_o)$$

$$+ \frac{1}{2}\left[\left.\frac{\partial^2 g}{\partial a^2}\right|_{\substack{a=a_o\\ m=m_o}} (a-a_o) + \left.\frac{\partial^2 g}{\partial m^2}\right|_{\substack{a=a_o\\ m=m_o}} (m-m_o)^2\right.$$

$$\left. + \left.\frac{\partial^2 g}{\partial a \partial m}\right|_{\substack{a=a_o\\ m=m_o}} (a-a_o)(m-m_o)\right] + \ldots .$$

Looking only at the linear terms, we obtain

$$g(a_o,m_o) + \left.\frac{\partial g}{\partial a}\right|_{\substack{a=a_o\\ m=m_o}} \Delta a + \left.\frac{\partial g}{\partial m}\right|_{\substack{a=a_o\\ m=m_o}} \Delta m = 0 \qquad (3.10)$$

where

$$\frac{\partial g}{\partial a} = \frac{\partial^2 L}{\partial a^2} = \frac{n}{a^2} - \frac{2m}{a^2} \sum_{i=1}^{n} \frac{x_i^2(3a^2+x_i^2)}{(a^2+x_i^2)^2}$$

$$\frac{\partial g}{\partial m} = \frac{\partial^2 L}{\partial a \partial m} = \frac{\partial h}{\partial m} = \frac{2}{a} \sum_{i=1}^{n} \frac{x_i^2}{a^2+x_i^2}$$

$$\Delta a = a - a_o$$

$$\Delta m = m - m_o$$

Similarly, from (3.9)

$$h(a_o,m_o) + \left.\frac{\partial h}{\partial a}\right|_{\substack{a=a_o\\m=m_o}} \Delta a + \left.\frac{\partial h}{\partial m}\right|_{\substack{a=a_o\\m=m_o}} \Delta m = 0 \tag{3.11}$$

where

$$\frac{\partial h}{\partial m} = n[\psi'(m) - \psi'(m-\tfrac{1}{2})].$$

In matrix notation, the equations (3.10) and (3.11) may be written as

$$\begin{pmatrix} \frac{\partial g}{\partial a} & \frac{\partial g}{\partial m} \\ \frac{\partial h}{\partial a} & \frac{\partial h}{\partial m} \end{pmatrix} \begin{pmatrix} \Delta a \\ \Delta m \end{pmatrix} = -\begin{pmatrix} g \\ h \end{pmatrix}$$

$$\therefore \begin{pmatrix} \Delta a \\ \Delta m \end{pmatrix} = -\begin{pmatrix} \frac{\partial g}{\partial a} & \frac{\partial g}{\partial m} \\ \frac{\partial h}{\partial a} & \frac{\partial h}{\partial m} \end{pmatrix}^{-1} \begin{pmatrix} g \\ h \end{pmatrix}$$

which on simple calculation reduces to

$$\left.\begin{aligned} \Delta a &= \frac{a_2 h - a_3 g}{a_1 a_3 - a_2^2} \\ \Delta m &= \frac{a_2 g - a_1 h}{a_1 a_3 - a_2^2} \end{aligned}\right\} \tag{3.12}$$

where

$$a_1 = \frac{n}{a^2} - \frac{2m}{a^2}[3a^2 S_3 + S_4]$$

$$a_2 = \frac{2}{a} S_1$$

$$a_3 = n[\psi'(m) - \psi'(m-\tfrac{1}{2})]$$

and

$$S_3 = \sum_1^n \frac{x_i^2}{(a^2+x_i^2)^2}$$

$$S_4 = \sum_1^n \left(\frac{x_i^2}{a^2+x_i^2}\right)^2 .$$

Finally,

$$\hat{a} = a_o + \Delta a$$

$$\hat{m} = m_o + \Delta m.$$

We continue the process till it converges.

The method of moments for type VII distribution leads to a solution for a and m in terms of population μ_2 and β_2 as

$$\left.\begin{aligned} a &= \sqrt{\frac{2\mu_2\beta_2}{\beta_2-3}} \\ m &= \frac{5\beta_2-9}{2(\beta_2-3)} \end{aligned}\right\} \qquad (3.13)$$

(Elderton and Johnson (1969)).

In (3.13), if m is known, then β_2 is known and as a result a is a function of μ_2 only. Similarly, if a is known, then m is a function of μ_2 only. This property is useful in numerical computations of a and m.

The simultaneous method of moment estimators for a and m is

$$\left.\begin{aligned}\tilde{a} &= \sqrt{\frac{2m_2 b_2}{b_2-3}} \\ \tilde{m} &= \frac{5b_2-9}{2(b_2-3)}\end{aligned}\right\} \qquad (3.14)$$

where

$$b_2 = \frac{m_4}{m_2^2}$$

and

m_k = k-th sample (corrected) moment.

3.2 DISTRIBUTION FUNCTION OF TYPE VII DISTRIBUTION

When $y \geq 0$, the distribution function of type VII distribution is

$$F(y) = \int_{-\infty}^{y} f(x|a,m)\,dx$$

$$= \frac{1}{2} + \int_0^y \frac{1}{aB(m-\frac{1}{2},\frac{1}{2})}\left(1+\frac{x^2}{a^2}\right)^{-m} dx.$$

Letting $1 + \frac{x^2}{a^2} = \frac{1}{Z}$, we obtain

$$F(y) = \frac{1}{2} + \frac{1}{2}\int_r^1 \frac{1}{B(m-\frac{1}{2},\frac{1}{2})} Z^{m-\frac{3}{2}}(1-Z)^{-\frac{1}{2}}dZ$$

where $r = \frac{a^2}{a^2+y^2}$,

$$= 1 - \frac{1}{2}\int_0^r \frac{1}{B(m-\frac{1}{2},\frac{1}{2})} Z^{m-\frac{1}{2}-1}(1-Z)^{\frac{1}{2}-1}dZ$$

$$= 1 - \frac{1}{2} I_r(m-\tfrac{1}{2},\tfrac{1}{2}).$$

Similarly, when $y < 0$, the distribution function of type VII distribution is

$$F(y) = \frac{1}{2} - \int_y^0 f(x|a,m)\,dx$$

$$= \frac{1}{2} I_r(m-\tfrac{1}{2},\tfrac{1}{2}).$$

Denoting $F(y)$ by u, since $F \sim U(0,1)$, we have

$$u = \begin{cases} 1 - \frac{1}{2} I_r(m-\tfrac{1}{2},\tfrac{1}{2}), & y \geq 0 \\ \frac{1}{2} I_r(m-\tfrac{1}{2},\tfrac{1}{2}), & y < 0, \end{cases}$$

which implies

$$I_r(m-\tfrac{1}{2},\tfrac{1}{2}) = \begin{cases} 2-2u, & y \geq 0 \\ 2u, & y < 0. \end{cases}$$

Thus, it follows that if

$$0 < u < \frac{1}{2},$$

we have

$$I_r(m-\tfrac{1}{2},\tfrac{1}{2}) = 2u \quad \text{and} \quad y < 0;$$

and if

$$\frac{1}{2} < u < 1,$$

we have

$$I_r\left(m-\tfrac{1}{2},\tfrac{1}{2}\right) = 2(1-u) \quad \text{and} \quad y \geq 0.$$

Using the above result, we generated a sample of size n = 100 from type VII distribution with a = 4.0 and m = 5.5. The observations are:

(3.15)

-1.59579	2.05225	0.11791	1.50780	1.52414	-0.67271
2.36340	-1.81101	-0.93933	-0.10608	1.11285	-2.46865
-0.51288	1.07251	-1.08069	-1.89978	-0.56204	0.47940
-0.66230	0.27973	-0.57666	-0.36237	-0.96488	1.88778
-0.10558	-0.33427	-1.60879	-0.18383	1.58983	-2.54929
-1.80244	0.10793	0.69716	0.57065	-0.71167	0.17098
-2.01476	0.55922	1.38929	1.50133	-1.71486	-0.35109
1.72985	1.37682	2.22678	-0.02336	0.23044	1.44962
0.79462	0.02846	4.49561	-0.79846	0.27090	0.43497
-0.63640	-0.17290	0.47428	2.33221	0.50090	0.91021
-2.47112	1.32663	-1.43906	0.21186	1.80208	0.54530
1.91001	-0.92331	-0.40029	1.03944	-1.60243	-1.37178
0.28463	0.24700	1.49400	0.55229	0.54978	-0.32458
-0.67128	-1.41091	0.85193	0.43047	0.18907	2.51570
2.37930	0.42337	0.94249	-0.05505	-0.20891	0.93837
-2.25304	-1.98386	-0.62076	-0.32529	0.19132	-1.71588
0.80764	1.83338	-0.28171	2.09435		

3.3 DERIVATION OF JIP FOR a AND m

Jeffreys' invariant prior (JIP) for a and m is given by

$$g(a,m) \propto |I(a,m)|^{\frac{1}{2}} \tag{3.16}$$

where

$$I(a,m) = \begin{pmatrix} -E\left(\dfrac{\partial^2 \log f}{\partial a^2}\right) & -E\left(\dfrac{\partial^2 \log f}{\partial a \partial m}\right) \\ -E\left(\dfrac{\partial^2 \log f}{\partial m \partial a}\right) & -E\left(\dfrac{\partial^2 \log f}{\partial m^2}\right) \end{pmatrix}$$

is the Fisher's information matrix based on a single observation. For the density (3.1),

$$\frac{\partial^2 \log f}{\partial a^2} = \frac{1}{a^2} - \frac{2mx^2(3a^2+x^2)}{a^2(a^2+x^2)^2}$$

$$\frac{\partial^2 \log f}{\partial m \partial a} = \frac{\partial^2 \log f}{\partial a \partial m} = \frac{2x^2}{a(a^2+x^2)}$$

$$\frac{\partial^2 \log f}{\partial m^2} = \frac{\partial^2}{\partial m^2}\left[\log\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}\right] = \psi'(m) - \psi'(m-\tfrac{1}{2})$$

where $\psi(p)$ is given in (3.6) and $\psi'(p)$ is the tri-gamma function can be evaluated by using the asymptotic formula

$$\psi'(p) = \frac{1}{p} + \frac{1}{2p^2} + \frac{1}{6p^3} - \frac{1}{30p^5} + \frac{1}{42p^7} - \frac{1}{30p^9} + \ldots \tag{3.17}$$

(Abramowitz and Stegum, 1964).

After some algebra, it can be shown that

$$E\left(\frac{\partial^2 \log f}{\partial a^2}\right) = -\frac{2m-1}{a^2(m+1)}$$

$$E\left(\frac{\partial^2 \log f}{\partial m \partial a}\right) = \frac{1}{am}$$

$$E\left(\frac{\partial^2 \log f}{\partial m^2}\right) = \psi'(m) - \psi'(m-\tfrac{1}{2}).$$

Thus,

$$|I(a,m)| = \begin{vmatrix} \dfrac{2m-1}{a^2(m+1)} & -\dfrac{1}{am} \\ & \\ -\dfrac{1}{am} & -\{\psi'(m) - \psi'(m-\frac{1}{2})\} \end{vmatrix}.$$

Using (3.17),

$$\psi'(m) - \psi'(m-\tfrac{1}{2})$$

$$= \{ \frac{1}{m} + \frac{1}{2m^2} + \frac{1}{6m^3} - \frac{1}{30m^5} + \ldots\} - \{(m-\tfrac{1}{2})^{-1}$$

$$+ \frac{1}{2} (m-\tfrac{1}{2})^{-2} + \frac{1}{6} (m-\tfrac{1}{2})^{-3} - \ldots\}$$

$$= \{ \frac{1}{m} + \frac{1}{2m^2} + \frac{1}{6m^3} - \frac{1}{30m^5} + \ldots\} - \{\frac{1}{m} (1+ \frac{1}{2m}$$

$$+ \frac{1}{4m^2} + \ldots) + \frac{1}{2m^2} (1+ \frac{2}{2m} + \frac{3}{4m^2} + \frac{4}{8m^3} + \ldots)$$

$$+ \frac{1}{6m^3} (1 + \frac{3}{2m} + \frac{6}{4m^2} + \frac{10}{8m^3} + \ldots) + \ldots\}$$

$$= \{\frac{1}{m} + \frac{1}{2m^2} + \frac{1}{6m^3} - \frac{1}{39m^5} + \ldots\} - \{\frac{1}{m} + \frac{1}{m^2} + o(\frac{1}{m^3})\}$$

$$\approx - \frac{1}{m^2} + o(\frac{1}{m^3}) .$$

Finally,

$$|I(a,m)| \doteq \frac{2m - 1}{a^2 (m+1) m^2} - \frac{1}{a^2 m^2} = \frac{1}{a^2 m^2} (\frac{m-2}{m+1}), \qquad (3.18)$$

giving

$$g(a,m) \propto \frac{1}{am} \text{ for large m.} \qquad (3.19)$$

That is, apriori a and m are independently distributed, and that the marginal JIP for a may be taken to be

$$g_1(a) \propto \frac{1}{a}$$

and that for m may be taken to be

$$g_2(m) \propto \frac{1}{m} .$$

3.4 BAYES ESTIMATORS OF a AND m

When m is known the likelihood (3.2) is

$$\ell(x|a,m) \propto \frac{1}{a^n} \prod_{i=1}^{n} \left(1 + \frac{x_i^2}{a^2}\right)^{-m} . \qquad (3.20)$$

Let us consider the family of NIP for a with density

$$g_1(a) \propto \frac{1}{a^t} , \quad t > 0 . \qquad (3.21)$$

Combining (3.20) and (3.21), the posterior density of a is

$$\pi_1(a|\underset{\sim}{x}) = k\, a^{-(n+t)} \prod_{i=1}^{n} \left(1 + \frac{x_i^2}{a^2}\right)^{-m} , \quad a > 0 \qquad (3.22)$$

where k is the constant of proportionality given by

$$k^{-1} = \int_0^{\infty} a^{-(n+t)} \prod_{i=1}^{n} \left(1 + \frac{x_i^2}{a^2}\right)^{-m} da .$$

Using the generated data (3.15), we plotted the density (3.22) for t = 1 (1) 5 in Figure 1. We note that as t deviates more and more from JIP (t = 1), the posteriors appear less and less robust.

Bayes estimator of a under squared error loss function is given by

$$a^* = E(a|\underset{\sim}{x}) = \frac{\int_0^{\infty} a^{-(n+t-1)} \prod_{1}^{n} \left(1 + \frac{x_i^2}{a^2}\right)^{-m} da}{\int_0^{\infty} a^{-(n-t)} \prod_{1}^{n} \left(1 + \frac{x_i^2}{a^2}\right)^{-m} da} . \qquad (3.23)$$

When a is known, the likelihood (3.2) is

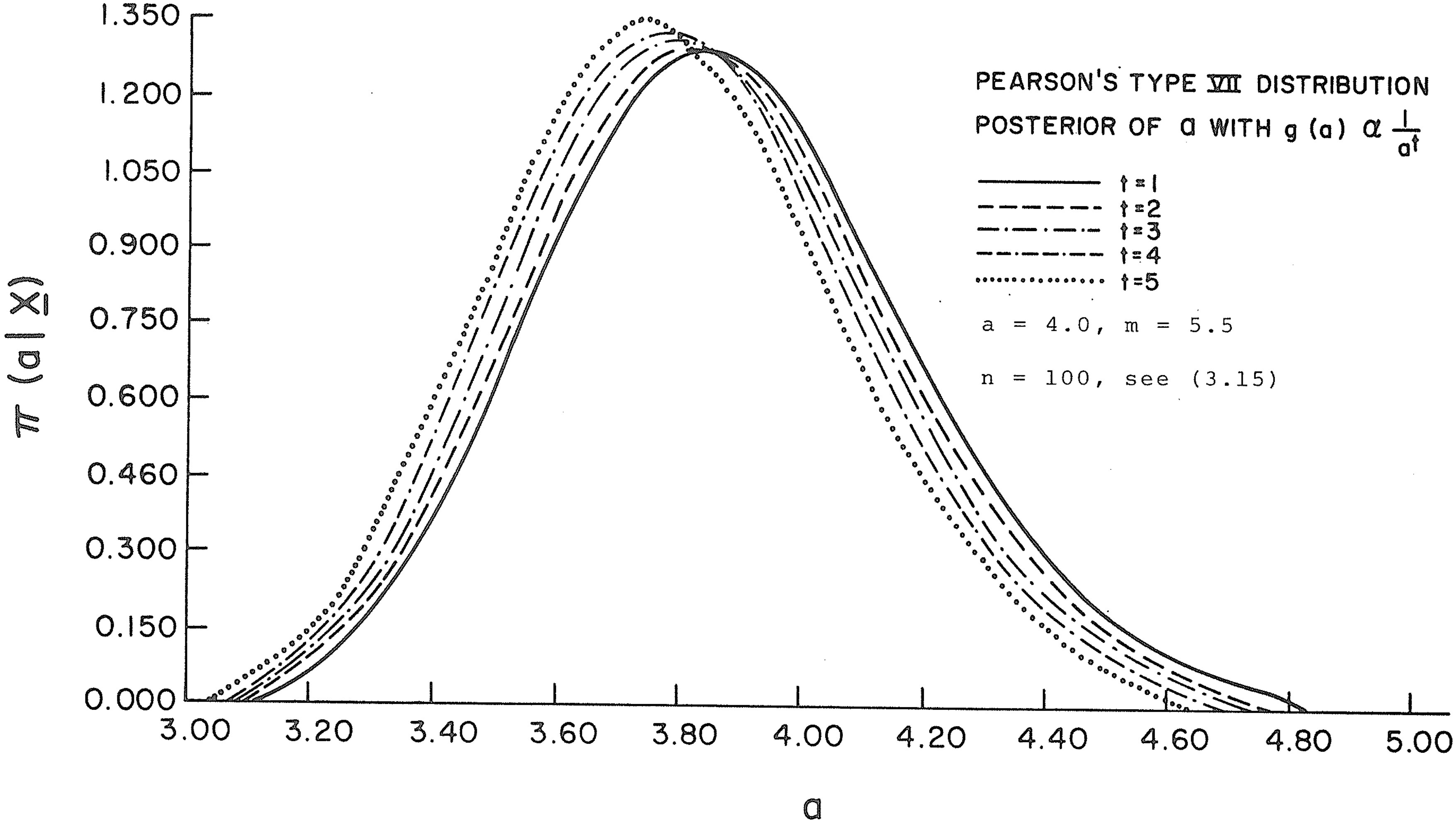

PEARSON'S TYPE VII DISTRIBUTION
POSTERIOR OF a WITH g (a) α 1/a^t
t=1
t=2
t=3
t=4
t=5
a = 4.0, m = 5.5
n = 100, see (3.15)
1.350
1.200
1.050
0.900
0.750
0.600
0.460
0.300
0.150
0.000
π (a | X)
3.00
3.20
3.40
3.60
3.80
4.00
4.20
4.40
4.60
4.80
5.00
a

Fig. 1

$$\ell(\underset{\sim}{x}|a,m) \propto \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m}. \tag{3.24}$$

Consider the family of NIP for m with density

$$g_2(m) \propto \frac{1}{m^t}, \; t > 0. \tag{3.25}$$

Combining (3.24) and (3.25), the posterior density of m is

$$\pi_2(m|x) = \frac{m^{-t}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m}}{\int_{\frac{1}{2}}^{\infty} m^{-t}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m} dm}, \; m > \tfrac{1}{2}. \tag{3.26}$$

Using generated data (3.15), we plotted (3.26) for $t = 1\ (1)\ 5$ in Figure 2. We observe the same pattern in the robustness of the posteriors.

The Bayes estimator of m under squared error loss is

$$m^* = E(m|\underset{\sim}{x}) = \frac{\int_{\frac{1}{2}}^{\infty} m^{-(t-1)}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m} dm}{\int_{\frac{1}{2}}^{\infty} m^{-t}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m} dm}. \tag{3.27}$$

When both a and m are unknown, consider Jeffreys' invariant prior

$$g(a,m) \propto \frac{1}{am}. \tag{3.28}$$

Combining the likelihood (3.2) and the prior (3.28), the joint posterior density of a and m is

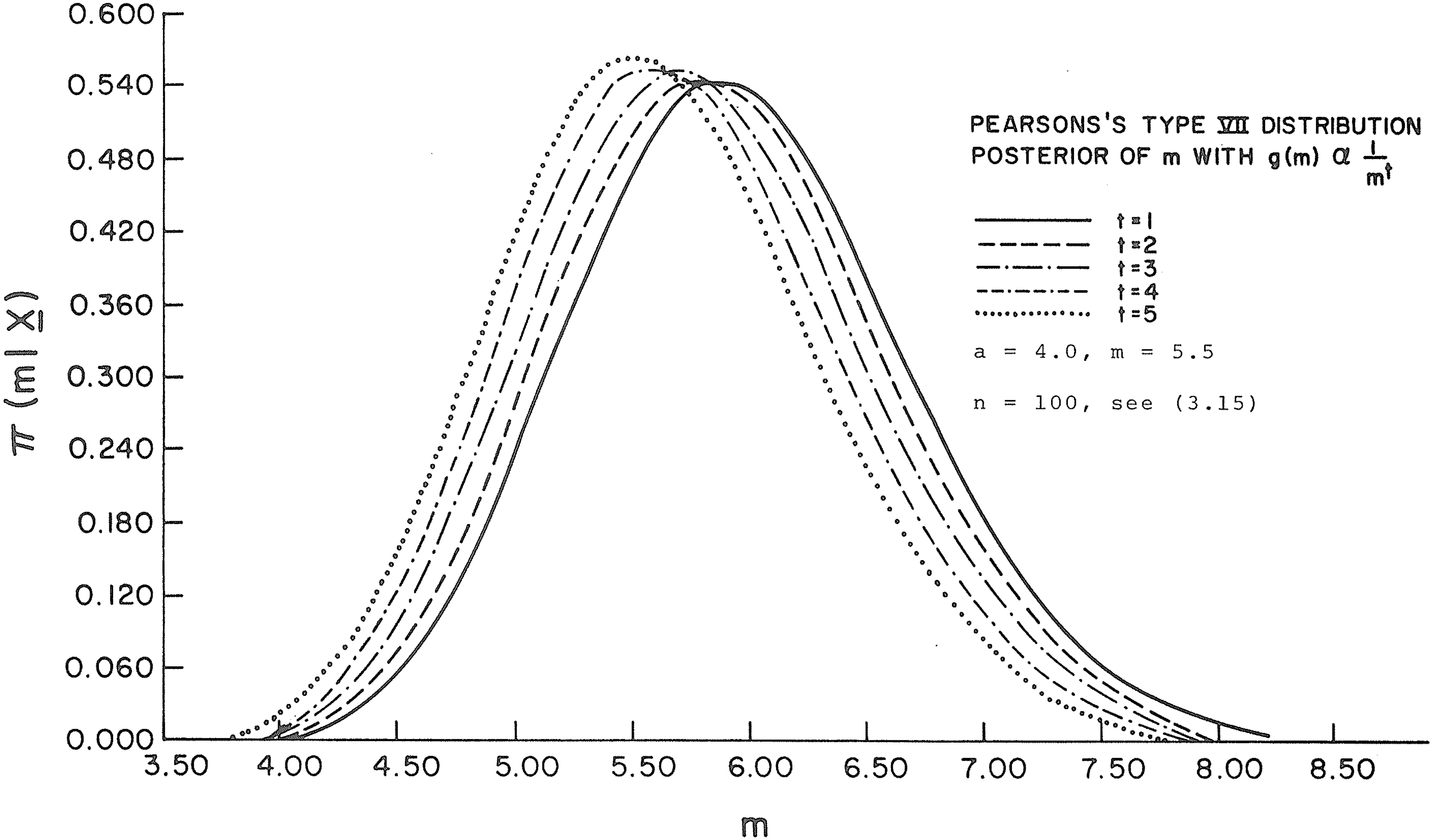

PEARSONS'S TYPE VII DISTRIBUTION
POSTERIOR OF m WITH g(m) α 1/m^t
t=1
t=2
t=3
t=4
t=5
a = 4.0, m = 5.5
n = 100, see (3.15)
π (m | X)
0.600
0.540
0.480
0.420
0.360
0.300
0.240
0.180
0.120
0.060
0.000
3.50
4.00
4.50
5.00
5.50
6.00
6.50
7.00
7.50
8.00
8.50
m

Fig. 2

$$\pi(a,m|\underset{\sim}{x}) \propto \frac{1}{m}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n a^{-(n+1)} \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m},$$

$$m > \tfrac{1}{2},\ a > 0. \qquad (3.29)$$

Integrating out a from (3.29), we obtain the marginal posterior density of m as

$$\pi(m|\underset{\sim}{x}) = \frac{\int_0^\infty \frac{1}{m}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n a^{-(n+1)} \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m} da}{\int_{\frac{1}{2}}^\infty \int_0^\infty \frac{1}{m}\left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n a^{-(n+1)} \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m} dadm}. \qquad (3.30)$$

The Bayes estimator of m under squared error loss function is

$$m^* = E(m|\underset{\sim}{x}) = \frac{\int_{\frac{1}{2}}^\infty \int_0^\infty \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n a^{-(n+1)} \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m} dadm}{\int_{\frac{1}{2}}^\infty \int_0^\infty \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n a^{-(n+1)} m^{-1} \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m} dadm}. \qquad (3.31)$$

Again, integrating out m from (3.29) and after obtaining the constant of proportionality, the marginal posterior density of a is

$$\pi_4(a|\underset{\sim}{x}) = \frac{\int_{\frac{1}{2}}^\infty a^{-(n+1)} m^{-1} \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m} dm}{\int_0^\infty \int_{\frac{1}{2}}^\infty a^{-(n+1)} m^{-1} \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1+\frac{x_i^2}{a^2}\right)^{-m} dmda},$$

$$a > 0. \qquad (3.32)$$

And the Bayes estimator under squared error loss function is

$$a^* = E(a|\underset{\sim}{x}) = \frac{\int_0^\infty \int_{\frac{1}{2}}^\infty a^{-n} m^{-1} \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m} dmda}{\int_0^\infty \int_{\frac{1}{2}}^\infty a^{-(n+1)} m^{-1} \left\{\frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}\right\}^n \prod_1^n \left(1 + \frac{x_i^2}{a^2}\right)^{-m} dmda} . \tag{3.33}$$

The integrals in (3.23), (3.27), (3.31), and (3.33) appear to be intractable. We will approximate them by using Lindley's method as discussed in Appendix A.1.

We denote the logarithm of the joint posterior density, except the normalizing constant, by $\Lambda(a,m)$, giving

$$\Lambda(a,m) = L(a,m) + \rho(a,m)$$

where

$L(a,m)$ = logarithm of the likelihood (3.2)

$\rho(a,m)$ = logarithm of the prior (3.28).

Then, the posterior expectation of m in (3.31) is approximately

$$m^* = E(m|\underset{\sim}{x}) \sim \tilde{\tilde{m}} + \frac{1}{2}\Lambda_{30}\tau_{20}^2 + \frac{3}{2}\Lambda_{21}\tau_{20}\tau_{12} + \frac{1}{2}\Lambda_{12}(\tau_{20}\tau_{02}+2\tau_{12}^2) + \frac{1}{2}\Lambda_{03}\tau_{02}\tau_{12}. \tag{3.34}$$

Similarly,

$$a^* = E(a|\underset{\sim}{x}) \sim \tilde{\tilde{a}} + \frac{1}{2}\Lambda_{03}\tau_{02}^2 + \frac{3}{2}\Lambda_{12}\tau_{02}\tau_{21} + \frac{1}{2}\Lambda_{21}(\tau_{02}\tau_{20}+2\tau_{21}^2) + \frac{1}{2}\Lambda_{30}\tau_{20}\tau_{21} \tag{3.35}$$

where τ's are given by the matrix

$$-\begin{pmatrix} \dfrac{\partial^2 \Lambda}{\partial \theta_1^2} & \dfrac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} \\ \\ \dfrac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_1} & \dfrac{\partial^2 \Lambda}{\partial \theta_2^2} \end{pmatrix}^{-1} = \begin{pmatrix} \tau_{20} & \tau_{12} \\ \tau_{21} & \tau_{02} \end{pmatrix}.$$

All the quantities are to be evaluated at the joint posterior mode of (a,m).

In order to apply Lindley's method we need to locate the joint posterior mode of (3.29). Working in the same line as for the joint mle of a and m, the iterative equations are:

$$\left.\begin{aligned} \tilde{\tilde{a}} &= a_o + \Delta a \\ \tilde{\tilde{m}} &= m_o + \Delta m \end{aligned}\right\} \tag{3.36}$$

where Δa, Δm are given by equation (3.12) with

$$h = -\frac{1}{m} + n[\psi(m) - \psi(m-\tfrac{1}{2})] - S_2, \qquad g = -\frac{n+1}{a} + \frac{2mS_1}{a}$$

$$a_1 = \frac{n+1}{a^2} - \frac{2m}{a^2}(3a^2 S_3 + S_4)$$

$$a_2 = \frac{2S_1}{a}, \qquad a_3 = n[\psi'(m) - \psi'(m-\tfrac{1}{2})] + \frac{1}{m^2}$$

where S_1, S_2, S_3 and S_4 are defined before.

The joint posterior mode of (3.29) using the iterative equations (3.36) and the generated data (3.15) is

$$\tilde{\tilde{m}} = 4.712$$

$$\tilde{\tilde{a}} = 3.474.$$

We now evaluate

$$\Lambda(a,m) = 100\{\log_e\Gamma(m) - \log_e\Gamma(m-\tfrac{1}{2})\} - \log_e m$$

$$- 101 \log_e a - m \sum_{1}^{100} \log_e\left(1 + \frac{x_i^2}{a^2}\right)$$

for different values of a and m around their posterior modes. The values of $\Lambda(a,m)$ are shown in Table 3.1.

Tables 3.2 to 3.5 give the values of the derivatives $\Lambda_{ij}(a,m)$.

TABLE 3.1

Values of the Logarithm of Posterior Density, $\Lambda(a,m)$, Around Posterior Mode

m \ a	3.274	3.374	3.474	3.574	3.674
4.512	-173.934022708	-173.860473835	-173.921819976	-174.102568849	-174.389081209
4.612	-174.018549712	-173.876022839	-173.873347897	-173.994577577	-174.225667919
4.712	-174.135732836	-173.924227963	-173.857531938	-173.919242423	-174.094910748
4.812	-174.283958958	-174.003476086	-173.872758977	-173.874950268	-173.995196576
4.912	-174.461731186	-174.112270313	-173.917532122	-173.860204218	-173.925028510

TABLE 3.2

Values of Λ_{20}

m \ a	3.274	3.374	3.474	3.574	3.674
4.512					
4.612	-3.2656120	-3.2656120	-3.2656120	-3.2656120	-3.2656120
4.712	-3.1042998	-3.1042998	-3.1042998	-3.1042998	-3.1042998
4.812	-2.9546106	-2.9546106	-2.9546106	-2.9546106	-2.9546106
4.912					

giving $\Lambda_{20}(\tilde{\tilde{a}},\tilde{\tilde{m}}) = -3.1042998$.

TABLE 3.3

Values of Λ_{30}

m \ a	3.274	3.374	3.474	3.574	3.674
4.512					
4.612					
	1.613122	1.613122	1.613122	1.613122	1.613122
4.712					
	1.496892	1.496892	1.496892	1.496892	1.496892
4.812					
4.912					

giving $\Lambda_{30}(\tilde{\tilde{a}},\tilde{\tilde{m}}) = 1.555007$.

TABLE 3.4

Values of Λ_{02}

m \ a	3.274	3.374	3.474	3.574	3.674
4.512		-13.4895014	-11.9402732	-10.5763487	
4.612		-13.9851931	-12.3904622	-10.9860662	
4.712		-14.4808848	-12.8406510	-11.3957840	
4.812		-14.9765763	-13.2908400	-11.8055017	
4.912		-15.4722682	-13.7410287	-12.2152196	

giving $\Lambda_{02}(a,m) = -12.8406510$.

TABLE 3.5

Values of Λ_{03}

m \ a	3.274	3.374	3.474	3.574	3.674
4.512			15.492282	13.639245	
4.612			15.947309	14.043960	
4.712			16.402338	14.448670	
4.812			16.857363	14.853383	
4.912			17.312395	15.258091	

giving $\Lambda_{03}(\tilde{\tilde{a}},\tilde{\tilde{m}}) = 15.425504$.

Due to local orthogonality of the parameters (a,m), τ_{ij} in equations (3.34) and (3.35) are zero for $i \neq j$. Also, $\Lambda_{21}(\tilde{\tilde{a}},\tilde{\tilde{m}})$ is zero, as a result from (3.34) and (3.35), we obtain

$$E(m|\underset{\sim}{x}) = m^* \sim \tilde{\tilde{m}} + \frac{1}{2}\Lambda_{30}\tau_{20}^2$$

and

$$E(a|\underset{\sim}{x}) = a^* \sim \tilde{\tilde{a}} + \frac{1}{2}\Lambda_{03}\tau_{02}^2$$

where

$$\tau_{20} = (-\Lambda_{20})^{-1} = 0.32213384$$

$$\tau_{02} = (-\Lambda_{02})^{-1} = 0.07787767.$$

Finally,

$$m^* = 4.7924255,$$

$$a^* = 3.5207773.$$

3.5 MONTE CARLO STUDY

We generated 620 samples each of size 100 from a type VII distribution with $a = 4.0$ and $m = 5.5$ to compare the performance of the various estimators. Since Cauchy is a particular case of type VII distribution, we expected that we could be generating a number of samples which would not satisfy the condition $b_2 > 3$. Finally, we were left with 400 samples to work with. The mean and mse of 400 estimators under different methods of estimation are tabulated. The entries in the parentheses represent the mse's.

Comparing Table 3.6, we observe that MLE has minimum MSE for the estimation of a when m is known as well as for the estimation of m when a is known than the corresponding Bayes estimator under JIP ($t = 1$). But for a more realistic situation when a and m are both unknown, we observe in Table 3.7 that, if mse is accepted as an index of precision, Bayesian method has the minimum mse compared to the moment and maximum likelihood methods and hence should be preferred.

3.6 SAMPLING DISTRIBUTIONS OF a* AND m*

The sampling distributions of a^* and m^* under squared error loss function and with the prior

$$g(a,m) \propto \frac{1}{am} ,$$

were obtained ($N = 400$, $n = 100$) and are represented in Figures 3 and 4, respectively. The moments, β_1 and β_2,

TABLE 3.6

Estimates of a and m under Different Methods of Estimation (Squared Error Loss Function for the Bayesian Method), N = 400, n = 100 (a = 4, m = 5.5)

Methods / Parameters	MM	MLE	BAYES				
			t = 1	t = 2	t = 3	t = 4	t = 5
a (m known)	3.9418836 (0.0981229)	3.9443444 (0.0959597)	4.0351195 (0.1141221)	4.0075800 (0.1077849)	3.8900281 (0.1030893)	3.9524824 (0.1000379)	3.9249368 (0.0986297)
m (a known)	5.6953786 (0.4811806)	5.6690282 (0.4685353)	5.7799368 (0.5364719)	5.6946073 (0.4783211)	5.6078668 (0.4353107)	5.5218318 (0.4074406)	5.4357968 (0.3947107)

TABLE 4.7

Estimates when Both a and m Are Unkown (Squared Error Loss Function for the Bayesian Method) N = 400, n = 100

Methods / Parameters	MM	MLE	BAYES
a	5.7882261 (17.9663613)	4.6106310 (8.1707970)	2.6456918 (2.0590924)
m	14.0877783 (917.0662471)	9.0598044 (219.7514231)	3.1711571 (5.8005475)

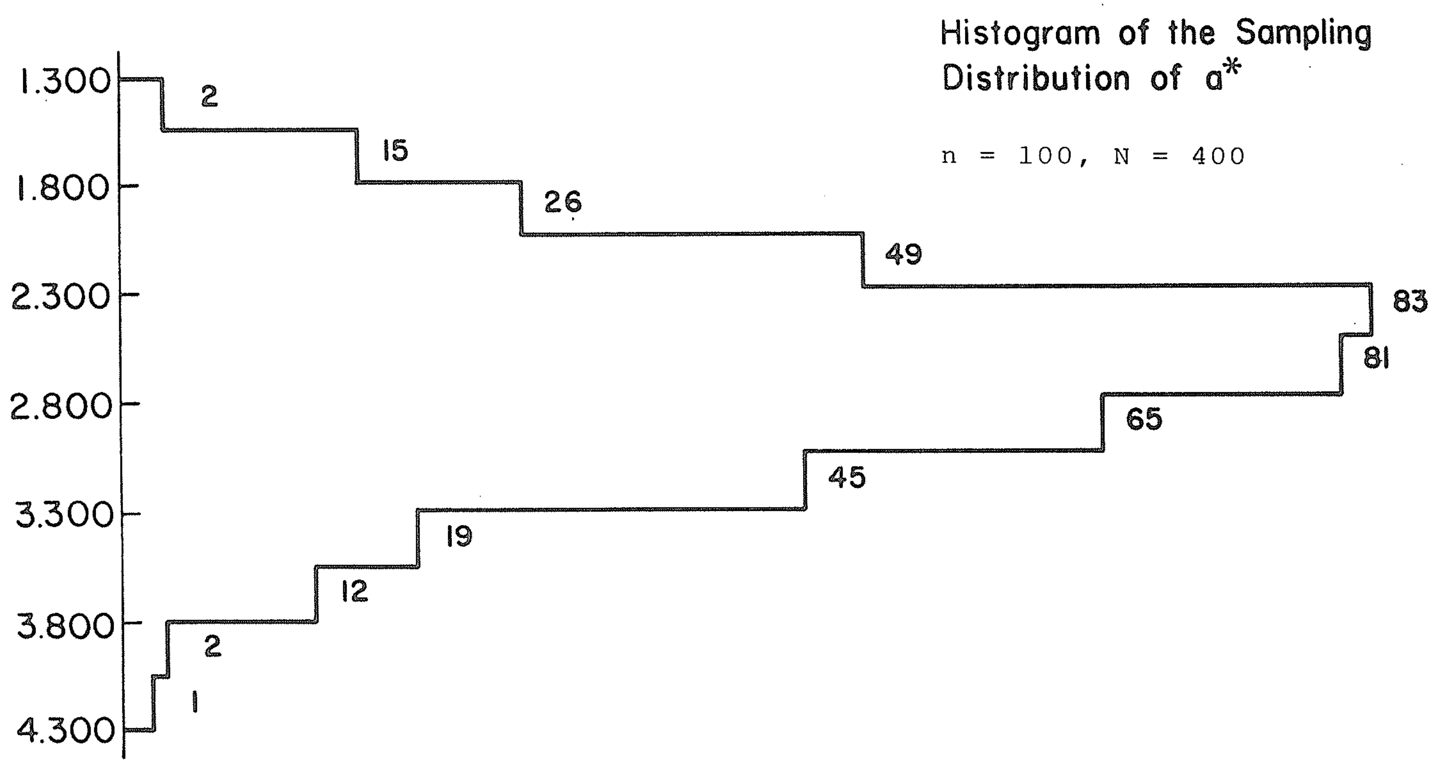

Histogram of the Sampling
Distribution of a*
n = 100, N = 400
1.300
1.800
2.300
2.800
3.300
3.800
4.300
2
15
26
49
83
81
65
45
19
12
2
1

Fig. 3

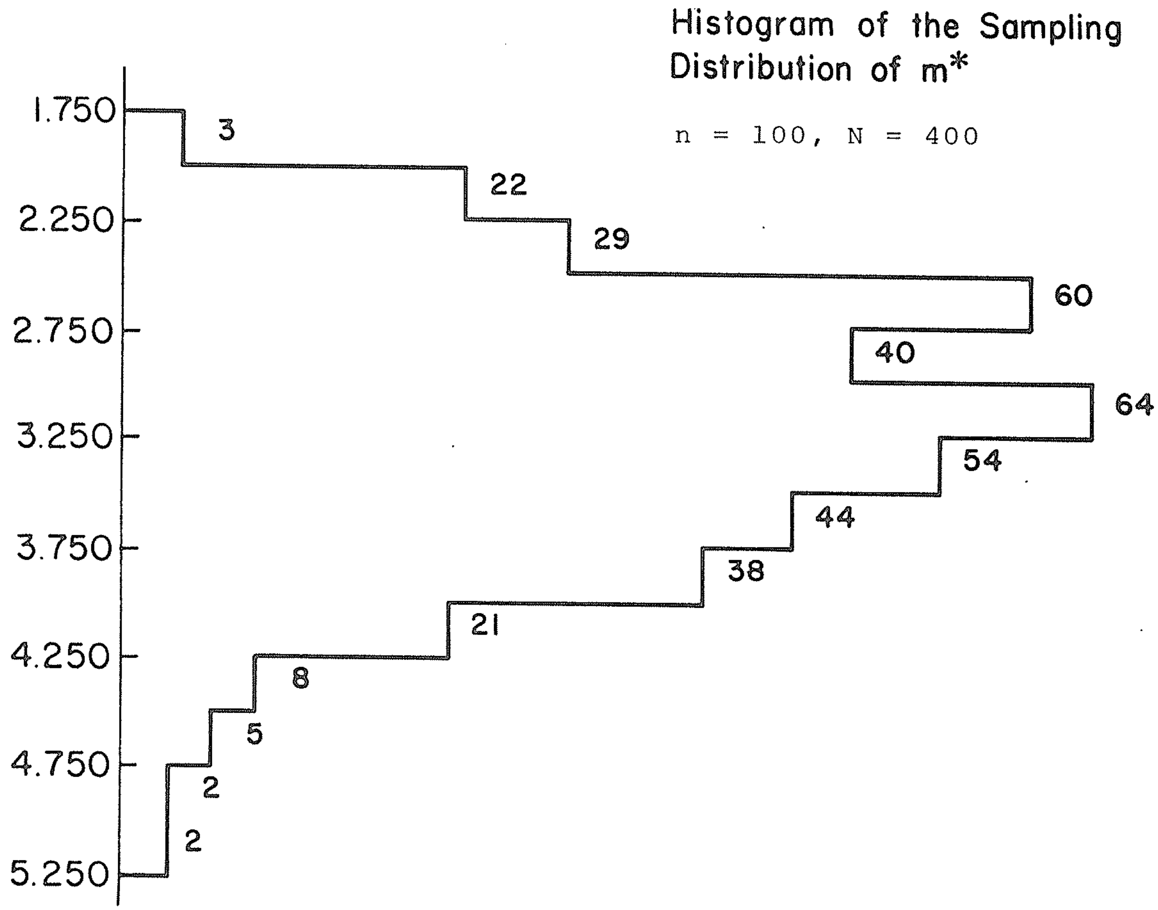

Histogram of the Sampling
Distribution of m*
n = 100, N = 400
1.750
2.250
2.750
3.250
3.750
4.250
4.750
5.250
3
22
29
60
40
64
54
44
38
21
8
5
2
2

Fig. 4

were computed using Sheppard's corrections for grouping for each of the distributions.

The sampling distribution of a* gives

$\mu_2 = 0.2276007$, $\mu_3 = 0.0147733$,

$\beta_1 = 0.0172964$, $\beta_2 = 2.8871637$.

The point (β_1, β_2) was plotted on the chart relating to the Pearson's system of curves (Pearson and Hartley, 1966, Vol. 1) which suggests that Pearson's type I curve

$$f(x) = y_o\left(1 + \frac{x}{a_1}\right)^{m_1}\left(1 - \frac{x}{a_2}\right)^{m_2}, \quad -a_1 < x < a_2$$

should fit the sampling distribution of a*, where

$$y_o = \frac{N}{a_1+a_2} \frac{m_1^{m_1} m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}} \frac{1}{B(m_1+1, m_2+1)}$$

(Elderton and Johnson (1969)), with

$$\frac{m_1}{a_1} = \frac{m_2}{a_2}$$

$$r = 6(\beta_2 - \beta_1 - 1)/(6 + 3\beta_1 - 2\beta_2)$$

$$a_1 + a_2 = \frac{1}{2}\sqrt{\mu_2}\sqrt{\{\beta_1(r+2)^2+16(r+1)\}}$$

m_1 and m_2 are given by

$$\frac{1}{2}\left\{r - 2 \pm r(r+2)\sqrt{\frac{\beta_1}{\beta_1(r+2)^2 + 16(r+1)}}\right\}.$$

If μ_3 is positive, m_2 is the positive root.

The origin of the curve is at the mode given by

$$\text{mode} = \text{mean} - \frac{1}{2}\,\frac{\mu_3}{\mu_2}\,\frac{r+2}{r-2} .$$

Thus

$$x = X - \text{mode}.$$

The constants of the curve are found to be

$r = 40.4205677,$ $\qquad m_1 = 14.9298006$

$m_2 = 23.4907671,$ $\qquad a_1 = 2.4416367$

$a_2 = 3.8417070$

$y_o = 329.0683086,$ $\qquad \text{Mode} = 2.60604197.$

Thus,

$$f(x) = 329.0683086\left(1 + \frac{x}{2.4416367}\right)^{14.9298006} \cdot \left(1 - \frac{x}{3.8417070}\right)^{23.4907671},$$

$$-2.4416367 < x < 3.8417070.$$

3.7 TESTS OF GOODNESS OF FIT

First, we apply the χ^2-goodness of fit. The observed and expected frequencies are shown in the table below. The expected frequencies were obtained by using Simpson's $\frac{1}{3}$ -rule; for details refer to Chapter Two.

TABLE 3.8

Class Boundaries	Observed Frequencies		Expected Frequencies	
1.30 = 1.55	2	=17	2.88	=13.7
1.55 - 1.80	15		10.82	
1.80 - 2.05	26		29.18	
2.05 - 2.30	49		54.86	
2.30 - 2.55	83		76.08	
2.55 - 2.80	81		80.65	
2.80 - 3.05	65		66.76	
3.05 - 3.30	45		43.58	
3.30 - 3.55	19		22.50	
3.55 - 3.80	12		9.08	
3.80 - 4.05	2	=15	2.83	=12.69
4.05 - 4.30	1		0.78	
	400		400	

We obtain $\chi^2 = 3.46$ with 5 degrees of freedom

$$\chi^2_{.05,5} = 11.07$$

which suggests a good fit.

There are some objections to the χ^2 goodness of fit tests when parameters are estimated by the method of moments.

We apply Kolmogorov-Smirnov goodness of fit and Cramér-Von Mises goodness of fit tests to the sampling distribution to see if it follows a type I curve. For this purpose the distribution function of a type I curve is

$$F_o(x) = \int_{-a_1}^{x} y_o\left(1 + \frac{x}{a_1}\right)^{m_1}\left(1 - \frac{x}{a_2}\right)^{m_2} dx,$$

$$y_o = \frac{1}{a_1+a_2} \frac{m_1^{m_1} m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}} \frac{1}{B(m_1+1, m_2+1)}.$$

Let

$$Z = \frac{a_1+x}{a_1+a_2}, \qquad dx = (a_1+a_2)\,dZ$$

$$= \int_0^{\frac{a_1+x}{a_1+a_2}} y_o\left(\frac{a_1+a_2}{a_1}\right)^{m_1} Z^{m_1}\left(\frac{a_1+a_2}{a_2}\right)^{m_2} (1-Z)^{m_2}(a_1+a_2)\,dZ$$

$$= \int_0^{\frac{a_1+x}{a_1+a_2}} \frac{m_1^{m_1} m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}} \frac{(a_1+a_2)^{m_1+m_2}}{a_1^{m_1} a_2^{m_2}} \frac{1}{B(m_1+1, m_2+1)}$$

$$\cdot Z^{m_1}(1-Z)^{m_2}\,dZ$$

$$= \left(\frac{m_1}{a_1}\right)^{m_1} \left(\frac{m_2}{a_2}\right)^{m_2} \left(\frac{a_1+a_2}{m_1+m_2}\right)^{m_1+m_2} I_{\frac{a_1+x}{a_1+a_2}} \left(m_1+1, m_2+1\right)$$

$$= I_{\frac{a_1+x}{a_1+a_2}} \left(m_1+1, m_2+1\right),$$

since

$$\frac{a_1+x}{a_1+a_2} < 1$$

and

$$\left(\frac{m_1}{a_1}\right)^{m_1} \left(\frac{m_2}{a_2}\right)^{m_2} \left(\frac{a_1+a_2}{m_1+m_2}\right)^{m_1+m_2} = 1.$$

To evaluate the Incomplete beta function $I_R(p,q)$, we will use the approximation

$$I_R(p,q) \sim \Phi(Z_i)$$

due to Peizer and Pratt (1966) and quoted in Pearson and Johnson (1968), where

$$Z_i = d_i \left\{ \frac{1 + R\ g\left(\frac{q-0.5}{n(1-R)}\right) + (1-R)\ g\left(\frac{p-0.5}{nR}\right)}{\left(n + \frac{1}{6}\right) R(1-R)} \right\}^{\frac{1}{2}}$$

$$R = \frac{a_1+x}{a_1+a_2}$$

$$n = p + q - 1,$$

and the function $g(u)$ is defined as

$$g(u) = (1-u)^{-2}(1 - u^2 + 2u \log_e u).$$

$i = 1$: $d_1 = q - 0.5 + \frac{1}{6} - (n + \frac{1}{3})(1-R)$ as a first approximation

$i = 2$: $d_2 = d_1 + \frac{1}{50}\{\frac{R}{q} - \frac{1-R}{p} + \frac{R-0.5}{p+q}\}$ as a second approximation.

In the sampling distribution of a*,

$p = 15.9298006$, $q = 24.4907671$

$n = 39.4205677$

$d_1 = 24.15743377 - 39.75390108(1-R)$

The test statistic

$$D_n = \sup_x \left|S_n(x) - F_o(x)\right|$$

is found to be 0.0229994.

For large n, the critical region is

$$P(D_{n,\alpha} > Z_\alpha | \sqrt{n}) = \alpha.$$

For $\alpha = .05$, $Z_\alpha = 1.36$, giving

$$P(D_{n,.05} > 0.06082) = 0.05.$$

Thus, we fail to reject the null hypothesis that the sampling distribution of a* follows a type I curve.

We now apply the Cramér-Von Mises test of goodness of fit, the test statistic of which is

$$n\omega_n^2 = \frac{1}{12n} + \sum_1^n [F_o(x_{(i)}) - \frac{2i-1}{2n}]^2.$$

TABLE 3.9

Computation of Empirical Distribution Function, $S_n(x)$ and Hypothesized Distribution Function, $F_o(x)$

X	x = X-Mode	R	$S_n(x)$	$F_o(x)$	$\lvert S_n(x)-F_o(x)\rvert$
1.45484	-1.15120	0.20537	0.0025	0.00352	0.00102
1.51299	-1.09305	0.21463	0.0050	0.00554	0.00054
1.60494	-1.00111	0.22926	0.0075	0.01068	0.00318
1.64444	-0.96160	0.23555	0.0100	0.01384	0.00384
1.68995	-0.91609	0.24279	0.0125	0.01839	0.00589
⋮	⋮	⋮	⋮	⋮	⋮
3.72604	1.12000	0.56684	0.9925	0.98656	0.00594
3.81177	1.20573	0.58048	0.9950	0.99155	0.00345
3.84481	1.23877	0.58574	0.9975	0.99299	0.00451
4.11223	1.50619	0.62830	1.0000	0.99872	0.00128

Details are shown in Appendix A.2.

The value of the statistic $n\omega_n^2$ is found to be 0.03223.

The 5% critical value of the statistic $n\omega_n^2$ is 0.461 (Conover, 1971), showing the evidence again that the true distribution function of a* is a type I curve.

The sampling distribution of m* gives

$\mu_2 = 0.3812276$, $\mu_3 = 0.0760759$,

$\beta_1 = 0.1044578$, $\beta_2 = 2.7619320$.

The joint (β_1, β_2) again suggests that Pearson's type I curve should fit the sampling distribution.

The constants of the type I curve, for the sampling distribution of m*, are

$r = 12.5962343$ $m_1 = 3.3794380$

$m_2 = 7.2167963$ $a_1 = 1.5246660$

$a_2 = 3.2559270$ $y_o = 248.0336847$

Mode = 3.0388077

giving

$$f(x) = 248.0336847\left(1 + \frac{x}{1.5246660}\right)^{3.3794380}$$

$$\cdot\left(1 - \frac{x}{3.2559270}\right)^{7.2167963},$$

$$-1.5246660 < x < 3.2559270.$$

The observed and expected frequences are shown in Table 3.10.

TABLE 3.10

Observed and Expected Frequencies of the

Sampling Distribution of m*

Class Boundaries	Observed Frequencies		Expected Frequencies	
1.75 - 2.00	3	=25	4.97	=21.99
2.00 - 2.25	22		17.02	
2.25 - 2.50	29		34.22	
2.50 - 2.75	60		50.16	
2.75 - 3.00	48		59.83	
3.00 - 3.25	64		61.21	
3.25 - 3.50	54		55.18	
3.50 - 3.75	44		44.38	
3.75 - 4.00	38		31.95	
4.00 - 4.25	21		20.52	
4.25 - 4.50	8		11.62	
4.50 - 4.75	5	=9	5.68	=8.94
4.75 - 5.00	2		2.32	
5.00 - 5.25	2		0.94	
	400		400	

We obtain $\chi^2 = 9.74$ with 7 degrees of freedom.

$\chi^2_{.05,7} = 14.07$,

suggesting a good fit.

Working in the same line as in Table 3.9, the Kolmogorov-Smirnov and Cramér-Von Mises statistics under a null hypothesis of a type I fit for the sampling distribution of m* gave

$$D_n = 0.0241816$$

and

$$n\omega_n^2 = 0.02868 .$$

respectively, showing again the evidence in both the tests that a type I curve is a good fit.

3.8 CONCLUDING REMARKS

For the joint estimation of the parameters a and m, the Monte Carlo study shows that the Bayesian method is by far the best with respect to the mse criterion. But for single parameter estimation, MLE of a when m is known as well as MLE of m when a is known have smaller MSE than their Bayesian counterparts. The method of moments again proved to be the most inefficient. The sampling distributions of the Bayesian estimates of a and m both follow a Pearson's type I curve.

CHAPTER FOUR

RAYLEIGH DISTRIBUTION

4.0 INTRODUCTION

We consider the Rayleigh probability density function (pdf) of the time to failure X as given by

$$f(x|\sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x, \sigma > 0. \tag{4.1}$$

The mean and variance of the distribution are

$$E(X) = \sqrt{\frac{\pi}{2}}\,\sigma \quad \text{and} \quad Var(X) = 2\sigma^2\left(1 - \frac{\pi^2}{4}\right)$$

In an expository paper Siddiqui [1962] discussed the origin and properties of Rayleigh distribution. Polovko [1968], and Dyer and Whisenand [1973] noted the importance of this distribution in electrovacuum devices and communication engineering. Cheng [1980] obtained the optimum spacing for the asymptotically best linear unbased estimator and an optimum t-test of the parameter of Rayleigh distribution based on order statistics selected from complete and censored samples when the sample size is large. The instantaneous failure rate

$$\mu(t) = \frac{f(t|\sigma)}{1-F(t|\sigma)} = \frac{t}{\sigma^2}$$

is a linearly increasing function of time and this property makes the model (4.1) specially suitable for life testing experiments of components which 'age' with time.

4.1 ESTIMATION OF σ^2

(i) MLE and UMVUE

It follows from (4.1) that

$$\frac{x^2}{\sigma^2} \sim \chi^2(2).$$

Given a random sample $\underset{\sim}{x} = (x_1,\ldots,x_n)$ from the pdf (4.1), we therefore have

$$\frac{S^2}{\sigma^2} \sim \chi^2(2n), \qquad S^2 = \sum_1^n x_i^2 \tag{4.1a}$$

and

$$E\left(\frac{S^2}{\sigma^2}\right) = 2n. \tag{4.2}$$

The likelihood function

$$\ell(\underline{x}|\sigma) = \left(\prod_{i=1}^n x_i\right)(\sigma^2)^{-n}\exp\left(-\frac{\sum_1^n x_i^2}{2\sigma^2}\right) \tag{4.3}$$

$$\log\ell(\underset{\sim}{x}|\sigma) = \sum_{i=1}^n \log x_i - n\log\sigma^2 - \frac{\sum_1^n x_i^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\sigma^2}\log\ell(\bar{x}|\sigma) = 0$$

yields the maximum likelihood estimator (mle)

$$\hat{\sigma}^2 = \frac{S^2}{2n}.$$

From (4.2),

$$E(\hat{\sigma}^2) = \sigma^2$$

$$g(S^2|\sigma^2) = \frac{1}{\Gamma(n)(2\sigma^2)^n}(S^2)^{n-1}\exp\left(-\frac{S^2}{2\sigma^2}\right), \quad 0 < S^2 < \infty$$

$$E(S) = \frac{1}{\Gamma(n)(2\sigma^2)^n}\int_0^\infty S^{2n-1}\exp\left(-\frac{S^2}{2\sigma^2}\right)dS^2$$

$$= \frac{1}{\Gamma(n)(2\sigma^2)^n} \int_0^\infty (S^2)^{\frac{2n+1}{2}} \exp\left(-\frac{S^2}{2\sigma^2}\right) d\sigma^2$$

$$= \frac{\Gamma(n+\frac{1}{2})}{\Gamma(n)} \sqrt{2}\ \sigma.$$

Using the asymptotic expansion

$$\frac{\Gamma(x+a)}{\Gamma(x)} \to x^a(1 + \frac{a^2+a}{2x}), \quad x \to \infty \tag{4.4}$$

we have

$$E(\hat{\sigma}) = \frac{E(S)}{\sqrt{2n}} = \frac{\sigma}{\sqrt{n}} \frac{\Gamma(n+\frac{1}{2})}{\Gamma(n)} \to \sigma(1 + \frac{3}{8n}).$$

Thus the mle of σ^2 is unbiased but that of σ is unbiased only asymptotically.

We may write (4.1) as belonging to the general exponential family,

$$f(x|\sigma) = \exp\left[\log x - \log \sigma^2 - \frac{x^2}{2\sigma^2}\right]$$

$$\equiv \exp[g(x) + q(\theta) + p(\theta)k(x)].$$

Consequently, $S^2 = \Sigma x_i^2$ is complete and sufficient by Lehmann-Scheffé theorem (Hogg and Craig, 1978). If we can find a function $\phi(S^2)$ such that

$$E[\phi(S^2)] = \sigma^2,$$

then $\phi(S^2)$ is the unique uniformly minimum variance unbiased estimator (umvue) of σ^2. From (4.2), we obtain

$$\phi(S^2) = \frac{S^2}{2n} = \hat{\sigma}^2.$$

We observe that for the Rayleigh distribution the mle and umvue of σ^2 are identical.

(ii) Bayes' Estimator

We will obtain Bayes' estimators of σ^2 and σ under Jeffreys' invariant prior (JIP)

$$p(\sigma) \propto \sqrt{I(\sigma)}$$

where

$$I(\sigma) = -E\left(\frac{\partial^2 \log f(x|\sigma)}{\partial\sigma^2}\right)$$

and the minimal information prior (MIP)

$$g(\sigma) \propto \exp[I_x(\sigma)]$$

where

$$I_x(\sigma) = E[\log f(x|\sigma)].$$

Taking logarithms of (4.1), we have

$$\log f = \log x - \log \sigma^2 - \frac{x^2}{2\sigma^2}$$

$$\frac{\partial^2}{\partial\sigma^2} \log f = \frac{2}{\sigma^2} - \frac{3x^2}{\sigma^4}$$

$$-E\left(\frac{\partial^2}{\partial\sigma^2} \log f\right) = -\frac{2}{\sigma^2} + \frac{3}{\sigma^2} E\left(\frac{x^2}{\sigma^2}\right)$$

$$= -\frac{2}{\sigma^2} + \frac{6}{\sigma^2} \qquad \text{from (4.2)}$$

$$= \frac{4}{\sigma^2}$$

$$= I(\sigma).$$

Hence

$$p(\sigma) \propto \frac{1}{\sigma}.$$

$$E(\log X) = \int_0^\infty \frac{x \log x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

$$= \int_0^\infty \left(\frac{1}{2}\log 2 + \log\sigma + \frac{1}{2}\log y\right)\exp(-y)\,dy$$

$$= \frac{1}{2}\log 2 + \log\sigma + \frac{1}{2}\psi(1)$$

where $-\psi(1)$ is Euler's constant

$$I_x = E[\log f(x|\sigma)] = \frac{1}{2}\log 2 + \log\sigma + \frac{1}{2}\psi(1)$$

$$- 2\log\sigma - 1.$$

$$= \log\left(\frac{1}{\sigma}\right) + \text{constant}.$$

The MIP $g(\sigma) \propto \frac{1}{\sigma}$ and we note that for Rayleigh distribution MIP and JIP are identical.

We will consider a family of improper priors

$$g(\sigma) \propto \frac{1}{\sigma^p}, \quad p > 0 \tag{4.5}$$

of which JIP is a particular case when $p = 1$.

Combining (4.3) and (4.5), the posterior distribution of σ is given by

$$\Pi(\sigma|\underline{x}) = K\,\ell(\underline{x}|\sigma)\,g(\sigma)$$

$$= \frac{K}{\sigma^{2n+p}} \exp\left(-\frac{S^2}{2\sigma^2}\right)$$

where

$$K^{-1} = \int_0^\infty \frac{\exp\left(-\frac{S^2}{2\sigma^2}\right)}{\sigma^{2n+p}}\,d\sigma$$

$$= \frac{1}{2}\,\Gamma\left(\frac{2n+p-1}{2}\right)\left(\frac{2}{S^2}\right)^{\frac{2n+p-1}{2}}$$

Restoring the normalizing constant we obtain

$$\Pi(\sigma|\underline{x}) = \frac{2\left(\frac{S^2}{2}\right)^{\frac{2n+p-1}{2}} \exp\left(-\frac{S^2}{2\sigma^2}\right)}{\Gamma\left(\frac{2n+p-1}{2}\right)\sigma^{2n+p}}, \quad \sigma > 0. \tag{4.6}$$

Under squared error loss function, Bayes estimator

$$\sigma^* = E(\sigma|\underline{x})$$

$$= \frac{\left(\frac{S^2}{2}\right)^{\frac{2n+p-1}{2}}}{\Gamma\left(\frac{2n+p-1}{2}\right)} \int_0^\infty \frac{\exp\left(-\frac{S^2}{2\sigma^2}\right)}{(\sigma^2)^{\frac{2n+p-2}{2}+1}} d\sigma^2$$

$$= \frac{S}{\sqrt{2}} \frac{\Gamma\left(\frac{2n+p-2}{2}\right)}{\Gamma\left(\frac{2n+p-1}{2}\right)}.$$

Using the expansion (4.4),

$$\sigma^* \to \frac{S}{\sqrt{2n}} = \text{mle of } \sigma \text{ as } n \to \infty.$$

Similarly, the posterior distribution of σ^2 and the corresponding Bayes estimators are obtained, viz.,

$$\Pi(\sigma^2|\underline{x}) = \frac{\left(\frac{S^2}{2}\right)^{\frac{2n+p-1}{2}} \exp\left(-\frac{S^2}{2\sigma^2}\right)}{\Gamma\left(\frac{2n+p-1}{2}\right)(\sigma^2)^{\frac{2n+p+1}{2}}} \tag{4.7}$$

and

$$\sigma^{*2} = \frac{S^2}{2n+p-3} \to \hat{\sigma}^2 = \frac{S^2}{2n}, \quad \text{as } n \to \infty$$

Let $T = CS^2$ where C is a constant, be the minimum mean square error (mse) estimator of σ^2.

$$M = E(T - \sigma^2)^2$$

$$= \sigma^4 - 2C\sigma^2 E(S^2) + C^2 E(S^4)$$

$$\frac{\partial M}{\partial C} = 0 = -\sigma^2 E(S^2) + CE(S^4)$$

$$\frac{\partial^2 M}{\partial C^2} = E(S^4) > 0.$$

For min (mse) estimator T, we have

$$C = \frac{\sigma^2 E(S^2)}{E(S^4)}$$

$$\frac{S^2}{\sigma^2} \sim \chi(2n).$$

Hence,

$$E(S^2) = 2n\sigma^2, \qquad V(S^2) = 4n\sigma^4,$$

$$E(S^4) = V(S^2) + \{E(S^2)\}^2$$

$$= 4n\sigma^4(n+1)$$

and

$$C = \frac{2n\sigma^4}{4n\sigma^4(n+1)} = \frac{1}{2n+2}$$

In the class of estimators, $T = CS^2$, the min (mse) estimator of σ^2 is given by

$$T = \frac{S^2}{2n+2},$$

mle and umvue

$$\hat{\sigma}^2 = \frac{S^2}{2n}$$

And the Bayes estimator of σ^2,

$$\sigma^{*2} = \frac{S^2}{2n+p-3} .$$

Note that the prior $g(\sigma) \propto \frac{1}{\sigma^3}$ leads to

$$\sigma^{*2} \equiv \text{the mle and umvue of } \sigma^2$$

and $g(\sigma) \propto \frac{1}{\sigma^5}$ yields

$$\sigma^{*2} \equiv \text{the min (mse) of } \sigma^2 .$$

4.2 RELIABILITY ESTIMATION

(i) MLE and UMVUE

A well-accepted definition of reliability function R_t is 'the probability of a device performing its purpose adequately for the time intended under the operating conditions encountered' (Bazovski [1961]). This probability is given by

$$R_t = P(X \geq t) = \int_t^\infty f(x)dx$$

where f(x) is the pdf of the lifetime X.

For Rayleigh distribution, the reliability function

$$R_t = \int_t^\infty \frac{x}{2} \exp\left(- \frac{x^2}{2\sigma^2}\right) dx$$

$$= \exp\left(- \frac{t^2}{2\sigma^2}\right) . \qquad (4.8)$$

For the exponential failure time distribution

$$p(x|\sigma) = \frac{1}{\sigma} \exp(-\frac{x}{\sigma}), \quad x, \sigma > 0$$

the reliability function

$$R_t = \exp(-\frac{t}{\sigma}). \tag{4.9}$$

Comparing (4.8) and (4.9), we observe that for small values of t, the reliabiity of a Rayleigh component changes more slowly than that of a component for which the failure time distribution is exponential with mean life σ (Dyer and Whisenand, [1973]).

(a) $\hat{R}_t$ = mle of R_t

$$= \exp\left(-\frac{t^2}{2\hat{\sigma}^2}\right)$$

$$= \exp\left(-\frac{nt^2}{S^2}\right).$$

(b) Following Basu [1964], we obtain $\tilde{R}_t$, the umvue of R_t.

The general method of finding the umvue of θ is to look for an unbiased estimator $T(x_1,x_2,\ldots,x_n)$ and a complete sufficient statistic $\tilde{\theta}$ if one exists. Then the umvue of θ is given by

$$E[T(x_1,x_2,\ldots,x_n)|\tilde{\theta}].$$

Consider a function $T(x_1,x_2,\ldots,x_n)$ such that

$$\begin{aligned} T(x_1) &= 1 \quad \text{if } x_1 \geq t \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

$$\begin{aligned} E[T(X_1)] &= 1\cdot P(X_1 \geq t) + 0\cdot P(X_1 < t) \\ &= R_t. \end{aligned}$$

Thus $T(x_1)$ is an unbiased estimator of R_t. Therefore,

$$\tilde{R}_t = E[T(x_1)|\underset{\sim}{\theta}] = \int_t^\infty f(x_1|\underset{\sim}{\theta})dx_1. \tag{4.10}$$

We write

$$\sum_1^n x_i^2 = x_1^2 + \sum_2^n x_j^2. \tag{4.11}$$

Let

$$\sum_1^n x_i^2 = y, \quad x_1^2 = y_1, \ \sum_2^n x_j^2 = (n-1)y_2.$$

Re-write (4.11) as

$$y = y_1 + (n-1)y_2. \tag{4.12}$$

We have noted earlier that $y = \sum_1^n x_i^2$ is complete and sufficient. From (4.10), we obtain

$$\tilde{R}_t = \int_t^\infty g(y_1|y, \sigma)dy_1. \tag{4.13}$$

By use of (4.12), (4.11) and using that $\frac{S^2}{\sigma^2} \sim \chi^2_{(2n)}$ from (4.1a), the joint density of (y_1, y_2) is given by

$$h(y_1, y_2|\sigma) = \frac{(n-1)^{n-1}}{(2\sigma^2)^n} \frac{y_2^{n-2}}{\Gamma(n-1)} \exp\left[-\frac{1}{2\sigma^2}\{y_1 + (n-2)y_2\}\right]$$

Making use of the relationship (4.12), we have the joint density of (y, y_1):

$$\xi(y_1, y|\sigma) = \frac{(n-1)^{n-2}}{\Gamma(n-1)(2\sigma^2)^n} \frac{1}{(n-1)} \left[\frac{y-y_1}{n-1}\right]^{n-2} \exp\left(-\frac{y}{2\sigma^2}\right)$$

$$\xi_1(y|\sigma) = \frac{1}{\Gamma(n)(2\sigma^2)^n} y^{n-1}\exp\left(-\frac{y}{2\sigma^2}\right).$$

Hence,

$$g(y_1|y,\sigma) = \frac{\xi(y_1,y|\sigma)}{\xi_1(y|\sigma)}$$

$$= \frac{(n-1)}{y}\left(1-\frac{y_1}{y}\right)^{n-2}, \quad 0 < y_1 < y.$$

From (4.13), we have

$$\tilde{R}_t = \frac{n-1}{y}\int_t^y \left(1-\frac{y_1}{y}\right)^{n-2} dy_1$$

$$= \frac{n-1}{S^2}\int_t^{S^2}\left(1-\frac{S_1^2}{S^2}\right)^{n-2} dS_1^2$$

$$= \frac{n-1}{S^2}\int_0^{1-\frac{t^2}{S^2}} u^{n-2}S^2 du$$

$$= \left(1-\frac{t^2}{S^2}\right)^{n-1}, \quad 0 < t < S.$$

Thus,

$$\left.\begin{aligned} \tilde{R}_t &= \left(1-\frac{t^2}{S^2}\right)^{n-1}, && 0 < t < S \\ &= 0 && t \geq S \end{aligned}\right\} \qquad (4.14)$$

Note that

$$\tilde{R}_t = \left(1-\frac{t^2}{2n\hat{\sigma}^2}\right)^{n-1}$$

$$\longrightarrow \exp\left(-\frac{t^2}{2\hat{\sigma}^2}\right) = \hat{R}_t \text{ as } n \to \infty.$$

(ii) Bayes estimator

$$R_t = \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

$$\log\left(\frac{1}{R_t}\right) = \frac{t^2}{2\sigma^2}.$$

Substituting

$$\sigma^2 = \frac{t^2}{2\log\left(\frac{1}{R_t}\right)}$$

in (4.7), we obtain the posterior distribution of R_t, viz.,

$$\Pi(R_t|\underline{x}) = \frac{\left(\frac{S^2}{t^2}\right)^{\frac{2n+p-1}{2}}}{\Gamma\left(\frac{2n+p-1}{2}\right)} (R_t)^{\frac{S^2}{t^2}-1} \left(\log\frac{1}{R_t}\right)^{\frac{2n+p-3}{2}},$$

$$0 < R_t < 1 \tag{4.15}$$

which agrees with Sinha and Guttman [1976] with

$$\left\{p,\ n,\ \frac{S^2}{t^2}\right\} \equiv \left\{1,\ r,\ \frac{S_r}{t}\right\}.$$

We have noted earlier that under squared-error loss function, Bayes estimator $\theta^* = E(\theta|\underline{x})$. Hence,

$$R_t^* = \int_0^1 R_t \Pi(R_t|\underline{x})\, dR_t$$

$$= \frac{\Gamma\left(\frac{2n+p-1}{2}\right)\left(\frac{S^2}{t^2}\right)^{\frac{2n+p-1}{2}}}{\left(\frac{S^2}{t^2}+1\right)^{\frac{2n+p-1}{2}} \Gamma\left(\frac{2n+p-1}{2}\right)}$$

$$= \frac{1}{\left(1+\frac{t^2}{S^2}\right)^{\frac{2n+p-1}{2}}} . \tag{4.16}$$

Comparing (4.14) and (4.16) we observe that $\tilde{R}_t = 0$ for $t \geq S$, which is realistically unacceptable, whereas Bayes estimator $R_t^* > 0$ for all $t > 0$.

We now obtain the posterior density and Bayes estimator for a Rayleigh component failure rate, $\mu(t) = \frac{t}{\sigma^2}$.

Substituting $\sigma^2 = \frac{t}{\mu(t)}$ in (4.7), the posterior of $\mu(t)$ is given by

$$\Pi\{\mu(t) \mid \underline{x}\} = \frac{\left(\frac{S^2}{2}\right)^{\frac{2n+p-1}{2}}}{\Gamma\left(\frac{2n+p-1}{2}\right)} \left[\frac{\exp\left\{-\frac{S^2\mu(t)}{2t}\right\}\left\{\mu(t)\right\}^{\frac{2n+p-3}{2}}}{(t)^{\frac{2n+p-1}{2}}}\right]$$

$$0 < \mu(t) < \infty, \tag{4.17}$$

and the corresponding estimator

$$\mu^*(t) = \frac{\left(\frac{S^2}{2}\right)^{\frac{2n+p-1}{2}}}{\Gamma\left(\frac{2n+p-1}{2}\right)} \frac{1}{t^{\frac{2n+p-1}{2}}} \int_0^\infty \exp\left\{-\frac{S^2\mu(t)}{2t}\right\}$$

$$\cdot \{\mu(t)\}^{\frac{2n+p-1}{2}} \, d\mu(t)$$

$$= \frac{(2n+p-1)t}{S^2}$$

$$\to \hat{\mu}(t) \quad \text{as } n \to \infty.$$

For JIP,

$$\hat{\mu}(t) \equiv \mu^*(t). \tag{4.18}$$

A sample of size $n = 20$ was generated from a Rayleigh pdf (4.1) with $\sigma = 2$ and the posterior distributions (4.6), (4.7), (4.15) and (4.17) were plotted in Figures 1 to 4 with $p = 1, (1), 4$. The posteriors tend to be less and less robust as the prior distribution of σ deviates from the JIP.

4.3 SIMULATION

In order to compare the Bayes estimators of σ, σ^2, $R(t)$ and $\mu(t)$ with their respective mle's $N = 500$ samples of size $n = 20$ were generated from the Rayleigh pdf (4.1) with $\sigma = 2$. The average of these estimates and their mse's are computed. We report the results in the following table. The entries within the parentheses represent the corresponding MSE.

TABLE 4.1

Bayes Estimates (*) and mle (^) under JIP,

$N = 500$, $n = 20$

Estimates / Parameter	True Value	Estimator: Bayes (*)	Estimator: MLE (^)	MSE(*) / MSE(^)
σ	2.0000	2.0295 (0.0557)	1.9912 (0.0529)	1.05
σ^2	4.0000	4.2291 (1.0125)	4.0177 (0.8667)	1.17
$R(5)$	0.0439	0.0585 (0.0014)	0.0483 (0.0011)	1.27
$\mu(5)$	1.2500	1.3128 (0.0998)	1.3128 (0.0998)	1.00

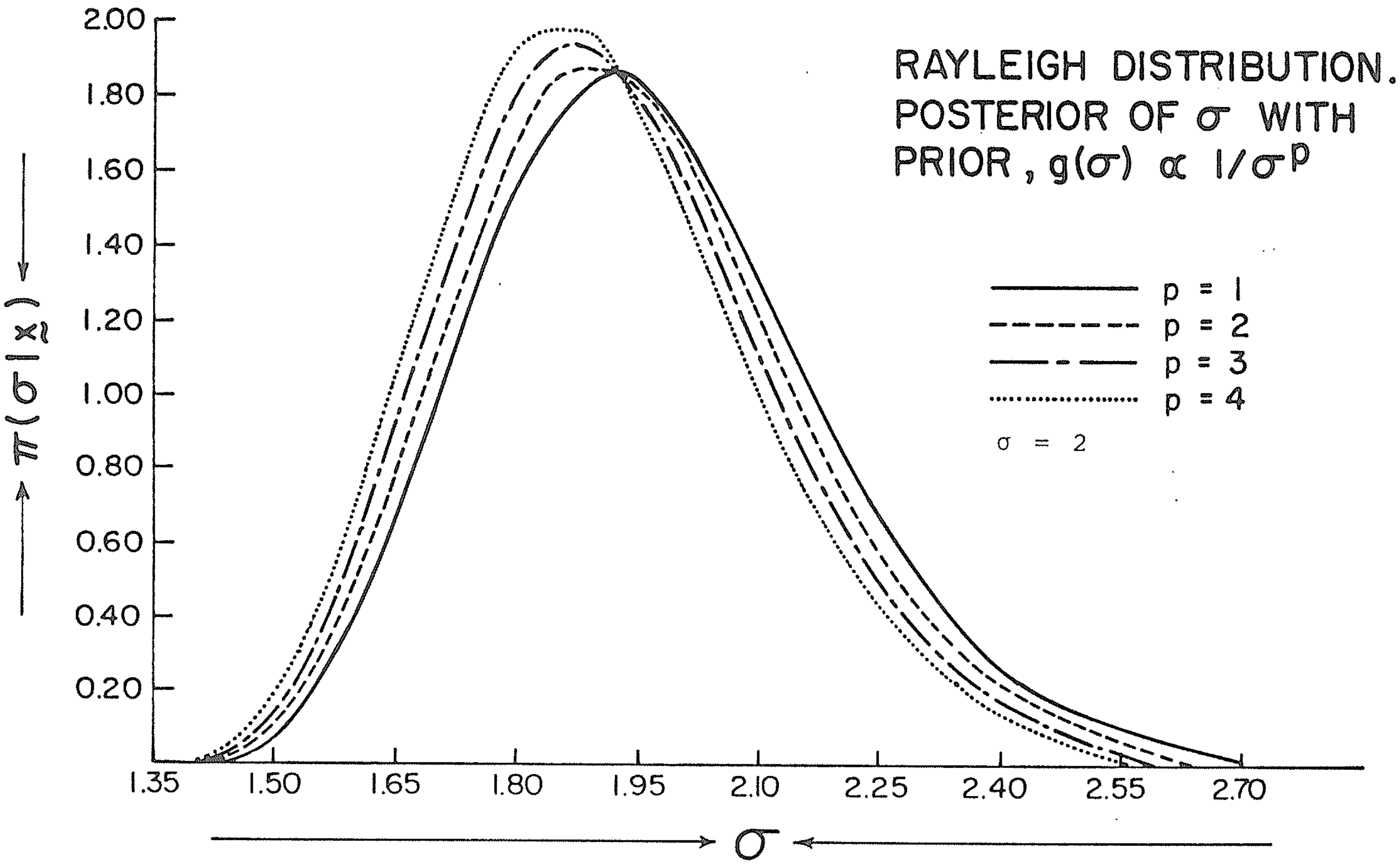

RAYLEIGH DISTRIBUTION.
POSTERIOR OF σ WITH
PRIOR, g(σ) ∝ 1/σ^p
p = 1
p = 2
p = 3
p = 4
σ = 2
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
1.35
1.50
1.65
1.80
1.95
2.10
2.25
2.40
2.55
2.70
π(σ | x)
σ

FIGURE 1

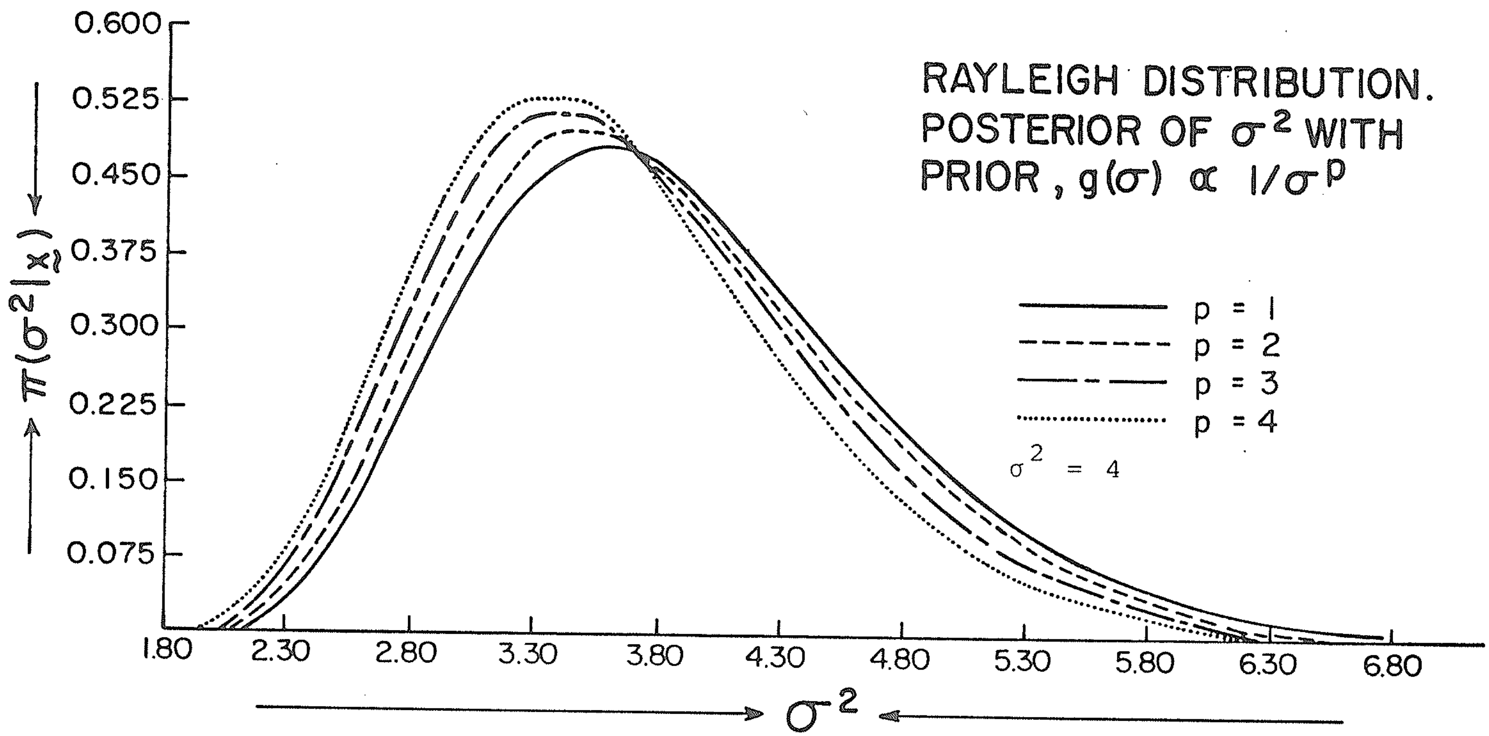

RAYLEIGH DISTRIBUTION.
POSTERIOR OF σ² WITH
PRIOR, g(σ) ∝ 1/σ^p
p = 1
p = 2
p = 3
p = 4
σ² = 4
0.600
0.525
0.450
0.375
0.300
0.225
0.150
0.075
π(σ²|x)
1.80
2.30
2.80
3.30
3.80
4.30
4.80
5.30
5.80
6.30
6.80
σ²

FIGURE 2

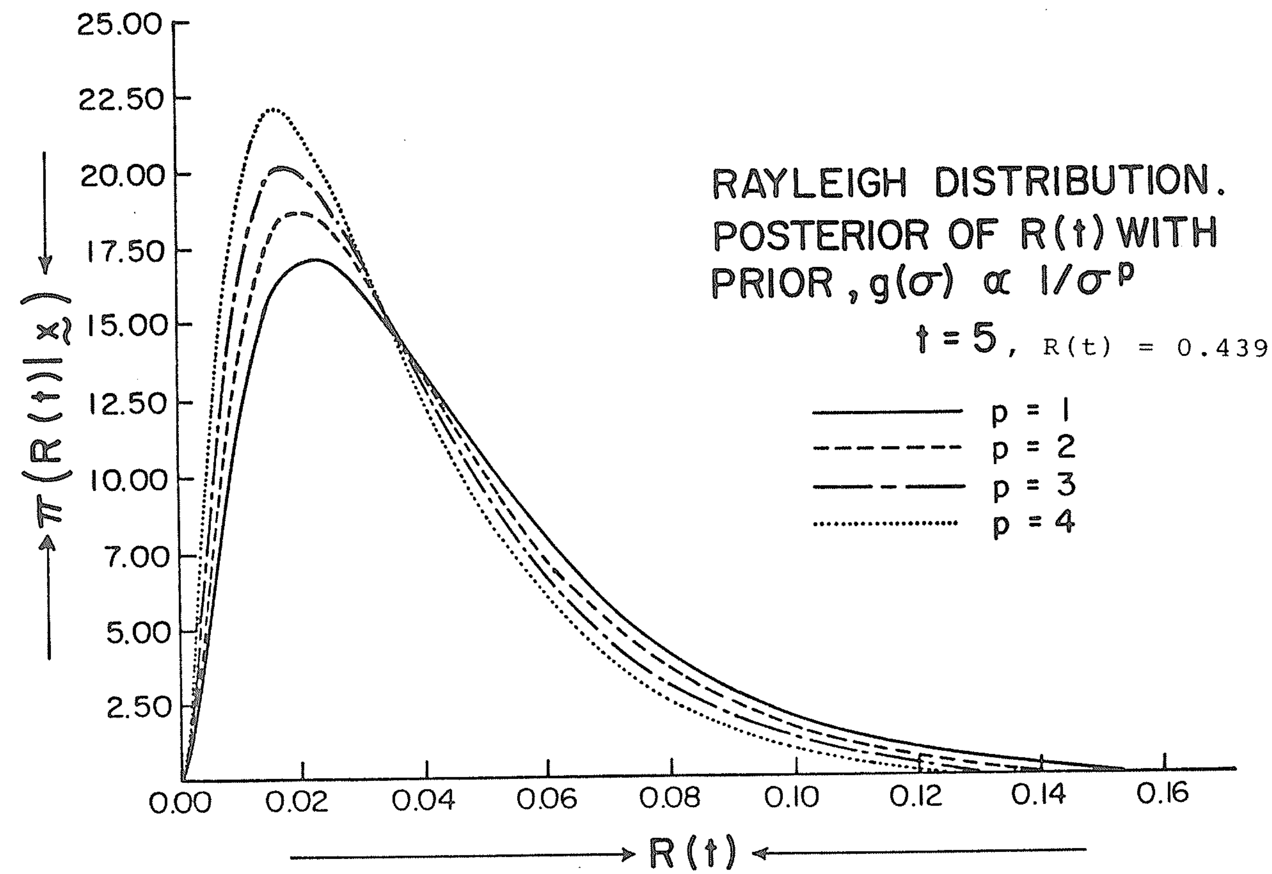

25.00
22.50
20.00
17.50
15.00
12.50
10.00
7.00
5.00
2.50
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
π (R(t)|x̰)
R(t)
RAYLEIGH DISTRIBUTION.
POSTERIOR OF R(t) WITH
PRIOR, g(σ) α 1/σ^p
t = 5, R(t) = 0.439
p = 1
p = 2
p = 3
p = 4

FIGURE 3

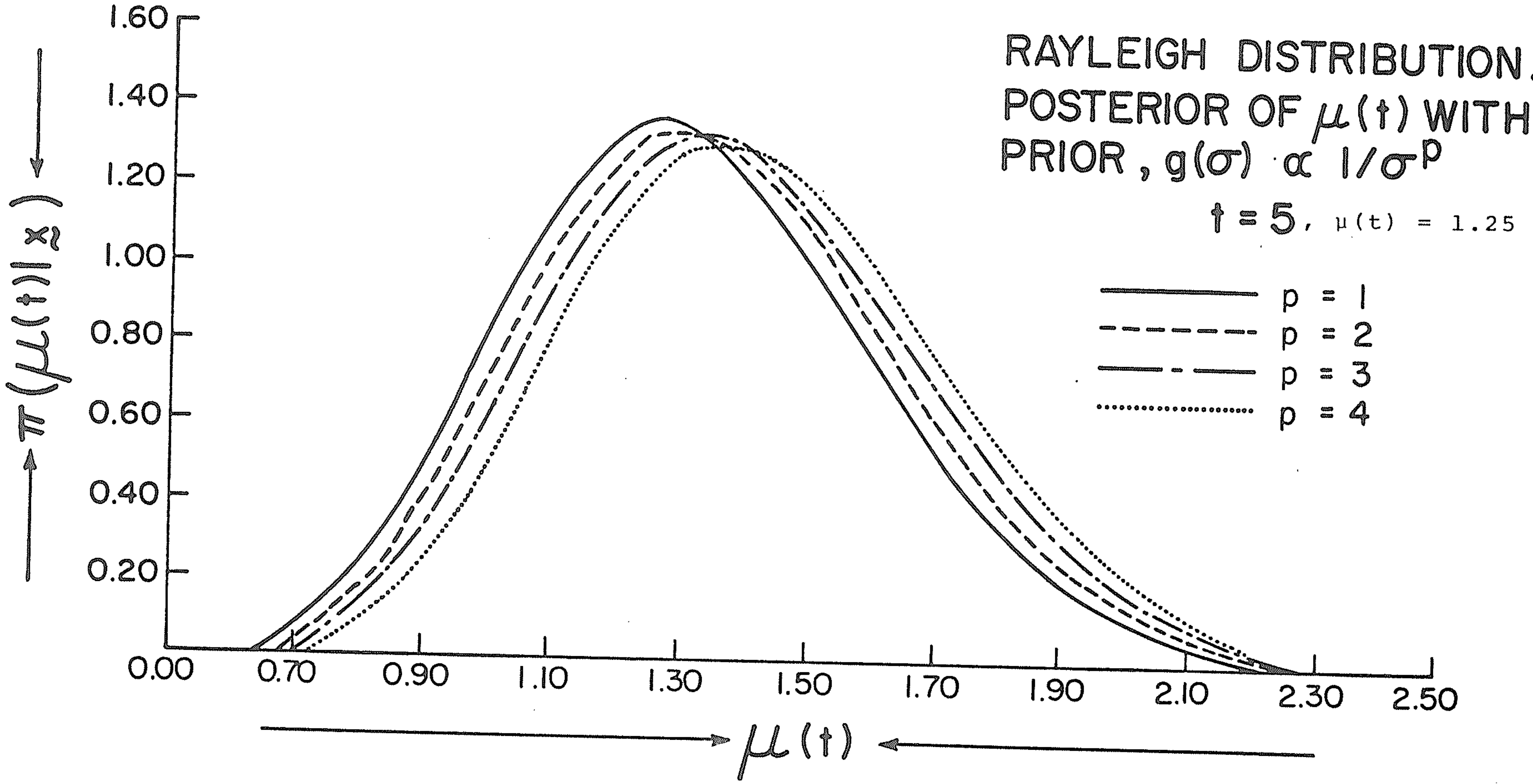

RAYLEIGH DISTRIBUTION.
POSTERIOR OF μ(t) WITH
PRIOR, g(σ) α 1/σ^p
t = 5, μ(t) = 1.25
p = 1
p = 2
p = 3
p = 4
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0.70
0.90
1.10
1.30
1.50
1.70
1.90
2.10
2.30
2.50
π(μ(t)|x)
μ(t)

FIGURE 4

Based on the above computations it appears that the mle of σ, σ^2 and R(5) have a slight edge over their Bayesian counterparts.

Using the same simulation, the UMVUE of R_t as given by (4.14) and the corresponding mse were also obtained. We have $\tilde{R}(5) = 0.0448$, MSE(~) = 0.0011.

True R(5): $\tilde{R}(5)$: $\hat{R}(5)$: $R^*(5)$:: 0.0439: 0.0448: 0.0483: 0.0583

MSE(~): MSE(^): MSE(*):: 0.0011: 0.0011: 0.0014

Thus the UMVUE and MLE of R(5) are equally efficient and are preferable to the Bayes estimate R*(5).

TABLE 4.2

Bayes Estimates under $g(\sigma) \propto \frac{1}{\sigma^p}$, N = 500, n = 20

p / Parameter	1	2	3	4
σ	2.0295 (0.0557)	2.0037 (0.0535)	1.9788 (0.0526)	1.9548 (0.0529)
σ^2	4.2291 (1.0125)	4.1207 (0.9259)	4.0177 (0.8667)	3.9197 (0.8311)
R(5)	0.0585 (0.0014)	0.0547 (0.0012)	0.0512 (0.0011)	0.0479 (0.0009)
$\mu(5)$	1.3128 (0.0998)	1.3456 (0.1099)	1.3784 (0.1222)	1.4112 (0.1368)

As expected, the more $g(\sigma)$ deviates from JIP, Table 4.2 reflects a similar pattern of lack of robustness of the posterior expectations as evidenced by the corresponding posteriors plotted in Figures 1 - 4.

4.4 HPD AND EQUAL-TAIL CREDIBLE INTERVALS

Having obtained the posterior distribution $\Pi(\theta|\underline{x})$, we ask 'how likely is it that the parameter θ lies within a specified interval $[c_1, c_2]$?' It may not be anywhere near certainty but it may not be highly unlikely either. Suppose we are $100(1-\alpha)\%$ confident that $\theta \nless c_1$ and $\ngtr c_2$. This is not the same as the classical confidence interval interpretation for θ where θ is a constant and it is meaningless to make a probability statement about a constant. Bayesians call this interval based on the posterior distribution a 'credible interval' (Edwards, Lindman and Savage [1963]) or 'Bayesian confidence interval' (Lindley [1965]) — an interval within which one assigns a fixed probability that an uncertain parameter θ is located.

It seems quite reasonable to require that such an interval should have the following properties:

(A) for a given probability content P, the interval should be as short as possible;

(B) the posterior density for every point inside the interval is greater than that for every point outside it so that the interval includes more probable values of the parameter and excludes less probable values.

Such an interval is called the highest posterior density or HPD-interval (Box and Tiao [1973]).

As earlier introduced, an interval $[c_1, c_2]$ is said to be a $100(1-\alpha)\%$ credible (or Bayesian confidence) interval for θ if

$$\int_{c_1}^{c_2} \Pi(\theta|\underline{x})\,d\theta = 1 - \alpha. \qquad (4.19)$$

For the shortest credible interval we have to minimize $I = c_2 - c_1$ subject to the condition (4.19) which requires

$$\Pi(c_1|\underline{x}) = \Pi(c_2|\underline{x}). \qquad \text{(Evans, 1976)} \qquad (4.20)$$

An interval $[c_1, c_2]$ which simultaneously satisfies (4.19) and (4.20) is called the 'shortest $100(1-\alpha)\%$ credible interval'.

For unimodal but not necessarily symmetrical posterior density, the shortest credible and the HPD-intervals are identical (Evans [1976]).

In general an HPD interval satisfying (A) and (B) has to be obtained numerically using innovative computer programming. To avoid laborious computations one may prefer a $100(1-\alpha)\%$ equal-tail credible interval $[a,b]$ for θ given by

$$\int_{-\infty}^{a} \Pi(\theta|\underline{x})\,d\theta = \int_{b}^{\infty} \Pi(\theta|\underline{x})\,d\theta = \frac{\alpha}{2}.$$

For a symmetrical, unimodal posterior distribution, the HPD, shortest credible and equal-tail credible intervals are one and the same (Evans, 1976).

(i) Equal-Tail and HPD Credible Intervals for σ and $R(t)$

We have shown earlier that

$$\frac{S^2}{\sigma^2} \sim \chi^2(2n)$$

$$1-\alpha = P\left[\chi^2_{1-\frac{\alpha}{2}}(2n) < \frac{S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(2n)\right]$$

$$= P\left[\frac{S}{\sqrt{\chi^2_{\frac{\alpha}{2}}(2n)}} < \sigma < \frac{S}{\sqrt{\chi^2_{1-\frac{\alpha}{2}}(2n)}}\right] \tag{4.21}$$

Thus

$$\left[\frac{S}{\sqrt{\chi^2_{\frac{\alpha}{2}}(2n)}}, \frac{S}{\sqrt{\chi^2_{1-\frac{\alpha}{2}}(2n)}}\right]$$

is the $100(1-\alpha)$% equal-tail credible (or Bayesian confidence) interval for σ where $S^2 = \sum_1^n x_i^2$ and $\chi^2_\beta(2n)$ = the upper 100β% point of a χ^2-distribution with 2n degrees of freedom.

The posterior distribution of σ as given by (4.6) is unimodal. For the HPD-interval for σ we must have the following equations simultaneously satisfied, viz.,

$$P(H_L < \sigma < H_U) = 1 - \alpha$$

$$\Pi(H_L|\underline{x}) = \Pi(H_U|\underline{x})$$

where H_L and H_U are the lower and upper $100(1-\alpha)$% HPD-credible limits of σ. From (4.21) and(4.6), the above equations lead to

$$P\left(\frac{S^2}{H_U^2} < \chi^2(2n) < \frac{S^2}{H_L^2}\right) = 1 - \alpha \tag{4.22}$$

$$\exp\left[-\frac{S^2}{2}\left(\frac{1}{H_L^2} - \frac{1}{H_U^2}\right)\right] = \left(\frac{H_L}{H_U}\right)^{2n+1} . \tag{4.23}$$

Using the data referred to earlier, we have $n = 20$, $S^2 = 150.14349$, $\chi^2(0.05,40) = 55.7585$, $\chi^2(0.95,40) = 26.5093$. From (4.21), the lower and upper equal-tail 90% credible limits for σ are

$$c_L = \sqrt{\frac{150.14349}{55.7585}} = 1.6410$$

$$c_U = \sqrt{\frac{150.14349}{26.5093}} = 2.3799.$$

Computations for H_L and H_U are a little more involved than those for (c_L, c_U) since (H_L, H_U) must satisfy the equations (4.22) and (4.23) simultaneously.

For the same set of data, we obtained the 90% HPD-credible limits

$H_L = 1.6061, \qquad H_U = 2.3336.$

Note that the 90% HPD-interval is shorter than the corresponding equal-tail credible interval which, of course, is expected.

From (4.15) it follows that

$$\frac{2S^2}{t^2}\log\left(\frac{1}{R_t}\right) \sim \chi^2(2n).$$

Hence $100(1-\alpha)\%$ equal-tail credible limits $c_L(t)$ and $c_U(t)$ for the reliability function $R(t)$ are the solutions of

$$\frac{2S^2}{t^2} \log \left\{\frac{1}{c_U(t)}\right\} = \chi^2_{1-\frac{\alpha}{2}}(2n) \tag{4.24}$$

$$\frac{2S^2}{t^2} \log \left\{\frac{1}{c_L(t)}\right\} = \chi^2_{\frac{\alpha}{2}}(2n). \tag{4.25}$$

The corresponding HPD-limits $H_U(t)$ and $H_L(t)$ are the simultaneous solutions of

$$P\left[\frac{2S^2}{t^2} \log \left\{\frac{1}{H_U(t)}\right\} < \chi^2(2n) < \frac{2S^2}{t^2} \log \left\{\frac{1}{H_L(t)}\right\}\right] = 1-\alpha \tag{4.26}$$

and

$$\{H_L(t)\}^{\frac{S^2}{t^2}-1} [\log\{H_L(t)\}]^{(n-1)} = \{H_U(t)\}^{\frac{S^2}{t^2}-1} [\log\{H_U(t)\}]^{(n-1)} \tag{4.27}$$

Using the data referred to in the preceding at $t = 5$ we obtain 90% equal-tail credible limits of $R(5)$ from (4.24) and (4.25), given by

$$c_U(5) = 0.110024, \qquad c_L(5) = 0.009636$$

and

$$H_U(5) = 0.090624, \qquad H_L(5) = 0.003507.$$

As expected, the HPD-interval is shorter than the corresponding equal-tail credible interval.

4.5 PREDICTIVE DISTRIBUTIONS

In recent years much attention has been given to the prediction problems of the life time models, especially to the two-parameter (sometimes called 'left truncated') exponential model both in sampling theory and Bayesian viewpoints (Dunsmore [1974], Likeš [1974], Lawless [1977], Evans and Nigm [1980]). In this section we use a Bayesian approach to the prediction problem when the underlying lifetime has a Rayleigh distribution. We consider the prediction of a future observation.

(i) HPD-Predictive Interval for a Future Observation

Given $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$, the conditional joint density of x_{n+1} and σ is

$$h(x_{n+1}, \sigma | \underset{\sim}{x}) = f(x_{n+1} | \sigma, \underset{\sim}{x}) \Pi(\sigma | \underset{\sim}{x})$$

$$= f(x_{n+1} | \sigma) \Pi(\sigma | \underset{\sim}{x}),$$

since x_{n+1} is independent of $\underset{\sim}{x}$.

From (4.1) and (4.6) with $p = 1$, we obtain

$$h(x_{n+1}, \sigma | \underset{\sim}{x}) = \frac{x_{n+1}}{\sigma^2} e^{-\frac{x_{n+1}^2}{2\sigma^2}} \frac{S^{2n}}{2^{n-1}\Gamma(n)} \frac{e^{-\frac{S^2}{2\sigma^2}}}{\sigma^{2n+1}} . \qquad (4.28)$$

We integrate out σ from (4.28) to get the predictive density of x_{n+1} as

$$\Pi(x_{n+1}|\underset{\sim}{x}) = 2n \frac{S^{2n} x_{n+1}}{(S^2+x_{n+1}^2)^{n+1}} . \tag{4.29}$$

We note that x_{n+1}^2 has an inverted beta distribution, $\mathrm{InBe}(1, n, S^2)$.

The inverted beta distribution with parameters α, β, and r, $\alpha, \beta, r > 0$ denoted $\mathrm{InBe}(\alpha,\beta,r)$, is defined by the pdf

$$\frac{y^{\alpha-1} r^{\beta}}{B(\alpha,\beta)(y+r)^{\alpha+\beta}}$$

for $y \geq 0$, $\alpha, \beta, r > 0$.

Bayes estimator of x_{n+1} under squared error loss function is given by

$$x_{n+1}^{*} = E(x_{n+1}|\underset{\sim}{x}) = \int_0^{\infty} 2n \frac{S^{2n} x_{n+1}^2}{(S^2+x_{n+1}^2)^{n+1}} \, dx_{n+1}$$

$$= \frac{\Gamma(n-\tfrac{1}{2})\Gamma(3/2)}{\Gamma(n)} S .$$

The reliability of the predictive density (4.29) is

$$R_{x_{n+1}}(t) = P(x_{n+1} > t) = \int_t^{\infty} 2n \frac{S^{2n} x_{n+1}}{(S^2+x_{n+1}^2)^{n+1}} \, dx_{n+1}$$

$$= \left(1 + \frac{t^2}{S^2}\right)^{-n}$$

which is precisely the expected value of the reliability function of the original Rayleigh distribution as obtained in (4.16) with $p = 1$.

The predictive density (4.29) is unimodal, the $100(1-\alpha)\%$ HPD-predictive interval for x_{n+1} has limits c_1 and c_2 satisfying the equations

$$\left.\begin{aligned} &P(c_1 < x_{n+1} < c_2) = 1 - \alpha \\ \text{and} \\ &\Pi(c_1|\underset{\sim}{x}) = \Pi(c_2|\underset{\sim}{x}). \end{aligned}\right\} \tag{4.30}$$

Since

$$y = x_{n+1}^2 \sim \mathrm{InBe}(1, n, S^2),$$

$$\Pi_{x_{n+1}}(c_1|\underset{\sim}{x}) = \Pi_y(c_1|\underset{\sim}{x}) \left|\frac{\partial y}{\partial x_{n+1}}\right| = \frac{S^{2n} 2\sqrt{c_1}}{B(1,n)(c_1+S^2)^{n+1}} .$$

Similarly,

$$\Pi_{x_{n+1}}(c_2|\underset{\sim}{x}) = \frac{S^{2n} 2\sqrt{c_2}}{B(1,n)(c_2+S^2)^{n+1}} .$$

Thus, the equations in (4.30) take the form

$$\left.\begin{aligned} &P(c_1^2 < y < c_2^2) = 1 - \alpha \\ \text{and} \\ &\left(\frac{c_2+S^2}{c_1+S^2}\right)^{n+1} = \sqrt{\frac{c_2}{c_1}} \end{aligned}\right\} \tag{4.31}$$

where

$$P(c_1^2 < y < c_2^2) = \int_{c_1^2}^{\infty} \Pi_y(y|\underset{\sim}{x})\,dy - \int_{c_2^2}^{\infty} \Pi_y(y|\underset{\sim}{x})\,dy.$$

Consider

$$\int_{c_1^2}^{\infty} \Pi_y(y|\underset{\sim}{x})\,dy = \int_{c_1^2}^{\infty} \frac{1}{S^2 B(1,n)} (1 + y/S^2)^{-(n+1)} dy$$

$$= \int_0^{\frac{S^2}{S^2+c_1^2}} \frac{1}{B(1,n)} Z^{n-1} dZ = I_{\frac{S^2}{S^2+c_1^2}}(n,1)$$

where

$$1 + \frac{y}{S^2} = \frac{1}{z} .$$

Similarly,

$$\int_{c_2^2}^{\infty} \Pi_y(y|\underset{\sim}{x})\,dy = I_{\frac{S^2}{S^2+c_2^2}}(n,1)$$

where

$$I_p(a,b) = \frac{1}{B(a,b)} \int_0^p x^{a-1}(1-x)^{b-1} dx$$

Hence, the first equation in (4.31) takes the form

$$P(c_1^2 < y < c_2^2) = I_{\frac{S^2}{S^2+c_1^2}}(n,1) - I_{\frac{S^2}{S^2+c_2^2}}(n,1) .$$

Since beta distribution is not as extensively tabulated as binomial distribution is (Weintraub [1963]), we use the relation between the incomplete beta function and binomial expansion, viz.,

$$I_p(a,b) = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} p^i (1-p)^{a+b-1-i} .$$

Thus,

$$\mathbb{P}(c_1^2 < y < c_2^2) = \left(\frac{s^2}{s^2+c_1^2}\right)^n - \left(\frac{s^2}{s^2+c_2^2}\right)^n .$$

Thus, the $100(1-\alpha)\%$ HPD-predictive interval is the simultaneous solution of the equations

$$\left.\begin{aligned} &\left(\frac{s^2}{s^2+c_1^2}\right)^n - \left(\frac{s^2}{s^2+c_2^2}\right)^n = 1 - \alpha \\ &\text{and} \\ &\left(\frac{c_2+s^2}{c_1+s^2}\right)^{n+1} = \sqrt{\frac{c_2}{c_1}} \ . \end{aligned}\right\} \tag{4.32}$$

Using the generated data with $n = 20$, $s^2 = 105.14349$, the solution of (4.32) is

$$c_1 = 0.8903$$

and

$$c_2 = 9.7457.$$

CHAPTER FIVE

BAYES ESTIMATORS OF THE TWO PARAMETERS OF A NORMAL DISTRIBUTION

5.0 INTRODUCTION

Bayesian treatment of the variance σ^2 of a normal distribution was considered by Goodman (1960) and Evans (1964). Lindley (1965), Zellner (1971), and Box and Tiao (1973) consider the normal distribution under uniform prior. Sinha (1980) also considers the normal distribution under JIP. In this dissertation we will consider Bayesian treatment of this distribution under the Natural Conjugate Priors (NCP) (DeGroot, 1970) for σ for the two parameter normal density function

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},$$

$$-\infty < x,\ \mu < \infty,\ \sigma > 0 \qquad (5.1)$$

and obtain Bayes estimators for μ and σ^2 under a variety of loss functions. We will further consider a Monte Carlo study to compare these estimators and search for the optimal loss function, if any, which yields minimum mean squared error estimators for all combinations of the prior parameter involved.

5.1 POSTERIOR DISTRIBUTIONS OF μ AND σ^2

We will use the NCP for σ given by

$$g(\sigma) = \frac{2c^{p/2}}{\Gamma(p/2)} \frac{\exp(-c/\sigma^2)}{\sigma^{p+1}}$$

and the diffuse prior

$$g(\mu) = \text{constant}$$

for μ. In the absence of any prior knowledge about μ and σ, it is not unreasonable, at least to start with, to assume stochastic independence of the parameters (Box and Tiao, 1973). Our "belief" about the priors will, of course, be influenced and revised in the light of the subsequent posterior analysis. Thus, the joint prior of μ and σ is

$$g(\mu,\sigma) \doteq g(\mu)g(\sigma) \propto \frac{\exp(-c/\sigma^2)}{\sigma^{p+1}} . \tag{5.2}$$

Given the sample $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$, we may write the likelihood function

$$\ell(\underset{\sim}{x}|\mu,\sigma^2) = \left[\frac{1}{\sqrt{2\pi}\,\sigma}\right]^n \exp\left\{-\frac{1}{2\sigma^2}\,\Sigma(x_i-\mu)^2\right\}. \tag{5.3}$$

Combining (5.2) and (5.3), the joint posterior of μ and σ is

$$\Pi(\mu,\sigma|\underset{\sim}{x}) = \frac{k}{\sigma^{\frac{n+p+2}{2}}} \exp\left[-\frac{1}{2\sigma^2}\{A + 2c + n(\mu-\bar{x})^2\}\right] \tag{5.4}$$

where $A = \sum_1^n (x_i-\bar{x})^2$ and k is a normalizing constant.

Integrating out μ and restoring the normalizing constant, we have the marginal posterior of σ as

$$\Pi(\sigma|\underset{\sim}{x}) = \frac{(A+2c)^{\frac{n+p-1}{2}}}{\Gamma(\frac{n+p-1}{2})\,2^{\frac{n+p-3}{2}}}\;\frac{\exp\left\{-\frac{1}{2\sigma^2}(A+2c)\right\}}{\sigma^{n+p}} ,$$

and the marginal posterior of σ^2 as

$$\Pi(\sigma^2|\underset{\sim}{x}) = \frac{\left(\frac{A+2c}{2}\right)^{\frac{n+p-1}{2}}}{\Gamma\left(\frac{n+p-1}{2}\right)} \frac{\exp\left\{-\frac{1}{2\sigma^2}(A+2c)\right\}}{(\sigma^2)^{\frac{n+p+1}{2}}}, \quad \sigma^2 > 0 \qquad (5.5)$$

which again is an inverted gamma distribution.

The marginal posterior of μ from (5.4) is given by

$$\Pi(\mu|\underset{\sim}{x}) = \frac{1}{\sqrt{\lambda}\, B\left(\frac{1}{2}, \frac{m}{2}\right)\left[1 + \frac{(\mu-\bar{x})^2}{\lambda}\right]^{\frac{m+1}{2}}}, \quad -\infty < \mu < \infty \qquad (5.6)$$

where

$$\lambda = \frac{A+2c}{n}, \quad m = n+p-1.$$

We now find the posterior moments. Consider an inverted gamma pdf

$$f(x|a,b) = \frac{a^b}{\Gamma(b)} \frac{e^{-a/x}}{x^{b+1}}, \quad x,\ a,\ b > 0 \qquad (5.7)$$

where

$$E(x) = \frac{a}{b-1}$$

$$\mu_r = E\left(x - \frac{a}{b-1}\right)^r$$

$$= \frac{a^b}{\Gamma(b)} \int_0^\infty \left(x - \frac{a}{b-1}\right)^r \frac{e^{-a/x}}{x^{b+1}}\, dx$$

$$= \frac{a^b}{\Gamma(b)} \int_0^\infty \sum_{k=0}^{r} (-1)^k \binom{r}{k} \left(\frac{a}{b-1}\right)^k \frac{x^{r-k} e^{-a/x}}{x^{b+1}}\, dx$$

$$= \frac{a^b}{\Gamma(b)} \sum_{k=0}^{r} (-1)^k \binom{r}{k} \frac{a^k}{(b-1)^k} \frac{\Gamma(b+k-r)}{a^{b+k-r}}$$

$$= \frac{a^r}{\Gamma(b)} \sum_{k=0}^{r} (-1)^k \binom{r}{k} \frac{\Gamma(b+k-r)}{(b-1)^k} .$$

Thus,

$$\mu_2 = \frac{a^2}{(b-1)^2(b-2)} , \qquad \mu_3 = \frac{4a^3}{(b-1)^3(b-2)(b-3)}$$

$$\mu_4 = \frac{3a^4(b+5)}{(b-1)^4(b-2)(b-3)(b-4)}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{16(b-2)}{(b-3)^2}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(b+5)(b-2)}{(b-3)(b-4)} .$$

Putting

$$a = \frac{A+2c}{2} , \quad b = \frac{n+p-1}{2}$$

for the posterior density of σ^2 in (5.5), we have

$$\beta_1 = \frac{32(n+p-5)}{(n+p-7)^2} \to 0 \quad \text{as} \quad n \to \infty$$

and

$$\beta_2 = \frac{3(n+p+9)(n+p-5)}{(n+p-7)(n+p-9)} \to 3 \quad \text{as} \quad n \to \infty$$

This implies that

$$\Pi(\sigma^2 | \underset{\sim}{x}) \to \text{Normality} \quad \text{as} \quad n \to \infty.$$

Again, $\Pi(\mu | \underset{\sim}{x})$ is symmetric in $(\mu - \bar{x})$ about zero, which implies

$$E(\mu-\bar{x})^{2r+1} = 0, \quad r = 0,1,2,...$$

$$\mu_{2r} = E(\mu-\bar{x})^{2r}$$

$$= \frac{1}{\sqrt{\lambda}} \frac{1}{B(\frac{1}{2}, \frac{m}{2})} \int_0^\infty \frac{(\mu-\bar{x})^{2r-1} d(\mu-\bar{x})^2}{\left\{1+\frac{(\mu-\bar{x})^2}{\lambda}\right\}^{\frac{m+1}{2}}}.$$

Putting

$$\frac{(\mu-\bar{x})^2}{\lambda} = \frac{t^2}{m},$$

we have

$$\mu_{2r} = \frac{1}{\sqrt{\lambda}} \frac{1}{B(\frac{1}{2}, \frac{m}{2})} \int_0^\infty \frac{(t^2 \frac{\lambda}{m})^{r-\frac{1}{2}} (\frac{\lambda}{m}) dt^2}{\left(1+\frac{t^2}{m}\right)^{\frac{m+1}{2}}}$$

$$= \frac{\lambda^r}{B(\frac{1}{2}, \frac{m}{2})} \int_0^\infty \frac{\left(\frac{t^2}{m}\right)^{(r+\frac{1}{2})-1} d(\frac{t^2}{m})}{\left(1+\frac{t^2}{m}\right)^{\frac{m+1}{2}}}$$

$$= \frac{\lambda^r}{B(\frac{1}{2}, \frac{m}{2})} B(r+\frac{1}{2}, \frac{m}{2}-r)$$

$$= \frac{\lambda^r \Gamma(r+\frac{1}{2}) \Gamma(\frac{m}{2}-r)}{\Gamma(\frac{1}{2}) \Gamma(\frac{m}{2})}.$$

Thus

$$\mu_2 = \frac{\lambda}{m-2}, \qquad \mu_4 = \frac{3\lambda^2}{(m-2)(m-4)}$$

$$\beta_1 = 0, \qquad \beta_2 = \frac{3(m-2)}{(m-4)} = \frac{3(n+p-3)}{(n+p-5)} \to 3 \text{ as } n \to \infty$$

showing that

$$\Pi(\mu|\underset{\sim}{x}) \to \text{Normality as } n \to \infty.$$

A random sample of $n = 20$ was generated from a normal population with $\mu = 20$ and $\sigma = 3$. The sample gave $\bar{x} = 20.50139$ and $A = \sum_{1}^{20}(x_i - \bar{x})^2 = 173.2500$, the posterior $\Pi(\sigma^2 | \underset{\sim}{x})$ was plotted for $c = 1, 2, 3, 4$ and fixed $p = 1\ (1)\ 4$ in Figures 1 to 4 and for $p = 1, 2, 3, 4$ and $c = 1\ (1)\ 4$ in Figures 5 to 8. The posteriors of σ^2 appear to be quite robust, almost symmetrical for fixed p and varying c but less so for fixed c and varying p. The posteriors of μ were found to be very much robust for variations in both p and c. The graphs came so close that it was difficult to distinguish among them and hence are not displayed.

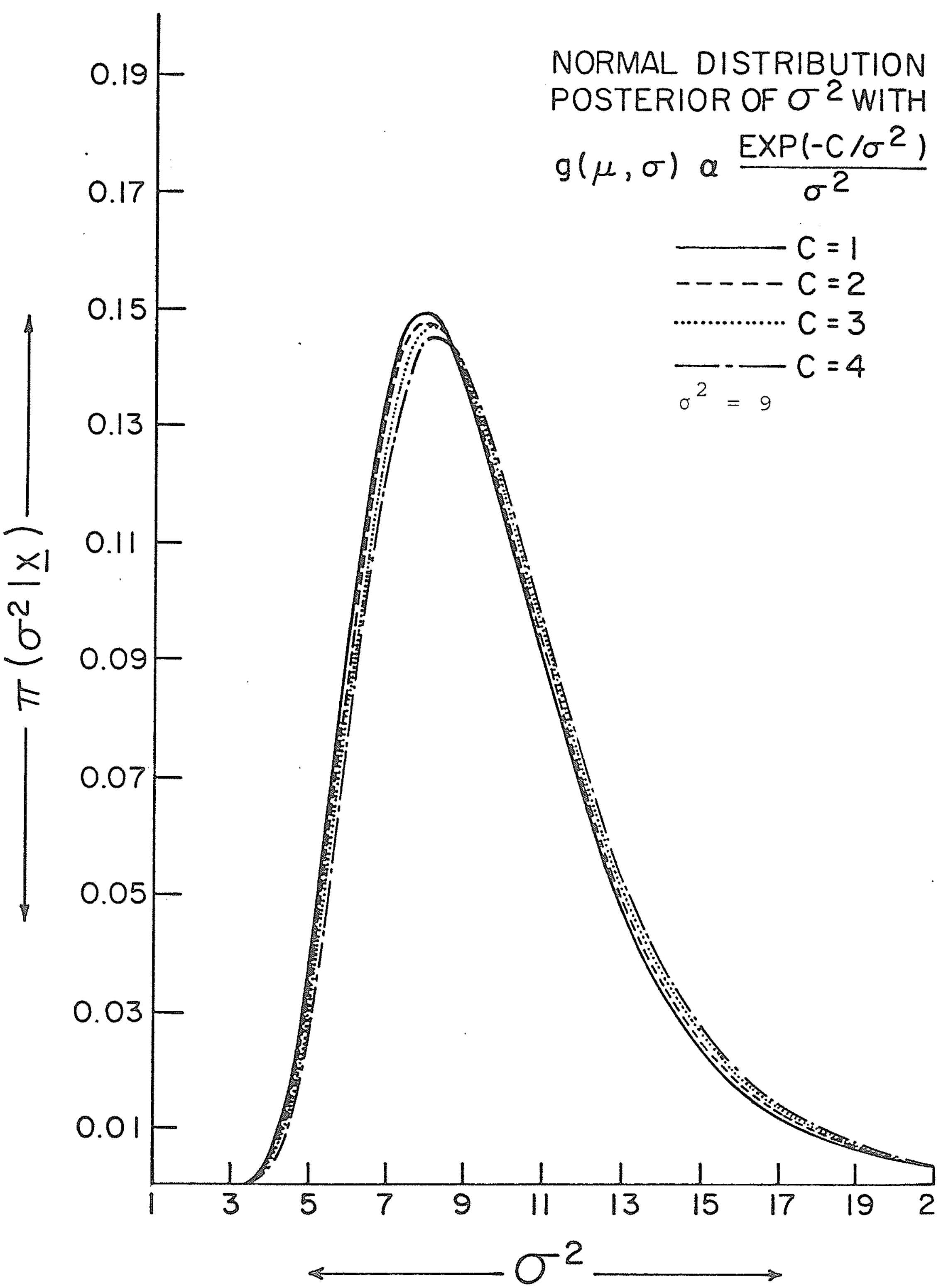
5.2 BAYES ESTIMATORS OF μ AND σ^2

In this section we will obtain the Bayes estimators of μ and σ^2 under the loss functions discussed in Section 2.3. We will also consider the zero-one loss (ZOL) function.

The ZOL function is given by

$$L(T, \theta | \underset{\sim}{x}) = \begin{cases} 0 & \text{if } |T-\theta| \le a \\ 1 & \text{if } |T-\theta| > a \end{cases}$$

and is appropriate when there is loss or no loss if the error does or does not exceed a certain constant a.



NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu, \sigma) \propto \frac{EXP(-C/\sigma^2)}{\sigma^2}$
C = 1
C = 2
C = 3
C = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2 | \underline{x})$
1
3
5
7
9
11
13
15
17
19
21
σ^2

FIGURE 1

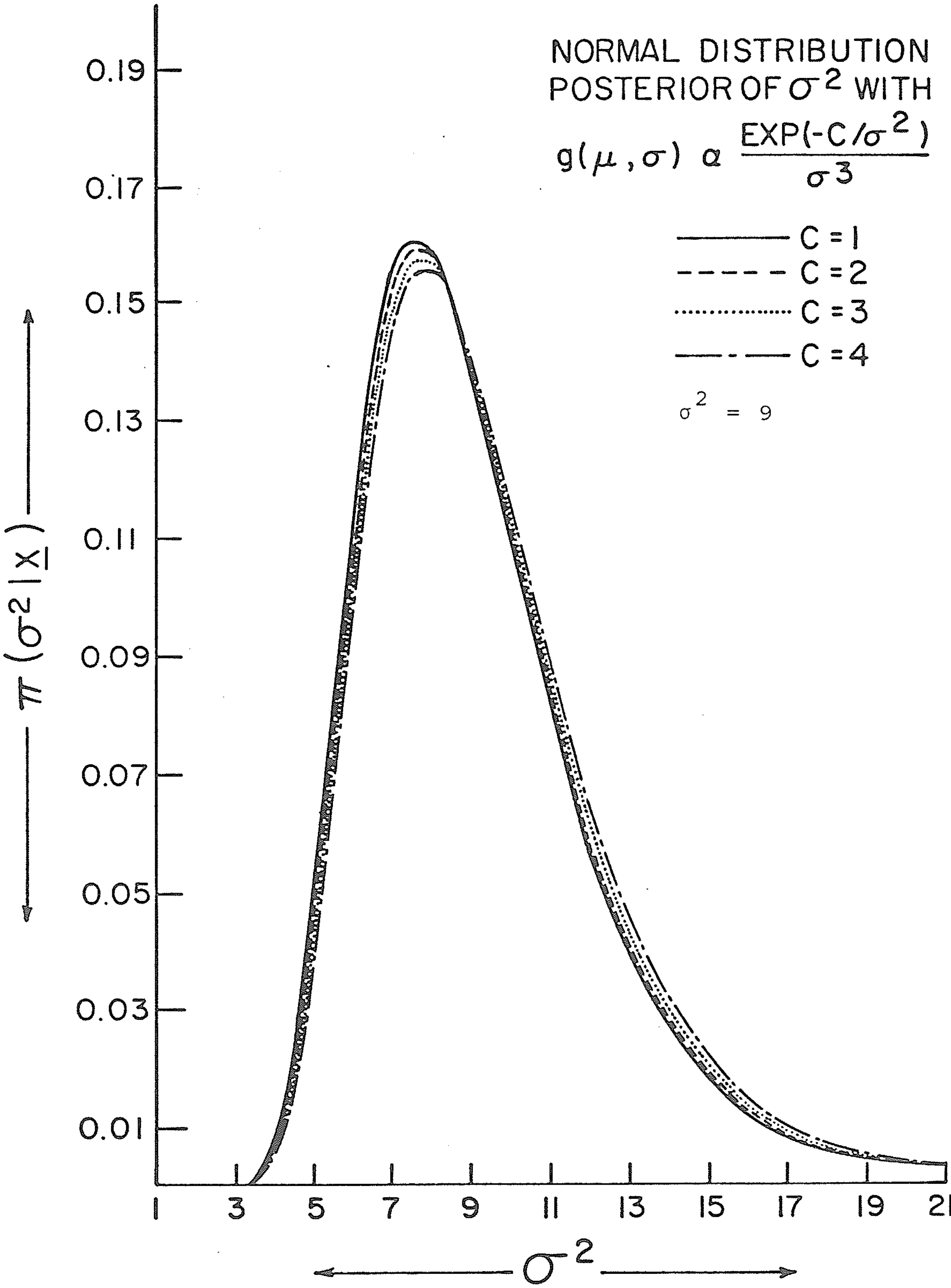

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu,\sigma) \propto \frac{EXP(-C/\sigma^2)}{\sigma^3}$
C = 1
C = 2
C = 3
C = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2 | \underline{X})$
1
3
5
7
9
11
13
15
17
19
21
σ^2

FIGURE 2

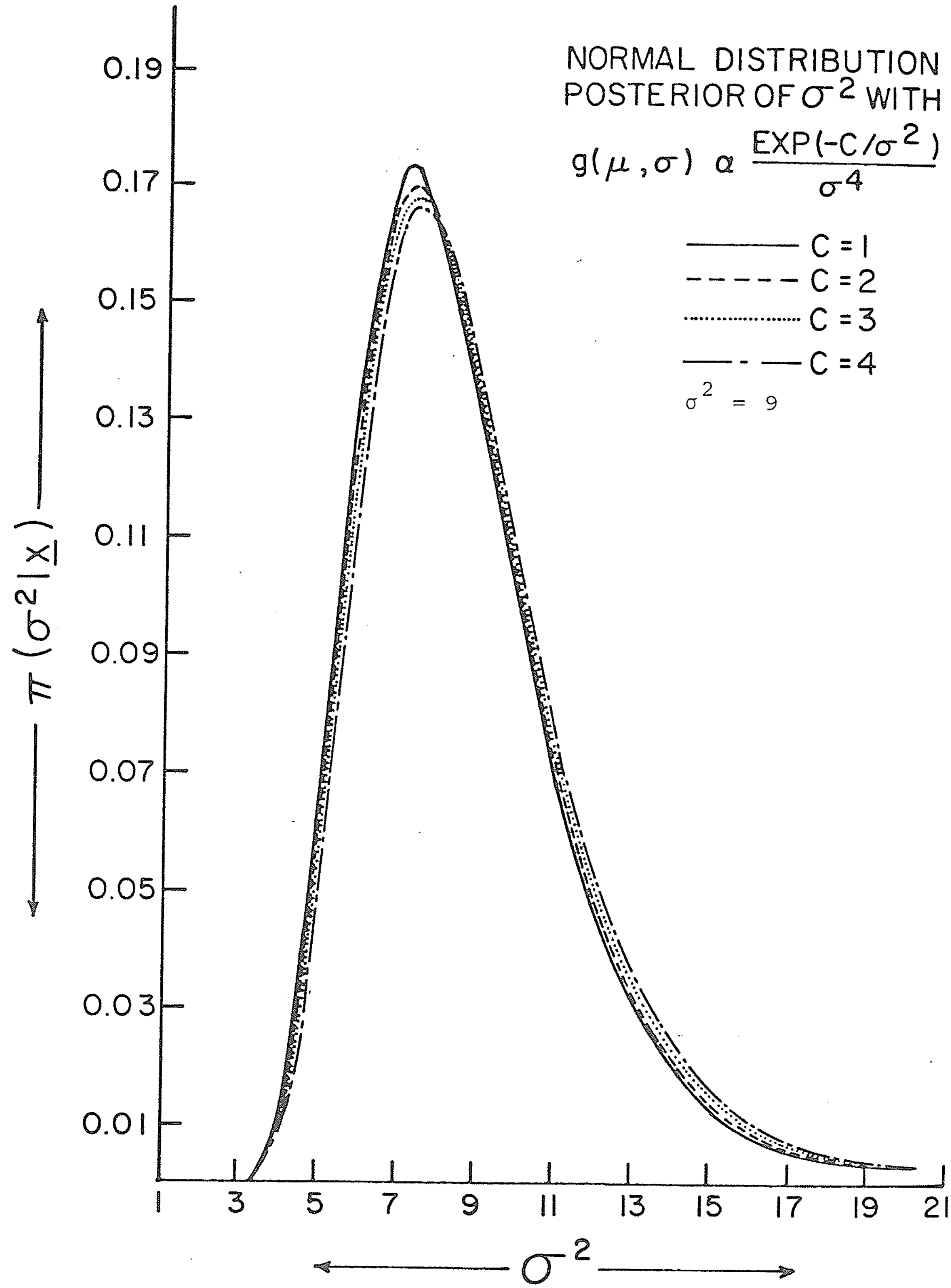

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu, \sigma) \; \alpha \; \frac{EXP(-C/\sigma^2)}{\sigma^4}$
C = 1
C = 2
C = 3
C = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2 | \underline{x})$
1 3 5 7 9 11 13 15 17 19 21
σ^2

FIGURE 3

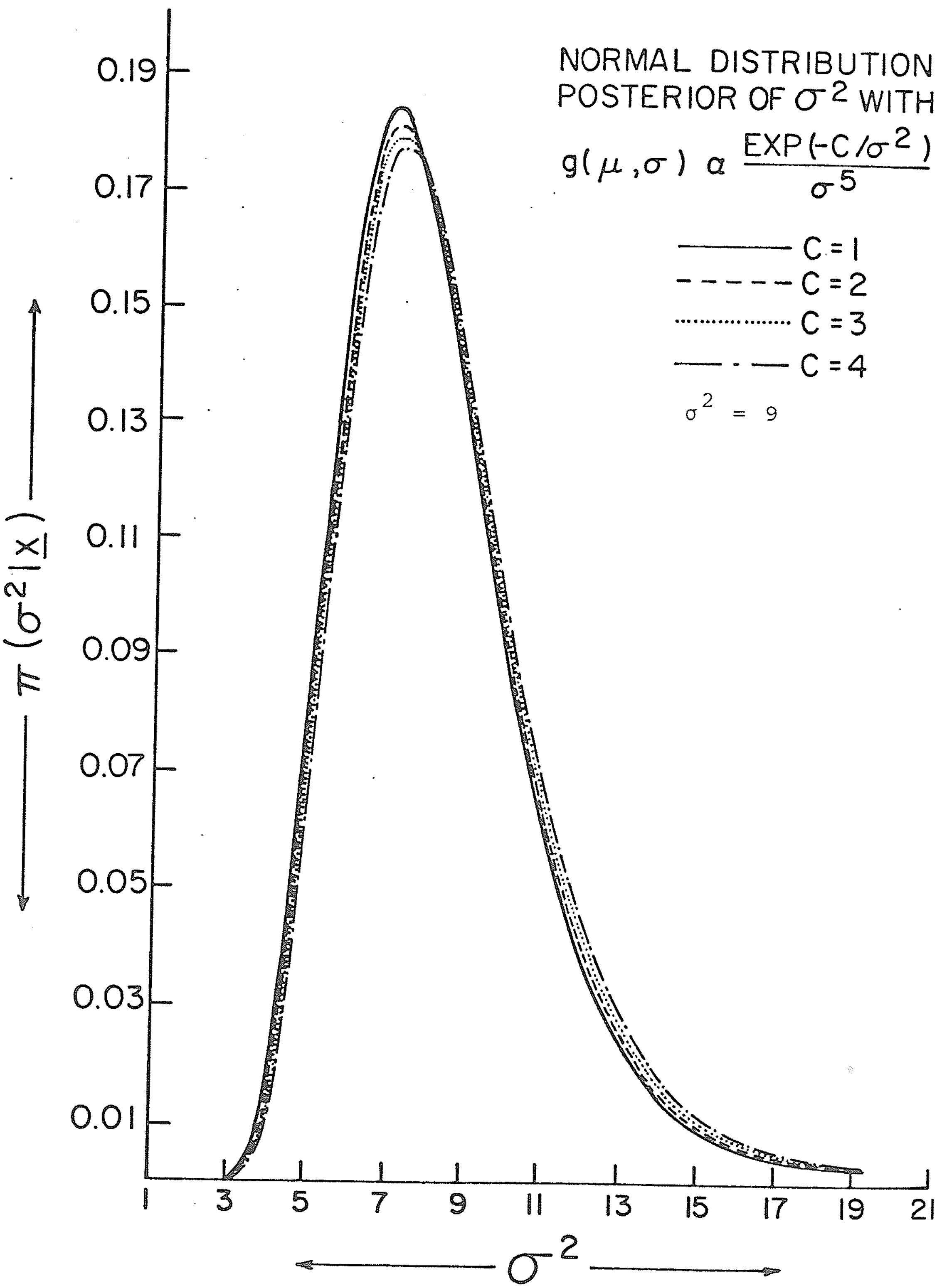

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu,\sigma)\ \alpha\ \frac{EXP(-C/\sigma^2)}{\sigma^5}$
C = 1
C = 2
C = 3
C = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2|\underline{x})$
1
3
5
7
9
11
13
15
17
19
21
σ^2

FIGURE 4

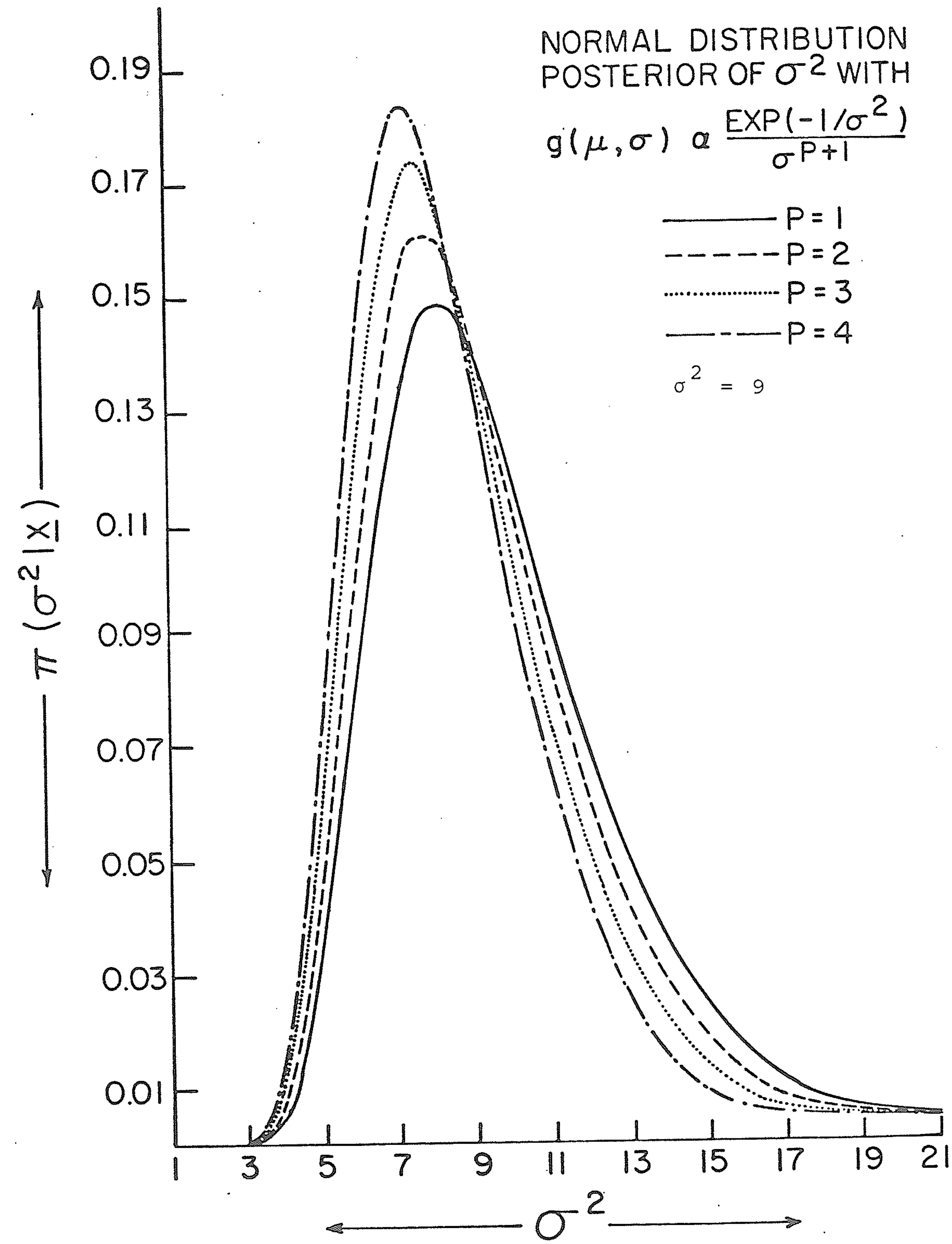

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu,\sigma) \; \alpha \; \frac{EXP(-1/\sigma^2)}{\sigma^{P+1}}$
P = 1
P = 2
P = 3
P = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2|\underline{X})$
1
3
5
7
9
11
13
15
17
19
21
σ^2

FIGURE 5

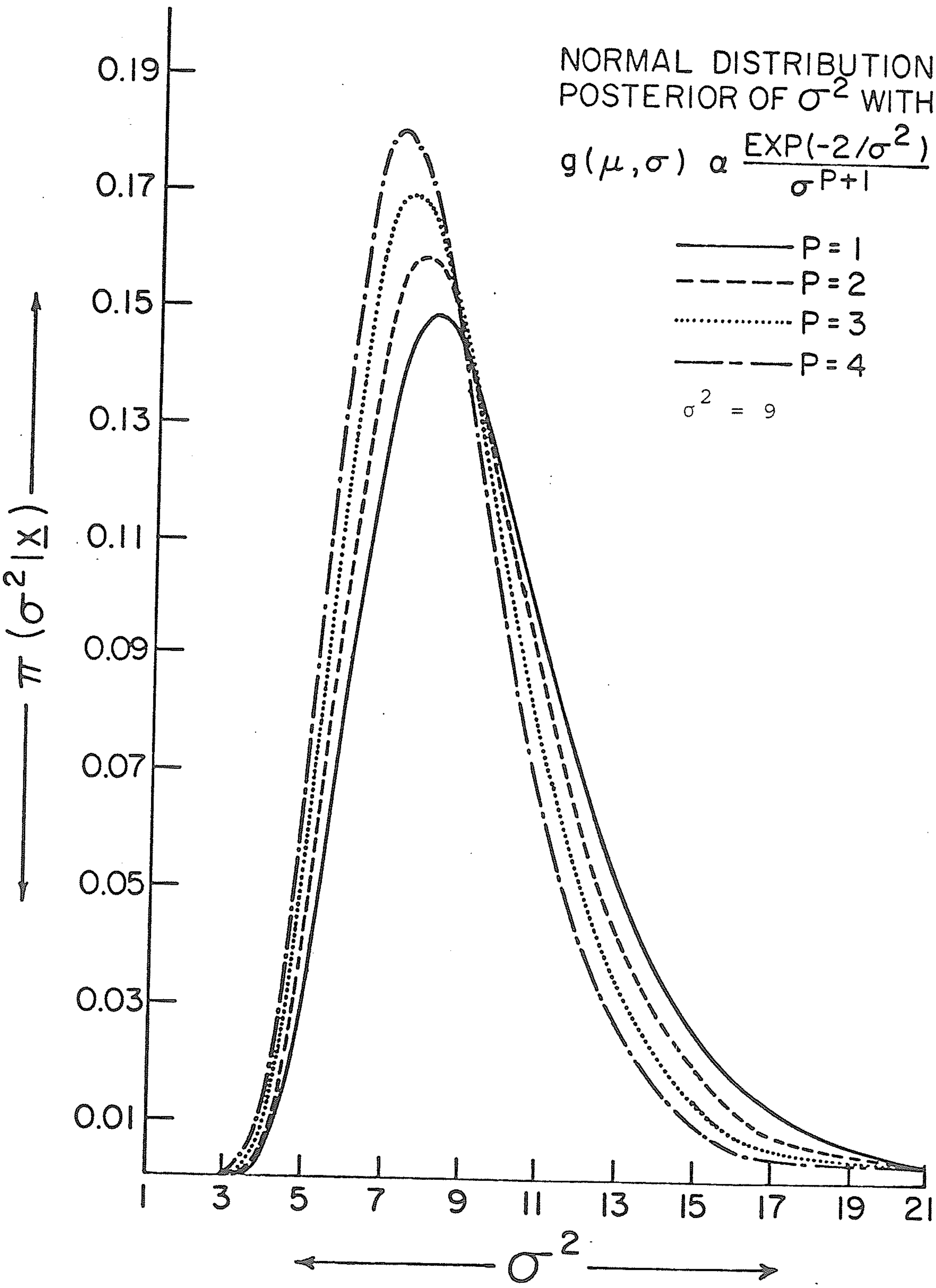

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu,\sigma)\ \alpha\ \frac{EXP(-2/\sigma^2)}{\sigma^{P+1}}$
P = 1
P = 2
P = 3
P = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2|\underline{X})$
1
3
5
7
9
11
13
15
17
19
21
σ^2

FIGURE 6

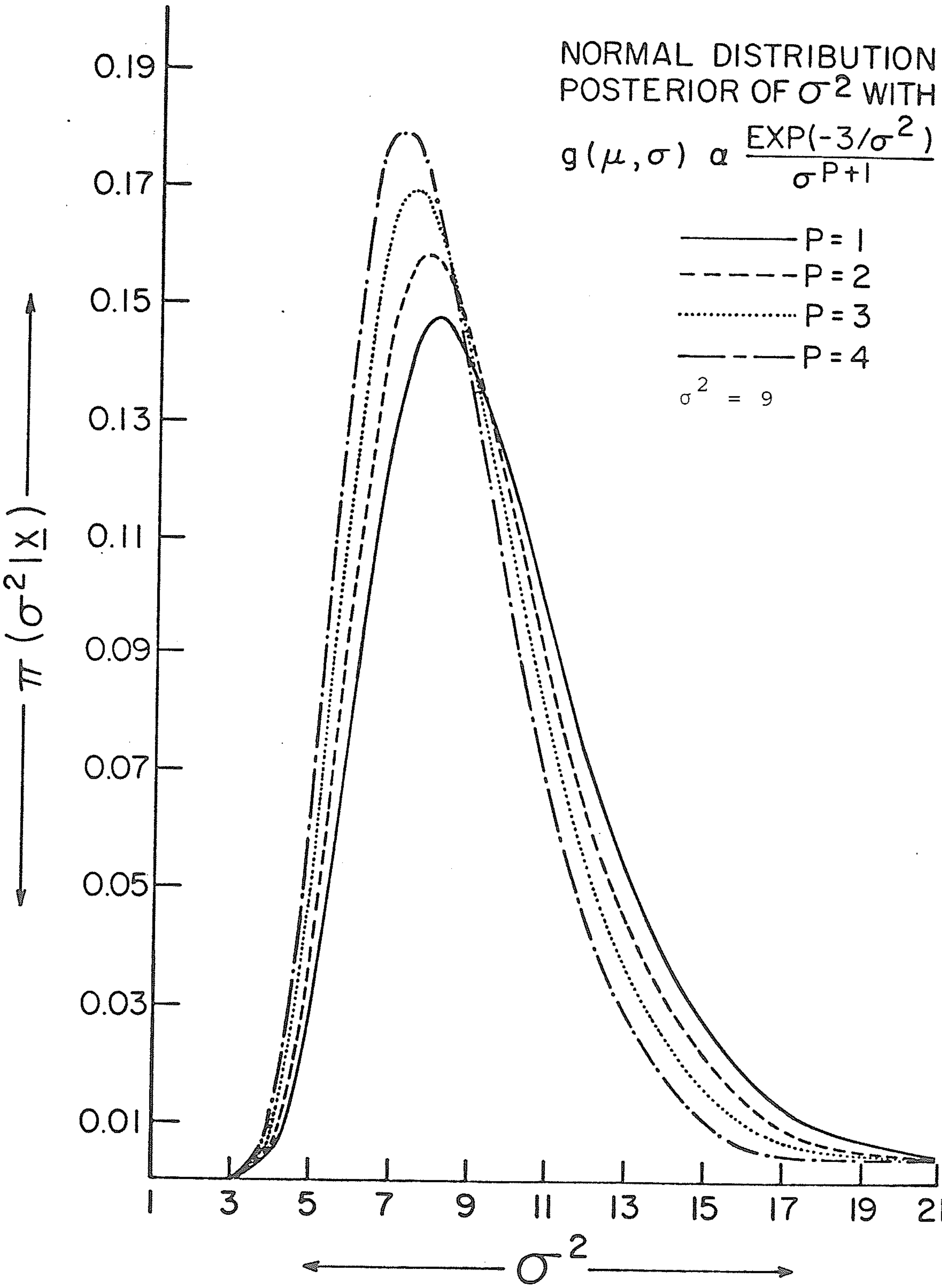

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu,\sigma)\ \alpha\ \frac{EXP(-3/\sigma^2)}{\sigma^{P+1}}$
P = 1
P = 2
P = 3
P = 4
$\sigma^2 = 9$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
$\pi(\sigma^2|\underline{x})$
1
3
5
7
9
11
13
15
17
19
21
σ^2

FIGURE 7

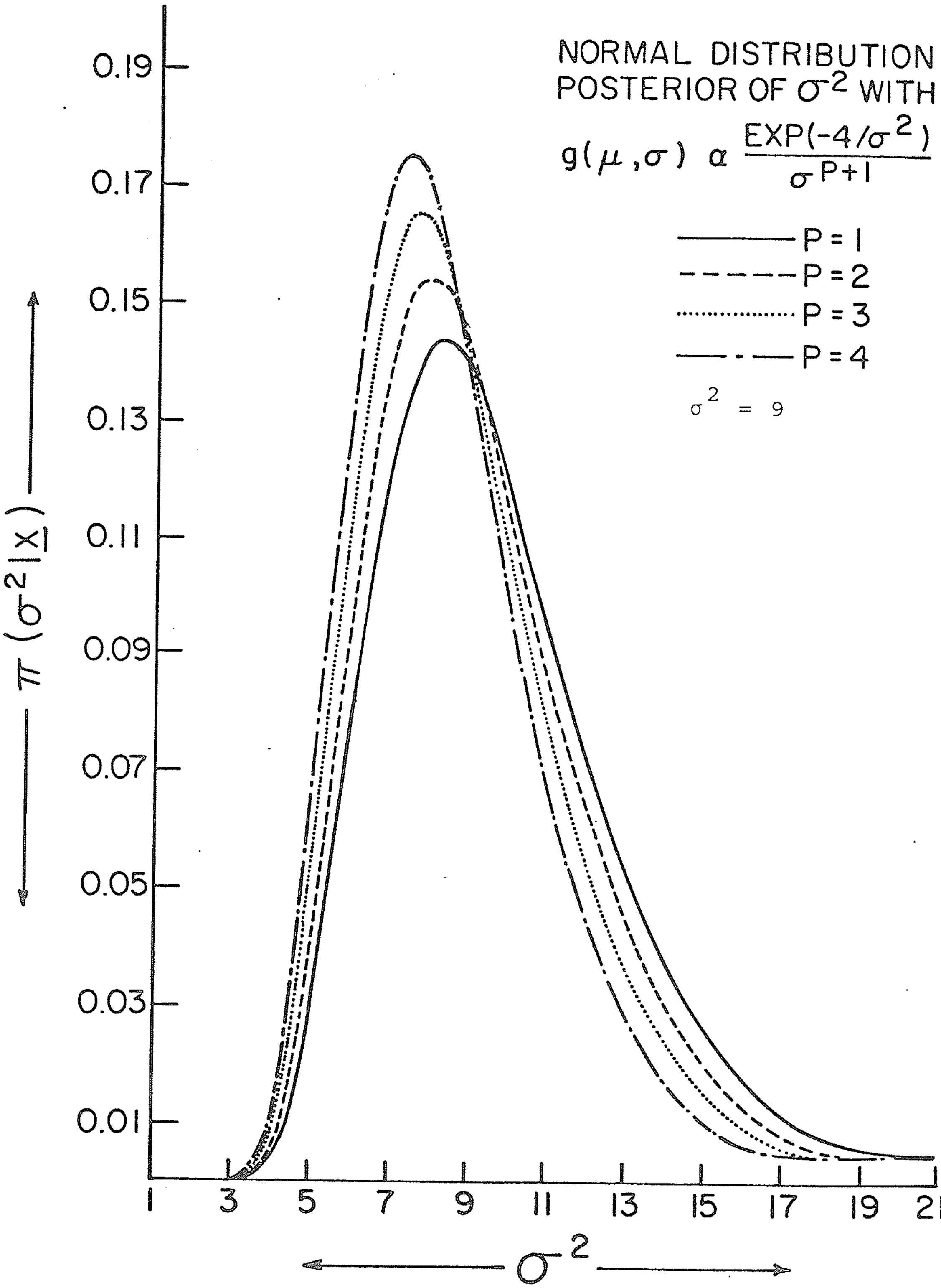

NORMAL DISTRIBUTION
POSTERIOR OF σ^2 WITH
$g(\mu,\sigma) \; \alpha \; \frac{EXP(-4/\sigma^2)}{\sigma^{P+1}}$
P = 1
P = 2
P = 3
P = 4
$\sigma^2 = 9$
$\pi(\sigma^2|\underline{x})$
0.19
0.17
0.15
0.13
0.11
0.09
0.07
0.05
0.03
0.01
1 3 5 7 9 11 13 15 17 19 21
σ^2

FIGURE 8

$$E[L(T,\theta)|\underset{\sim}{x}] = P[|T-\theta| > a|\underset{\sim}{x}]$$

$$= 1 - P(T-a \le \theta \le T+a)$$

$$= 1 - \int_{T-a}^{T+a} \Pi(\theta|\underset{\sim}{x})\,d\theta$$

which is minimized as

$$I = \int_{T-a}^{T+a} \Pi(\theta|\underset{\sim}{x})\,d\theta$$

is maximized. This implies that

$$\Pi(T+a|\underset{\sim}{x}) = \Pi(T-a|\underset{\sim}{x}).$$

If the posterior distribution is unimodal and symmetric, the point estimator T is the posterior mode under ZOL. It may be noted that T is the centre of the highest posterior density (HPD) interval of length 2a.

Bayes estimator of σ^2 under the SEL function (2.12)

$$\sigma^{*2} = E(\sigma^2|\underset{\sim}{x}) = \int_0^\infty \sigma^2 \Pi(\sigma^2|\underset{\sim}{x})\,d\sigma^2$$

$$= \frac{A+2c}{n+p-3}\,. \tag{5.8}$$

The posterior distribution of σ^2 as given by (5.5) is an inverted gamma which is unimodal but not symmetric. Under ZOL the estimator T must satisfy $\Pi(T+a|\underset{\sim}{x}) = \Pi(T-a|\underset{\sim}{x})$. Let

$$T - a = \theta_1$$

$$T + a = \theta_2$$

$$2a = \theta_2 - \theta_1, \qquad \theta_2 = 2a + \theta_1.$$

We want a θ such that $\Pi(\theta) = \Pi(2a+\theta)$, in that case $T = \theta+a$. We proceed to obtain T by the following iterative process:

(i) For known values of the prior parameters (c,p) start with $\theta = \theta_1 <$ mode $= \frac{A+2c}{n+p+1}$ (known, since for the given data $\underset{\sim}{x}$, n and $A = \sum_1^n (x-\bar{x})^2$ are known).

(ii) Compute $\Pi(\theta_1)$ and $\Pi(\theta_1+2a)$ for an arbitrary a. If $\Pi(\theta_1) < \Pi(\theta_1+2a)$, choose a $\theta_2 = \theta_1+\delta_1$ and if $\Pi(\theta_1) > \Pi(\theta_1+2a)$, choose a $\theta_2 = \theta_1-\delta_1$, $\delta_1 > 0$.

(iii) Compute $\Pi(\theta_2)$ and $\Pi(\theta_2+2a)$. If $\Pi(\theta_2) < \Pi(\theta_2+2a)$, choose a $\theta_3 = \theta_2+\delta_2$ and if $\Pi(\theta_2) > \Pi(\theta_2+2a)$, choose a $\theta_3 = \theta_2-\delta_2$, $\delta_2 > 0$.

(iv) Continue till we obtain a θ_k such that $\Pi(\theta_k) = \Pi(\theta_k+2a)$ and we have $T = \theta_k+a$. For a fairly symmetrical posterior, as the arbitrary value a decreases T tends to the posterior mode.

The Bayes estimator of σ^2 under LSEL (2.13) is

$$\log \sigma^{*2} = \int_0^\infty \log \sigma^2 \frac{\left(\frac{A+2c}{2}\right)^{\frac{n+p-1}{2}}}{\Gamma\left(\frac{n+p-1}{2}\right)} \frac{\exp\left\{-\frac{1}{2\sigma^2}(A+2c)\right\}}{(\sigma^2)^{\frac{n+p+1}{2}}} d\sigma^2 .$$

Substituting $\frac{A+2c}{2\sigma^2} = q$, we have

$$\log \sigma^{*2} = \int_0^\infty \frac{1}{\Gamma(\frac{n+p-1}{2})} [\log(\frac{A+2c}{2}) - \log q] e^{-q} q^{\frac{n+p-1}{2} - 1} dq$$

$$= \log(\frac{A+2c}{2}) - \frac{1}{\Gamma(m)} \int_0^\infty \log q e^{-q} q^{m-1} dq,$$

putting $m = \frac{n+p-1}{2}$,

$$= \log(\frac{A+2c}{2}) - \frac{d}{dm} \log \Gamma(m).$$

Thus

$$\sigma^{*2} = (\frac{A+2c}{2}) \exp\{- \frac{d}{dm} \log \Gamma(m)\}$$

$$= (\frac{A+2c}{2}) \exp\{-\psi(m)\},$$

where

$$\psi(m) = \frac{d}{dm} \log \Gamma(m);$$

the digamma function is extensively tabulated in Abramowitz and Stegun [1964].

From (2.22),

$$\exp\{\psi(m)\} = m \exp(- \frac{1}{2m}) \approx m-\frac{1}{2}.$$

Thus

$$\sigma^{*2} \approx \frac{A+2c}{2m-1} = \frac{A+2c}{n+p-2}$$

For the GLF, let us choose $\lambda(\theta) = \sigma^{2\alpha}$; then

$$\sigma^{*2} = [E(\sigma^{2(\alpha+\beta)} | \underset{\sim}{x}) / (E\sigma^{2\alpha} | \underset{\sim}{x})]^{1/\beta}$$

where

$$E(\sigma^{2(\alpha+\beta)}|\underset{\sim}{x}) = \int_0^\infty \sigma^{2(\alpha+\beta)}\,\Pi(\sigma^2|\underset{\sim}{x})\,d\sigma^2$$

$$= \left(\frac{A+2c}{2}\right)^{\alpha+\beta}\Gamma\left(\frac{n+p-2\alpha-2\beta-1}{2}\right)/\Gamma\left(\frac{n+p-1}{2}\right).$$

Thus,

$$\sigma^{*2} = \left(\frac{A+2c}{2}\right)\left\{\Gamma\left(\frac{n+p-2\alpha-2\beta-1}{2}\right)/\Gamma\left(\frac{n+p-2\alpha-1}{2}\right)\right\}^{1/\beta}.$$

Using asymptotic expansion of gamma functions, namely $\frac{\Gamma(x+a)}{\Gamma(x)} \sim x^a\left(1 + \frac{a^2-a}{2x}\right)$ (Erdelyi, 1953), the Bayes estimator of σ^2 under the GLF takes the form

$$\sigma^{*2} \approx \left(\frac{A+2c}{n+p-2\alpha-1}\right)\left\{1 + \frac{\beta(\beta+1)}{n+p-2\alpha-1}\right\}^{1/\beta}$$

$$= \frac{A+2c}{n+p-2\alpha-1}\left(1 + \frac{\beta + 1}{n+p-2\alpha-1}\right).$$

If we assume $c = 0$, $p = \nu-1$ in (5.2), the joint prior of μ and σ is $g(\mu,\sigma) \propto \frac{1}{\sigma^\nu}$ which is improper. The corresponding estimators turn out to be

	Estimator	Loss Function
$\sigma^{*2} =$	$\frac{A}{n+\nu-4}$	SEL
	An estimator T $\ni$ $\Pi(T+a) = \Pi(T-a)$	ZOL
	$\frac{A}{n+\nu-3}$ (asymptotically)	LSEL
	$\frac{A}{n+\nu-2\alpha-2}\left(1 + \frac{\beta + 1}{n+\nu-2\alpha-2}\right)$ (asymptotically)	GLF

With suitable choice of ν, σ^{*2} under SEL, and LSEL identify with well-known classical estimators.

For example,

$\nu = 4$ under SEL => MLE

$\nu = 3$ under LSEL => MLE

$\nu = 3$ under SEL => UMVUE

$\nu = 2$ under LSEL => UMVUE

$\nu = 5$ under SEL => min MSE

$\nu = 4$ under LSEL => min MSE

Bayes estimator of μ under SEL function, using (5.6), is

$$\mu^* = \frac{1}{\sqrt{\lambda}\, B(\frac{1}{2}, \frac{m}{2})} \int_{-\infty}^{\infty} \frac{\mu}{\left\{1 + \frac{(\mu-\bar{x})^2}{\lambda}\right\}^{\frac{m+1}{2}}} d\mu .$$

Let

$$\frac{\mu-\bar{x}}{\sqrt{\lambda}} = \frac{t}{\sqrt{m}}, \qquad d\mu = \sqrt{\frac{\lambda}{m}}\, dt$$

$$\mu^* = \frac{1}{B(\frac{1}{2}, \frac{m}{2})\sqrt{m}} \int_{-\infty}^{\infty} \frac{\left[\bar{x} + \sqrt{\frac{\lambda}{m}}\, t\right]}{\left(1 + \frac{t^2}{m}\right)^{\frac{m+1}{2}}} dt$$

$$= \bar{x},$$

the second integral vanishes, since $E(t) = 0$. Thus, the Bayes estimator of μ is independent of the prior parameters c and p, which, of course, is expected since the priors for μ and σ are independent.

The Bayes estimator of μ under ZOL is also $\bar{x}$.

We now apply the loss function GLF (2.14),

$$L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^{2}$$

where μ^{α} is a positive weight function of μ and $T(\cdot)$, an estimate of μ.

The Bayesian estimate for μ is the solution of the equation

$$\frac{\partial}{\partial T} \int L(T,\mu)\Pi(\mu|\underset{\sim}{x})d\mu = 0,$$

giving

$$T = [E(\mu^{\alpha+\beta}|\underset{\sim}{x})/E(\mu^{\alpha}|\underset{\sim}{x})]^{1/\beta}$$

where

$$E(\mu^{s}|\underset{\sim}{x}) = \frac{1}{B(\frac{1}{2}, \frac{m}{2})\sqrt{\lambda}} \int_{-\infty}^{\infty} \frac{\mu^{s}}{\left[1 + \frac{(\mu-\bar{x})^{2}}{\lambda}\right]^{\frac{m+1}{2}}} d\mu .$$

Let

$$\frac{\mu-\bar{x}}{\sqrt{\lambda}} = \frac{t}{\sqrt{m}}, \qquad d\mu = \sqrt{\frac{\lambda}{m}}\, dt$$

$$E(\mu^{s}|\underset{\sim}{x}) = \frac{1}{B(\frac{1}{2}, \frac{m}{2})\sqrt{\lambda}} \int_{-\infty}^{\infty} \frac{\left(\bar{x} + t\sqrt{\frac{\lambda}{m}}\right)^{s}}{\left(1 + \frac{t^{2}}{\lambda}\right)^{\frac{m+1}{2}}} dt$$

$$= \frac{1}{B(\frac{1}{2}, \frac{m}{2})\sqrt{\lambda}} \sum_{k=0,2,4,\ldots}^{s} \binom{s}{k} \bar{x}^{s-k} \left(\frac{\lambda}{m}\right)^{\frac{k}{2}} \int_{-\infty}^{\infty} \frac{(t^{2})^{\frac{k-1}{2}}}{\left(1+ \frac{t^{2}}{m}\right)^{\frac{m+1}{2}}} dt^{2}$$

$$= \frac{1}{B(\frac{1}{2}, \frac{m}{2})} \sum_{k=0}^{s} \binom{s}{k} \lambda^{\frac{k}{2}} \bar{x}^{s-k} \int_0^{\infty} \frac{(\frac{t^2}{m})^{\frac{k+1}{2}-1} d(\frac{t^2}{m})}{\left(1 + \frac{t^2}{m}\right)^{\frac{m+1}{2}}}$$

$$= \frac{1}{B(\frac{1}{2}, \frac{m}{2})} \sum_{k=0}^{s} \binom{s}{k} \lambda^{\frac{k}{2}} \bar{x}^{s-k} B(\frac{k+1}{2}, \frac{m-k}{2}) . \quad (5.9)$$

Putting s = 1, we have

$$E(\mu | \underset{\sim}{x}) = \frac{1}{B(\frac{1}{2}, \frac{m}{2})} \sum_{k=0}^{1} \binom{1}{k} \lambda^{\frac{k}{2}} \bar{x}^{1-k} B(\frac{k+1}{2}, \frac{m-k}{2})$$

$$= \bar{x}$$

which has been obtained earlier.

Put $s = \alpha$ and $s = \alpha+\beta$ in (5.9) to get $E(\mu^{\alpha} | \underset{\sim}{x})$ and $E(\mu^{\alpha+\beta} | \underset{\sim}{x})$, respectively.

5.3 MONTE CARLO STUDY

We generated 5000 samples each of size 20 from a normal distribution with $\mu = 20$ and $\sigma^2 = 9$ to compare the performance of the various estimators. The mean and mse of 5000 estimators under various loss functions are computed and tabulated in Tables 5.1 to 5.8. The entries in the parentheses represent the mse's.

TABLE 5.1

Bayes Estimates of σ^2 under Squared Error Loss Function with n = 20, N = 5000

c \ p	1	2	3
1	9.616674 (9.933476)	9.110434 (8.584599)	8.654841 (7.854102)
2	9.727773 (10.083230)	9.215700 (8.619310)	8.754810 (7.795399)
3	9.838876 (10.257680)	9.320970 (8.676115)	8.854854 (7.756692)

TABLE 5.2

Bayes Estimates of σ^2 under $L(T,\sigma^2) = \sigma^{2\alpha}(T^{\beta}-\sigma^{2\beta})^2$ with n = 20, N = 5000, α = 1, β = 1

c \ p	1	2	3
1	10.822353 (15.411305)	10.185744 (12.115766)	9.619869 (9.937095)
2	10.947353 (15.882518)	10.303391 (12.408605)	9.730980 (10.087190)
3	11.072353 (16.384981)	10.421038 (12.729126)	9.842091 (10.261975)

TABLE 5.3

Bayes Estimates of σ^2 under $L(T,\sigma^2) = \sigma^{2\alpha}(T^{\beta}-\sigma^{2\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 1$, $\beta = 2$

c \ p	1	2	3
1	11.569582 (20.420276)	10.843552 (15.536433)	10.203412 (12.195165)
2	11.703212 (21.124883)	10.968797 (16.013910)	10.321263 (12.492701)
3	11.836843 (21.865204)	11.094042 (16.522760)	10.439114 (12.818014)

TABLE 5.4

Bayes Estimates of σ^2 under $L(T,\sigma^2) = \sigma^{2\alpha}(T\ -\sigma^{2\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 1$, $\beta = 3$

c \ p	1	2	3
1	12.453705 (27.938109)	11.613073 (20.749755)	10.879314 (15.749758)
2	12.597548 (28.952378)	11.747206 (21.468745)	11.004972 (16.237850)
3	12.741390 (30.008028)	11.881339 (22.223718)	11.130629 (16.757520)

TABLE 5.5

Bayes Estimates of σ^2 under $L(T,\sigma^2) = \sigma^{2\alpha}(T^\beta - \sigma^{2\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 2$, $\beta = 1$

c \ p	1	2	3
1	12.368403 (27.137598)	11.543843 (20.227251)	10.822353 (15.411305)
2	12.511260 (28.120407)	11.677176 (20.923387)	10.947353 (15.882518)
3	12.654117 (29.144032)	11.810510 (21.655079)	11.072353 (16.384981)

TABLE 5.6

Bayes Estimates of σ^2 under $L(T,\sigma^2) = \sigma^{2\alpha}(T^\beta - \sigma^{2\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 2$, $\beta = 2$

c \ p	1	2	3
1	13.359402 (37.427756)	12.400077 (27.432960)	11.569582 (20.420276)
2	13.513706 (38.796906)	12.543299 (28.427411)	11.703212 (21.124883)
3	13.668009 (40.213676)	12.686522 (29.462887)	11.836843 (21.865204)

TABLE 5.7

Bayes Estimates of σ^2 under $L(T,\sigma^2) = \sigma^{2\alpha}(T^{\beta}-\sigma^{2\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 3$, $\beta = 1$

c \ p	1	2	3
1	14.429804 (50.976697)	13.319819 (36.975188)	12.368403 (27.137598)
2	14.596470 (52.814410)	13.473665 (38.328032)	12.511260 (28.120407)
3	14.763137 (54.707677)	13.627511 (39.728213)	12.654117 (29.144032)

TABLE 5.8

Bayes Estimates of σ^2 under $L(T,\sigma^2) = (\log T - \log \sigma^2)^2$

with $n = 20$, $N = 5000$

c \ p	1	2	3
1	9.110434 (8.584599)	8.654841 (7.854102)	8.242619 (7.587810)
2	9.215700 (8.619310)	8.754810 (7.795399)	8.337867 (7.452951)
3	9.320970 (8.676115)	8.854854 (7.756692)	8.433069 (7.336213)

The Bayes estimators of μ are all identical under the three standard loss functions, namely, $\bar{x}$ which on simulated data of 5000 samples each of size 20 is 20.00507 with a common mse of 0.445827, while the variance of the mle of μ is $\sigma^2/n = 0.45$. It may be noted here that the mle and the Bayes estimator of μ under the standard loss functions are the same.

Table 5.9 shows the Bayes estimates of σ^2 under ZOL and Table 5.10 gives the modal estimates of σ^2.

TABLE 5.9

Solution of $\Pi(T-a) = \Pi(T+a)$, $a = 2$, with $n = 20$, $N = 5000$

c \ p	1	2	3
1	8.769383 (6.441417)	8.427862 (6.166241)	8.114651 (6.142349)
2	8.860092 (6.409856)	8.514219 (6.080545)	8.197585 (6.005903)
3	8.951001 (6.392682)	8.601175 (6.003622)	8.280318 (5.885377)

TABLE 5.10

Modal Estimator $\sigma^{*2} = \dfrac{A+2c}{n+p+1}$, with $n = 20$, $N = 5000$

c \ p	1	2	3
1	8.655662 (7.856342)	8.243487 (7.590700)	7.868783 (7.674505)
2	8.755662 (7.797474)	8.338725 (7.455673)	7.959692 (7.477094)
3	8.855662 (7.758607)	8.433963 (7.338786)	8.050601 (7.296212)

Tables 5.11 to 5.16 show the Bayes estimates of μ under the GLF.

TABLE 5.11

Bayes Estimates of μ under $L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 1$, $\beta = 1$

c \ p	1	2	3
1	20.041060 (0.446579)	20.039794 (0.446513)	20.038655 (0.446457)
2	20.041338 (0.446589)	20.040058 (0.446522)	20.038906 (0.446465)
3	20.041616 (0.446600)	20.040321 (0.446532)	20.039156 (0.446474)

TABLE 5.12

Bayes Estimates of μ under $L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 1$, $\beta = 2$

c \ p	1	2	3
1	20.053049 (0.447380)	20.051154 (0.447232)	20.049449 (0.447106)
2	20.053465 (0.447405)	20.051548 (0.447255)	20.049823 (0.447127)
3	20.053881 (0.447432)	20.051943 (0.447278)	20.050198 (0.447147)

TABLE 5.13

Bayes Estimates of μ under $L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 1$, $\beta = 3$

c \ p	1	2	3
1	20.065015 (0.448496)	20.062492 (0.448235)	20.060222 (0.448011)
2	20.065569 (0.448544)	20.063017 (0.448277)	20.060720 (0.448049)
3	20.066122 (0.448592)	20.063542 (0.448320)	20.061219 (0.448088)

TABLE 5.14

Bayes Estimates of μ under $L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 2$, $\beta = 1$

c \ p	1	2	3
1	20.065046 (0.448498)	20.062521 (0.448236)	20.060248 (0.448012)
2	20.065600 (0.448546)	20.063046 (0.448279)	20.060747 (0.448051)
3	20.066155 (0.448594)	20.063572 (0.448322)	20.061246 (0.448089)

TABLE 5.15

Bayes Estimates of μ under $L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 2$, $\beta = 2$

c \ p	1	2	3
1	20.077009 (0.449931)	20.073857 (0.449523)	20.071020 (0.449175)
2	20.077701 (0.450008)	20.074513 (0.449592)	20.071643 (0.449236)
3	20.078393 (0.450085)	20.075169 (0.449661)	20.072266 (0.449298)

TABLE 5.16

Bayes Estimates of μ under $L(T,\mu) = \mu^{\alpha}(T^{\beta}-\mu^{\beta})^2$

with $n = 20$, $N = 5000$, $\alpha = 3$, $\beta = 1$

c \ p	1	2	3
1	20.088980 (0.451682)	20.085201 (0.451095)	20.081798 (0.450594)
2	20.089810 (0.451793)	20.085987 (0.451195)	20.082545 (0.450683)
3	20.090640 (0.451906)	20.086773 (0.451296)	20.083293 (0.450774)

5.4 CONCLUDING REMARKS

We make the following observations from this study:

(1) From Tables 5.1 and 5.2 to 5.7, it follows that within the class of loss functions

$$L(T,\theta) = \theta^{\alpha}(T^{\beta} - \theta^{\beta})^{2},$$

the squared error loss function ($\alpha = 0$, $\beta = 1$) yields the minimum mse estimates of σ^2 for all combinations of (c,p). Further, from Table 5.1, the mse's are fairly robust for varying c and a fixed p but less so for varying p and a fixed c, which supports the observation we made on studying the graphs of $\Pi(\sigma^2|\underset{\sim}{x})$.

(2) From Table 5.8, we observe that the logarithmic squared error loss function

$$L(T,\theta) = (\log T - \log \theta)^2$$

results in the minimum mse estimates for all combinations of (c,p) compared to the loss functions in (1) above.

(3) The zero-one loss function

$$\begin{aligned} L(T,\theta) &= 0 \quad \text{if } |T-\theta| \le a \\ &= 1 \quad \text{if } |T-\theta| > a \end{aligned}$$

with arbitrary a, the estimate T depends on the choice of a. With a = 2 (Table 5.9), we find that the ZOL estimates have uniformly smaller mse for all combinations of (c,p) than the modal estimates

(Table 5.10) or those obtained under the loss functions in (1) and (2) above. The same pattern was observed for $a = .5$, 1.0, and 1.5.

(4) The Bayes estimates of μ are quite robust for the choice of the various loss functions we have considered while those of σ^2 are less so (Tables 5.11 to 5.16).

(5) For large n, the posterior distributions of σ^2 and μ tend to normality (Section 5.1).

CHAPTER SIX

ESTIMATION OF REGRESSION PARAMETERS — INDEPENDENT AND CORRELATED CASES

6.0 INTRODUCTION

In this chapter we will consider the relation

$$E(Y|x) = \alpha + \beta(x - \bar{x}) \tag{6.1}$$

studying the dependence of the random variable Y on the non-random variable X. The variation in Y is expected to be at least partly explained by X, the other part is due to unobservable random errors assumed to be distributed as $N(0,\sigma^2)$.

Tiao and Zellner (1964) considered the Bayesian analysis of the general regression model

$$Y = M\beta + \epsilon,$$

where Y is a $(q\times 1)$ vector of observations, M is a $(q\times p)$ matrix of fixed elements with rank p and $\beta' = (\beta_1,\beta_2,\ldots,\beta_p)$. The main objective of their work was to develop procedures for using information from one sample as prior knowledge in the analysis of the next independent sample. These two samples were assumed to be independently drawn from normal populations with unequal variances. Theil (1963) considered the case within the sampling theory framework, using prior beliefs about regression coefficients when the error variance is unknown — requiring a large sample solution of the problem.

Some discussion of the Bayesian analyses of the simple linear regression model using uniform priors can be seen in Zellner (1971), Jeffreys (1961), and Lindley (1965). Maddala (1977) considers the Bayesian analysis of the regression model (at $\alpha = 0$)

$$Y_i = \beta x_i + \epsilon_i$$

under the natural conjugate priors when ϵ_i are independently and identically distributed normal random variables with 0 mean and variance σ^2. He considers the cases of σ known and σ unknown.

In this dissertation, we will study the posteriors of $(\alpha, \beta, \sigma^2)$ using the model (6.1), known as the orthogonal model, under the following situations:

I. When little is known about α, β and σ, Jeffreys (1961) suggested assuming the prior distribution of α, β and $\log \sigma$ to be locally uniform and independent.

The prior independence of the parameters may be justified as follows: Since the parameter β measures the dependence of Y on X, in the absence of any information about this dependence, it is not unwise to assume the prior distribution of β to be constant over the effective range of the likelihood. Again, σ^2 is the variance and α is the expectation of Y averaged over the values of X, i.e.,

$$\alpha = \frac{1}{n} \sum_{1}^{n} E(y_i | X).$$

The knowledge of the expectation and/or of the variance of Y will not change the knowledge of the dependence of Y on X which explains the prior independence of α and σ^2. In this study we will consider the prior distribution of σ as the inverted gamma density which is also the NCP for σ

$$g(\sigma) \propto \frac{\exp(-c/\sigma^2)}{\sigma^{p+1}}, \quad c,\ p,\ \sigma > 0$$

and treat the uniform prior for σ as a special case when $c = p = 0$. The joint prior density of (α, β, σ) may be written as

$$g(\alpha, \beta, \sigma) \propto \frac{\exp(-c/\sigma^2)}{\sigma^{p+1}} \tag{6.2}$$

where

$$p_1(\alpha) \propto \text{constant}$$

and

$$p_2(\beta) \propto \text{constant}.$$

II. When α and β may not be assumed independent, we will consider a bivariate normal prior for (α, β)

and without any loss of generality, we take

$E(\alpha) = 0$ and $E(\beta) = 0$ and

$Var(\alpha) = \tau_\alpha$, $Var(\beta) = \tau_\beta$

and a coefficient of correlation, ρ between α and β, σ^2 being assumed known.

III. Finally, we will consider a special case of II, where along with $\mu_\alpha = \mu_\beta = 0$, we further assume $\tau_\alpha = \tau_\beta = \tau$.

6.1 BAYES ESTIMATORS UNDER I

Let $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ and $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)$. The likelihood of $\underset{\sim}{y}$ given $\underset{\sim}{x}$ is

$$\ell(\underset{\sim}{y}, \underset{\sim}{x} | \alpha, \beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{- \sum_{i=1}^{n} [y_i - \alpha - \beta(x_i - x)]\right\}^2 / 2\sigma^2$$

$$\propto (\sigma^2)^{-n/2} \exp\{-[S^2 + n(\alpha - a)^2 + S_{xx}(\beta - b)^2]\}, \tag{6.3}$$

where

$$a = \bar{y}, \quad b = S_{xy}/S_{xx}, \quad S^2 = S_{yy} - S_{xy}^2/S_{xx} = \text{Residual S.S.}$$

Using Bayes theorem, the likelihood function (6.3) is combined with the prior density (6.2) to yield a joint posterior density, $\pi(\alpha, \beta, \sigma^2 | \underset{\sim}{x}, \underset{\sim}{y})$ for the parameters α, β and σ^2, i.e.,

$$\pi(\alpha, \beta, \sigma^2 | \underset{\sim}{\ }, \underset{\sim}{\ }) \propto (\sigma^2)^{-(\frac{n+p}{2}+1)} e^{-c/\sigma^2} \exp\left[-\frac{S^2}{2\sigma^2}\right]$$

$$\cdot \exp\left\{-\frac{n(\alpha - a)^2}{2\sigma^2}\right\} \exp\left\{-\frac{S_{xx}(\beta - b)^2}{2\sigma^2}\right\}. \tag{6.4}$$

Integrating out α from (6.4), we have

$$\pi(\beta,\sigma^2|\underset{\sim}{x},\underset{\sim}{y}) \propto (\sigma^2)^{-(\frac{n+p+1}{2})} e^{-c/\sigma^2} \exp\left\{-\frac{s^2}{2\sigma^2}\right\}$$

$$\cdot\exp\left\{-\frac{S_{xx}(\beta-b)^2}{2\sigma^2}\right\}. \qquad (6.5)$$

Using the fact that $\int_{-\infty}^{\infty} \exp\left\{-\frac{n(\alpha-a)^2}{2\sigma^2}\right\}d\alpha = \sqrt{2\pi\frac{\sigma^2}{n}}$.

Integrating out σ^2 from (6.5), we have the marginal posterior of β as

$$\pi_1(\beta|\underset{\sim}{X},\underset{\sim}{Y}) \propto \int_0^{\infty} (\sigma^2)^{-(\frac{n+p+1}{2})} \exp\left\{-\frac{1}{2\sigma^2}[2c + s^2 + S_{xx}(\beta-b)^2]\right\}d\sigma^2.$$

Substituting

$$\frac{2c + s^2 + S_{xx}(\beta-b)^2}{2\sigma^2} = \quad ,$$

$$d\sigma^2 = -\frac{2c + s^2 + S_{xx}(\beta-b)^2}{2y^2}\, dy$$

$$\pi_1(\beta|\underset{\sim}{x},\underset{\sim}{y}) \propto \int_0^{\infty} e^{-y} \left\{\frac{2y}{2c+s^2+S_{xx}(\beta-b)^2}\right\}^{\frac{n+p+1}{2}}$$

$$\cdot\frac{2c+s^2+S_{xx}(\beta-b)^2}{2y^2}\, dy$$

$$\propto \{2c + s^2 + S_{xx}(\beta-b)^2\}^{-\frac{1}{2}(n+p+1)}.$$

Thus,

$$\pi_1(\beta|\underset{\sim}{x},\underset{\sim}{y}) = \frac{\left\{1 + \frac{S_{xx}}{2c+S^2}(\beta-b)^2\right\}^{-\frac{1}{2}(n+p-1)}}{\int_{-\infty}^{\infty}\left\{1 + \frac{S_{xx}}{2c+S^2}(\beta-b)^2\right\}^{-\frac{1}{2}(n+p-1)} d\beta} .$$

To evaluate the denominator, let

$$\sqrt{\frac{S_{xx}}{2c+S^2}}\,(\beta-b) = \tan\,\theta$$

$$\beta = \sqrt{\frac{2c+S^2}{S_{xx}}}\,\tan\,\theta + b$$

$$d\beta = \sqrt{\frac{2c+S^2}{S_{xx}}}\,\mathrm{Sec}^2\theta d\theta,$$

giving

$$\int_{-\pi/2}^{\pi/2} \mathrm{Sec}\theta^{-(n+p-1)}\sqrt{\frac{2c+S^2}{S_{xx}}}\,\mathrm{Sec}^2\theta d\theta$$

$$= \sqrt{\frac{2c+S^2}{S_{xx}}}\int_{-\pi/2}^{\pi/2} \mathrm{Cos}^2\theta^{\left(\frac{n+p-1}{2}\right)-2} d\theta$$

$$= \sqrt{\frac{2c+S^2}{S_{xx}}}\int_{-\pi/2}^{\pi/2} \mathrm{Sin}\,\theta^{2(\frac{1}{2})-1}\mathrm{Cos}^2\theta^{\left(\frac{n+p-1}{2}\right)-1} d\theta$$

$$= \sqrt{\frac{2c+S^2}{S_{xx}}}\,B\left(\frac{1}{2}\,,\,\frac{n+p-2}{2}\right).$$

Thus,

$$\pi_1(\beta|\underset{\sim}{x},\underset{\sim}{y}) = \sqrt{\frac{S_{xx}}{2c+S^2}}\,\frac{1}{B(\frac{1}{2},\frac{n+p-2}{2})}$$

$$\cdot\left\{1+\frac{S_{xx}}{2c+S^2}(\beta-b)^2\right\}^{-\frac{1}{2}(n+p-1)} . \qquad (6.6)$$

$c = p = 0$ in (6.2) implies that the prior distributions of α, β and $\log\sigma^2$ are independent and uniform, i.e.,

$$g(\alpha,\beta,\sigma^2) \propto \frac{1}{\sigma^2} . \qquad (6.7)$$

Under (6.7), the posterior density of β is

$$\pi_1(\beta|\underset{\sim}{x},\underset{\sim}{y}) = \sqrt{\frac{S_{xx}}{S^2}}\,\frac{1}{B(\frac{1}{2},\frac{n-2}{2})}\left\{1+\frac{S_{xx}}{S^2}(\beta-b)^2\right\}^{-\frac{1}{2}(n-1)},$$

and that of

$$(\beta-b)\sqrt{\frac{S_{xx}(n-2)}{S^2}}$$

is, as was pointed out by Lindley (1965), a Student's t with (n-2) degrees of freedom.

Under squared-error loss function, the Bayes estimator of β is

$$\beta^* = E(\beta|\underset{\sim}{x},\underset{\sim}{y}) = \sqrt{\frac{S_{xx}}{2c+S^2}}\,\frac{1}{B(\frac{1}{2},\frac{n+p-2}{2})}$$

$$\cdot\int_{-\infty}^{\infty}\beta\left\{1+\frac{S_{xx}}{2c+S^2}(\beta-b)^2\right\}^{-\frac{1}{2}(n+p-1)} d\beta .$$

By the same substitution,

$$\beta^* = \frac{1}{B(\frac{1}{2}, \frac{n+p-2}{2})} \int_{-\pi/2}^{\pi/2} \left[b + \sqrt{\frac{2c+S^2}{S_{xx}}} \tan\theta \right]$$

$$\text{Sec}\ \theta^{-(n+p-1)} \text{Sec}^2\theta d\theta$$

$$= b + \frac{1}{B(\frac{1}{2}, \frac{n+p-2}{2})} \sqrt{\frac{2c+S^2}{S_{xx}}} \int_{-\pi/2}^{\pi/2} \text{Sin}\theta \text{Cos}^2\theta^{(\frac{n+p-3}{2})-1} d\theta$$

$$= b \tag{6.8}$$

since

$$\int_{-\pi/2}^{\pi/2} \text{Sin}\ \theta\ \text{Cos}^2\theta^{(\frac{n+p-3}{2})-1} = 0,$$

showing that Bayes estimator of β is independent of the prior parameters c and p. Integrating out β from (6.5), the marginal posterior of σ^2 is

$$\pi_2(\sigma^2 | \underset{\sim}{x}, \underset{\sim}{y}) = \frac{(\sigma^2)^{-(\frac{n+p}{2})} \exp\left\{-\frac{1}{2\sigma^2}(2c+S^2)\right\}}{\int_0^\infty (\sigma^2)^{-(\frac{n+p}{2})} \exp\left\{-\frac{1}{2\sigma^2}(2c+S^2)\right\} d\sigma^2}$$

$$= \left(\frac{2c+S^2}{2}\right)^{\frac{n+p-2}{2}} \frac{(\sigma^2)^{-(\frac{n+p}{2})}}{\Gamma(\frac{n+p-2}{2})} \exp\left\{-\frac{1}{2\sigma^2}(2c+S^2)\right\}. \tag{6.9}$$

Thus,

$$\sigma^{*2} = E(\sigma^2 | \underset{\sim}{x}, \underset{\sim}{y}) = \left(\frac{2c+S^2}{2}\right)^{\frac{n+p-2}{2}} \frac{1}{\Gamma(\frac{n+p-2}{2})} \int_0^\infty (\sigma^2)^{-\frac{n+p}{2}+1}$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}(2c+S^2)\right\} d\sigma^2$$

$$= \left(\frac{2c+S^2}{2}\right)^{\frac{n+p-2}{2}} \frac{1}{\Gamma(\frac{n+p-2}{2})} \left(\frac{2}{2c+S^2}\right)^{\frac{n+p}{2}-2} \Gamma(\frac{n+p}{2}-2)$$

$$= \frac{2c+S^2}{2}\,\frac{2}{n+p-4} = \frac{2c+S^2}{n+p-4}\ . \tag{6.10}$$

Integrating out β from (6.4), the joint posterior of α and σ^2 is

$$\pi(\alpha,\sigma^2|\underset{\sim}{x},\underset{\sim}{y}) \propto (\sigma^2)^{-\frac{n+p+1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\,[2c+S^2+n(\alpha-a)^2]\right\}. \tag{6.11}$$

Again, integrating out σ^2 from (6.11), the marginal posterior of α is

$$\pi_3(\alpha|\underset{\sim}{x},\underset{\sim}{y}) = \frac{\left\{1+\frac{n}{2c+S^2}\,(\alpha-a)^2\right\}^{-\frac{1}{2}(n+p-1)}}{\int_{-\infty}^{\infty}\left\{1+\frac{n}{2c+S^2}\,(\alpha-a)^2\right\}^{-\frac{1}{2}(n+p-1)}d\alpha}$$

$$= \sqrt{\frac{n}{2c+S^2}}\;\frac{1}{B(\frac{1}{2},\frac{n+p-2}{2})} \cdot\left\{1+\frac{n}{2c+S^2}\,(\alpha-a)^2\right\}^{-\frac{1}{2}(n+p-1)}. \tag{6.12}$$

Hence, under squared error loss function, the Bayes estimator is

$$\alpha^* = E(\alpha|\underset{\sim}{x},\underset{\sim}{y}) = \sqrt{\frac{n}{2c+S^2}}\;\frac{1}{B(\frac{1}{2},\frac{n+p-2}{2})} \cdot\int_{-\infty}^{\infty}\alpha\left\{1+\frac{n}{2c+S^2}\,(\alpha-a)^2\right\}^{-\frac{1}{2}(n+p-1)}d\alpha$$

$$= a,$$

as in β^*, the Bayes estimator of α is independent of the prior parameters c and p.

It may be noted here that under (6.7), the posterior density of

$$(\alpha - a)\sqrt{\frac{n(n-2)}{S^2}}$$

is a Student's-t with (n-2) degrees of freedom (Lindley, 1965).

6.2 BAYES ESTIMATORS UNDER CASE II

Assume that σ^2 is known and that apriori (α,β) has a bivariate normal distribution with

$$\mu_\alpha = \mu_\beta = 0, \quad V(\alpha) = \tau_\alpha, \quad V(\beta) = \tau_\beta$$

and coefficient of correlation ρ, i.e.,

$$g(\alpha,\beta) = \frac{1}{\sqrt{2\rho}\,\sqrt{\tau_\alpha \tau_\beta}\,\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \cdot \left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha \tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}, \quad -\infty < \alpha,\ \beta < \infty. \tag{6.14}$$

Combining (6.3) and (6.14), the joint posterior density of α and β is

$$\pi(\alpha,\beta|\underset{\sim}{x},\underset{\sim}{y}) \propto \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha \tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2}[S^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2]\right\}. \tag{6.15}$$

Integrating out β from (6.15), we obtain

$$\pi(\alpha|\underset{\sim}{x},\underset{\sim}{y}) = \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha\tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}\right.$$

$$\left.\cdot\exp\left\{-\frac{1}{2\sigma^2}[s^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2]\right\}d\beta\right] \Big/$$

$$\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha\tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}\right.$$

$$\left.\cdot\exp\left\{-\frac{1}{2\sigma^2}[s^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2]\right\}d\beta d\alpha\right]. \tag{6.16}$$

Therefore,

$$\alpha^* = E(\alpha|\underset{\sim}{x},\underset{\sim}{y}) = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \alpha\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha\tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}\right.$$

$$\left.\cdot\exp\left\{-\frac{1}{2\sigma^2}[s^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2]\right\}d\beta d\alpha\right] \Big/$$

$$\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha\tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}\right.$$

$$\left.\cdot\exp\left\{-\frac{1}{2\sigma^2}[s^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2]\right\}d\beta d\alpha\right]. \tag{6.17}$$

Similarly, integrating out α from (6.15), we obtain the posterior density of β and, hence,

$$\beta^* = E(\beta|\underset{\sim}{x},\underset{\sim}{y}) = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha\tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}\right.$$

$$\left.\cdot\exp\left\{-\frac{1}{2\sigma^2}[s^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2]\right\}d\alpha d\beta\right] \Big/$$

$$\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\alpha^2}{\tau_\alpha} - \frac{2\rho\alpha\beta}{\sqrt{\tau_\alpha\tau_\beta}} + \frac{\beta^2}{\tau_\beta}\right)\right\}\right.$$

$$\left.\cdot\exp\left\{-\frac{1}{2\sigma^2}\quad S^2 + n(\alpha-a)^2 + S_{xx}(\beta-b)^2\right\}d\alpha d\beta\right]. \tag{6.18}$$

The integrals in (6.17) and (6.18) do not seem to take simple form and we will evaluate them by using Lindley's (1980) technique.

For two parameter case (see Appendix A1), applying Lindley's technique, the equation (6.17) takes the form

$$\begin{aligned}\alpha^* = E(\alpha|\underset{\sim}{x},\underset{\sim}{y}) &= \hat{\alpha} + \rho_\alpha\sigma_{20} + \rho_\beta\sigma_{12} + \tfrac{1}{2}L_{30}\sigma_{20}^2 \\ &\quad + \frac{3}{2}L_{21}\sigma_{20}\sigma_{12} + \frac{1}{2}L_{12}(\sigma_{20}\sigma_{02} + 2\sigma_{12}^2) \\ &\quad + \frac{1}{2}L_{03}\sigma_{12}\sigma_{02}.\end{aligned} \tag{6.19}$$

Similarly, (6.18) takes the form

$$\begin{aligned}\beta^* = E(\beta|\underset{\sim}{x},\underset{\sim}{y}) &= \hat{\beta} + \rho_\beta\sigma_{02} + \rho_\alpha\sigma_{21} + \frac{1}{2}L_{03}\sigma_{02}^2 \\ &\quad + \frac{3}{2}L_{12}\sigma_{02}\sigma_{21} + \frac{1}{2}L_{21}(\sigma_{02}\sigma_{20} + 2\sigma_{21}^2) \\ &\quad + \frac{1}{2}L_{30}\sigma_{21}\sigma_{20}\end{aligned} \tag{6.20}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are mle of α and β respectively.

ρ_α and ρ_β are the first derivatives of the logarithm of the joint prior density with respect to α and β respectively.

$\sigma_{ij} = \left(-\dfrac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right)^{-1}$, and is equal to 0 if $\dfrac{\partial^2 L}{\partial\theta_i\partial\theta_j} = 0$.

L_{ij} = i-th derivative with respect to α and j-th derivative with respect to β of the log-likelihood function.

For example,

$$L_{12} = \frac{\partial^3 L}{\partial\alpha\partial\beta^2} .$$

From (6.14), we obtain

$$\rho_\alpha = -\frac{1}{1-\rho^2}\left(\frac{\alpha}{\tau_\alpha} - \frac{\rho\beta}{\sqrt{\tau_\alpha\tau_\beta}}\right)$$

$$\rho_\beta = -\frac{1}{1-\rho^2}\left(\frac{\beta}{\tau_\beta} - \frac{\rho\alpha}{\sqrt{\tau_\alpha\tau_\beta}}\right).$$

ρ_α and ρ_β are to be evaluated at $\hat{\alpha}$ and $\hat{\beta}$.

To work out Lindley's method numerically, we generate a sample from the model

$$E(Y_i|x) = \alpha + \beta(x_i-\bar{x}), \quad i = 1,2,\ldots,n$$

for $\alpha = 5$, $\beta = 2$, and $\sigma = 5$ with the relation

$$Y_i = 5 + 2(x_i-\bar{x}) + 5Z,$$

where

$$Z \sim N(0,1).$$

We take the fixed values of x as 5, 10, 15, 20, 25, and 30 and generate 5 values of Y for each x to give n = 30. The generated data are shown in Table 6.1. We then evaluated the logarithm of the likelihood (6.3) except for the normalizing constant, around mle of α and β in Table 6.2. Tables 6.3 and 6.4 give the values of the derivatives L_{20} and L_{02} respectively.

TABLE 6.1

x	y	x	y	x	y
5	-21.2710226	15	0.2610639	25	22.3605800
5	-19.8939294	15	4.1783792	25	19.5565206
5	-17.8635520	15	- 4.2414939	25	27.3651361
5	-17.6278672	15	- 7.6268387	25	19.9234593
5	-22.5216764	15	7.4637699	25	15.6214675
10	-10.0369930	20	10.6308830	30	33.6543295
10	-14.5834488	20	11.6564479	30	30.0276557
10	- 9.4530529	20	1.5779400	30	35.4171181
10	-12.0320082	20	9.3923527	30	30.0694993
10	-10.2099758	20	18.8509417	30	39.6734619

The data give: $\hat{\alpha} = 5.6773049$, $\hat{\beta} = 2.1435446$, $SSE = S^2 = 502.4790579$.

TABLE 6.2

$$L(\alpha,\beta) = -[502.4790579 + 30(\alpha-5.677304)^2 + 2187.5(\beta-2.143544)^2]$$

α \ β	1.9435	2.0435	2.1435	2.2435	2.3435
5.4773	-591.217773438	-525.573486328	-503.678955078	-525.533447266	-591.137695312
5.5773	-590.317626953	-524.673339844	-502.778808594	-524.633300781	-590.237548828
5.6773	-590.017822266	-524.373535156	-502.479003906	-524.333496094	-589.937744141
5.7773	-590.317626953	-524.673339844	-502.778808694	-524.633300781	-590.237548828
5.8773	-591.217529297	-525.573242187	-503.678710937	-525.533203125	-591.137451172

TABLE 6.3

Values of L_{20}

α \ β	1.9435	2.0435	2.1435	2.2435	2.3435
5.4773					
5.5773	-60.0341797	-60.0341797	-60.0341797	-60.0341797	-60.0341797
5.6773	-59.9609374	-59.9609374	-59.9609374	-59.9609374	-59.9609374
5.7773	-60.0341797	-60.0341797	-60.0341797	-60.0341797	-60.0341797
5.8773					

$L_{20}(\hat{\alpha},\hat{\beta}) = -59.9609374.$ $\quad L_{21}(\hat{\alpha},\hat{\beta}) = 0.$

TABLE 6.4

Values of L_{02}

α \ β	1.9435	2.0435	2.1435	2.2435	2.3435
5.4773		-4374.975586	-4374.902344	-4374.947786	
5.5773		-4374.975586	-4374.902344	-4374.947786	
5.6773		-4374.975586	-4374.902344	-4374.947786	
5.7773		-4374.975586	-4374.902344	-4374.947786	
5.8773		-4374.975586	-4374.902344	-4374.947786	

$L_{02}(\hat{\alpha},\hat{\beta}) = -4374.902344.$ $\quad L_{12}(\hat{\alpha},\hat{\beta}) = 0.$

The above tables show that all the mixed second and third derivatives are zero and hence, we will use two one-parameter formulae for $E(\alpha)$ and $E(\beta)$ given by

$$\left.\begin{aligned} \alpha^* &= E(\alpha \mid \underset{\sim}{x}, \underset{\sim}{y}) = \hat{\alpha} + \rho_\alpha \sigma_\alpha^2 + \frac{1}{2} L_{30} \sigma_\alpha^4 \\ \beta^* &= E(\beta \mid \underset{\sim}{x}, \underset{\sim}{y}) = \hat{\beta} + \rho_\beta \sigma_\beta^2 + \frac{1}{2} L_{03} \sigma_\beta^4 \end{aligned}\right\} \qquad (6.21)$$

(Lindley, 1980), where

$$\sigma_\alpha^2 = -L_{20}^{-1} = 0.0166775$$

$$\sigma_\beta^2 = -L_{02}^{-1} = 0.0002286$$

$$L_{30} = 0$$

$$L_{03} = 0.$$

Since L_{30}, L_{03} are zero at joint mle, (6.21) takes the form

$$\left.\begin{aligned} \alpha^* &= E(\alpha \mid \underset{\sim}{x}, \underset{\sim}{y}) = \hat{\alpha} + \rho_\alpha \sigma_\alpha^2 \\ \beta^* &= E(\beta \mid \underset{\sim}{x}, \underset{\sim}{y}) = \hat{\beta} + \rho_\beta \sigma_\beta^2 \end{aligned}\right\} \qquad (6.22)$$

ρ_α, ρ_β involve the prior parameters τ_α, τ_β and ρ. For reasonable choice of τ's and ρ we can obtain the Bayes estimator of α and β. For instance, $\tau_\alpha = \tau_\beta = 1$ and $\rho = .5$, give

$$\alpha^* = 5.6773049 - 0.1024117 = 5.5748932$$

$$\beta^* = 2.1435446 + 0.0002119 = 2.1437565.$$

Monte Carlo study for α^* and β^* in (6.22) are shown in Tables 6.7 to 6.12.

We note that under uniform and independent prior for (α,β), ρ_α and ρ_β in (6.22) are zero, giving

$$\left.\begin{aligned} \alpha^* &= \hat{\alpha} \\ \beta^* &= \hat{\beta} \end{aligned}\right\} \qquad (6.23)$$

6.3 MONTE CARLO STUDY

In order to compare the performance of the different estimators, we generated 1000 (N) samples each of size 30 (n) from (6.1) with $\alpha = 5$, $\beta = 2$ and $\sigma = 5$. The mean of the estimators and the corresponding MSE (in parentheses) are shown in the tables below:

TABLE 6.5

MLE (6.23) of α and β: $N = 1000$, $n = 30$

Parameters	Estimates
α	5.0224503 (0.8787140)
β	1.9956223 (0.0116023)

TABLE 6.6

Bayes Estimates (6.10) of σ^2 under Squared Error Loss Function: N = 1000, n = 30

c \ p	1	2	3	4	5
1	25.7961857 (48.4236235)	24.8748934 (44.4527689(	24.0171384 (42.3913438)	23.2165671 (41.8902994)	22.4676456 (42.6653681)
2	25.8702598 (48.5470639)	24.9463219 (44.4399986)	24.0861039 (42.2605329)	23.2832338 (41.6569528)	22.5321618 (42.3427750)
3	25.9443339 (48.6814783)	25.0177505 (44.4374323)	24.1550695 (42.1392345)	23.3499005 (41.4324951)	22.5966779 (42.0285066)
4	26.0184079 (48.8268666)	25.0891791 (44.4450702)	24.2240350 (42.0274486)	23.4165671 (41.2169262)	22.6611940 (41.7225629)
5	26.0924820 (48.9832288)	25.1606077 (44.4629121)	24.2930005 (41.9251752)	23.4832338 (41.0102463)	22.7257101 (41.4249438)

For estimation of σ^2, c = 5 and p = 4 seem to be the optimum choice for this study.

TABLE 6.7

Bayes Estimates (6.22) of α under Squared Error Loss Function and Different Combination of $(\tau_\alpha, \tau_\beta)$ and $\rho = 0.1$, $N = 1000$, $n = 30$

τ_α \ τ_β	1	50	100
1	4.9412569 (0.8523491)	4.9383724 (0.8526887)	4.9382333 (0.8527055)
50	5.0212344 (0.8780708)	5.0208265 (0.8780526)	5.0208068 (0.8780517)
100	5.0219408 (0.8783966)	5.0216523 (0.8783832)	5.0216384 (0.8783826)

TABLE 6.8

Bayes Estimates (6.22) of α under Squared Error Loss Function and Different Combination of $(\tau_\alpha, \tau_\beta)$ and $\rho = 0.5$, $N = 1000$, $n = 30$

τ_α \ τ_β	1	50	100
1	4.9330139 (0.8441587)	4.9139762 (0.8470204)	4.9130577 (0.8471769)
50	5.0233540 (0.8779833)	5.0206616 (0.8778576)	5.0205317 (0.8778519)
100	5.0235516 (0.8783803)	5.0216478 (0.8782891)	5.0215560 (0.8782849)

TABLE 6.9

Bayes Estimates (6.22) of α under Squared Error Loss Function and Different Combination of $(\tau_\alpha, \tau_\beta)$ and $\rho = 0.9$, $N = 1000$, $n = 30$

τ_α \ τ_β	1	50	100
1	5.0337996 (0.8778552)	5.6041653 (0.8876365)	4.5976394 (0.8928288)
50	5.0359198 (0.8764824)	5.0167900 (0.8754216)	5.0158671 (0.8753889)
100	5.0337996 (0.8778552)	5.0202728 (0.8770869)	5.0196202 (0.8770591)

TABLE 6.10

Bayes Estimates (6.22) of β under Squared Error Loss Function and Different Combination of $(\tau_\alpha, \tau_\beta)$ and $\rho = 0.1$, $N = 1000$, $n = 30$

τ_α \ τ_β	1	50	100
1	1.9952774 (0.0116002)	1.9956294 (0.0116021)	1.9956292 (0.0116022)
50	1.9951778 (0.0116010)	1.9956153 (0.0116022)	1.9956192 (0.0116022)
100	1.9951730 (0.0116011)	1.9956146 (0.0116022)	1.9956187 (0.0116022)

TABLE 6.11

Bayes Estimates (6.22) of β under Squared Error Loss Function and Different Combination of $(\tau_\alpha, \tau_\beta)$ and $\rho = 0.5$, $N = 1000$, $n = 30$

τ_α \ τ_β	1	50	100
1	1.9957793 (0.0115947)	1.9957183 (0.0116014)	1.9956926 (0.0116017)
50	1.9951222 (0.0115999)	1.9956253 (0.0116021)	1.9956269 (0.0116022)
100	1.9950905 (0.0116002)	1.9956201 (0.0116021)	1.9956238 (0.0116022)

TABLE 6.12

Bayes Estimates (6.22) of β under Squared Error Loss Function and Different Combination of $(\tau_\alpha, \tau_\beta)$ and $\rho = 0.9$, $N = 1000$, $n = 30$

τ_α \ τ_β	1	50	100
1	1.9937652 (0.0115947)	1.9963432 (0.0115967)	1.9961420 (0.0115983)
50	1.9939905 (0.0115922)	1.9956829 (0.0116013)	1.9956751 (0.0116016)
100	1.9937652 (0.0115947)	1.9956511 (0.0116015)	1.9956526 (0.0116018)

The entries in the principal diagonals of the Tables 6.7 to 6.12 give the estimates under the priors in III.

It is apparent from these tables that the estimates of (α,β) based on the bivariate normal prior (6.14) have minimum mse for all combinations of prior variances (τ_α,τ_β) and the prior coefficient of correlation, ρ under consideration, than the ml estimates. The changes in the values of the estimates and their mse's are very small for variation in the prior parameters. While the ml estimators of (α,β) are identical with Bayes estimators under the prior (6.2), they have uniformly higher mse's than the Bayesian estimators based on the bivariate normal prior for (α,β).

6.4 CONCLUSION

Bayes estimators of (α,β) based on the uniform prior for (α,β) and independent inverted gamma prior for σ^2 is independent of the prior parameters c and p, as was expected. Bayes estimators, in this situation, are identical with the ml estimators.

For the particular choice of α, β and σ, the Monte Carlo study indicates that the Bayes estimators of (α,β) based on the bivariate normal prior have considerably lower mse than their ml counterparts. Also, the estimators are very much robust for the choice of the prior parameters (τ_α,τ_β) and ρ.

APPENDIX A.1

(i) Lindley's Method for One Parameter Case

Lindley (1980) developed a technique whereby one may evaluate the ratio of integrals of the form

$$I = \frac{\int w(\theta) e^{L(\theta)} d\theta}{\int v(\theta) e^{L(\theta)} d\theta} \tag{1}$$

where $L(\theta)$ is the logarithm of the likelihood.

Clearly, if $w(\theta) = u(\theta) \cdot v(\theta)$ and $v(\theta)$ is the prior density of θ, then I yields the posterior expectation of $u(\theta)$.

Expanding $w(\theta)$ and $L(\theta)$ by Taylor-series expansion about $\hat{\theta}$, the mle of θ,

$$\begin{aligned} w(\theta) e^{L(\theta)} d\theta = & \left[w(\hat{\theta}) + (\theta-\hat{\theta}) w_1(\hat{\theta}) + \frac{(\theta-\hat{\theta})^2}{2!} w_2(\hat{\theta}) \right. \\ & \left. + \frac{(\theta-\hat{\theta})^3}{3!} w_3(\hat{\theta}) + \ldots \right] \left[\exp\left\{ L(\hat{\theta}) + (\theta-\hat{\theta}) L_1(\hat{\theta}) \right. \right. \\ & \left. \left. + \frac{(\theta-\hat{\theta})^2}{2!} L_2(\hat{\theta}) + \frac{(\theta-\hat{\theta})^3}{3!} L_3(\hat{\theta}) + \ldots \right\} \right] \end{aligned} \tag{2}$$

where

$$w_k(\hat{\theta}) = \left. \frac{\partial^k w(\theta)}{\partial \theta^k} \right|_{\theta=\hat{\theta}}$$

$$L_k(\hat{\theta}) = \left. \frac{\partial^k L(\theta)}{\partial \theta^k} \right|_{\theta=\hat{\theta}} .$$

It is assumed that $w(\hat{\theta}) \neq 0$. Clearly, $L_1(\hat{\theta}) = 0$.

Let

$$W_i(\hat{\theta}) = \frac{w_i(\hat{\theta})}{w(\hat{\theta})},$$

then the RHS of (2) is

$$w(\hat{\theta})e^{L(\hat{\theta})} \int \left[1 + W_1(\hat{\theta})(\theta-\hat{\theta}) + W_2(\hat{\theta}) \frac{(\theta-\hat{\theta})^2}{2!} + W_3(\hat{\theta}) \frac{(\theta-\hat{\theta})^3}{3!} + \ldots\right] e^{\frac{(\theta-\hat{\theta})^2}{2} L_2(\hat{\theta})} \cdot \left[1 + L_3(\hat{\theta}) \frac{(\theta-\hat{\theta})^3}{6} + L_4(\hat{\theta}) \frac{(\theta-\hat{\theta})^4}{24} + \ldots\right] d\theta$$

$$= w(\hat{\theta})e^{L(\hat{\theta})} \int e^{\frac{(\theta-\hat{\theta})^2}{2} L_2(\hat{\theta})} \left[1 + W_1(\hat{\theta})(\theta-\hat{\theta}) + W_2(\hat{\theta}) \frac{(\theta-\hat{\theta})^2}{2} + W_3(\hat{\theta}) \frac{(\theta-\hat{\theta})^3}{6} + L_3(\hat{\theta}) \frac{(\theta-\hat{\theta})^3}{6} + W_1(\hat{\theta})L_3(\hat{\theta}) \frac{(\theta-\hat{\theta})^4}{6} + L_4(\hat{\theta}) \frac{(\theta-\hat{\theta})^4}{24} + \ldots\right] d\theta. \quad (3)$$

Using $e^{\frac{(\theta-\hat{\theta})^2}{2} L_2(\hat{\theta})}$ proportional to a univariate normal density with variance

$$\sigma^2 = -L_2^{-1}(\hat{\theta}), \quad (3a)$$

the integrals in (3) involve moments of a univariate normal density and we have

$$w(\hat{\theta})e^{L(\hat{\theta})}\sqrt{2\pi}\ \sigma\left[1 + \frac{\sigma^2}{2} W_2(\hat{\theta}) + \frac{\sigma^4}{2} L_3(\hat{\theta})W_1(\hat{\theta})\right.$$

$$\left. + \frac{\sigma^4}{8} L_4(\hat{\theta}) + \ldots.\right].$$

Similarly for the denominator of (1), we have

$$v(\hat{\theta})e^{L(\hat{\theta})}\sqrt{2\pi}\ \sigma\left[1 + \frac{\sigma^2}{2} V_2(\hat{\theta}) + \frac{\sigma^4}{2} L_3(\hat{\theta})V_1(\hat{\theta})\right.$$

$$\left. + \frac{\sigma^4}{8} L_4(\hat{\theta}) + \ldots\right]$$

where

$$V_i(\hat{\theta}) = \frac{v_i(\hat{\theta})}{v(\hat{\theta})},$$

assuming $v(\hat{\theta}) \neq 0$, and

$$v_k(\hat{\theta}) = \left.\frac{\partial^k v(\hat{\theta})}{\partial \theta^k}\right|_{\theta=\hat{\theta}}$$

Hence, the ratio (1)

$$\frac{w(\hat{\theta})}{v(\hat{\theta})}\left[1 + \frac{\sigma^2}{2} W_2(\hat{\theta}) + \frac{\sigma^4}{2} L_3(\hat{\theta})W_1(\hat{\theta}) + \frac{\sigma^4}{8} L_4(\hat{\theta}) + \ldots\right]$$

$$\cdot\left[1 - \left\{\frac{\sigma^2}{2} V_2(\hat{\theta}) + \frac{\sigma^4}{2} L_3(\hat{\theta})V_1(\hat{\theta}) + \ldots\right\}\right.$$

$$\left. + \left\{\frac{\sigma^2}{2} V_2(\hat{\theta}) + \frac{\sigma^4}{2} L_3(\hat{\theta})V_1(\hat{\theta}) + \ldots\right\}^2 + \ldots\right]$$

$$= \frac{w(\hat{\theta})}{v(\hat{\theta})}\left[1 + \frac{\sigma^2}{2}\{W_2(\hat{\theta}) - V_2(\hat{\theta})\}\right.$$

$$\left. + \frac{\sigma^4}{2} L_3(\hat{\theta})\{W_1(\hat{\theta}) - V_1(\hat{\theta})\} + \ldots\right]. \qquad (4)$$

Let $w(\theta) = u(\theta)\cdot v(\theta)$ where $v(\theta)$ is the prior density of θ.

$$w_1(\hat{\theta}) = u_1(\hat{\theta})v(\hat{\theta}) + u(\hat{\theta})v_1(\hat{\theta})$$

$$w_2(\hat{\theta}) = u_2(\hat{\theta})v(\hat{\theta}) + 2u_1(\hat{\theta})v_1(\hat{\theta}) + u(\hat{\theta})v_2(\hat{\theta})$$

$$W_2(\hat{\theta}) - V_2(\hat{\theta}) = \frac{w_2(\hat{\theta})}{w(\hat{\theta})} - \frac{v_2(\hat{\theta})}{v(\hat{\theta})}$$

$$= \frac{u_2(\hat{\theta})}{u(\hat{\theta})} + \frac{2u_1(\hat{\theta})v_1(\hat{\theta})}{u(\hat{\theta})v(\hat{\theta})} \tag{5}$$

$$W_1(\hat{\theta}) - V_1(\hat{\theta}) = \frac{u_1(\hat{\theta})}{u(\hat{\theta})} . \tag{6}$$

Let

$$\rho(\theta) = \log v(\theta),$$

$$v_1(\theta) = v(\theta)\rho_1(\theta). \tag{7}$$

Substituting (7) in (5),

$$W_2(\hat{\theta}) - V_2(\hat{\theta}) = \frac{u_2(\hat{\theta}) + 2u_1(\hat{\theta})\rho_1(\hat{\theta})}{u(\hat{\theta})} . \tag{8}$$

Substituting (8) and (6) in (4), we obtain

$$E[u(\theta)|\underset{\sim}{x}] = \frac{\int u(\theta)v(\theta)e^{L(\theta)}d\theta}{\int v(\theta)e^{L(\theta)}d\theta}$$

$$\approx u(\hat{\theta}) + \frac{\sigma^2}{2}\{u_2(\hat{\theta}) + 2u_1(\hat{\theta})\rho_1(\hat{\theta})\}$$

$$+ \frac{\sigma^4}{2} L_3(\hat{\theta})u_1(\hat{\theta}).$$

The terms other than the first are referred to as correction terms.

Thus, if we are interested in the posterior mean of θ, i.e., $u(\theta) = \theta$, we have $u_1(\theta) = 1$, $u_2(\theta) = 0$ and

$$E(\theta|\underset{\sim}{x}) \approx \hat{\theta} + \sigma^2 \rho_1(\hat{\theta}) + \frac{\sigma^4}{2} L_3(\hat{\theta}) \tag{9}$$

where

$$\rho_1(\hat{\theta}) = \frac{\partial \log g(\hat{\theta})}{\partial \theta}$$

$$\sigma^2 = -L_2^{-1}(\hat{\theta})$$

are evaluated at the mle of θ.

In numerical work, Lindley did not use the differential calculus to evaluate $L_i(\theta)$; instead he computed the log-likelihood function $L(\theta)$ for a set of values of θ around mle $\hat{\theta}$ and used the differences from difference tables. The relation between the derivative D and difference Δ is

$$\frac{d^k L(\theta)}{d\theta^k} = L_k(\theta) = \lim_{\Delta\theta \to 0} \frac{\Delta^k L(\theta)}{(\Delta\theta)^k} \qquad \text{(Neilson, 1965)}$$

where $\Delta\theta$ is the increment and $\Delta^k L(\theta)$ is the k-th difference.

An alternative approach of evaluating the ratio of integrals of the form (1) is by denoting

$$\Lambda(\theta) = L(\theta) + \rho(\theta),$$

i.e., $\Lambda(\theta)$ is the logarithm of the posterior density of θ except for the constant of proportionality. Following the same argument as above and expanding $\Lambda(\theta)$ about the posterior mode of θ, the posterior mean of θ is approximately

$$E(\theta|\underset{\sim}{x}) \simeq \tilde{\theta} + \frac{1}{2}\Lambda_3\tau^4 \tag{10}$$

where

$$\tilde{\theta} = \text{posterior mode}$$

$$\tau^2 = -\Lambda_2^{-1}(\tilde{\theta})$$

and

$$\Lambda_k(\theta) = \frac{\partial^k \Lambda(\theta)}{\partial\theta^k} .$$

$\Lambda_k(\theta)$ will be evaluated about the posterior mode while $L_k(\theta)$ is evaluated about the mle of θ.

The form (10) would be useful in the cases in which posterior mode is readily available or easier to compute than the mle $\hat{\theta}$.

We now apply the method for known cases to see how good the approximation is.

(i) Suppose $X \sim N(\theta,1)$ and the prior density of θ:

$$g(\theta) \propto e^{-\frac{1}{2}(\theta-\mu)^2}, \quad -\infty < \mu,\ \theta < \infty.$$

It is easy to show that the posterior density of

$$\Pi(\theta|\underset{\sim}{x}) \sim N\left(\frac{\mu/n+\bar{x}}{1+1/n}, \frac{1/n}{1+1/n}\right)$$

and the posterior mean under squared error loss function

$$\theta^* = \frac{\mu/n+\bar{x}}{1+1/n} = \frac{n\bar{x}+\mu}{n+1} .$$

The log-likelihood

$$L(\theta) = c - \frac{n}{2}(\bar{x}-\theta)^2$$

where c is a constant independent of θ.

$$L_1(\theta) = n(\bar{x}-\theta), \qquad L_2(\theta) = -n, \qquad L_3(\theta) = 0$$

$$\hat{\theta} = \bar{x}, \qquad \sigma^2 = -L_2^{-1}(\hat{\theta}) = \frac{1}{n} .$$

$$\rho_1(\hat{\theta}) = -(\bar{x}-\mu) .$$

Hence, the equation (9) takes the form

$$E(\theta \mid \underset{\sim}{x}) = \bar{x} - \frac{\bar{x}-\mu}{n} = \frac{(n-1)\bar{x}+\mu}{n}$$

$$= \theta^* - \left[\frac{n\bar{x}+\mu}{n+1} - \frac{(n-1)\bar{x}+\mu}{n}\right]$$

$$= \theta^* - \frac{\bar{x}-\mu}{n(n+1)} = \theta^* - 0(\frac{1}{n^2}) .$$

Thus, the first term neglected is of order $\frac{1}{n^2}$.

(ii) $X \sim b(n,p)$

$$\hat{p} = \frac{x}{n} .$$

Let the prior density of P be uniform, i.e., $g(p) \propto$ constant, then

$$p^* = \frac{x+1}{n+2}$$

$$L(p) = c + x \log_e p + (n-x)\log_e(1-p)$$

$$L_1(p) = \frac{x}{p} - \frac{n-x}{1-p}, \qquad L_2(p) = \frac{-x}{p^2} - \frac{n-x}{(1-p)^2}$$

$$L_3(p) = \frac{2x}{p^3} - \frac{2(n-x)}{(1-p)^3}, \quad L_3(\hat{p}) = \frac{2n^4(n-2x)}{x^2(n-x)^2}$$

$$\rho_1(p) = 0, \qquad \sigma^2 = -L_2^{-1}(\hat{p}) = \frac{x(n-x)}{n^3} .$$

The equation (9) takes the form

$$E(p|\underset{\sim}{x}) = \frac{x}{n} + \frac{x^2(n-x)^2}{2n^6} \frac{2n^4(n-2x)}{x^2(n-x)^2}$$

$$= \frac{x}{n} + \frac{n-2x}{n^2}$$

$$= p^* + \left(\frac{x}{n} + \frac{n-2x}{n^2} - \frac{x+1}{n+2}\right)$$

$$= p^* + \frac{2(n-2x)}{n^2(n+2)}$$

$$= p^* + 0(\frac{1}{n^2}).$$

(ii) <u>Lindley's Method for Two-Parameter Case</u>

Notation:

$\underset{\sim}{\theta} = (\theta_1, \theta_2)$

$L_{ijk} = \dfrac{\partial^3 L(\underset{\sim}{\theta})}{\partial\theta i \partial\theta j \partial\theta k}$, third derivative w.r.t. θ_i, θ_j and θ_k.

that is,

$L_{222} = \dfrac{\partial^3 L(\underset{\sim}{\theta})}{\partial\theta_2^3}$, third derivative w.r.t. θ_2.

L_{ijk} are unaffected by permutation of its suffixes.

$L_{ij} = \dfrac{\partial^2 L(\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j}$, second derivative w.r.t. θ_i and θ_j.

The Taylor series expansion for $L(\underset{\sim}{\theta})$ about $\hat{\underset{\sim}{\theta}}$ may be given by

$$L(\underset{\sim}{\theta}) = L(\hat{\underset{\sim}{\theta}}) + \sum_{i=1}^{2} L_i(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)$$

$$+ \frac{1}{2!}\sum_{i,j=1}^{2} L_{ij}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)$$

$$+ \frac{1}{3!}\sum_{i,j,k=1}^{2} L_{ijk}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k) + \ldots$$

Similarly, the Taylor series expansions for $w(\underset{\sim}{\theta})$ and $v(\underset{\sim}{\theta})$.

Working in the same line as for one-parameter case, the numerator of (1),

$$\int w(\underset{\sim}{\theta})e^{L(\underset{\sim}{\theta})}d\underset{\sim}{\theta} = w(\hat{\underset{\sim}{\theta}})e^{L(\hat{\underset{\sim}{\theta}})}\int\Big[1 + \sum_{i=1}^{2} W_i(\hat{\theta})(\theta_i-\hat{\theta}_i)^2$$

$$+ \frac{1}{2}\sum_{i,j=1}^{2} W_{ij}(\hat{\theta})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j) + \ldots\Big]$$

$$\cdot \exp\left[\frac{1}{2}\sum_{i,j=1}^{2} L_{ij}(\hat{\theta})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)\right]$$

$$\cdot\Big[1 + \frac{1}{6}\sum_{i,j,k=1}^{2} L_{ijk}(\hat{\theta})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k)$$

$$+ \frac{1}{24}\sum_{i,j,k,\ell=1}^{2} L_{ijk\ell}(\hat{\theta})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k)(\theta_\ell-\hat{\theta}_\ell)$$

$$+ \ldots\Big]d\theta, \qquad (11)$$

where

$$W_i(\hat{\underset{\sim}{\theta}}) = \frac{w_i(\hat{\underset{\sim}{\theta}})}{w(\hat{\underset{\sim}{\theta}})}, \qquad L_i(\hat{\underset{\sim}{\theta}}) = 0$$

and assumed that $w(\hat{\underset{\sim}{\theta}})$ does not vanish. Collecting terms of like order together, the RHS of (11) is

$$w(\hat{\underset{\sim}{\theta}})e^{L(\hat{\underset{\sim}{\theta}})} \int e^{\frac{1}{2}\sum_{i,j=1}^{2} L_{ij}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)}$$

$$\cdot\Bigg[1 + \sum_{i=1}^{2} W_i(\hat{\theta})(\theta_i-\hat{\theta}_i) + \frac{1}{6}\sum_{i,j,k=1}^{2} L_{ijk}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)$$

$$\cdot(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k) + \frac{1}{2}\sum_{i,j=1}^{2} W_{ij}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)$$

$$+ \frac{1}{6}\left\{\sum_{i=1}^{2} W_i(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)\right\}\sum_{i,j,k=1}^{2} L_{ijk}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)$$

$$\cdot(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k) + R\Bigg]d\theta. \qquad (12)$$

R does not involve W or its derivatives.

Considering

$$\exp\left\{\frac{1}{2}\sum_{i,j=1}^{2} L_{ij}(\hat{\underset{\sim}{\theta}})(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)\right\}$$

as the kernel of a bivariate normal distribution with precision matrix having elements $-L_{ij}(\hat{\underset{\sim}{\theta}})$, and the variance-covariance matrix,

$$\Sigma = -\begin{pmatrix} \dfrac{\partial^2 L}{\partial\theta_1^2} & \dfrac{\partial^2 L}{\partial\theta_1\partial\theta_2} \\ \dfrac{\partial^2 L}{\partial\theta_2\partial\theta_1} & \dfrac{\partial^2 L}{\partial\theta_2^2} \end{pmatrix}^{-1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

and using the results of the multivariate normal distribution, Anderson (1958, equation (26) of §2.6), namely,

$$E(\theta_i-\hat{\theta}_i) = 0, \qquad E(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j) = \sigma_{ij}$$

$$E(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k) = 0,$$

$$E(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k)(\theta_\ell-\hat{\theta}_\ell) = \sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk},$$

the integrations in (12) takes the form

$$w(\hat{\underset{\sim}{\theta}})e^{L(\hat{\underset{\sim}{\theta}})}(2\pi)|\Sigma|^{\frac{1}{2}}\Bigg[1 + \frac{1}{2}\sum_{i,j=1}^{2} W_{ij}(\hat{\underset{\sim}{\theta}})\sigma_{ij}$$

$$+ \frac{1}{6}\sum_{i,j,k=1}^{2} L_{ijk}W_\ell \ (\sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}) + R^*\Bigg] \qquad (13)$$

and R^* arises from R not involving W or its derivatives.

Similarly, the denominator of (1) is

$$v(\hat{\underset{\sim}{\theta}})e^{L(\hat{\underset{\sim}{\theta}})}(2\pi)|\Sigma|^{\frac{1}{2}}\Bigg[1 + \frac{1}{2}\sum_{i,j=1}^{2} V_{ij}(\hat{\underset{\sim}{\theta}})\sigma_{ij}$$

$$+ \frac{1}{6}\sum_{i,j,k=1}^{2} L_{ijk}(\theta)V_\ell(\hat{\theta})(\sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}) + R^*\Bigg] \qquad (14)$$

where

$$V_{ij}(\underset{\sim}{\hat{\theta}}) = \frac{v_{ij}(\underset{\sim}{\hat{\theta}})}{v(\underset{\sim}{\hat{\theta}})} .$$

Note that all three terms in the second summation in (13) are equal since L_{ijk} is unaffected by the permutation of its suffixes. Permuting j and k in the first term gives

$$\Sigma L_{ikj}(\underset{\sim}{\hat{\theta}}) W_\ell(\underset{\sim}{\hat{\theta}}) \sigma_{ij}\sigma_{k\ell},$$

and then interchanging roles of j and k makes this equal to $\Sigma L_{ijk} W_\ell \sigma_{ik}\sigma_{j\ell}$; similarly the second and third terms. This applies to the second summation in (14).

Finally, the ratio of integrals in (1) is appproximately,

$$\frac{w(\underset{\sim}{\hat{\theta}})}{v(\underset{\sim}{\hat{\theta}})} \Bigg[1 + \frac{1}{2} \sum_{i,j=1}^{2} W_{ij}(\underset{\sim}{\hat{\theta}})\sigma_{ij} + \frac{1}{2} \sum_{i,j,k=1}^{2} L_{ijk}(\underset{\sim}{\hat{\theta}})$$

$$W_\ell(\underset{\sim}{\hat{\theta}})\sigma_{ij}\sigma_{k\ell}\Bigg]\Bigg[1 + \frac{1}{2} \sum_{i,j=1}^{2} V_{ij}(\underset{\sim}{\hat{\theta}})\sigma_{ij}$$

$$\frac{1}{2} \sum_{i,j,k=1}^{2} L_{ijk}(\underset{\sim}{\hat{\theta}}) V_\ell(\underset{\sim}{\hat{\theta}})\sigma_{ij}\sigma_{k\ell}\Bigg]^{-1}$$

$$= \frac{w(\underset{\sim}{\hat{\theta}})}{v(\underset{\sim}{\hat{\theta}})} \Bigg[1 + \frac{1}{2} \sum_{i,j=1}^{2} (W_{ij}(\underset{\sim}{\hat{\theta}}) - V_{ij}(\underset{\sim}{\hat{\theta}}))\sigma_{ij}$$

$$+ \frac{1}{2} \sum_{i,j,k=1}^{2} L_{ijk}(\underset{\sim}{\hat{\theta}})(W_\ell(\underset{\sim}{\hat{\theta}}) - V_\ell(\underset{\sim}{\hat{\theta}}))\sigma_{ij}\sigma_{k\ell} + \ldots\Bigg] \quad (15)$$

As in one parameter case, it is easy to show that

$$W_\ell - V_\ell = u_\ell/u$$

and

$$W_{ij} - V_{ij} = \frac{u_{ij}}{u} + (u_i g_j + u_j g_i)/ug.$$

Finally, we obtain (1) as

$$\int u(\underset{\sim}{\theta})e^{L(\underset{\sim}{\theta})+\rho(\underset{\sim}{\theta})}d\underset{\sim}{\theta}\Big/\int e^{L(\underset{\sim}{\theta})+\rho(\underset{\sim}{\theta})}d\underset{\sim}{\theta} \approx u(\hat{\underset{\sim}{\theta}})$$

$$+ \frac{1}{2}\sum_{i,j}(u_{ij}(\hat{\underset{\sim}{\theta}}) + 2u_i(\hat{\underset{\sim}{\theta}})\rho_j(\hat{\underset{\sim}{\theta}}))\sigma_{ij}$$

$$+ \frac{1}{2}\sum_{i,j,k} L_{ijk}(\hat{\underset{\sim}{\theta}})u_l(\hat{\underset{\sim}{\theta}})\sigma_{ij}\sigma_{k\ell}. \tag{16}$$

The second and third terms are known as correction terms. The first term neglected is $O(n^{-2})$.

If we are interested in the posterior mean of θ_s, $1 \leq s \leq 2$, then from (16)

$$E(\theta_s|\underset{\sim}{x}) \sim \hat{\theta}_s + \sum_{i=1}^{2}\rho_i(\hat{\underset{\sim}{\theta}})\sigma_{is} + \frac{1}{2}\sum_{i,j,k=1}^{2} L_{ijk}(\hat{\underset{\sim}{\theta}})\sigma_{ij}\sigma_{ks}$$

since $u(\underset{\sim}{\theta}) = \theta_s$ gives

$$u_1(\underset{\sim}{\theta}) = 1$$

$$u_2(\underset{\sim}{\theta}) = 0,$$

$$u_{ij}(\underset{\sim}{\theta}) = 0.$$

For $X \sim N(\mu,\sigma^2)$ and under Jeffreys' invariant prior for σ,

$$\sigma^{*2} = \frac{ns^2}{n-2}.$$

Using Lindley's method

$$E(\sigma^2|\underset{\sim}{x}) = \sigma^{*2} - \frac{s^2}{n(n-2)}$$

$$= \sigma^{*2} + O(\frac{1}{n^2}).$$

An alternative way of evaluating (1) is denoted by

$$\Lambda(\underset{\sim}{\theta}) = L(\underset{\sim}{\theta}) + \rho(\underset{\sim}{\theta}),$$

the logarithm of the posterior density of $\underset{\sim}{\theta}$, except the normalizing constant and expand $\Lambda(\underset{\sim}{\theta})$ instead of $L(\underset{\sim}{\theta})$, about the posterior mode of $\underset{\sim}{\theta}$. Following the same argument as above, the posterior mean of θ_s, $1 \le s \le 2$, is

$$\theta_s^* = E(\theta_s|\underset{\sim}{x}) \sim \tilde{\tilde{\theta}}_s + \frac{1}{2} \sum_{i,j,k=1}^{2} \Lambda_{ijk}(\tilde{\underset{\sim}{\theta}})\tau_{ij}\tau_{ks}$$

where $\tilde{\tilde{\theta}}_s$ = posterior mode of the s-th parameter, and τ_{ij} are given by

$$-\begin{pmatrix} \dfrac{\partial^2\Lambda}{\partial\theta_1^2} & \dfrac{\partial^2\Lambda}{\partial\theta_1\partial\theta_2} \\ \dfrac{\partial^2\Lambda}{\partial\theta_2\partial\theta_1} & \dfrac{\partial^2\Lambda}{\partial\theta_2^2} \end{pmatrix}^{-1} = \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{pmatrix}$$

and all the quantities are to be evaluated at the joint posterior mode, $\tilde{\underset{\sim}{\theta}}$.

For type VII distribution

$m = \theta_1$ (i.e., $s = 1$)

$$m^* = E(m|\underset{\sim}{x}) \sim \tilde{\tilde{m}} + \frac{1}{2} \sum_{ijk=1}^{2} \Lambda_{ijk}\tau_{ij}\tau_{kl}$$

$$= \tilde{\tilde{m}} + \frac{1}{2} \sum_{ij=1}^{2} (\Lambda_{ij1}\tau_{ij}\tau_{11} + \Lambda_{ij2}\tau_{ij}\tau_{21})$$

$$= \tilde{\tilde{m}} + \frac{1}{2} \sum_{i=1}^{2} (\Lambda_{i11}\tau_{i1}\tau_{11} + \Lambda_{i12}\tau_{i1}\tau_{21}$$

$$+ \Lambda_{i21}\tau_{i2}\tau_{11} + \Lambda_{i22}\tau_{i2}\tau_{21})$$

$$= \tilde{\tilde{m}} + \frac{1}{2} (\Lambda_{111}\tau_{11}\tau_{11} + \Lambda_{112}\tau_{11}\tau_{21} + \Lambda_{121}\tau_{12}\tau_{11}$$

$$+ \Lambda_{122}\tau_{12}\tau_{21} + \Lambda_{211}\tau_{21}\tau_{11} + \Lambda_{212}\tau_{21}\tau_{21}$$

$$+ \Lambda_{221}\tau_{22}\tau_{11} + \Lambda_{222}\tau_{22}\tau_{21})$$

$$= \tilde{\tilde{m}} + \frac{1}{2} \Lambda_{30}\tau_{11}^2 + \frac{3}{2} \Lambda_{21}\tau_{11}\tau_{12} + \frac{1}{2} \Lambda_{12}(\tau_{11}\tau_{22} + 2\tau_{12}^2)$$

$$+ \frac{1}{2} \Lambda_{03}\tau_{22}\tau_{12}.$$

Using

$$\Lambda_{111} = \Lambda_{30}, \qquad \Lambda_{222} = \Lambda_{03}$$

$$\tau_{12} = \tau_{21}, \qquad \tau_{11} = \tau_{20}, \qquad \tau_{22} = \tau_{02},$$

$$= \tilde{\tilde{m}} + \frac{1}{2} \Lambda_{30}\tau_{20}^2 + \frac{3}{2} \Lambda_{21}\tau_{20}\tau_{12} + \frac{1}{2} \Lambda_{12}(\tau_{20}\tau_{02} + 2\tau_{12}^2)$$

$$+ \frac{1}{2} \Lambda_{03}\tau_{02}\tau_{12}.$$

Similarly,

$$a^* \sim \tilde{\tilde{a}} + \frac{1}{2}\Lambda_{03}\tau_{02}^2 + \frac{3}{2}\Lambda_{12}\tau_{02}\tau_{12} + \frac{1}{2}\Lambda_{21}(\tau_{02}\tau_{20} + 2\tau_{12}^2)$$

$$+ \frac{1}{2}\Lambda_{30}\tau_{20}\tau_{12}.$$

If the second and third mixed derivatives are zero or negligible,

$$m^* \sim \tilde{\tilde{m}} + \frac{1}{2}\Lambda_{30}\tau_{20}^2$$

and

$$a^* \sim \tilde{\tilde{a}} + \frac{1}{2}\Lambda_{03}\tau_{02}^2.$$

APPRENDIX A.2

Sample No.	$F_o(x)$	D_n	Sample No.	$F_o(x)$	D_n
1	0.0035157	0.0010157	41	0.0912401	0.0112599
2	0.0055410	0.0005410	42	0.1056272	0.0006272
3	0.0106764	0.0031762	43	0.1057445	0.0017555
4	0.0138433	0.0038433	44	0.1129161	0.0029161
5	0.0183926	0.0058926	45	0.1o65284	0.0040284
6	0.0187089	0.0037089	46	0.1195963	0.0045963
7	0.0190500	0.0015500	47	0.1220076	0.0045076
8	0.0159665	0.0004335	48	0.1273093	0.0073093
9	0.0198867	0.0026133	49	0.1338856	0.0113856
10	0.0214283	0.0035717	50	0.1381915	0.0131915
11	0.0217040	0.0057960	51	0.1417659	0.0142659
12	0.0256500	0.0043500	52	0.1427845	0.0127845
13	0.0257993	0.0067007	53	0.1449732	0.0124732
14	0.0266066	0.0083934	54	0.1488667	0.0138667
15	0.0268463	0.0106537	55	0.1506227	0.0131227
16	0.0328711	0.0071289	56	0.1536317	0.0136317
17	0.0338303	0.0086697	57	0.1549835	0.0124835
18	0.0368928	0.0081072	58	0.1562855	0.0112855
19	0.0404547	0.0070453	59	0.1636022	0.0161022
20	0.0442766	0.0057234	60	0.1642380	0.0142380
21	0.0449742	0.0075258	61	0.1642919	0.0117919
22	0.0449858	0.0100142	62	0.1694649	0.0144649
23	0.0546658	0.0118342	63	0.1747696	0.0172696
24	0.0468468	0.0131532	64	0.1761673	0.0161673
25	0.0471552	0.0153448	65	0.1781278	0.0155278
26	0.0478761	0.0171239	66	0.1789645	0.0139645
27	0.0563380	0.0111620	67	0.1802092	0.0127092
28	0.0599394	0.0100606	68	0.1834876	0.0134876
29	0.0614459	0.0110541	69	0.1847226	0.0122226
30	0.0623654	0.0126346	70	0.1868037	0.0118037
31	0.0659553	0.0115447	71	0.1937984	0.0162984
32	0.0672731	0.0127269	72	0.1938027	0.0138027
33	0.0714146	0.0110854	73	0.1939073	0.0114073
34	0.0714983	0.0135017	74	0.1971699	0.0121699
35	0.0716407	0.0158593	75	0.2003505	0.0128505
36	0.0828287	0.0071713	76	0.2017199	0.0117199
37	0.0837708	0.0087292	77	0.2052242	0.0127242
38	0.0866263	0.0083737	78	0.2128502	0.0178502
39	0.0871944	0.0103056	79	0.2150518	0.0175518
40	0.0875719	0.0124281	80	0.2188576	0.0188576

Sample No.	$F_o(x)$	D_n	Sample No.	$F_o(x)$	D_n
81	0.2245925	0.0220925	121	0.3044198	0.0019198
82	0.2279994	0.0229994	122	0.3069517	0.0019517
83	0.2294847	0.0219847	123	0.3075683	0.000683
84	0.2320881	0.0220881	124	0.3092059	0.0007941
85	0.2341417	0.0216417	125	0.3105328	0.0019672
86	0.2356749	0.0206749	126	0.3202865	0.0052865
87	0.2367458	0.0192458	127	0.3241293	0.0065293
88	0.2402956	0.0202956	128	0.3247457	0.0047457
89	0.2412129	0.0187129	129	0.3254904	0.0029904
90	0.2431484	0.0181484	130	0.3272295	0.0022295
91	0.2441505	0.0166505	131	0.3328491	0.0053491
92	0.2445074	0.0145074	132	0.3330126	0.0030126
93	0.2471016	0.0146016	133	0.3331809	0.0006809
94	0.2501127	0.0151127	134	0.3365925	0.0015925
95	0.2543882	0.0168882	135	0.3417141	0.0042141
96	0.2559190	0.0159190	136	0.3449517	0.0049517
97	0.2620241	0.0195241	137	0.3464641	0.0039641
98	0.2631393	0.0181393	138	0.3467712	0.0017712
99	0.2636092	0.0161092	139	0.3495864	0.0020864
100	0.2642724	0.0142724	140	0.3504609	0.0004609
101	0.2654969	0.0129969	141	0.3523156	0.0001844
102	0.2665557	0.0115557	142	0.3553945	0.0003945
103	0.2676567	0.0101567	143	0.3614223	0.0039223
104	0.2767125	0.0167125	144	0.3670005	0.0070005
105	0.2776943	0.0151943	145	0.3764848	0.0139848
106	0.2782799	0.0132799	146	0.3782275	0.0132275
107	0.2786090	0.0111090	147	0.3798501	0.0123501
108	0.2810290	0.0110290	148	0.3846814	0.0146814
109	0.2866665	0.0141665	149	0.3855980	0.0130980
110	0.2898238	0.0148238	150	0.3858136	0.0108136
111	0.2905603	0.0130603	151	0.3888272	0.0113272
112	0.2907652	0.0107652	152	0.3892293	0.0092293
113	0.2908947	0.0083947	153	0.3895865	0.0070865
114	0.2910615	0.0060615	154	0.3913579	0.0063579
115	0.2942501	0.0067501	155	0.3931365	0.0056365
116	0.2994746	0.0094746	156	0.3948342	0.0048342
117	0.2998662	0.0073662	157	0.3977903	0.0052903
118	0.3004278	0.0054278	158	0.3981058	0.0031058
119	0.3012651	0.0037651	159	0.3988128	0.0013128
120	0.3017896	0.0017896	160	0.4026585	0.0026585

Sample No.	$F_o(x)$	D_n	Sample No.	$F_o(x)$	D_n
161	0.4030791	0.0005791	201	0.5094243	0.0069243
162	0.4037060	0.0012940	202	0.5133484	0.0083484
163	0.4037089	0.0037911	203	0.5152620	0.0077620
164	0.0405337	0.0054663	204	0.5158925	0.0058925
165	0.4058686	0.0066314	205	0.5175584	0.0050584
166	0.4078403	0.0071597	206	0.5196161	0.0046161
167	0.4121469	0.0053531	207	0.5225709	0.0050709
168	0.4167129	0.0032871	208	0.5281929	0.0081929
169	0.4178250	0.0046750	209	0.5317260	0.0092260
170	0.4193732	0.0056258	210	0.5347342	0.0097342
171	0.4209239	0.0065761	211	0.5362149	0.0087149
172	0.4257221	0.0042779	212	0.5377890	0.0077890
173	0.4290001	0.0034999	213	0.5396907	0.0071907
174	0.4303250	0.0046750	214	0.5417909	0.0067909
175	0.4327928	0.0047072	215	0.5428038	0.0053038
176	0.4373702	0.0026298	216	0.5450272	0.0050272
177	0.4391247	0.0033753	217	0.5485203	0.0060203
178	0.4491979	0.0041979	218	0.5493450	0.0043450
179	0.4518848	0.0043848	219	0.5523401	0.0048401
180	0.4546774	0.0046774	220	0.5588662	0.0088662
181	0.4556355	0.0031355	221	0.5626119	0.0101119
182	0.4559420	0.0009420	222	0.5641837	0.0091837
183	0.4622120	0.0047120	223	0.5677715	0.0102715
184	0.4639142	0.0039142	224	0.5727065	0.0127065
185	0.4644565	0.0019565	225	0.5732370	0.0107370
186	0.4682753	0.0052753	226	0.5749108	0.0099108
187	0.4698648	0.0023648	227	0.5787606	0.0112606
188	0.4724046	0.0024046	228	0.5788353	0.0088353
189	0.4743903	0.0018903	229	0.5809619	0.0084619
190	0.4837491	0.0087491	230	0.5834717	0.0084717
191	0.4845458	0.0070458	231	0.5847804	0.0072804
192	0.4872596	0.0072596	232	0.5877081	0.0077081
193	0.4940922	0.0115922	233	0.5877979	0.0052979
194	0.4942332	0.0092332	234	0.5884555	0.0034555
195	0.4946148	0.0071148	235	0.5900311	0.0025311
196	0.4948201	0.0048201	236	0.5917932	0.0017932
197	0.5030714	0.0105714	237	0.5922953	0.0002047
198	0.5045453	0.0095453	238	0.5963668	0.0013668
199	0.5046222	0.0071222	239	0.5970457	0.0004543
200	0.5052192	0.0052192	240	0.5972661	0.0027339

Sample No.	$F_o(x)$	D_n	Sample No.	$F_o(x)$	D_n
241	0.5983895	0.0041105	281	0.6968535	0.0056465
242	0.6008054	0.0041946	282	0.6980414	0.0069586
243	0.6063070	0.0011930	283	0.6992188	0.0082812
244	0.6063361	0.0036639	284	0.7032399	0.0067601
245	0.6070373	0.0054627	285	0.7071993	0.0053007
246	0.6072513	0.0077487	286	0.7104792	0.0045208
247	0.6086111	0.0088889	287	0.7147435	0.0027565
248	0.6098490	0.0101510	288	0.7147652	0.0052348
249	0.6103441	0.0121559	289	0.7200382	0.0024618
250	0.6178719	0.0071281	290	0.7201566	0.0048434
251	0.6209279	0.0065721	291	0.7266538	0.0008462
252	0.6211284	0.0088716	292	0.7317853	0.0017853
253	0.6222794	0.0102206	293	0.7320083	0.0004917
254	0.6297374	0.0052626	294	0.7343152	0.0006848
255	0.6313393	0.0061607	295	0.7438764	0.0016236
256	0.6335197	0.0064803	296	0.7362705	0.0037295
257	0.6384551	0.0040449	297	0.7410640	0.0014360
258	0.6419962	0.0030038	298	0.7440907	0.0009093
259	0.6420085	0.0054915	299	0.7460752	0.0014248
260	0.6433911	0.0066089	300	0.7466261	0.0033739
261	0.6435337	0.0089663	301	0.7515113	0.0009887
262	0.6435695	0.0114305	302	0.7526775	0.0023225
263	0.6449623	0.0125377	303	0.7558809	0.0016191
264	0.6462676	0.0137324	304	0.7565354	0.0034646
265	0.6482838	0.0142162	305	0.7663382	0.0038382
266	0.6569043	0.0080957	306	0.7685265	0.0035265
267	0.6572998	0.0102002	307	0.7685595	0.0010595
268	0.6590119	0.0109881	308	0.7724933	0.0024933
269	0.6623902	0.0101098	309	0.7755067	0.0030067
270	0.6657451	0.0092549	310	0.7758984	0.0008984
271	0.6700020	0.0074980	311	0.7761641	0.0013359
272	0.6746837	0.0053163	312	0.7795399	0.0004601
273	0.6770985	0.0054015	313	9.7803536	0.0021464
274	0.6775931	0.0074069	314	0.7804666	0.0046334
275	0.6790295	0.0084705	315	0.7850998	0.0024002
276	0.6832159	0.0067841	316	0.7857293	0.0042707
277	0.6837620	0.0087380	317	0.7917811	0.0007189
278	0.6887329	0.0062671	318	0.7949117	0.0000883
279	0.6925477	0.0049523	319	0.7970051	0.0004949
280	0.6958687	0.0041313	320	0.7973830	0.0026170

Sample No.	$F_o(x)$	D_n	Sample No.	$F_o(x)$	D_n
321	0.7994036	0.0030964	361	0.8981896	0.0043104
322	0.8057349	0.0007349	362	0.8993798	0.0056202
323	0.8091610	0.0016610	363	0.9069900	0.0005100
324	0.8110680	0.0010680	364	0.9073136	0.0026864
325	0.8148891	0.0023891	365	0.9082895	0.0042105
326	0.8156408	0.0006408	366	0.9086437	0.0063563
327	0.8186834	0.0011834	367	0.9147861	0.0027139
328	0.8188545	0.0011455	368	0.9228775	0.0028775
329	0.8289555	0.0064555	369	0.9309966	0.0084966
330	0.8337360	0.0087360	370	0.9337227	0.0087227
331	0.8362234	0.0087234	371	0.9349351	0.0074351
332	0.8368088	0.0068088	372	0.9367099	0.0067099
333	0.8373189	0.0048189	373	0.9399195	0.0074195
334	0.8376979	0.0026979	374	0.9418746	0.0068746
335	0.8377122	0.0002122	375	0.9428443	0.0053443
336	0.8384657	0.0015343	376	0.9428534	0.0028534
337	0.8385233	0.0039767	377	0.9429186	0.0004186
338	0.8473793	0.0023793	378	0.9429726	0.0020274
339	0.8478281	0.0003281	379	0.9455487	0.0019513
340	0.8540507	0.0040507	380	0.9518900	0.0018900
341	0.8561283	0.0036283	381	0.9525326	0.0000326
342	0.8598576	0.0048576	382	0.9594612	0.0044612
343	0.8599110	0.0024110	383	0.9641373	0.0066373
344	0.8602048	0.0002048	384	0.9659398	0.0059398
345	0.8617998	0.0007002	385	0.9661687	0.0036687
346	0.8628978	0.0021022	386	0.9684954	0.0034954
347	0.8637082	0.0037918	387	0.9691284	0.0016284
348	0.8641409	0.0058591	388	0.9723981	0.0023981
349	0.8682020	0.0042980	389	0.9740395	0.0015395
350	0.8697443	0.0052557	390	0.9760053	0.0010053
351	0.8701251	0.0073749	391	0.9776249	0.0001249
352	0.8717375	0.0082625	392	0.9788956	0.0011044
353	0.8739559	0.0085441	393	0.9817964	0.0007036
354	0.8876835	0.0026835	394	0.9827256	0.0022744
355	0.8912151	0.0037151	395	0.9827616	0.0047383
356	0.8920560	0.0020560	396	0.9865517	0.0034483
357	0.8922736	0.0002264	397	0.9865580	0.0059420
358	0.8929356	0.0020144	398	0.9915465	0.0034535
359	0.8952355	0.0022645	399	0.9929894	0.0045106
360	0.8978117	0.0021883	400	0.9987203	0.0012797

APPENDIX A.3

TABLE: Bayes Estimates of β Under NCP and SEL Function

(An Extension of Table 2.5), N = 500, n = 100

c \ p	1	2	3	4	5	6	7	8	9	10
1	.71394 (.00524)	.70889 (.00499)	.70389 (.00478)	.69897 (.00464)	.69413 (.00455)	.68934 (.00450)	.68463 (.00451)	.67998 (.00456)	.67539 (.00465)	.67086 (.00479)
2	.72101 (.00560)	.71590 (.00526)	.71086 (.00498)	.70590 (.00477)	.70100 (.00460)	.69617 (.00449)	.69141 (.00443)	.68671 (.00442)	.68208 (.00445)	.67751 (.00453)
3	.72808 (.00603)	.72292 (.00563)	.71783 (.00528)	.71281 (.00499)	.70787 (.00500)	.70299 (.00458)	.69818 (.00445)	.69344 (.00437)	.68876 (.00434)	.68415 (.00435)
4	.73515 (.00659)	.72994 (.00610)	.72480 (.00567)	.71974 (.00531)	.71474 (.00500)	.70982 (.00475)	.70496 (.00455)	.70017 (.00441)	.69545 (.00431)	.69079 (.00426)
5	.74221 (.00724)	.73695 (.00667)	.73177 (.00617)	.72666 (.00572)	.72161 (.00534)	.71684 (.00502)	.71174 (.00476)	.70691 (.00454)	.70214 (.00438)	.69743 (.00427)
6	.74928 (.00799)	.74397 (.00734)	.73874 (.00676)	.73358 (.00524)	.72849 (.00578)	.72347 (.00539)	.71852 (.00505)	.71364 (.00477)	.70882 (.00454)	.70407 (.00436)
7	.75635 (.00884)	.75099 (.00811)	.74571 (.00744)	.74050 (.00685)	.73536 (.00632)	.73029 (.00585)	.72530 (.00544)	.72037 (.00508)	.71551 (.00478)	.71072 (.00454)
8	.76342 (.00980)	.75801 (.00898)	.75268 (.00828)	.74742 (.00755)	.74223 (.00694)	.73712 (.00640)	.73208 (.00591)	.72710 (.00549)	.72220 (.00512)	.71736 (.00481)
9	.77049 (.01085)	.76503 (.00994)	.75965 (.00911)	.75434 (.00836)	.74910 (.00767)	.74394 (.00704)	.73886 (.00649)	.73384 (.00599)	.72889 (.00555)	.72400 (.00517)
10	.77756 (.01200)	.77205 (.01101)	.76661 (.01010)	.76126 (.00926)	.75598 (.00849)	.75077 (.00779)	.74563 (.00715)	.74057 (.00658)	.73557 (.00607)	.73064 (.00561)

In this table minimum MSE may be observed for the following combinations of c and p:

c = 1, p = 6   c = 2, p = 8   c = 3, p = 9

BIBLIOGRAPHY

1. Abramowitz, M. and Stegun, I.A. (1964), "Handbook of Mathematical Functions", National Bureau of Standard Applied Mathematics, Series 35.

2. Allais, M. (1968), "Pareto, Vilfredo: I. Contributions to economics", Vol. 11. Pp. 399-441 in International Encyclopedia of Social Sciences. Edited by David L. Sills, Macmillan and Free-Press, New York.

3. Anderson, T.W. (1958), "An Introduction to Multivariate Analysis", Wiley & Sons, New York.

4. Basu, A.B. (1964), "Estimates of reliability for some distributions useful in life test", Technometrics 6, 215-219.

5. Bayes, Thomas (1958), "An essay towards solving a problem in the doctrine of chances", Biometrika 45, 293-315. (Reprinted from Philos. Trans. Roy. Soc. London, 53 (1763), 370-418, with biographical note by G.A. Barnard.)

6. Bazovski, I. (1961), "Reliability Theory and Practice", Prentice Hall.

7. Box, G.E.P. and Tiao, G.C. (1973), "Bayesian Inference in Statistical Analysis", Addison-Wesley.

8. Cheng, S.W. (1980), "On the ABLUE and the optimum t-test of the parameter of Rayleigh distributions", Tamkang Journal of Mathematics, Vol. 11, No. 1, 11-17.

9. Conover, W.J. (1971), "Practical Non Parametric Statistics", John Wiley and Sons, Inc.

10. Copson, E.T. (1971), "Asymptotic Expansions", Cambridge University Press.

11. Dawid, A.P., Stone, M. and Zidek, J.V. (1973), "Marginalization paradoxes in Bayesian and structural inference (with discussion)", J. Roy. Statist. Soc. (B), Vol. 35, No. 2, 189-233.

12. De Finetti, Bruno (1964), "Foresight: its logical laws, its subjective sources, studies in subjective probability", Henry E. Kyburg, Jr. and Howard E. Smokler, eds., Wiley, New York, 93-158. (Translation of La prevision: ses lois logiques, ses sources subjectives, Ann. Inst. H. Poincare, 7 (1937), I-68.)

13. De Finetti, Bruno (1974), "Theory of Probability", Vol. 1, Wiley, London.

14. De Finetti, Bruno (1975), "Theory of Probability", Vol. 2, Wiley, London.

15. DeGroot, M.H. (1970), "Optimal statistical decisions", McGraw-Hill Book Co., New York.

16. Dunsmore, I.R. (1974), "The Bayesian predictive distribution in life testing models", Technometrics, 16, 455-460.

17. Dyer, D.D. and Whisenand, C.W. (1973), "Best linear unbiased estimator of the parameter of the Rayleigh distribution", Part I and II, IEEE Transactions on Reliability, Vol. R-22, Nos. 1 and 4.

18. Edwards, W., Lindman, H. and Savage, L.J. (1963), "Bayesian Statistical Inference for Psychological Research", Psychological Review, Vol. 70, 193-242.

19. Elderton, W.P. and Johnson, N.L. (1969), "Systems of frequency curves", Cambridge University Press.

20. El-Sayyad, G.M. (1967), "Estimation of the parameters of an exponential distribution", Journal of the Royal Statistical Society, Series B, Vol. 29, 525-532.

21. Erdelyi, A. et al (1953), "Higher Transcendental Functions", Vol. II. McGraw-Hill, Inc.

22. Evans, I.G. (1964), "Bayesian estimation of the variance of a normal distribution", J. Roy. Statist. Soc. (B), Vol. 26, 63-68.

23. Evans, I.G. (1976), Unpublished Lecture Notes, University College of Wales, Aberystwyth, U.K.

24. Evans, I.G. and Nigm, A.H.M. (1980), "Bayesian prediction for the left-truncated exponential distribution", Technometrics, 22, 201-204.

25. Fisher, R.A. (1922), "On the mathematical foundations of theoretical statistics", Philos. Trans. of the Royal Society of London, Series A, 222, 309-368.

26. Goodman, L.A. (1953), "A simple method for improving some estimators", Ann. Math. Statist., Vol. 24, 114-117.

27. Goodman, L.A. (1960), "A note on the estimation of variance", Sankhya, Vol. 22, 221-228.

28. Hewitt, E. and Savage, L.J. (1955), "Symmetric measures on cartesian products", Trans. Amer. Math. Soc., 80, 470-501.

29. Hogg, R.V. and Craig, A.T. (1978), "Introduction to Mathematical Statistics", Third Edition, Macmillan, New York.

30. Jeffreys, Sir Harold (1961), "Theory of Probability", Oxford, Clarendon Press.

31. Kendall, M.G. and Stuart, A. (1973), "The Advanced Theory of Statistics", Vol. 2, Third Edition, Hafner Publishing Co., New York.

32. Lawless, J.F. (1977), "Prediction intervals for the two-parameter exponential distribution", Technometrics, 19, 469-472.

33. Likes, J. (1974), "Prediction of s^{th} ordered observation for the two-parameter exponential distribution", Technometrics, 16, 241-244.

34. Lindley, D.V. (1965), "Introduction to Probability and Statistics, Part 2", Cambridge University Press.

35. Lindley, D.V. (1971), "The estimation of many parameters". In foundation of Statistical Inference, (V.P. Godambe and D.A. Sprott, eds.) 435-455, Toronto: Holt, Rinehart and Winston.

36. Lindley, D.V. (1972), "Bayesian Statistics, A Review", SIAM, Philadelphia, Pennsylvania.

37. Lindley, D.V. (1975), "The future of Statistics - A Bayesian 21st Century", Supp. Adv. Appl. Prob., Vol. 7, 106-115. Applied Probability Trust.

38. Lindley, D.V. (1978), "The Bayesian Approach", Scandinavian Journal of Statistics, Vol. 5, 1-26.

39. Lindley, D.V. (1980), "Approximate Bayesian methods", Trabajos Estadistica, 31, 223-237.

40. Lindley, D.V. and Novick, M.R. (1981), "The role of exchangeability in inference", Ann. Statist., Vol. 9, No. 1, 45-58.

41. Lindley, D.V. and Smith, A.F.M. (1972), "Bayes estimates for the linear model (with discussion)", Journal of the Royal Statistical Society, Series B, 34, 1-42.

42. Maddala, G.S. (1977), "Econometrics", McGraw-Hill, New York.

43. Mosteller, R., Rourke, R.E.K., Thomas, Jr., G.B. (1970), "Probability with Statistical Applications", Second Edition, Addison-Wesley Publishing Co., Reading, Massachusetts.

44. Nielson, K.L. (1965), "Methods in Numerical Analysis", Second Edition, MacMillan and Co., New York.

45. Pearson, E.S. and Hartley, H.O. (1966), "Biometrika tables for statisticians, Vol. 1", Cambridge University Press.

46. Pearson, E.S. and Johnson, N.L. (1968), "Tables of the incomplete beta function", Cambridge University Press.

47. Peizer, D.B. and Pratt, J.W. (1966), "Approximating the binomial, F and commonly used related distributions, I", Technical Report No. 12, Department of Statistics, Harvard University.

48. Polovko, A.M. (1968), "Fundamentals of reliability theory", Academic Press, New York.

49. Raiffa, H. and Schlaifer, R. (1961), "Applied Statistical Decision Theory", Harvard University,

50. Savage, L.J. et al (1962), "The foundations of Statistical Inference: A Symposium", Wiley, New York.

51. Savage, L.J. (1962), "Bayesian statistics in decision and information processes", New York, MacMillan and Co.

52. Siddiqui, M.M. (1962), "Some problems connected with Rayleigh distributions", Journal of Research of the National Bureau of Standards-D, Vol. 66D, No. 2.

53. Sinha, S.K. (1980), "Robustness of posteriors under improper priors in some life testing models", Journal of the Indian Statistical Association, Vol. 18, 127-144.

54. Sinha, S.K. and Guttman, I. (1976), "Bayesian inference about the reliability for the exponential distributions", Communications in Statistics - Theory. Meth. A5(5), 471-479.

55. Sinha, S.K. and Howlader, H.A. (1980), "On the sampling distributions of the Bayesian estimators of the Pareto-parameter with proper and improper priors and associated goodness of fit", Journal of the Society of Management Sciences and Applied Cybernetics, Vol. 9, No. 1, 31-60.

56. Theil, H. (1963), "On the use of incomplete prior information in regression analysis", Journal of the American Statistical Association, Vol. 58, 401-414.

57. Tiao, G.C. and Zellner, A. (1964), "Bayes' theorem and the use of the prior knowledge in regression analysis", Biometrika, Vol. 51, 219-230.

58. Wald, A. (1950), "Statistical Decision Functions", Wiley, New York.

59. Weintraub, S. (1963), "Tables of the cumulative binomial probability distribution for small values of p", Free Press of Glencoe, New York.

60. Wetherill, G.B. (1961), "Bayesian Sequential Analysis", Biometrika, 48, 281-292.

61. Zellner. A. (1971), "An Introduction to Bayesian Inference in Econometrics", Wiley, New York.

BIOGRAPHICAL DATA

Hatem Ali Howlader was born in Charkaya, Jurkathi, Barisal, Bangladesh, on December 31, 1946. He attended A.K. institution, Barisal, and B.M. College, Barisal, to obtain Secondary School Certificate (S.S.C.) and Higher Secondary Certificate (H.S.C.) in 1963 and 1965, respectively. He obtained his B.Sc. Honours and M.Sc. degrees in Statistics from the University of Dacca, Bangladesh in 1968 and 1970, respectively. Prior to his joining the Department of Statistics, University of Chittagong as a lecturer in December, 1970, he worked for a brief period of nine months at the Institute of Statistical Research and Training (ISRT), University of Dacca, as a Research Associate. In April, 1973, he was promoted to the rank of Assistant Professor of Statistics in the same Department and worked there till his departure for Aberystwyth, Wales in October 1976. In 1977 he obtained his Diploma in Statistics from the University College of Wales, Aberystwyth, and joined the University of Manitoba, Winnipeg, Canada, as a Ph.D. student in September 1977. While working on his Ph.D. Programme, he worked as a teaching assistant, Department of Statistics, University of Manitoba, in 1977-78, and as a Sessional Lecturer from 1978-1981. He is now working as a Lecturer in the Department of Statistics, University of Winnipeg, Winnipeg, Manitoba, Canada.