CERTAIN NON-CLASSICAL INFERENCE PROCEDURES APPLIED TO THE INVERSE GAUSSIAN DISTRIBUTION

By

JOHN MARK PALMER

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1967

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1970

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1973

CERTAIN NON-CLASSICAL INFERENCE PROCEDURES

APPLIED TO THE INVERSE GAUSSIAN

DISTRIBUTION

Thesis Approved:

Lyle D. Broemeling
Thesis Adviser

J. Leroy Folks

P. L. Claypool

J P Chandler

John R. Bosworth

D D Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

The author would like to express his sincere appreciation to the statistics faculty for their interest, advice, and direction in the preparation of this study. In particular, thanks must go to Professors Folks, Chandler, Claypool, and Bosworth for serving on my advisory committee, and to Professor Broemeling for directing the study.

In addition, the author wishes to express his gratitude to Drs. Funk, Holbert, Browne, Brandt, and Murphy, who provided encouragement and contributed to the author's understanding of many facets of statistics.

The author desires to extend his thanks to Mrs. Ann Waughtal for her conscientious efforts in typing this material and assisting in presenting this study in its final form.

It is difficult to judge the impact of different personalities on one's life, but the author feels that his father and family can take credit for providing a home atmosphere conducive to the development of an inquisitive personality.

Finally, no tribute is really sufficient to express the author's gratitude to his wife, Linda, for her steadfast love and support. Truly, John Riley had no better.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

Table | Page

LIST OF FIGURES

Figure | Page

CHAPTER I

INTRODUCTION

The usual parametrization of the distribution of a (random variable) X which follows an inverse Gaussian probability law results in the representation of the density as:

$$f(x;\mu,\lambda) = \begin{cases} \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right] & x > 0 \\ 0 & \text{elsewhere} \end{cases} \qquad (1.1)$$

where both μ and λ are positive real numbers. A considerable body of knowledge concerning this distribution is available. Its characteristic function and distribution function are known in closed form as are its moments. The usual moments of interest are the mean (μ) and variance (μ^3/λ). The density is positively skewed.

This probability law apparently was first encountered by Schrodinger (23) in 1915 in the study of Brownian motion. He determined that the time required for a particle under one dimensional Brownian motion with constant velocity v to travel a distance d follows (1.1) with parameters $\mu = d/v$ and $\lambda = d^2/\beta$ where β is the diffusion constant. Following this Wald (26) encountered the inverse Gaussian distribution as a limiting form of the distribution of the sample size of certain procedures in his development of sequential analysis. These densities are of the same form as (1.1) with $\mu = 1$ and are frequently referred to as "Wald distributions."

The emergence of this probability law as a distribution of general practical import was due to Tweedie (25) who noted a relationship between the cumulant generating functions of the Gaussian and inverse Gaussian distributions and so named the latter. It is, as Schrodinger observed, the waiting-time distribution for a Gaussian process as the gamma distribution is the waiting time distribution for a Poisson process. Tweedie studied this distribution extensively and sparked sufficient interest in it to attract the attention of Wasan and Roy (27) and Shuster (24) among others. More recently Chhikara developed UMP and UMPU tests for the parameters of this distribution and studied its usage in reliability where an object is subjected to a stress that can be characterized by a Gaussian process. A summary of the classical procedures available for dealing with this distribution is given by Johnson and Kotz (12) and Chhikara (4).

There appears to be little in the way of positive results concerning the inverse Gaussian that have been developed through Bayesian procedures. Hunt (10) in 1971 considered a Bayesian approach to not only the (μ, λ) parametrization but also several others with the objective of making inferences about the mean of the distribution. He established that the natural conjugate prior for the mean failed to exist and that the posterior distribution obtained with at least one type of diffuse prior failed to exist. The author has independently verified these assertions. Nevertheless, this does not preclude the use of nonclassical techniques. This study is divided into two sections the first of which is concluded in Chapter IV and which pertains to the usual Bayesian approach as applied to this distribution. Chapter II constitutes the a priori analysis and modifications of the usual Bayesian techniques. Chapter III is concerned with point estimation and posterior probabilities. Chapter IV presents some results and algorithms for determining Highest Posterior Density regions for one of the parameters when the other is either known or of no interest. The second section deals with an empirical Bayesian approach to the analysis of the inverse Gaussian probability law. Chapter V provides an introductory framework for the Empirical Bayes analysis which follows. Chapter VI is directed toward point estimation and Chapter VII introduces EB test procedures and related them to those used in the non-Bayesian framework. Chapter VIII attempts to assess the utility of the procedures derived and considers what areas would seem profitable to pursue further.

CHAPTER II

A PRIORI ANALYSIS

The Bayesian approach to inference is based on the concept of regarding some or all of the parameters of a population to be random variables. The probability structure associated with these variates usually has a different interpretation than that of classical analysis. In the Bayesian analysis the probability law describing the random parameter does not necessarily represent a frequency distribution but rather a distribution of beliefs held by the user. The latter point of view will be adopted by the author in the following.

Definition 2.1: A prior distribution $\Pi_0(\theta)$ on the parameter θ of the distribution $F(x|\theta)$ is a nonnegative measure defined on the measurable space (Θ, S) which represents an experimenter's subjective measure of confidence that $\theta \epsilon A$ for any $A \epsilon S$ prior to the performance of an experiment resulting in a realization(s) of the random variable X.

One may visualize that for a given parameter under consideration the Bayesian would have a limitless number of prior distributions to select from in representing his beliefs. In point of fact the literature indicates that only two types of prior distributions have found much prominence. These are the natural conjugate prior and the diffuse prior distributions. The author will likewise limit the analysis to these two classes of prior distributions.

<u>Definition 2.2</u>: Let X be a r.var. with c.d.f. $F(x;\theta)$ where $\theta\epsilon\Theta$. Let $t(x)$ be sufficient for θ with $t(x)\epsilon A$. Then $dF(x;\theta) = h(x)dG(t(x);\theta)$ by the Neymann-Fisher factorization theorem as referenced in Hogg and Craig (9). The factor $dG(t(x);\theta)$ is a kernel of $F(x;\theta)$. Suppose $\int G(t(x);d\theta) < \infty$ then the distribution function defined by $\pi_o(\theta') = G(\alpha;\theta')/\int G(\alpha,d\theta)$ for $\alpha\epsilon A$ and $\theta\epsilon\Theta$ is a natural conjugate prior of $F(x;\theta)$. A natural conjugate distribution may not exist and may not be unique. If $\pi_o(\theta)$ exists it may be "enriched" either by increasing the range of α or by introducing additional parameters as explained in Raiffa and Schlaifer (19).

<u>Definition 2.3</u>: Let X be a r.var. with c.d.f. $F(x;\theta)$. Let $I(\theta)=(i_{mn})$ be the Fisher information matrix where $i_{m,n}=-E\{D_{\theta m\theta n}\log D_x F(x;\theta)\}$. If $K=\iint F(x;\theta)\ \sqrt{\det I(\theta)}\ dxd\theta<\infty$ then $\pi_J(\theta)=K^{-1}(\det I(\theta))^{\frac{1}{2}}$ is the Jeffrey's invariant prior distribution for $F(x;\theta)$. The use of the Jeffrey prior is partially based on the assumption that $\pi_J(\theta)$ is not θ-integrable.

<u>Definition 2.4</u>: Assume $\Pi_o(\theta)$ is a prior distribution on parameter θ of the distribution $F(x;\theta)$ such that $\int_{-\infty}^{\theta} F(x;\theta_o)\Pi_o(d\theta_o)$ is a probability measure on $\Theta\times X$. Assume an experiment is conducted resulting in realizations of the random vectors (x_i,θ_j) $i=1,2,\ldots,m$; $j=1,2,\ldots,n$; $n\leq m$; and where the θ_j's are i.i.d.$\pi_o(\theta)$ and $X_i \mid \theta_1,\theta_2\cdots{}_n$ and $X_{i'}|\theta_1,\theta_2\ldots\theta_n$ are independent for $i\neq i'$. Then the posterior distribution of $(\theta_1,\theta_2,\ldots\theta_n)$ is the conditional distribution of $\theta_1,\theta_2,\ldots\theta_n \mid X_1,X_2,\ldots X_m$.

It is usually assumed that θ is fixed throughout the experiment so that n+1 and the posterior distribution is defined on a spece whose dimension is the same as that of Θ. For example, suppose X is absolutely continuous with density $f(x;\theta)$. Suppose also that Θ is thought to be described by a prior density $\pi_o(\theta)$. An experiment is conducted resulting in realizations $x_1, x_2, \ldots x_n$. Then the posterior density of Θ is given by

$$\pi(\theta) = \frac{\left(\prod_{i=1}^{n} f(x_i;\theta)\right) \pi_o(\theta)}{\int \left(\prod_{i=1}^{n} f(x_i;\theta)\right) \pi_o(\theta) d\theta}$$

The posterior distribution represents the experimenter's subjective measure of confidence regarding the possible values of θ in light of the experimental data.

Before applying these concepts to the inverse Gaussian distribution the author will briefly discuss the two types of prior distributions previously introduced.

The popularity of using the natural conjugate prior distribution appears to be based on the property that the posterior distribution is usually of the same form as the prior distribution. This has appeal in two respects. First, the posterior distribution is easy to find since its functional form is known. Second, there is an appeal to a consistency of beliefs of the experimenter. It is felt that the experimenter's beliefs should not be altered so drastically by the data as to result in a change of distributional families.

The Jeffrey's prior is used to represent a prior state of no knowledge in that no probability statement about beliefs regarding the value of the parameter can be derived from such a prior distribution. This type of diffuse prior distribution is used because of its invariance properties. Hartigan has shown that the Jeffrey's prior possesses six invariance properties and the interested reader is referred to Zellner (28) for an enumeration of these results. The author will limit his remarks to that property originally discovered by Jeffrey (11). Briefly this property is that all posterior probability statements about the parameter of a distribution are invariant under any bijective differentiable transformations of the parameter. This would seem to be a requisite property of a diffuse prior distribution.

1. Conjugate Prior Distributions

The first case to be considered is that for which both μ and λ are unknown. The existence of the natural conjugate distribution for this situation will now be discussed.

The density function of the inverse Gaussian can be written in the form:

$$f(x;\mu,\lambda)=(2\pi)^{-1/2}x^{-3/2}\left[\lambda^{1/2}\exp\left(-\frac{\lambda}{2}\mu^{-2}x+\lambda\mu^{-1}-\frac{\lambda}{2}x^{-1}\right)\right] \qquad (2.1)$$

The statistic (x,x^{-1}) is a joint sufficient statistic for the parameters (μ,λ) and a kernel of $f(x;\mu,\lambda)$ is the second factor within brackets. If a natural conjugate prior exists it will necessarily be of the same form as that of a kernel so that:

$$\pi_o^*(\mu,\lambda)\propto\lambda^d\exp\left(-\lambda(a\mu^{-2}-b\mu^{-1}+c)\right) \qquad o<\mu<\infty \quad (2.2)$$
$$o<\lambda<\infty$$
$$=o \qquad \text{elsewhere.}$$

Since $\frac{\lambda(x-\mu)^2}{2\mu^2x}>o$ it is necessary that $a\mu^{-2}-b\mu^{-1}+c>o$.

To accomplish this it is required that $a>o$, $b^2-4ac<o$.

In addition it is necessary that $d > -1$ in order that $\pi_o(\mu,\lambda)$ be λ integrable. The corresponding posterior distribution would then be of the same form with parameters

$$a' = a + \frac{n\bar{x}}{2},\ b' = b + n,\ c' = c + \frac{1}{2}\sum_{i=1}^{n} X_i^{-1} \text{ and } d' = d + \frac{n}{2}.$$

Let $\lambda = \lambda_o$ and consider $\lim_{\mu\to\infty} \pi_o(\mu,\lambda_o)$. The limit of the exponent of $\pi_o(\mu,\lambda_o)$ is zero so that $\lim_{\mu\to\infty} \pi_o(\mu,\lambda) = \text{constant}$. Consequently, $\pi_o(\mu,\lambda)$ is not μ-integrable for any λ and therefore the natural conjugate prior does not exist.

Since the conjugate prior fails to exist one may: (1) just discard the concept for the inverse Gaussian; (2) reparametrize to a parametrization for which a natural conjugate exists; (3) force integrability on the functional form already obtained. The first is not constructive nor in general is the second. A large body of knowledge is available about the (μ,λ) parametrization and it will be beneficial, in the author's opinion, to remain with this representation as much as possible.

One may retain the essential feature of the natural conjugate prior by truncating (2.2) with respect to μ. This gives:

$$\pi_o(\mu,\lambda) \propto \lambda^d \exp - \lambda(a\mu^{-2} - b\mu^{-1} + c) \qquad o < \mu < p,\ o < \lambda < \infty \tag{2.3}$$

$$= o \qquad \text{elsewhere}$$

It is the author's belief that if one possesses sufficient knowledge concerning μ to use the functional form above and specify the required four parameters then it seems reasonable that one will be able to place an upper bound p on μ.

Let $Q(\mu) = a\mu^{-2} - b\mu^{-1} + c$. For any $\mu\varepsilon(o,p)$ $\pi_o(\mu,\lambda)$ is seen to be in the form of a gamma density so that:

$$I(\mu) = \int_0^{\infty} \pi_o(\mu,\lambda)d\lambda \propto \frac{\Gamma(d + 1)}{[Q(\mu)]^{d+1}} \propto \frac{\mu^{2(d+1)}}{(a - b\mu + c\mu^2)^{d+1}} \qquad (2.3.1)$$

Therefore $\lim_{\mu \to 0} I(\mu) = 0$ so that $I(\mu)$ is bounded at zero and since $b^2 - 4ac < 0$ the denominator of $I(\mu)$ is nonzero for real μ. Hence $I(\mu)$ is bounded and continuous on (o,p) and therefore integrable. By Tonelli's theorem one can then conclude that $\pi_o(\mu,\lambda)$ is (μ,λ) integrable and therefore a proper density.

It should be recognized that $\pi_o(\mu,\lambda)$ as defined differs from the natural conjugate prior also in that the parameter p does not change in response to experimental data. This can be rectified by taking $p' = \max(\bar{x}, p)$.

Now suppose λ is known and μ is the parameter of interest. Examining equation (2.1) it is easy to see that x is sufficient for μ and that the functional form of the natural conjugate is:

$$\pi_o^*(\mu) \propto \exp - \mu^{-1}(a\mu^{-1} - b) \qquad o < \mu < \infty$$
$$= o \qquad \text{elsewhere}$$

where $a > o$ and $b > o$.

The posterior distribution is of the same form with $a' = a + \frac{\lambda \Sigma X_i}{2}$ and $b' = b + n\lambda$. As in the case with both parameters unknown, $\pi_o^*(\mu)$ is not a proper density since it is not integrable.

Therefore the same approach will be used here as in the preceding case. The range of μ is assumed to be truncated thereby defining

$$\pi_o(\mu) \propto \exp[- \mu^{-1}(a\mu^{-1} - b)] \qquad o < \mu < p \qquad (2.4)$$
$$o \qquad \text{elsewhere}$$

The last case considered is the development of the natural conjugate prior when μ is known and λ is unknown. If one writes the inverse Gaussian density in its usual form as

$$f(x;\mu,\lambda) = \frac{\lambda^{\frac{1}{2}}}{\sqrt{2\pi x^3}} \exp - \lambda \frac{(x-\mu)^2}{2\mu^2 x}$$

then a kernel of the Neyman-Fisher factorization is $\lambda^{\frac{1}{2}} \exp - \frac{\lambda(x-\mu)^2}{2\mu^2 x}$ so that the natural conjugate is of the form $\pi_o^*(\lambda) \propto \lambda^p \exp(-a\lambda)$. This is recognized as a gamma density. Therefore take

$$\pi_o(\lambda) = \frac{a^p \lambda^{p-1} e^{-a\lambda}}{\Gamma(p)} \qquad o < \lambda < \infty \tag{2.5}$$

$$= o \qquad \text{elsewhere}$$

where $a > o$, $p > -1$. The posterior distribution of λ is then gamma distributed with parameters $a' = a + \frac{1}{2}\mu^{-2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{x_i}$ and $p' = p + \frac{n}{2}$.

2. Jeffrey's Prior Distributions

It may be the case that an experimenter's prior knowledge is not sufficient to warrant the use of an informative prior such as those previously developed. In such cases one may then consider the Jeffrey prior distribution to represent a lack of knowledge regarding the parameters of interest. Accordingly such representations will now be developed.

As stated in definition 2.3 it is necessary to build the Fisher information matrix.

Letting f represent the inverse Gaussian density function then apart from terms of no consequence $\log f = \frac{1}{2}\log\lambda - \frac{\lambda(x-\mu)^2}{2x\mu^2}$ giving:

$$D_\lambda \log f = \frac{1}{2\lambda} - \frac{1}{2x}\frac{(x-\mu)^2}{\mu^2},$$

$$D^2_{\lambda\lambda} \log f = -\frac{1}{2\lambda^2},$$

$$D_\mu \log f = \frac{-\lambda}{2x}\left\{\frac{-2\mu^2(x-\mu) - 2\mu(x-\mu)^2}{\mu^4}\right\} = \frac{\lambda(x-\mu)}{\mu^3},$$

$$D^2_{\mu\mu} \log f = \lambda[-3\mu^{-4}(x-\mu) - \mu^{-3}] = \frac{-\lambda(3x-2\mu)}{\mu^4},$$

$$D^2_{\mu\lambda} \log f = \mu^{-3}(x-\mu).$$

Taking expectations with respect to X produces:

$$-E_x D^2_{\lambda\lambda} \log f = (2\lambda^2)^{-1} \qquad -E_x D^2_{\mu\mu} \log f = \frac{3\lambda}{\mu^4} E_x X - \frac{2\lambda}{\mu^3} = \lambda\mu^{-3}$$

$$-E_x D^2_{\mu\lambda} \log f = \frac{-1}{\mu^3} E_x X + \frac{1}{\mu^2} = 0 \qquad \text{so that}$$

$$I(\mu,\lambda) = \begin{pmatrix} (2\lambda^2)^{-1} & 0 \\ 0 & \lambda\mu^{-3} \end{pmatrix} \tag{2.6}$$

Therefore, $\pi_J(\mu,\lambda) \propto (\det I(\mu,\lambda))^{\frac{1}{2}} = (2\lambda\mu^3)^{-\frac{1}{2}}$ or equivalently

$$\pi_J(\mu,\lambda) \propto \begin{cases} \lambda^{-1/2}\mu^{-3/2} & \mu > 0 \quad \lambda > 0 \\ 0 & \text{elsewhere} \end{cases}$$

This distribution is improper since it is not λ-integrable.

To determine the posterior distribution the likelihood is formed giving:

$$L(\mu,\lambda;\vec{x}) = (2\pi)^{-n/2}\lambda^{n/2}\prod_{i=1}^{n} x_i^{-3/2} \exp - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}$$

from which the posterior is obtained through the relationship $\pi(\mu,\lambda) \propto L(\mu,\lambda;\vec{x})\pi_J(\mu,\lambda)$ which apart from factors which do not affect the normalizing constant is

$$\pi(\mu,\lambda) \propto \lambda^{\frac{n-1}{2}} \mu^{-3/2} \exp\left[-\lambda\left(\frac{\Sigma X_i}{2}\mu^{-2} - n\mu^{-1} + \tfrac{1}{2}\Sigma X_i^{-1}\right)\right] \tag{2.7}$$

It is necessary to demonstrate that the above can be normed and is a proper density. Substitute $a = \frac{\Sigma X_i}{2}$ and $b = \frac{1}{2} \Sigma X_i^{-1}$ in (2.7) to simplify notation. Since $\pi(\mu, \lambda)$ is an algebraic composition of continuous functions it is itself continuous on $T = (0, \infty) \times (0, \infty)$ and therefore measurable over the open first quadrant. Since $\pi(\mu, \lambda)$ is nonnegative in this region Tonelli's theorem as referenced in Royden (19) may be applied to conclude that:

$$\iint \pi(\mu, \lambda) d(\mu, \lambda) = \int_0^\infty \left\{ \int_0^\infty \pi(\mu, \lambda) d\lambda \right\} d\mu$$

The first iterated integral can be integrated by the definition of the gamma function giving:

$$\int_0^\infty \pi(\mu, \lambda) d\lambda \propto \mu^{-3/2} \int_0^\infty \lambda^{\frac{n-1}{2}} \exp - \lambda(a\mu^{-2} - n\mu^{-1} + b) d\lambda$$
$$= \frac{\Gamma\left(\frac{n+1}{2}\right) \mu^{n-1/2}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}} \qquad (2.8)$$

It is now sufficient to establish that the above is μ-integrable.

Write the relevant portion of (2.8) as

$$\int_0^\infty \frac{\mu^{n-1/2}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}} d\mu = \int_0^h \frac{\mu^{n-1/2} d\mu}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}} + \int_h^\infty \frac{\mu^{n-1/2} d\mu}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}} \qquad (2.9)$$

where $h > 0$. Recall that $a\mu^{-2} - n\mu^{-1} + b$ being an expansion of $\frac{\Sigma(x_i - \mu)^2}{2\mu^2 x_i}$ is nonnegative. Therefore if $a\mu^{-2} - n\mu^{-1} + b = 0$ the root must be double. Examining the discriminant yields as a necessary consequence $n^2 - (\Sigma X_i)(\Sigma X_i^{-1}) = 0$ which occurs only if

$x_1 = x_2 = \cdots = x_n = 1$ an event with probability zero. Therefore, $a - n\mu + b\mu^2 > o$ almost surely and the integrand of (2.8) has no discontinuity point in the range of integration. The first summand on the right in (2.9) clearly exists since the integrand is bounded and continues in (o, h). The second integral exists by the limit comparison test. The following justifies this statement.

$$\lim_{\mu \to \infty} \frac{\dfrac{\mu^{n-1/2}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}}}{\dfrac{1}{\mu^{3/2}}} = \lim_{\mu \to \infty} \frac{\mu^{n+1}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}}$$

$$= \lim_{v \to o} \frac{v^{n+1}}{v^{n+1}(av^2 - nv + b)} = \frac{1}{b^{\frac{n+1}{2}}} > o$$

Therefore, since $\int_h^\infty \frac{1}{\mu^{3/2}}\, d\mu$ exists so also does $\int_h^\infty \frac{\mu^{n-\frac{1}{2}}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}}\, d\mu$.

Applying Tonelli's theorem establishes that $\pi(\mu, \lambda)$ is proper density function.

The foregoing is appropriate when both parameters are of interest. Consider now the case when μ is known and λ is unknown. If one refers to definition 2.3 and equation (2.6) then the first element of the principal diagonal of $I(\mu, \lambda)$ indicates the form of the Jeffrey's prior. That is $\pi_J(\lambda) \propto \frac{1}{\lambda}$. Forming the product of the likelihood and prior gives:

$$\pi(\lambda) \propto \left(\frac{\lambda}{2\pi x^3}\right)^{n/2} \exp\left[\left(-\frac{\lambda}{2\mu_o^2} \frac{(x_i - \mu)^2}{x_i}\right)\right] \frac{1}{\lambda} \qquad (2.10)$$

Inspection of (2.10) shows that it is λ-integrable being of the gamma family. In fact λ is gamma distributed with parameters $a = a\mu_o^{-2} - n\mu_o^{-1} + b$ and $p = \frac{n}{2}$ or equivalently $2(a\mu_o^{-2} - n\mu_o^{-1} + b)\lambda \sim \chi^2(n)$.

Likewise when $\lambda = \lambda_o$ and μ is unknown the form of Jeffrey's non-informative prior on μ is proportional to the square root of the second element of the principal diagonal of the information matrix in (2.6). This gives $\pi_J(\mu) \propto \lambda_o\mu^{-3/2}$ from which the posterior is determined as above to be of the form:

$$\pi(\mu) \propto \mu^{-3/2} \exp\left(-\lambda_o\mu^{-1}(a\mu_o^{-1} - n)\right) \quad o < \mu < \infty \qquad (2.11)$$

Compare (2.11) with (2.7) and one observes they are of the same form with respect to μ-integrability. It has been demonstrated that (2.7) is a μ-integrable so likewise is (2.11).

This concludes the a priori analysis as limited to this study. The determination of normalizing constants is deferred to the chapter concerning point estimation.

CHAPTER III

POINT ESTIMATION

A posterior distribution of a parameter is the Bayesian's basis for inferences concerning that parameter. It is the purpose of this chapter to refine the work of Chapter II by determining normalizing constants for the sundry posterior distributions therein derived and developing point estimates for these parameters. The author will develop these quantities in closed form to the extent of his abilities. However, it will be apparent that even in cases for which such closed expression exist numerical procedures would be more versatile. The reason for this is that the requisite numerical procedures are to a much greater degree independent of the sample size.

1. Point Estimates of μ Averaged Over λ With an Informative Prior Distribution

Suppose now that the experimenter's prior knowledge is expressed by the truncated conjugate prior given by (2.3) and that although λ is unknown it is of no interest to the experimenter. The investigator will then be concerned with the marginal of μ given by (2.3.1) as

$$\pi(\mu) \propto \frac{\mu^{2(d+1)}}{(a - b\mu + c\mu^2)^{d+1}} \quad o < \mu < p \qquad (3.1)$$

$$= \quad o \quad \text{elsewhere}$$

and where $b^2 - 4ac < o$, $a > o$, and $d > -1$. It is of interest to determine $\int_o^p \mu^k \pi(\mu)d\mu$ for $k = o$ (normalizing constant), $k = 1$ (mean), and $k = 2$ (variance). These integrals can be determined in terms of elementary functions. The author will assume that d is a nonnegative integer assuming that the evaluations so obtained are reasonable approximations for other real permissable values of d.

Applying the transformation $x = \frac{1}{\mu}$ to the definite integral $\int_o^p \mu^k \pi(\mu)d\mu$ results in:

$$\int_{1/p}^{\infty} \frac{dx}{x^{k+2}(ax^2 - bx + c)^{d+1}} \tag{3.2}$$

Let $X = ax^2 - bx + c$, $g = 4ac - b^2$ and utilize the following:

$$\int \frac{dx}{X} = \frac{2}{\sqrt{g}} \text{Arctan } \frac{2ax + b}{\sqrt{g}} , \tag{3.3}$$

$$\int \frac{dx}{xX} = \frac{1}{2c} \log \frac{x^2}{X} - \frac{b}{2c} \int \frac{dx}{X} , \tag{3.4}$$

$$\int \frac{dx}{X^{n+1}} = \frac{2ax + b}{ngX^n} + \frac{2(2n - 1)}{gn} \int \frac{dx}{X^n} , \tag{3.5}$$

$$\int \frac{dx}{xX^n} = \frac{1}{2c(n - 1)X^{n-1}} - \frac{b}{2c} \int \frac{dx}{X^n} + \frac{1}{c} \int \frac{dx}{xX^{n-1}} , \tag{3.6}$$

$$\begin{aligned} \int \frac{dx}{x^m X^{n+1}} = &- \frac{1}{(m - 1)cx^{m-1}X^n} \\ &- \frac{n + m - 1}{m - 1} \frac{b}{c} \int \frac{dx}{x^{m-1}X^{n+1}} \\ &- \frac{2n + m - 1}{m - 1} \frac{a}{c} \int \frac{dx}{x^{m-2}X^{n+1}} . \end{aligned} \tag{3.7}$$

Note that (3.2) is of the form (3.7). Equations (3.5) and (3.6) will be used to reduce (3.7) to a manageable form and equations (3.3) and (3.4) will be used for the final evaluation.

The author deems it expedient to simplify the present notation. Therefore in formula (3.6) put $\alpha_n = \frac{1}{2cnX^n}$, $\beta = \frac{b}{2c}$, $\gamma = \frac{1}{c}$, and $\varphi_n = \int \frac{dx}{xX^n}$. Applying the reduction formula in (3.6) n - 1 times gives:

$$\varphi_n = \sum_{k=o}^{n-2} \alpha_{n-k-1} \gamma^k - \beta \sum_{k=1}^{n-1} \gamma^{k-1} \theta_{n-k+1} + \gamma^{n-1} \varphi_1 \tag{3.8}$$

where $\theta_n = \int \frac{dx}{X^n}$ in formula (3.5). In addition put $\delta_n = \frac{2ax + b}{ngX^n}$ and $\epsilon_n = \frac{2(2n - 1)a}{ng}$ in equation (3.5) Repeated application of the recurrence relation (3.5) yields after k iterations that:

$$\theta_{n+1} = \delta_n + \sum_{j=1}^{k-1} \left(\prod_{i=o}^{j-1} \epsilon_{n-i}\right) \delta_{n-j} + \left(\prod_{i=o}^{k-1} \epsilon_{n-i}\right) \theta_{n-k+1} \tag{3.9}$$

and for k = 1 the second summand vanishes. One may now substitute equation (3.9) in equation (3.8) which results in the equation for φ_n below.

$$\varphi_n = \sum_{k=o}^{n-2} \alpha_{n-k-1} \gamma^k - \beta \sum_{k=1}^{n-1} \gamma^{k-1} \left\{\delta_{n-k} + \sum_{j=1}^{n-k-1} \left(\prod_{i=o}^{j-1} \epsilon_{n-k-i}\right)\delta_{n-k-j} + \left(\prod_{i=o}^{n-k-1} \epsilon_{n-k-i}\right) \theta_1 \right\} + \gamma^{n-1} \left\{\frac{1}{2c} \log \frac{x^2}{X} - \frac{b}{2c} \theta_1\right\} \tag{3.10}$$

Note that by equation (3.3) $\theta_1 = \frac{2}{\sqrt{g}} \text{Arctan} \frac{2ax + b}{\sqrt{g}}$

Now in equation (3.7) put $\psi_m = \int \frac{dx}{x^m X^{n+1}}$ $\pi_m = \frac{-1}{mcx^m X^n}$ $\tau_m = - \frac{n + m}{m} \frac{b}{c}$, and $\omega_m = - \frac{2n + m}{m} \frac{a}{c}$.

The reader will please observe that $\psi_0 = \theta_n$ and $\psi_1 = \varphi_{n+1}$.

The evaluation of the integral in (3.2) for k = 0, k = 1, and k = 2 can be accomplished by determining ψ_2, ψ_3, and ψ_4 respectively. Repeatedly using the reduction formula (3.7) generates:

$$\psi_2 = \pi_1 + \tau_1\varphi_{n+1} + \omega_1\theta_n,$$

$$\psi_3 = \pi_2 + \tau_2\pi_1 + (\tau_1\tau_2 + \omega_2)\varphi_{n+1} + \omega_1\tau_2\theta_n,$$

$$\psi_4 = \pi_3 + \tau_3\pi_2 + \tau_2\tau_3\pi_1 + \omega_3\pi_1 + (\tau_1\tau_2\tau_3 + \tau_3\omega_2 + \tau_1\omega_3)\varphi_{n+1} + (\tau_2\tau_3\omega_1 + \omega_1\omega_3)\theta_n.$$

Each term in the expansions above has been previously determined through definition or evaluation and represents a constant or finite sum of anti-derivatives defined in (3.9) and (3.10). Therefore,

$$\int_0^p \pi(\mu)d\mu = \psi_2 \Big|_{x=1/p}^{x=\infty} = K, \tag{3.11}$$

$$\int_0^p \mu\pi(\mu)d\mu = \psi_3 \Big|_{x=1/p}^{x=\infty} = \Sigma_1,$$

$$\int_0^p \mu^2\pi(\mu)d\mu = \psi_4 \Big|_{x=1/p}^{x=\infty} = \Sigma_2,$$

so that $E(\mu \mid \vec{x}) = \Sigma_1$, and (3.12)

$$\mathrm{Var}(\mu \mid \vec{x}) = \Sigma_2 - \Sigma_1^2 . \tag{3.13}$$

It may be the case that the experimenter is not interested in the posterior mean but rather the posterior mode. One reason for this is that the mode is computationally much easier to calculate. Another is that a knowledge of the mode may be required to determine the existence and construction of the HPD region for μ. A discussion of these regions is presented in Chapter IV.

To determine the mode of the marginal posterior distribution of μ write (3.1) as:

$$\pi(\mu) \propto (a\mu^{-2} - b\mu^{-1} + c)^{-(d+1)} \quad o < \mu < p \qquad (3.12)$$
$$= o \qquad\qquad \text{elsewhere}$$

Noting that the value of K in (3.11) determines the magnitude of the density at the mode but not the location of the mode; so one may differentiate (3.12) and set $D_\mu \pi(\mu) = o$. This results in the equation $-2a + b\mu_m = o$ or $\mu_m = \frac{2a}{b}$ where μ_m represents the mode. An examination of (3.12) in light of the fact that $d > -1$ indicates an interior extremum exists and that it must be located at the previously derived μ_m.

It is irrelevant to the Bayesian to consider the sampling properties of a Bayesian estimator since this is fundamentally at odds with the premises of Bayesian analysis. The author is not in total agreement. First some experiments using a Bayesian analysis may attach a frequency interpretation to the posterior distribution. Second some sort of comparison meaningful to the frequency oriented experimenter may be obtained by a study of the sampling attributes of Bayesian estimators.

Therefore the author will make remarks concerning sampling properties of Bayesian estimators. First it is intuitively clear that Bayesian estimators will be biased in general since they are influenced by the bias of the prior distribution. The bias in the previously derived modal estimator can be easily determined. The posterior distribution of μ has parameters $a' = a + \frac{n\bar{x}}{2}$ and $b' = b + n$ so that:

$$\mu_m = \frac{2a + n\bar{x}}{b + n}$$

$$E_{x \mid \mu_o}(\mu_m) = \frac{2a + n\mu_o}{b + n}$$ where a and b are the

parameters of the prior distribution.

$$\text{However, } \lim_{n\to\infty} \mu_n \overset{p}{=} \lim_{n\to\infty} \frac{2a}{b+n} + \lim_{n\to\infty} \frac{1}{1+b/n} \cdot \bar{x}$$

$$\overset{p}{=} \mu_o$$

by Slutsky's theorem as referenced by Fisz (6) so that μ_m is a consistent estimator of μ_o.

The third estimate of central tendency an experimenter may consider is the median of the posterior distribution of μ. The median m_e is of course the solution to $\psi_2 \Big|_{1/m_e}^{\infty} = \frac{1}{2}\psi_2 \Big|_{1/p}^{\infty}$. The author has not been able to present an expression for m_e explicitly and assumes the determination of m_e must be accomplished by numerical quadrature.

2. Point Estimation of λ Averaged Over μ

With an Informative Prior

From equation (2.3) describing the joint prior distribution of μ and λ the marginal prior or posterior is of the form

$$\pi(\lambda) \propto \lambda^d \int \exp\left\{-\lambda(a\mu^{-2} - b\mu^{-1} + c)\right\} d\mu \qquad o < \lambda < \infty \qquad (3.13)$$
$$= \quad o \qquad\qquad \text{elsewhere.}$$

As far as the author is aware the integral in (3.13) can be presented only in tabular form. The author will develop an approximation to (3.13) as an alternative to tabular representation.

Rewriting (3.13) as $\pi(\lambda) \propto \lambda^d e^{-\lambda c} \int_o^p \exp\left(-\frac{\lambda}{\mu}\left(\frac{a}{\mu} - b\right)\right) d\mu$ and making the transformation $v = \frac{1}{\mu}$ gives

$$\int_o^p \exp\left[-\frac{\lambda}{\mu}\left(\frac{a}{\mu} - b\right)\right] d\mu = \int_{1/p}^{\infty} \frac{e^{-\lambda v(av-b)}}{v^2}\, dv. \qquad (3.14)$$

Let $h = \frac{1}{p}$ and $h' = N(1 + h)$. As v becomes unbounded the integrand of (3.14) approaches zero so that the right side of (3.14) is approximated by

$$\int_h^{h'} \frac{1}{v^2} e^{-\lambda v(av-b)} \, dv \qquad (3.15)$$

For N sufficiently large the error of trunction will be bounded by $\frac{1}{N(1+h)}$. Since the integral of (3.15) is now proper one has more freedom in attempting a term by term integration. Accordingly (3.15) can be written as:

$$\int_h^{h'} \frac{1}{v^2} e^{-\lambda v(av-b)} \, dv = \int_h^{h'} \sum_{n=o}^{\infty} \frac{(-1)^n \lambda^n}{n!} \sum_{k=o}^{n} (-1)^k a^{n-k} b^k v^{2n-k-2} \, dv \quad (3.16)$$

Since the series is convergent and the partial sums are bounded over the interval (h,h') the series of (3.16) may be integrated term by term. To simplify the notation of the resultant integration put

$$k_o = \frac{1}{h} - \frac{1}{h'}$$

$$k_1 = -a(h' - h) + b \, \ell n\left(\frac{h'}{h}\right)$$

$$k_n = \frac{(-1)^n}{n!} \sum_{k=o}^{n} (-1)^k \frac{a^{n-k} b^k \left[h'^{2n-k-1} - h^{2n-k-1}\right]}{2n - k - 1} \quad (n \geq 2)$$

to obtain:

$$\int_h^{h'} \frac{1}{v^2} e^{-\lambda v(av-b)} \, dv \approx \sum_{j=o}^{\infty} k_j \lambda^j . \qquad (3.17)$$

Since the series of (3.17) is alternating the error of truncation is bounded by the first term neglected. The final approximation to $\pi(\lambda)$ is:

$$\pi_A(\lambda) = \sum_{j=o}^{m} H_j \lambda^{j+d} e^{-\lambda c} \qquad o < \lambda < \infty \tag{3.18}$$

$$= \quad o \qquad\qquad \text{elsewhere,}$$

where $H_j = \dfrac{k_j}{\sum_{i=j}^{m} \dfrac{\Gamma(i + d + 1)\, k_i}{c^{i+d+1}}}$ is a norming constant.

The posterior mean and variance of λ can be approximated by

$$E(\lambda \mid \vec{x}) \approx \frac{1}{c^{d+2}} \sum_{j=o}^{m} \frac{H_j \Gamma(j + d + 2)}{c^j}$$

$$\mathrm{Var}(\lambda \mid \vec{x}) \approx \frac{1}{c^{d+3}} \sum_{j=o}^{m} \frac{H_j \Gamma(j + d + 3)}{c^j} - (E(\lambda \mid \vec{x}))^2$$

The approximation of $\pi(\lambda)$ by $\pi_A(\lambda)$ can be circumvented in determining the posterior mean and variance of λ when d is an integer. Recall that the previously defined function ψ_2 is implicitly dependent upon the sample size n. One may therefore write:

$$\psi_2(n) = \pi_1 + \tau_1 \varphi_{n+1} + \omega_1 \theta_n = \int (a\mu^{-2} - b\mu^{-1} + c)^{-(n+1)}\, d\mu$$

Therefore from (2.3.1)

$$\int_o^p \int_o^\infty \lambda^d \exp(-\lambda(a\mu^{-2} - b\mu^{-1} + c))\, d\lambda\, d\mu$$

$$= \Gamma(d + 1) \int_o^p (a\mu^{-2} - b\mu^{-1} + c)^{-(d+1)}\, d\mu$$

$$= d!\ \psi_2(d) \tag{3.19}$$

Space will be conserved if the limits of the ψ functions are suppressed. They are from $\frac{1}{p}$ to ∞ as specified above.

Equation (3.19) represents the normalizing constant for the bivariate prior distribution given in (2.3). In a like manner one may obtain the posterior mean and variance of λ:

$$E(\lambda \mid \vec{x}) = \frac{(d+1)!\ \psi_2(d+1)}{d!\ \psi_2(d)} = \frac{(d+1)\ \psi_2(d+1)}{\psi_2(d)}$$

and $$\mathrm{Var}(\lambda \mid \vec{x}) = \frac{(d+1)}{\psi_2(d)} \left[(d+2)\ \psi_2(d+2) - \frac{(d+1)(\psi_2(d+1))^2}{\psi_2(d)} \right]$$

An expression for the posterior mode of λ is not readily obtainable when μ is unknown and the experimenter is not interested in making statements about μ. The author will give no expression for the posterior mode but indicate what approaches might be used in obtaining a numerical solution. First, if one is satisfied with the approximation (3.18) one could attempt to maximize (3.18) with respect to λ. Second, one could maximize $\int_0^p \pi(\mu,\lambda)d\mu$ where the integral is determined by numerical quadrature. These are the obvious procedures. One should note that $\pi_A(\lambda)$ given in (3.18) is a linear combination of gamma densities and one can be reasonably confident that it possesses at least one mode. Assuming this only indicates that $\pi(\lambda)$ possesses a mode. Therefore, it would be useful to know beforehand if a solution exists. The author will leave this as an open question since its practical import is negligible.

The posterior median of λ is most readily obtained by numerical quadrature of either $\pi_A(\lambda)$ or $\int_0^p \pi(\mu,\lambda)d\mu$. One can substitute a series expansion of the incomplete gamma function in the equation $\int_0^{m_e} \pi_A(\lambda)d\lambda = \frac{1}{2}$ to obtain the relation:

$$\sum_{j=o}^{m} \sum_{t=o}^{\infty} H_j(m_e)^{j+d+1} \frac{(-cm_e)^t}{(j+d+t+1)t!} = \tfrac{1}{2}\,.$$

It is the author's opinion that attempting to extract the solution m_e in the above is not an efficient program to follow.

The inverse Gaussian distribution is normally not parametrized in terms of its variance σ^2. In some cases some inferences concerning σ^2 may be required. These conclusions can be obtained by simultaneous inferences about μ and λ. As an alternative the author will develop the posterior distribution of the variance and investigate its utility.

3. Posterior Distribution of σ^2 With Both μ and λ Unknown

If X follows an inverse Gaussian probability law with parameters μ and λ then it is known that $\sigma^2 = \frac{\mu^3}{\lambda}$. If one assumes a prior distribution of the form specified in (2.3) then the density of (μ,λ) has been determined to be

$$\pi(\mu,\lambda) = \frac{\lambda^d \exp - \lambda(a\mu^{-2} - b\mu^{-1} + c)}{d!\ \psi_2(d)} \tag{3.19.1}$$

for nonnegative integer d. Put $X = \frac{\mu^3}{\lambda}$, $Y = \lambda^{1/3}$. The Jacobian of the transformation is:

$$J = \begin{vmatrix} \frac{y}{3x^{2/3}} & x^{1/3} \\ o & 3y^2 \end{vmatrix} = \frac{Y^3}{X^{2/3}}$$

The support of $\pi(\mu,\lambda)$ is $\{(\mu,\lambda):\ o < \mu < p,\ o < \lambda < \infty\}$. The support of $f(x,y)$ is $\{(x,y):\ o < x,\ o < y,\ x < p^3/y\}$. The transformed distribution becomes:

$$f(x,y) = \frac{y^{3(d+1)}}{d!\ \psi_2(d)x^{2/3}} \exp - y(ax^{-2/3} - bx^{-1/3}y + cy^2) \quad (3.20)$$

The marginal of σ^2 is obtained by integrating (3.20) over y. To simplify notation consider those factors of x to be absorbed in a and b and temporarily ignore coefficients not involving the variable of integration y. That is take (3.20) as:

$$f(x,y) = y^{3(d+1)} \exp - y(a - by + cy^2)$$

The exponential factor can be written as $\sum_{n=o}^{\infty} \frac{(-1)^n y^n}{n!} (a - by + cy^2)^n$

and

$$(a - by + cy^2)^n = \sum_{j=o}^{n} \sum_{k=o}^{j} \binom{n}{j}\binom{j}{k}(-1)^k a^k b^{j-k} c^{n-j} y^{2n-j-k}$$

Since the range of integration of y is the finite interval $(o, \frac{p}{x^{1/3}})$ the sequence of partial sums of the exponential series is uniformly bounded and consequently term by term integration is permissible. For all values of the indices the power rule will suffice. So incorporating the normalizing constant, ignored and absorbed factors of x, substituting σ^2 for x yields:

$$\pi(\sigma^2) = \frac{1}{d!\ \psi_2(d)} \sum_{n=o}^{\infty} J_n(\sigma^2)^{-(n+d+2)} \qquad o < \sigma^2 < \infty$$

where $J_n = \sum_{j=o}^{n} \sum_{k=o}^{j} \frac{(-1)^{n+k}\binom{n}{j}\binom{j}{k} a^k b^{j-k} c^{n-k} p^{3(n+d+1)-j-k+1}}{n!\ (3n + 3d + 4 - j - k)}$

4. Estimation of μ When λ is Known With an Informative Prior

From (2.4) the appropriate functional form of the prior and posterior distribution is

$$\pi(\mu) \propto \exp - \frac{1}{\mu}\left(\frac{a}{\mu} - b\right) \qquad 0 < \mu < c \qquad (3.21)$$

Actual determination of the mean, variance, and median are by numerical quadrature. These quantities are dependent upon the normalizing constant from equation (3.21) and presumably this would be computed by numerical procedures also. The author will now develop a series expansion of the integral $\int_0^c \exp -\left(\frac{a}{\mu^2} + \frac{b}{\mu}\right) d\mu$ which some readers may find more palatable than numerical quadrature.

Let $t = \frac{1}{\mu}$ in the above integral and perform the change of variables to obtain

$$\int_0^c \exp - \frac{1}{\mu}\left(\frac{a}{\mu} - b\right) d\mu = \int_{1/c}^{\infty} \frac{1}{t^2} e^{-a(t-\frac{b}{2a})^2} dt \cdot e^{-\frac{b^2}{4a}}$$

Let $I = \int_{1/c}^{\infty} \frac{1}{t^2} e^{-a(t-\frac{b}{2a})^2} dt$ and utilize integration by parts by letting

$$\mu = \exp\left(-a(t - \frac{b}{2a}\right)^2 \qquad dv = \frac{dt}{t^2}$$

then:

$$d\mu = -2a\left(t - \frac{b}{2a}\right) \exp\left(-a(t - \frac{b}{2a})^2\right) \qquad v = \frac{-1}{t}$$

From this comes the equation:

$$I = \lim_{t \to \infty} - \frac{1}{t} \exp - a(t - \frac{b}{2a})^2 + c \exp - a(\frac{1}{c} - \frac{b}{2a})^2$$

$$-2a \int_{1/c}^{\infty} \exp - a(t - \frac{b}{2a})^2 dt + b \int_{1/c}^{\infty} \exp - a(t - \frac{b}{2a})^2 dt$$

Taking the limit and evaluating the third summand above gives:

$$I = ce^{-a(\frac{1}{c} - \frac{b}{2a})^2} - 2\sqrt{a\pi}\{1 - \Phi(\sqrt{2a}(\frac{1}{c} - \frac{b}{2a}))\}$$

$$+ b\int_{1/c}^{\infty} \frac{1}{t} \exp - a(t - \frac{b}{2a})^2 \, dt$$

where $\Phi(\cdot)$ is the cumulative normal distribution function.

Let $I_1 = \int_{1/c}^{\infty} \frac{1}{t} \exp - a(t - \frac{b}{2a})^2 \, dt$ and perform another change of variable by letting $t = \frac{x}{\sqrt{2a}} + \frac{b}{2a}$, $dt = \frac{dx}{\sqrt{2a}}$.

The equation for I_1 now becomes

$$I_1 = \int_{\sqrt{2a}(\frac{1}{c} - \frac{b}{2a})}^{\infty} \frac{1}{\frac{x}{\sqrt{2a}} + \frac{b}{2a}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2a}} \, dx$$

$$= \int_{\frac{\sqrt{2a}}{c} - \frac{b}{\sqrt{2a}}}^{\infty} \frac{e^{-\frac{x^2}{2}}}{x + \frac{b}{\sqrt{2a}}} \, dx$$

Use $d = \frac{b}{\sqrt{2a}}$ to obtain:

$$I_1 = \int_{-d + \frac{\sqrt{2a}}{c}}^{\infty} \frac{e^{-\frac{x^2}{2}}}{x + d} \, dx$$

$$= \int_{-d + \frac{\sqrt{2a}}{c}}^{d} \frac{e^{-\frac{x^2}{2}}}{x + d} \, dx + \int_{d}^{\infty} \frac{e^{-\frac{x^2}{2}}}{x + d} \, dx$$

$$= I_2 + I_3$$

with obvious substitutions being made. To continue the development in this manner requires that c be chosen large enough so that $-d + \frac{\sqrt{2a}}{c} < d$. If this is not desirable then it is not necessary to decompose I_1 and the choice of c is not crucial to a series development of I.

Referring to I_2 it is true that for $|x| \leq d$ the expansion

$$\frac{1}{x + d} = \frac{1}{d} \sum_{k=0}^{\infty} (-d)^{-k} x^k$$

is convergent. The partial sums of this series are bounded on $(-d,d)$ and Riemann integrable so by Arzela's theorem Apostal (2):

$$I_2 = \sum_{k=o}^{\infty} \int_{-d+\frac{\sqrt{2a}}{c}}^{d} (-d)^{-k} x^k e^{-x^2/2} dx$$

In a like manner $c^{-x^2/2} = \sum_{j=o}^{\infty} \frac{(-1)^j x^{2j}}{2^j j!}$ for all real x.

Again conditions for term by term integration are met so that

$$I_2 = \sum_{k=o}^{\infty} \sum_{j=o}^{\infty} \int_{-d+\sqrt{\frac{2a}{c}}}^{d} (-d)^{-k} \frac{x^k (-1)^j x^{2j}}{2^j j!} dx$$

therefore

$$I_2 = \sum_{k=o}^{\infty} [\sum_{j=o}^{\infty} \frac{(-1)^{k+j}}{d^k 2^j j! (k + 2j + 1)} x^{k+2j+1} \Big|_{x=\frac{\sqrt{2a}}{c} -d}^{d}]$$

This leaves only the integral I_3 to be expanded. Within the range of integration for I_3 the relation $|x| > d$ holds and the expansion

$$\frac{1}{x + d} = \sum_{k=o}^{\infty} \frac{(-d)^k}{x^{k+1}} \quad \text{is convergent.}$$

Again applying Arzela's theorem generates the series expansion:

$$I_3 = \sum_{k=o}^{\infty} \int_{d}^{\infty} \frac{(-d)^k}{x^{k+1}} e^{-x^2/2} dx$$

$$= \int_{d}^{\infty} \frac{e^{-x^2/2}}{x} dx + \sum_{k=1}^{\infty} \int_{d}^{\infty} (-d)^k \frac{e^{-x^2/2}}{x^{k+1}} dx \quad (3.22)$$

At this point it is necessary to begin developing asymptotic approximations by truncation. The first integral in (3.22) is approximated by

$$\int_{d}^{\omega} \frac{e^{-x^2/2}}{x} dx \quad (3.23)$$

where $$\left| \int_d^{\infty} \frac{e^{-x^2/2}}{x}\, dx - \int_d^{\omega} \frac{e^{-x^2/2}}{x}\, dx \right| < \sqrt{2\pi}\,(1 - \Phi(\omega))$$

provided $\omega \geq 1$. It is apparent that ω need not be large in order for this approximation to be quite accurate.

Since one has a finite range of integration in (3.23) and since the partial sums of the series expansion of $\frac{e^{-x^2/2}}{x} = \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j-1}}{2^j j!}$

are Riemann integrable and uniformly bounded then term by term integration can be performed.

This gives $$\int_d^{\omega} \frac{e^{-x^2/2}}{x}\, dx = \sum_{j=0}^{\infty} \frac{(-1)^j}{2^j j!(2j)}\, x^{2j} \Big|_d^{w} \qquad (3.24)$$

Consider the second summand of (3.22). Those terms of the series involving even-powered terms in X can be evaluated by use of the formula:

$$\int_d^{\infty} \frac{e^{-x^2/2}}{x^{2n+2}}\, dx = \frac{\sqrt{2\pi}\,\left\{1 - \Phi(d) - \frac{e^{-d^2/2}}{\sqrt{2\pi}}\, Q(d)\right\}}{(-1)^{n+1}\; 1\cdot3\cdot5\ldots(2n+1)} \qquad n \geq 0$$

where $Q(d) = 1 - \frac{1}{d^2} + \frac{1.3}{d^4} + \ldots + \frac{(-1)^n\, 1.3\ldots(2n-1)}{d^{2n}}$

The above formula is obtained from Abramowitz and Segun (1).

The odd powers occurring in (3.22) can be evaluated by the same procedure after repeated application of the reduction formula:

$$\int_d^{\infty} \frac{e^{-x^2/2}}{x^{2n+1}}\, dx = \frac{-1}{2n} \left(\int_d^{\infty} \frac{e^{-x^2/2}}{x^{2n}}\, dx + \int_d^{\infty} \frac{e^{-x^2/2}}{x^{2n-1}}\, dx \right) \qquad (3.25)$$

The use of (3.25) will result in a term involving $\int_d^{\infty} \frac{e^{-x^2/2}}{x}\, dx$ for which (3.23) was developed.

The previous development indicates the difficulties one may expect in attempting to produce simple expression for normalizing constants, means, etc. for the posterior distribution contained in this chapter. Such illustrative developments are always productive in that they indicate the utility of a particular procedure. The reader is not encouraged to think that those numerical examples in this study were accomplished through series developments. Professor Chandler kindly provided a numerical quadrature procedure called SQANK (15). Because of certain shortcomings in the error control of SQANK and economic disadvantages the author developed the program RINT. The interested reader will find a listing of this program in the appendix. In light of these difficulties the author is forced to state that the posterior mean, variance, and median of μ have not been determined in explicit form.

The mode of the posterior distribution is determined in the usual manner to be

$$\mu_m = \begin{cases} \frac{2a}{b} & b > o \\ p & b < o \end{cases}$$

If $\lambda = \lambda_o$ then $\mu_m = \dfrac{2a + \lambda_o \sum x_i}{b + n\lambda_o}$ for $b > o$ and it is easily verified that since

$$\mu_{max} = \frac{\frac{2a}{n} + \lambda_o \bar{x}}{\frac{b}{n} + \lambda_o}$$

then $\mu_{max} \xrightarrow{P} \mu$.

5. Estimation of λ When μ is Known With an Informative Prior

The prior distribution of λ is taken as in (2.5) to be

$$\pi_o(\lambda) = \frac{a^p \lambda^{p-1} e^{-a\lambda}}{\Gamma(p)} \qquad \lambda > o \qquad (3.26)$$

$$o \qquad \text{elsewhere}$$

The mean and variance of X are well known to be $\frac{p}{a}$ and $\frac{p}{a^2}$, respectively. The mode occurs at $\frac{a}{p-1}$ for $p > 1$ and $\lambda = o$ otherwise. The posterior median can be formulated in several ways. The median is the solution to $F(x) = \int_o^x \frac{a^p t^{p-1} e^{-at}}{\Gamma(p)} dt = \frac{1}{2}$. The integral F(x) can be simplified by transformation of variables.

The function F(x) can then be rewritten as $H(x) = \int_o^{ax} \frac{z^{p-1} e^{-z}}{\Gamma(p)} dz$.

The equation $H(x) = \frac{1}{2}$ can then be solved and the resultant solution divided by a to yield the posterior median m_e. The reason H(x) is preferable to F(x) is that tables and programs are available for evaluating the incomplete χ^2 distribution. It is known from Abramowitz and Segun (1) that $H(x) = P(\chi_o^2|\nu)$ where $P(\chi_o^2|\nu) = \Pr\{\chi^2 \leq \chi_o^2\}$ and where χ^2 has ν degrees of freedom. One obtains χ_o^2 and ν by the translation $\chi_o^2 = 2ax$ and $\nu = 2p$. The IBM subroutine CDTR in the scientific subroutine series evaluates $\Pr(\chi_o^2|\nu)$ for continuous . This does not change the problem of using an iterative procedure on H(x) but does circumvent the necessity of employing numerical quadrature. One may use AINVRT developed by the author for solving $H(x) = \frac{1}{2}$. A listing of AINVRT is supplied in the appendix for the interested reader.

6. Posterior Distribution of the Variance When μ is Known With an Informative Prior

Definition 3.1: A random variable X follows a gamma-2 distribution with parameters $\psi > 0$ and $\nu > 0$ if X can be shown to have a density of the form

$$f_{\gamma_2}(x;\psi,v) = \frac{v\psi}{2}\left[\frac{v\psi x}{2}\right]^{\frac{v}{2}-1} \exp\left(-\frac{v\psi x}{2}\right)\Big/\Gamma\left(\frac{v}{2}\right)$$

for $x > 0$ and zero elsewhere.

Definition 3.2: The r.var. Y follows an inverted gamma distribution with parameters $\psi > 0$, $v > 0$ if Y possesses a density function of the form

$$f_{i\gamma}(y;\psi,v) = \left[\frac{v\psi}{2y}\right]^{\frac{v}{2}+1} \frac{\exp\left(-\frac{\psi v}{2y}\right)}{\frac{v\psi}{2}\Gamma\left(\frac{v}{2}\right)}$$

for $y > 0$ and zero elsewhere.

If X has a standard gamma distribution with parameter $\frac{v}{2}$ then $Y = \frac{2x}{v\psi}$ follows a gamma-2 distribution with parameters v and ψ as shown in LaValle (14). Further $\frac{1}{Y}$ follows an inverted gamma distribution with parameters v and ψ.

Therefore since λ is assumed to follow a gamma distribution with parameters a and p (denoted $\lambda \sim Ga(a,p)$) then $\frac{1}{\sigma^2} = \frac{\lambda}{\mu_o^3} \sim Ga(\mu_o^3 a,p)$. From this it follows that $\frac{\mu_o^3 a}{\sigma^2} \sim Ga(1,p)$. Letting $v = 2p$ and $\psi = \frac{\mu_o^3 a}{p}$ then from the above discussion one can conclude that $\frac{\sigma^2 \psi v}{2\mu_o^3 a} = \sigma^2$ follows an inverted gamma distribution with parameters v and ψ. Since the mean and variance of an inverted gamma distributed variate are given in LaValle (14) as $\frac{v\psi}{v-2}$ and $\frac{2v^2\psi^2}{[(v-2)(v-4)]}$, respectively, the appropriate expressions for the posterior mean and variance of σ^2 are:

$$E(\sigma^2 \mid \vec{x}) = \frac{\mu_0{}^3 a}{p-1}$$

and

$$Var(\sigma^2 \mid \vec{x}) = \frac{(E\sigma^2)^2}{p-2}.$$

7. Simultaneous Estimation of μ and λ With an Informative Prior

The only point estimators of μ and λ that may be of value that have not been previously discussed are the simultaneous modal estimators of μ and λ. Take the posterior distribution of the form (3.19.1). That is take $\pi(\mu,\lambda) \propto \lambda^d \exp - \lambda(a\mu^{-2} - b\mu^{-c} + c)$ and form the equations $D_\mu \pi = o$ and $D_\lambda \pi = o$. These equations give

$$D_\mu \pi \propto \lambda^d \exp[-\lambda(a\mu^{-2} - b\mu^{-1} + c)](-2a\mu^{-3} + b\mu^{-2}) = o$$

which gives the solution:

$$\mu_{max} = \frac{2a}{b}$$

which is the same as that obtained from the marginal of μ given in the paragraph following (3.12).

Letting $Q(\mu,\lambda) = -\lambda(a\mu^{-2} - b\mu^{-1} + c)$ then

$$D_\lambda \pi(\mu,\lambda) \propto d\lambda^{d-1} \exp Q(\mu,\lambda) - \lambda^d \exp Q(\mu,\lambda)(a\mu^{-2} - b\mu^{-1} + c) = o$$

which gives the modal estimator:

$$\lambda_{max} = \frac{d}{a\mu_{max}^{-2} - b\mu_{max}^{-1} + c}$$

or upon substitution

$$\lambda_{max} = \frac{4ad}{4ac - b^2}.$$

The nervous reader may recall that $4ac - b^2 > o$ almost surely.

8. Estimation of μ Averaged Over λ Using a Diffuse Prior Distribution

The functional form of the density of μ with the Jeffrey diffuse prior was given in (2.8) as:

$$\pi(\mu) \propto \frac{\mu^{n-\frac{1}{2}}}{(a - n\mu + b\mu^2)^{n+1}} \tag{3.27}$$

where $a = \frac{\sum X_i}{2}$ and $b = \frac{1}{2}\sum X_i^{-1}$. It has been shown in Chapter II that the function of (3.27) possesses a convergent integral for $\mu \epsilon (o,\infty)$. However, it can be shown that $E(\mu)$ does not exist. Without going into great detail the reason for this is that for large μ the numerator of $\mu\pi(\mu)$ is of order $n + \frac{1}{2}$ while the denominator is of order $n + 1$. Therefore, the integrand is then of order $-\frac{1}{2}$ for large μ. Since $\int_o^\infty \mu^{-\frac{1}{2}} d\mu$ diverges one likewise can conclude $E(\mu) = \infty$ also. Consequently, no higher moment exists for μ.

To compute the median of this distribution or to develop probability statements requires the evaluation of $\int_o^\infty \frac{\mu^{n-\frac{1}{2}}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}} d\mu$ to obtain the normalizing constant. Since the integrand in the above integral can be written as $\mu^{-3/2}(a\mu^{-2} - n\mu^{-1} + b)^{-\frac{n+1}{2}}$ then one finds that:

$$I = \int_0^\infty \frac{\mu^{n-\frac{1}{2}}}{(a - n\mu + b\mu^2)^{\frac{n+1}{2}}} \, d\mu = \int_0^\infty \frac{1}{\sqrt{v}} \cdot \frac{1}{(av^2 - nv + b)^{\frac{n+1}{2}}} \, dv$$

Completing the square on v gives

$$av^2 - nv + b = a\left(v - \frac{n}{2a}\right)^2 + b^2 - \frac{n^2}{4a}$$

Substitute $c = \frac{n}{2a}$ and $d = b - \frac{n^2}{4a}$ in the above and the quadratic becomes $a(v - c)^2 + d$. Now let $x = v - c$ to give:

$$I = \int_{-c}^\infty \frac{dx}{\sqrt{x + c}\,(ax^2 + d)^{\frac{n+1}{2}}} \qquad (3.28)$$

Finally make the change of variable defined by $y^2 = x + c$ and (3.28) is then represented by:

$$I = \int_0^\infty \frac{dy}{[a(y^2 - c)^2 + d]^{\frac{n+1}{2}}} = \int_0^\infty f(y)dy$$

Although the present development is directed toward the determination of the normalizing constant for (3.27) it would also be of value to be able to determine $I(p) = \int_0^p f(y)dy$. As far as the author is aware this integral is not expressible in terms of elementary functions or elliptic integrals and therefore must be computed by numerical quadrature. However, the normalizing constant can be represented explicitly for particular n by the application of results in the theory of functions of a complex variable.

The author will assume that $n = 2k - 1$ for some $k \geq 1$ and let

$$f(z) = \frac{1}{[a(z^2 - c) + d]^{\frac{n+1}{2}}} = \frac{1}{[a(z - c)^2 + d]^k}$$

where z is complex.

Noting that f(y) is an even function of y then it follows that:

$$\int_{-\infty}^{\infty} f(y)dy = \tfrac{1}{2} \int_{o}^{\infty} f(y)dy$$

Therefore, from Pennisi (18) $\int_{o}^{\infty} f(y)dy = \pi i \sum_{i=1}^{N} Re(f(z);z_i)$ where the z_i are the poles of f in the upper half plane and $Res(f(z),z_i)$ is the residue of f at the pole z_i. It is a tacit assumption of this procedure that f(z) has no real zeroes and it is evident that this is the case.

The poles of f(z) are the values z_i; such that $a(z_i - c)^2 + d = o$. The roots of this equation are $z^2 = c \pm i\sqrt{\frac{d}{a}}$. One will note that $a = \frac{\sum X_i}{2} > o$ and since $4ab - n > o$ then $d = b - \frac{n^2}{4a}$ is also positive. Therefore $\sqrt{\frac{d}{a}}$ is real.

Let $e = \sqrt{\frac{d}{a}}$ and $r = \sqrt{c^2 + e^2}$. From these define $\theta_2 = Tan^{-1}(\frac{e}{c})$ and $\theta_2 = Tan^{-1}(-\frac{e}{c})$. Then the four roots of $a(z^2 - c)^2 + d = o$ are:

$$z_1 = \sqrt{r}\,[\cos \frac{\theta_1}{2} + i\sin \frac{\theta_1}{2}]$$

$$z_2 = \sqrt{r}\,[\cos(\frac{\theta_1}{2} + \pi) + i\sin (\frac{\theta_1}{2} + \pi)]$$

$$z_3 = \sqrt{r}\,[\cos \frac{\theta_2}{2} + i\sin \frac{\theta_2}{2}]$$

$$z_4 = \sqrt{r}\,[\cos(\frac{\theta_2}{2} + \pi) + i\sin (\frac{\theta_2}{2} + \pi)]$$

Of these zeroes the only ones of interest are z_1 and z_4 since the remainder fall in the lower half plane. Each pole is of order k so that

$$I = \pi i\,[Res(f,z_1) + Res(f,z_4)] \qquad (3.29)$$

This is a reasonably compact expression for the normalizing constant. A problem may arise in determining the required residues. The author is not aware of numerical procedures for computing residues but this of course does not preclude the existence of such techniques. Therefore, the author will complete the discussion of the evaluation of I by giving explicit formulae for the computation of the required residues.

Again from Pennisi (18) one obtains the equation

$$\operatorname{Res}(f(z),z_1) = \frac{1}{(k-1)!} \quad D_z^{(k-1)}[(z-z_2)^{-k}(z-z_3)^{-k}(z-z_4)^{-k}\Big| \; z = z_1 \tag{3.30}$$

Let $u(z) = (z - z_1)^{-k}$, $v(z) = (z - z_2)^{-k}$, $w(z) = (z - z_3)^{-k}$ and $x(z) = (z - z_1)^{-k}$. One may then laboriously expand (3.30) which yields the following expression for $\operatorname{Res}(f,z_1)$.

$$\operatorname{Res}(f,z_4) = \frac{1}{(k-1)!} \sum_{m=o}^{k-1} \sum_{j=o}^{m} \binom{k-1}{m}\binom{m}{j} D_z^{(j)} \; v \; D_z^{(m-j)} \; w \; D_z^{(k-1-m)} \; x\Big| \; z = z_1$$

Similarly

$$\operatorname{Res}(f,z_4) = (k-1)! \sum_{m=o}^{k-1} \sum_{j=o}^{m} \binom{k-1}{m}\binom{m}{j} D_z^{(j)} \; u \; D_z^{(m-j)} \; v \; D_z^{(k-1-m)} \; w \Big| \; z = z_4$$

where $D_z^{(1)}(z - z_t)^{-k} = (-1)^1 \, k \, (k+1) \cdot \cdots (k+1-1)(z - z_t)^{-(k+1)}$

One may further simplify the computations by substituting $A = \sqrt{\frac{r+c}{2}}$, $B = \sqrt{\frac{r-c}{2}}$ then it follows that $z_1 = A + iB$, $z_2 = -z_1$, $z_3 = \bar{z}_1$, and $z_4 = -\bar{z}_1$. It is apparent that the computation of I for large sample sizes would be tedious since the number of terms in each summation increases at approximately the square of the sample size.

Once I is determined then the normalizing constant for the joint Jeffrey's posterior is from (2.8)$[\Gamma(\frac{n+1}{2})I]^{-1}$, provided n is odd.

Since the mean fails to exist and the computation of the median appears to be a numerical process, the only alternative for this case is to consider the modal estimator.

The density $\pi(\mu) \propto \mu^{n-\frac{1}{2}}(a - n\mu + b\mu^2)^{-\frac{n}{2}-\frac{1}{2}}$ $\quad \mu > 0$

0 elsewhere

Forming the equation $D_\mu \pi(\mu) = 0$ gives:

$$(n - \tfrac{1}{2})\ \mu^{n-3/2}\ (a - n\mu + b\mu^2)^{-\frac{n}{2}-\frac{1}{2}} - (\tfrac{n+1}{2})\ \mu^{n-\frac{1}{2}}\ (2b\mu - n)\ \cdot$$
$$(a - n\mu + b\mu^2)^{-\frac{n}{2}-\frac{3}{2}} = 0$$

Multiply both sides of the above by $2\mu^{-n+3/2}\ (a - n\mu + b\mu^2)^{\frac{n}{2}+\frac{3}{2}}$ to obtain:

$$(2n - 1)(a - n\mu + b\mu^2) - \mu(n + 1)(2b\mu - n) = 0$$

or

$$-3b\mu^2 + (2n - n^2)\mu + (2n - 1)a = 0$$

Taking the positive root yields:

$$\mu_{max} = \frac{1}{6b}\left[-(n^2 - 2n) + \sqrt{(n^2 - 2n)^2 + 12ab(n - 1)}\right]$$

Now substitute $a = \frac{\Sigma X_i}{2}$, $b = \frac{1}{2}\Sigma X_i^{-1}$, and assume $n \neq 2$ to write:

$$\mu_{max} = \frac{n(n-2)}{3\ \Sigma X_i^{-1}} \left\{-1 + \sqrt{\frac{1 + 3\Sigma X_i\ \Sigma X_i^{-1}\ (2n - 1)}{n(n - 2)}}\right\} \qquad (3.31)$$

Denote the harmonic mean of the sample: $x_1, x_2, \ldots, x_n$ by $\bar{x}_h$. The harmonic mean is defined by $\bar{x}_h = \frac{n}{\Sigma X_i^{-1}}$. Then equation (3.31) can be written in terms of the arithmetic mean $\bar{X}$ and harmonic mean $\bar{X}_h$. This equation is:

$$\mu_{max} = \frac{n-2}{3}\left\{-\bar{X}_h + \sqrt{\bar{X}_h^2 + \frac{3(2n-1)\bar{X}\bar{X}_h}{(n-2)^2}}\right\}$$

The author will demonstrate that μ_{max} viewed as a sampling estimator is consistent. Recall that $X_i > o$ almost surely. Therefore, it can be shown that $\bar{X}_h \leq \bar{X}$ with equality occurring only if $X_1 = X_2 = \ldots X_n = 1$. Hence, $\bar{X}_h\bar{X} \leq \bar{X}^2$. Since $\bar{X}$ is a consistent estimator of μ, $\bar{X}$ is stochastically bounded, i.e. given $\varepsilon > o$ there exists an M such that $\varepsilon > \Pr(\bar{X} > \sqrt{M}) = \Pr(\bar{X}^2 > M)$ as referenced in Feller (5). Since $\bar{X}^2 \geq \bar{X}\bar{X}_h$ then $\Pr(\bar{X}^2 > M) \geq \Pr(\bar{X}\bar{X}_h > M)$ so that $\bar{X}\bar{X}_h$ is stochastically bounded. Therefore, $\frac{3(2n-1)\bar{X}\bar{X}_h}{(n-2)^2}$ converges to zero in probability. Now for small k the approximation $\sqrt{y+k} \approx \sqrt{y} + \frac{k}{2\sqrt{y}}$ is valid. Putting $y = \bar{X}_h$ and $k = \frac{3(2n-1)\bar{X}\bar{X}_h}{(n-2)^2}$ one obtains through this approximation that

$$\lim_{n\to\infty} \mu_{max} = \lim_{n\to\infty} \frac{n-2}{3}\left[-\bar{x}_h + \bar{x}_h + \frac{1}{2\bar{x}_h}\frac{3(2n-1)\bar{X}\bar{X}_h}{(n-2)^2}\right]$$

$$= \lim_{n\to\infty} \frac{2n-1}{2n-4}\ \bar{X} \overset{p}{=} \mu$$

9. Estimation of μ When λ is Known With a Diffuse Prior Distribution

When $\lambda = \lambda_o$ the Jeffrey prior on μ is proportional to $\mu^{-3/2}$ so that the posterior distribution of μ is:

$$\pi(\mu) \propto \mu^{-3/2} \exp - \lambda_o(a\mu^{-2} - n\mu^{-1}) \qquad o < \mu < \infty \quad (3.32)$$

$$= \qquad o \qquad\qquad \text{elsewhere.}$$

As in the previous case $\pi_o(\mu)$ is integrable but fails to possess a first moment since

$$E(\mu) \propto \int_o^{\infty} \mu^{-\frac{1}{2}} \exp - \lambda_o(a\mu^{-2} - n\mu^{-1})d\mu =$$

$$\int_o^{\infty} v^{-3/2} \exp(-\lambda_o v(av - n))dv > \int_o^{n/a} v^{-3/2}\, dv$$

$$+ \int_{n/a}^{\infty} v^{-3/2} \exp^{-\lambda_o v(av - n)}\, dv$$

The first integral summand above fails to exist so likewise does $E(\mu)$.

The author's opinion is that a search for a normalizing constant and explicit determination of the median would be, as past experience has indicated, fruitless from the practical standpoint. The author has provided sufficient numerical procedures in the appendix to determine both for any of the distributions in this paper.

The modal estimator for this distribution will be discussed since it is independent of the proportionality constant in (3.32). As μ becomes unbounded the numerator of (3.32) remains bounded so that $\lim_{\mu \to o} \pi(\mu) = o$. This is of course necessary for integrability. At $\mu = o$ one has that:

$$o \leq \lim_{\mu \to o} \pi_o(\mu) \propto \exp \left\{ \lim_{\mu \to} - \lambda_o(a\mu^{-2} - n\mu^{-1} - 3/2 \log \mu)\right\} =$$

$$\exp \left\{ \lim_{v \to \infty} - \lambda_o(av^2 - nv + 3/2 \log v) \leq \exp \left\{ \lim_{v \to \infty} - \lambda_o(av^2 - nv + 3/2v) = \right.\right.$$

$\exp \left\{ \lim_{v \to \infty} - \lambda_o v(av - n + 3/2) \right\}$. Since $a > o$ this limit is zero.

Since $\pi(\mu) > o$ it is evident that the mode(s) are interior points. Forming the equation $D_\mu \pi(\mu) = o$ results in:

$$D_\mu \pi(\mu) \propto -\lambda_o(-2a\mu^{-3} + n\mu^{-2} + 3/2\ \mu^{-1}) \exp(-\lambda_o(a\mu^{-2} - n\mu^{-1} + 3/2 \log \mu)$$

$$= o,$$

Solving the above gives the solutions determined by $2a\mu^{-2} - n\mu^{-1} - 3/2 = o$. Since μ is constrained to be positive the only solution to this equation is

$$\mu_{max} = \frac{4a}{n + \sqrt{n^2 + 12a}}$$

Now since $a = \frac{n\bar{x}}{2}$ then the above may be rewritten as:

$$\mu_{max} = \frac{2\bar{x}}{1 + \sqrt{1 + 6\frac{\bar{x}}{n}}} \qquad (3.33)$$

Since $\bar{x}$ is a consistent estimator of μ then from (3.33) it is evident that $\lim_{n \to \infty} \mu_{max} = \mu$ so that μ_{max} is a consistent estimator of μ.

In this particular instance the author has been able to derive the sampling distribution of μ_{max}. Inverting (3.33) gives $\bar{x} = \mu_{max} + \frac{6}{4n}\ \mu^2_{max}$. Note that (3.33) is a one-to-one transformation so that if $Q_{\mu_{max}}(\cdot)$ represent the c.d.f. of μ_{max} one has the relationship that

$$Q_{\mu_{max}}(\mu_o) = F_{\bar{x}}(\mu_o^2 + \frac{3\mu^2}{2n})$$

where $F_{\bar{x}}(\cdot)$ is the c.d.f. of the mean of a sample of size n from an inverse Gaussian r.var. with parameters μ and λ. From Chhikara (4) $\bar{x}$ follows an inverse Gaussian distribution with parameters μ and $n\lambda$. Also from Chhikara (5) one can obtain an expression for evaluating the c.d.f. of an inverse Gaussian. Utilizing this one obtains

$$F_{\mu_{max}}(\mu_o) = 1 - \Phi\left[-\sqrt{\frac{n\lambda}{v_o}}\left(1 - \frac{v_o}{\mu}\right)\right] + e^{\frac{2n\lambda}{2}}\left[1 - \Phi\left(-\sqrt{\frac{n\lambda}{v_o}}\left(1 + \frac{v_o}{\mu}\right)\right)\right]$$

where $\Phi(\cdot)$ is the c.d.f. of a standard normal and $v_o = \mu_o(1 + \frac{3\mu_o}{2})$.

10. Estimation of λ Averaged Over μ With a Diffuse Prior Distribution

The joint posterior density of μ and λ in this case is

$$\pi(\mu,\lambda) = \frac{\lambda^{\frac{n-1}{2}}}{I\mu^{3/2}\,\Gamma(\frac{n+1}{2})} \exp\left[-\lambda(a\mu^{-2} - n\mu^{-1} + b)\right] \quad \mu > o,\ \lambda > o$$

where I is determined elsewhere by (3.29). It was assumed in that development that the sample size n was odd. For large n this distinction is probably of no consequence.

The first estimator of interest is the posterior mean defined by:

$$E(\lambda) = \int_o^\infty\int_o^\infty \lambda\pi(\mu,\lambda)\,d\mu\,d\lambda = \int_o^\infty\int_o^\infty \lambda\pi(\mu,\lambda)\,d\lambda\,d\mu$$

$$= \int_o^\infty\int_o^\infty \frac{\lambda^{\frac{n+1}{2}}\exp - \lambda(a\mu^{-2} - n\mu^{-1} + b}{\Gamma\left(\frac{n+1}{2}\right)\, I\mu^{3/2}}\,d\lambda\,d\mu$$

$$= \frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right) I}\int_o^\infty \frac{1}{\mu^{3/2}(a\mu^{-2} - n\mu^{-1} + b)^{\frac{n+3}{2}}}\,d\mu \qquad (3.34)$$

This last integral is of the same form as that of (3.27). The evaluation of (3.34) can be accomplished in the same manner as that of (3.27) the only difference being that the poles z_1 and z_4 are of order k + 1 rather than k. Denoting the integral of (3.34) by I' and reducing the gamma coefficient gives

$$E(\lambda \mid \vec{x}) = \frac{(n+1)I'}{2I}. \tag{3.35}$$

In a like manner one concludes that the second moment is:

$$E(\lambda^2 \mid \vec{x}) = \frac{(n+3)}{2}\frac{(n+1)}{2}\frac{I''}{I} \tag{3.36}$$

where I" is likewise of the form (3.27) with poles at z_1 and z_4 of order $k + 2$. From this the variance of λ is determined to be:

$$\operatorname{Var}(\lambda \mid \vec{x}) = \frac{n+1}{2I} \left\{ \left(\frac{n+3}{2}\right) I'' - \frac{(I')^2}{I} \right\}$$

The author is not aware of any technique for obtaining the marginal distribution of λ in closed form. The evaluation of the moments of the distribution was accomplished by interchanging the order of integration of μ and λ which is permissible by Tonelli's theorem as referenced in Royden (21). Since $\pi(\lambda)$ is unavailable quantile estimates and probability statements concerning λ are not available. The author can suggest but two approaches to the problem of quantile estimation in this case. First one may obtain $\pi(\lambda)$ in tabular form by computing $\int_0^\infty \frac{\lambda^{\frac{n-1}{2}}}{\Gamma\left(\frac{n+1}{2}\right) I\mu^{3/2}} \exp - \lambda(a\mu^{-2} - n\mu^{-1} + b)d\mu$ by numerical quadrature. Numerical quadrature could then be applied to these tabled function values. Secondly $E(\lambda^k \mid \vec{x})$ may be readily calculated either by the techniques of complex variables utilized in obtaining (3.35) or preferably by numerical quadrature. An approximating distribution from the Pearson family could then be fitted. The author is not aware of how satisfactory either of these approaches would be to the approximation of $\pi(\lambda)$.

Even modal estimation in this situation is somewhat cumbersome. First $\pi(\lambda)$ is known to be continuous and integrable. Therefore $\pi(\lambda)$ is not monotone increasing on (o,∞). Also the integrability of $\pi(\lambda)$ implies $\pi(\lambda)$ is not constant. Hence $\pi(\lambda)$ must possess a mode. Since $\pi(\lambda = o) = o$ the mode does not occur at zero and therefore $\pi(\lambda)$ has a mode in the interior of (o,∞).

Again since $\pi(\lambda)$ is continuous and integrable, $\pi(\lambda)$ is bounded so that the mode is a solution to $D_\lambda \pi(\lambda) = o$. Since no formula exists for $\pi(\lambda)$ the process of determining $D_\lambda \pi(\lambda)$ will involve approximation of $\pi(\lambda)$ and then approximation of $D_\lambda \pi(\lambda)$. This would seem to be an undesirable procedure. A more desirable procedure would be to evaluate $\int_o^\infty D_\lambda \pi(\mu,\lambda)d\mu$ for successive values of λ thereby replacing one numerical process with an exact procedure. The justification of the above interchange of limits is the following topic.

In order to show that $\pi(\lambda)$ is differentiable and that the aforementioned interchange is permissible it is sufficient to show the following Apostal (2):

(i) $D_\lambda \pi(\mu,\lambda)$ is continuous for $o < \lambda < \infty$, $o < \mu < \infty$

(ii) $\int_o^\infty \pi(\mu,\lambda)d\mu$ converges pointwise for $\lambda \varepsilon (o,\infty)$

(iii) $\int_o^\infty D_\lambda \pi(\mu,\lambda)d\mu$ converges uniformly for $\lambda \varepsilon (o,\infty)$

Apart from the normalizing constant;

$$D_\lambda \pi(\mu,\lambda) = [\frac{n-1}{2} - \lambda(a\mu^{-2} - n\mu^{-1} + b)]\lambda^{\frac{n-3}{2}} \mu^{-3/2} \exp - \lambda(a\mu^{-2} - n\mu^{-1} + b)$$

The derivative $D_\lambda \pi(\mu,\lambda)$ is composed of continuous functions of μ and λ and therefore continuous for $o < \mu < \infty$ and $o < \lambda < \infty$. Condition (ii) will be established by use of the Weirstrass M test given by Apostol

(2). This will show that $\int_0^\infty \mu^{-3/2}\lambda^{\frac{n-1}{2}} \exp - \lambda(a\mu^{-2} - n\mu^{-1} + b)d\mu$ is uniformly convergent. This can be accomplished by maximizing the integrand above with respect to λ for each value of μ. The resulting function of μ which bounds the preceding integrand is

$$\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}} e^{\frac{-n-1}{2}} \frac{\mu^{n-1}}{\mu^{3/2}(a - n\mu + b\mu^2)^{\frac{n-1}{2}}} \qquad (3.37)$$

For $n > 1$ the function in (3.37) is integrable on (o,a) where $a > o$. For large μ (3.37) behaves as $\mu^{-3/2}$ on (a,∞) and could be shown to be integrable over that region by the limit comparison test. One therefore concludes that $\int_o^\infty \frac{\mu^{n-1}}{\mu^{3/2}(a - n\mu + b\mu^2)^{\frac{n-1}{2}}} d\mu$ is finite which implies, by the Weirstrass M test, that $\int_o^\infty \pi(\mu,\lambda)d\mu$ converges uniformly and therefore pointwise.

Finally consider (iii). Apart from a constant

$$\int_o^\infty D_\lambda \pi(\mu,\lambda)d\mu = \frac{n-1}{2}\int_o^\infty \mu^{-3/2}\lambda^{\frac{n-3}{2}} \exp(-\lambda(a\mu^{-2} - n\mu^{-1} + b))d\mu$$
$$- \int_o^\infty \mu^{-3/2}(a\mu^{-2} - n\mu^{-1} + b)\lambda^{\frac{n-3}{2}} \exp(-\lambda(a\mu^{-2} - n\mu^{-1} + b))d\mu \quad (3.38)$$

provided both integrals exist. It has already been established that integrals of the form of the first summand of the right hand side of (3.40) are uniformly convergent for $\lambda\varepsilon(o,\infty)$. Applying the Weirstrass M test to the second integral by maximizing the integrand with respect to λ for each value of μ gives a bounding function (apart from a constant) to be

$$f(\mu) \propto \frac{\mu^{n - \frac{13}{2}}}{(a - n\mu + b\mu^2)^{\frac{n-5}{2}}} \tag{3.39}$$

The function $f(\mu)$ specified in (3.39) possesses a convergent integral on the interval (o,a) for $n \geq 6$ and the integrand is of order $\mu^{-3/2}$ on the interval (a,∞) for large μ. One therefore concludes that for $n \geq 6$ $\int_o^\infty f(\mu)d\mu < \infty$ and consequently $\int_o^\infty D_\lambda \pi(\mu,\lambda)d\mu$ is uniformly convergent. It is reasonable to assume that a sample will contain at least 6 observations and in such cases it has been established that: (a) $\pi(\lambda)$ is differentiable and (b) $D_\lambda \pi(\lambda) = \int_o^\infty D_\lambda \pi(\mu,\lambda)d\mu$.

11. Estimation of λ When μ is Known With a Diffuse Prior Distribution

As determined following (2.10) the posterior distribution when $\mu = \mu_o$ of λ is gamma distributed or $\lambda(a\mu_o^{-2} - n\mu_o^{-1} + b) \sim \chi^2(n)$. From this one concludes that

$$E(\lambda \mid \vec{x}) = \frac{n}{2(a\mu_o^{-2} - n\mu_o^{-1} + b)}$$

$$Var(\lambda \mid \vec{x}) = \frac{n}{2(a\mu_o^{-2} - n\mu_o^{-1} + b)^2}$$

$$\lambda_{max} = \frac{n - 2}{2(a\mu_o^{-2} - n\mu_o^{-1} + b)}$$

where $a = \frac{1}{2}\sum X_i$ and $b = \frac{1}{2}\sum X_i^{-1}$.

The posterior median can be determined by iterative evaluation of the incomplete χ^2 distribution. Procedures for evaluating the median in a similar case were discussed following formula (3.26) on page 31.

For $n > 30$ the posterior median of a χ^2 variate with n degrees of freedom is approximately $n\,(1 - \frac{2}{9n})^3$.

12. Simultaneous Modal Estimation of μ and λ With a Diffuse Prior Distribution

It is sufficient to maximize the log of the density or $f(\mu,\lambda) = \frac{n-1}{2}\log\lambda - \frac{3}{2}\log\mu - \lambda(a\mu^{-2} - n\mu^{-1} + b)$. Setting $D_\mu f = o$ and $D_\lambda f = o$ gives the intermediate results:

$$\lambda = \frac{-3\mu^2}{2(n\mu - 2a)} \quad \text{and} \quad \lambda = \frac{(n-1)\mu^2}{2(a - n\mu + b\mu^2)}$$

respectively.

Equating these expressions for λ gives

$$\mu_{max} = \frac{n-4}{3}\left\{-\bar{x}_h + \sqrt{\bar{x}_h^2 + 3\bar{x}\bar{x}_h\frac{(2n-5)}{(n-4)^2}}\right\}$$

$$\mu_{max} = \frac{(n-1)\mu^2}{2(a - n\mu + b\mu^2)}\bigg|_{\mu = \mu_{max}}$$

In the same manner as was employed in the case of the marginal mode of μ one can show that μ_{max} is a consistent estimator of μ. In fact for large n, $\mu_{max} \approx \frac{(2n-5)\bar{x}}{2n-8}$.

Substituting this asymptotic value of μ_{max} into the equation for λ_{max} results in the asymptotic expression:

$$\lambda_{max} \approx \frac{\bar{x}\bar{x}_h}{\left(\frac{n}{n-1}\right)\bar{x} - \frac{2n(2n-8)}{(2n-5)^2}\,\bar{x}_h}$$

The maximum likelihood estimator $\hat{\lambda}$ for λ is obtained from Chhikara (4):

$$\hat{\lambda} = \frac{\bar{x}x_h}{\bar{x} - x_h}$$

and it is known that $\frac{n\lambda}{\hat{\lambda}} \sim \chi^2(n - 1)$. From this one concludes that $\hat{\lambda}$ follows an inverted gamma distribution with parameters $\psi = \frac{n\lambda}{n - 1}$ and $v = n - 1$. Therefore $E(\hat{\lambda}) = \frac{n\lambda}{n - 3}$ and Var $\hat{\lambda} = \frac{2n^2\lambda^2}{(n - 3)^2(n - 5)}$. Consequently $\hat{\lambda} \xrightarrow{p} E(\hat{\lambda})$ by Chebyshev's Law of Large Numbers. Since $\lim_{n\to\infty} E(\hat{\lambda}) = \lambda$ pointwise then $\hat{\lambda} \xrightarrow{p} \lambda$. Now since λ_{max} converges pointwise to $\hat{\lambda}$ then given $\epsilon > o$ there exist natural numbers N_1 and N_2 such that for $n > \max\{N_1, N_2\}$ the following hold simultaneously

$$|\lambda_{max} - \hat{\lambda}| < \epsilon \text{ and } \Pr\{|\hat{\lambda} - \lambda| < \epsilon\} > 1 - \epsilon$$

For any such n $|\lambda_{max} - \lambda| - \epsilon \leq |\hat{\lambda} - \lambda|$ by the triangle inequality. Hence, $\Pr(|\lambda_{max} - \lambda| - \epsilon < \epsilon) \geq \Pr(|\hat{\lambda} - \lambda| < \epsilon) > 1 - \epsilon$ or $\Pr(|\lambda_{max} - \lambda| < 2\epsilon) > 1 - \epsilon > 1 - 2\epsilon$. One then concludes that λ_{max} is a consistent estimator of λ.

CHAPTER IV

HPD REGIONS

A region of credibility is the Bayesian analogue of the classical confidence interval and may be used to provide an interval estimate of a parameter or to provide a test of hypotheses on a simple null hypothesis. The difference between the classical confidence interval and a credibility region is that the confidence coefficient associated with a confidence interval arises from considerations of the sampling distribution of which it is a realization while the confidence coefficient of a credibility region is a subjective probability associated with that particular interval.

Definition 4.1: Let $F(x, \theta)$ denote the distribution of the r.var. X. Assume a priori that Θ is distributed according to $G(\theta)$ such that the posterior distribution of Θ is absolutely continuous with density $p(\theta)$. Then a $(1 - \alpha)$ HPD region is a subset H of the support of p such that:

(i) $\int_H p(\theta)d(\theta) = 1 - \alpha$

(ii) If $\theta_1 \in H$ and $\theta_2 \notin H$ then $p(\theta_1) \geq p(\theta_2)$

If a subset A of the parameter space satisfies (i) above then A is said to be a $(1 - \alpha)$ credible region for θ. An experimenter may not desire to construct an HPD region but some other type of credible region. For example, one may desire to bound the parameter above or below which might not lead to a region that satisfies (ii).

In general there are several reasons for constructing HPD regions in preference to other credible regions. By (ii) of definition 4.1 an HPD region contains the most likely values of the parameter. A second facet of property (ii) is that an HPD region is, of all credible regions with equal probability content, the smallest in terms of generalized volume. In the univariate case this amounts to possessing minimum width.

Unfortunately HPD regions have some shortcomings. The most important of these from the practical standpoint is that HPD regions are more difficult to construct for asymetrical distribution than other common credible regions. Of lesser practical importance but worthy of comment are the following deficiencies. First, an HPD region may not exist. For example, if the posterior distribution were of the form $p(\theta) = \frac{|\sin\theta|}{4}$, $0 < \theta < 2\pi$ and zero elsewhere then no HPD region could be developed for θ. Second, HPD regions are not unique for r. variables following a uniform probability law. Last, HPD regions are not invariant under bijective transformations of the parameters but are invariant under nonsingular linear transformations as referenced in Zellner (28).

Box and Tiao (3) have developed an equivalent definition of HPD regions which is sometimes useful in construction of HPD regions.

<u>Definition 4.2</u>: Under the same distributional assumption as in definition 4.1 the $(1 - \alpha)$ HPD region H can be defined as:

$$H = \left\{\theta_o : \Pr\left\{p(\theta) > p(\theta_o)\right\} \leq 1 - \alpha\right\}$$

The boundary B(H) is represented by

$$B(h) = \left\{\theta_o : \Pr\left\{p(\theta) > p(\theta_o)\right\} = 1 - \alpha\right\}$$

Without dwelling on the equivalence of 4.1 and 4.2 one may note that $\Pr\left\{p(\theta) > p(\theta_o)\right\}$ is non-increasing in $p(\theta_o)$. Therefore, if $\theta_1 \epsilon H$ and $p(\theta_2) > p(\theta_1)$ then $\Pr\left\{p(\theta) > p(\theta_2)\right\} \leq \Pr\left\{p(\theta) > p(\theta_1)\right\} \leq 1 - \alpha$ so that $\theta_2 \epsilon H$. Therefore, definition 4.2 implies constraint (ii) of definition 4.1.

Some of the techniques for constructing HPD regions will now be discussed. This discussion will be principally limited to the development of univariate HPD regions. Multivariate HPD regions can be constructed but the computational difficulties encountered are disproportionately greater than those for the one dimensional case. One procedure that is fairly obvious is to employ Newton's iterative technique. Given a unimodal differentiable density $p(\theta)$, the HPD region will be of the form $H = (s,t)$ where $\int_s^t p(\theta)d\theta = 1 - \alpha$ and $p(s) = p(t)$. Define $P(s,t) = \int_s^t p(\theta)d\theta - 1 + \alpha$ and $Q(s,t) = p(s) - p(t)$. Then the boundary of the $(1 - \alpha)$ HPD region H is the simultaneous solution of:

$$\begin{aligned} P(s,t) &= o \\ Q(s,t) &= o \end{aligned} \tag{4.1}$$

If one employs the bivariate Taylor's expansions of P and Q to obtain approximating equations and substitutes these in (4.1) one obtains:

$$\begin{aligned} P(s_o,t_o) + D_sP(s_o,t_o)(s - s_o) + D_tP(s_o,t_o)(t - t_o) &= o \\ Q(s_o,t_o) + D_sQ(s_o,t_o)(s - s_o) + D_tQ(s_o,t_o)(t - t_o) &= o \end{aligned} \tag{4.2}$$

Here (s_o,t_o) is arbitrary and represents a trial solution for H. Solving the system (4.2) yields the iterative equations required.

They are:

$$s = s_o + \frac{p'(t_o)P(s_o,t_o) + p(t_o)Q(s_o,t_o)}{p(s_o)p'(t_o) - p'(s_o)p(t_o)}$$

$$t = t_o + \frac{p'(s_o)P(s_o,t_o) + p(s_o)Q(s_o,t_o)}{p(s_o)p'(t_o) - p'(s_o)p(t_o)} \tag{4.3}$$

The convergence of (4.3) to the desired solution depends upon the proximity of the trial value to the true value and the magnitude of the Jacobian at the point of intersection as referenced in Hildebrand (8).

Due to the author's ignorance of the general stability of the aforementioned method it was decided to approach the construction problem in a different manner. Noting that the magnitude of the density plays a crucial role in the definition of an HPD region and consideration of graphical displays motivated the author to solve the problem in two stages. First, define $F(c) = \int_{p(\theta) \geq c} p(\theta)d\theta$. Then given α solve the univariate equation $F(c) = 1 - \alpha$ for c_o. Finally solve the equation $p(\theta) = c_o$ to obtain the endpoints of the HPD region. The author was not able to locate a numerical procedure for finding $h^{-1}(c)$ for arbitrary h and c. In each case h was required to be differentiable or a polynomial. Therefore, the author developed the subroutine AINVRT listed in the appendix. This procedure is somewhat more general than is required since the author intended to display a bivariate HPD region. The procedure AINVRT is easy to use; it will handle any univariate continuous function h that is defined everywhere; it is reasonably stable and is therefore slow. Its speed is a function of the mode in which it is employed and whether there are multiple solutions indicated. The numerical examples on page 62 were obtained by using the routine HPDU which uses AINVRT and the previously mentioned numerical quadrature program SQANK. This procedure was developed independently of Box and Tiao definition but is closely linked with it in that it amounts to determining c such that $\Pr\{p(\theta) > c\} = 1 - \alpha$ and letting $B(H) = \{p^{-1}(c)\}$.

Box and Tiao have used definition 4.2 in another way to construct HPD regions. Starting from the same point, i.e. finding a solution to $\Pr\{p(\theta) > c\} = 1 - \alpha$ they attempt to determine the distribution of $p(\theta)$ and determining the $1 - \alpha$ quantile or by inverting the expression in definition 4.2 and using the distribution of Θ to find

$$H = \{\theta;\ \Pr\{\theta \epsilon p^{-1}(p(\theta) > c)\} = 1 - \alpha\} \qquad (4.4)$$

A complication in the last representation is that the form of p may make the form of $p^{-1}(p(\theta) > c)$ impossible to represent in terms of algebraic expressions. The author encountered this repeatedly in this study. Box and Tiao apparently prefer to generate the distribution of $p(\theta)$ or to approximate the distribution of $p(\theta)$. The author will demonstrate the procedure to be followed in fitting a member of the Pearson's family of distribution to that of $p(\theta)$ and then discuss those cases where (4.4) may be of value.

Let the density of the r.var. Y be $f(y)$ and suppose one desires to approximate the distribution of $f(y)$ by a member of the Pearson family. In general the moments of $f(y)$ can be approximated by expanding $f(y)$ about its mean μ_y', taking expectations and dropping all terms greater than a predetermined order. Since all members of the Pearson family are determined by their first four moments it is sufficient to approximate the first four moments of $f(y)$. The author assumes that these moments may be adequately approximated by the first four terms of the expansion of f(y) about μ'_y. In the following $\mu^n_y = E(y - \mu'_y)^n$. One then obtains the following formulae for approximating the moments of f(y).

To simplify notation, the substitution $z = \mu'_y$ is made.

$$E(f(y)) \doteq f(z) + \frac{f''(z)\mu^2_y}{2} + \frac{f'''(z)\mu^3_y}{6} + \frac{f''''(z)\mu^4_y}{24}$$

$$E(f^2(y)) \doteq f^2(z) + [f(z)f''(z) + (f'(z))^2]\mu^2_y + [f(z)f'''(z) + 3f'(z)f''(z)]\frac{\mu^3_y}{3} + [f(z)f''''(z) + 4f'(z)f'''(z) + 3(f''(z))^2]\frac{\mu^4_y}{24}$$

$$E(f^3(y)) \doteq f^3(z) + 3[2f(z)(f'(z))^2 + f^2(z)f''(z)]\frac{\mu^2_y}{2} + [2(f'(z))^3 + 6f(z)f'(z)f''(z) + (f(z))^2f'''(z)]\frac{\mu^3_y}{2} + [23(f'(z))^2f''(z) + 12f(z)(f''(z))^2 + 14f(z)f'(z)f'''(z) + 6(f'(z))^2f''(z) + [(f(z))^2f''''(z)]\frac{\mu^4_y}{8}$$

$$E(f^4(y)) \doteq f^4(z) + 2[3f^2(z)(f'(z))^2 + f^3(z)f''(z)]\mu^2_y + 2[6f(z)(f'(z))^3 + 9(f(z))^2f'(z)f''(z) + f^3(z)f'''(z)]\frac{\mu^3_y}{3} + [6(f'(z))^4 + 54f(z)(f'(z))^2f''(z) + 21(f(z))^2f'(z)f'''(z) + 18(f(z))^2(f''(z))^2 + (f(z))^3f''''(z)]\frac{\mu^4_y}{6}$$

Let $\mu_k' = Ef^k(y)$. Then the author has displayed above μ_1', μ_2', μ_3', and μ_4'.

The second, third, and forth central moments can then be determined by

$$\mu_2 = \mu_2' - (\mu_1')^2 \tag{4.5}$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3 \tag{4.6}$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2' - 3\mu_1'^4 \tag{4.7}$$

The particular member of the Pearson family which is appropriate for approximating the distribution of f(y) is determined by calculating

$$K = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} \tag{4.8}$$

where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\beta_2 = \frac{\mu_4}{\mu_2}$. The interested reader is referred to Kendall and Stuart (13) for a more comprehensive discussion regarding this procedure.

The above procedure cannot be applied if the r.var. Y possesses no moments. This is the case with the posterior distribution of μ when the Jeffrey prior distribution is utilized. One particular case that can be examined in more detail is that in which λ is unknown, μ is known, and the Jeffrey prior is chosen as representative of the prior knowledge on λ. The reader may recall that the posterior distribution of λ is such that $Y = k\lambda \sim \chi^2(n)$ where $k = 2(a\mu_o^{-2} - n\mu_o^{-1} + b)$. One may then find a $(1 - \alpha)$ HPD region for Y and due to the aforementioned invariance thereby obtain a $(1 - \alpha)$ HPD region for λ.

In this case one need not use the approximating formulae of page 55 since $\int_0^\infty [f(y)]^k dy$ is readily integrable.

If $Y \sim \chi^2(n)$ then the (k - 1)st moment about zero of the density of Y is:

$$E[f^{k-1}(y)] = \int_0^\infty \frac{y^{k(n/2-1)} e^{-k/2^y}}{2^{\frac{kn}{2}} [\Gamma(n/2)]^k} dy$$

$$= \frac{\Gamma(\frac{kn}{2} - (k-1))}{2^{\frac{kn}{2}} [\Gamma(n/2)]^k (k/2)^{\frac{kn}{2} - (k-1)}}$$

This after some simplification becomes:

$$E[f^{k-1}(y)] = \frac{\Gamma(k(\frac{n}{2} - 1) + 1)}{2^{k-1} k^{k(\frac{n}{2} - 1) + 1} [\Gamma(\frac{n}{2})]^k}$$

The first four moments about zero of f(y) are:

$$\mu_1' = \frac{\Gamma(n - 1)}{2^n [\Gamma(\frac{n}{2})]^2},$$

$$\mu_2' = \frac{9\Gamma(\frac{3n}{2} - 2)}{(\sqrt{3})^{3n} [\Gamma(\frac{n}{2})]^3},$$

$$\mu_3' = \frac{8\Gamma(2n - 3)}{16^n [\Gamma(\frac{n}{2})]^4},$$

$$\mu_4' = \frac{\Gamma(\frac{5n}{2} - 4)}{16 \cdot 5^{21} [\Gamma(\frac{n}{2})]^5}.$$

The author will not present a complete example of this process but will only indicate the direction in which the analysis leads.

If, for example, a sample of size ten is obtained then the above formula for the moments of a χ^2 density are calculated:

$$\mu_1' = 6.8359 \times 10^{-2}$$
$$\mu_2' = 5.4333 \times 10^{-3}$$
$$\mu_3' = 4.5884 \times 10^{-4}$$
$$\mu_4' = 4.0047 \times 10^{-5}\,.$$

Using the above in formulae (4.5), (4.6) and (4.7) gives:

$$\mu_2 = 7.6034 \times 10^{-4}$$
$$\mu_3 = -1.6528 \times 10^{-5}$$
$$\mu_4 = 1.4119 \times 10^{-6}$$

from which we note that the distribution of a $\chi^2(10)$ density is compact and skewed to the left.

Finally, from μ_2, μ_3, and μ_4 above one can compute:

$$\beta_1 = .6215$$
$$\beta_2 = 1.8570 \times 10^{-3}$$

which gives:

$$k = 9.5910 \times 10^{-2}\,.$$

Referring again to Kendall and Stuart (13) one determines that since $o < k < 1$ the appropriate approximating density is:

$$g(t) = k(1 + a^{-2}t^2)^{-m} \exp\left[-v \operatorname{Tan}^{-1}(a^{-1}t)\right] \qquad (4.9)$$
$$\infty < t < \infty$$

where t represents the r.var. defined by the $\chi^2(10)$ density function applied as a transformation on a χ^2 variate.

The constants a, v, and m can be determined by formulae given in Kendall and Stuart. The factor c must be determined by numerical quadrature. Since the author has already expended considerable funds on numerical integration this example will not be continued numerically. The essential idea is that after fitting (4.9) one would then determine the $(1 - \alpha)$ quantile of t and then solve the equation

$$\frac{y^4 e^{-y/2}}{2^5 \Gamma(5)} = t$$

for values y_1 and y_2. The values y_1 and y_2 constitute the $(1 - \alpha)$ HPD region for the χ^2 variate Y. The corresponding approximate HPD region for λ would then be obtained by setting $\lambda_i = y_{i/k}$ where

$$k = 2\left(\frac{\Sigma X_i}{2}\, \mu_o^{-2} - 10\mu_o^{-1} + \tfrac{1}{2}\, \Sigma X_i^{-1}\right) .$$

Corresponding to each case of point estimation treated in Chapter III there is associated the problem of interval estimation. As discussed any of these univariate cases can be handled in a direct numerical fashion by programs referenced in the appendix. The question arises, however, as to which of the estimation situations of Chapter III can be handled directly by (4.4). The author has not been particularly successful in this regard. This is because the distribution is not known or the inversion procedure is not algebraic. Even in those cases for which an expression for H is determined the reader will note that the endpoints are not specified. This limits the direct use of these expressions to testing a simple null hypothesis.

1. An HPD Region for μ Averaged Over λ With an Informative Prior Distribution

In this case the posterior distribution of μ is

$$\pi(\mu) \propto \frac{\mu^{2(d+1)}}{(a - b\mu + c\mu^2)^{d+1}}$$

The HPD region for μ is defined by $H(\mu) = \{\mu_o: \Pr\{\pi(\mu) < \pi(\mu_o)\} \geq \alpha$. Let $k(\mu_o) = \pi(\mu_o)^{1/d+1}$ then (4.5) is the above is equivalently expressed as:

$$H(\mu) = \left\{ \mu_o: \ \Pr\left\{ \frac{\mu^2}{a - b\mu + c\mu^2} < k(\mu_o) \right\} \geq \alpha \right\}$$

$$\text{or } H(\mu) = \left\{ \mu_o: \ \Pr\left\{ \frac{(ck(\mu_o) - 1)\mu^2}{k(\mu_o)} - b\mu + a > o \right\} \geq \alpha \right\}.$$

$$\text{Let } h = \sqrt{b^2 - \frac{4a[k(\mu_o)c - 1}{k(\mu_o)}}$$

$$\text{and let } R = \left[\frac{(b - h)k(\mu_o)}{2(ck(\mu_o) - 1)}, \frac{(b + h)k(\mu_o)}{(2ck(\mu_o) - 1)} \right] \cap [o,p]$$

then

$$H(\mu) = \left\{ \mu_o: \ [k(\mu_o) < \tfrac{1}{c}) \wedge (\Pr\{\mu \epsilon R\} \geq \alpha)] \vee [(k(\mu_o) > \tfrac{1}{c}) \wedge (\Pr\{\mu \epsilon R\} \geq \alpha)] \right\}. \qquad (4.10)$$

From (4.10) one can determine whether or not a given θ_σ falls in the $(1 - \alpha)$ HPD region.

2. An HPD Region for μ When λ Is Known With an Informative Prior Distribution

Referring to Chapter II one finds that $H(\mu)$ has the form:

$$H(\mu) = \left\{ \mu_o : \; Pr\left\{ \frac{1}{\mu}\left(\frac{a}{\mu} - b\right) > k(\mu_o) \right\} \geq \alpha \right\}$$

where $k(\mu_o) = - \; n \, \pi(\mu_o)$

$$\text{Let } h = \sqrt{b^2 + 4ak(\mu_o)}$$

$$\text{and } R = \left[\frac{-b-h}{2k(\mu_o)}, \frac{-b+h}{2k(\mu_o)} \right] \cap [o,p]$$

$$\text{then } H(\mu) = \{ \mu_o : \; (k(\mu_o) > o \wedge (Pr(\mu_o \epsilon R) \geq \alpha) \vee (k(\mu_o) < o \wedge (Pr(\mu_o \epsilon R) \geq \alpha))$$

The author has considered the other cases with no success.

In the author's opinion these representations are of little value in general because the endpoints are not displayed. All but one of the univariate cases can be handled by HPDU, however, so in practice these regions could be determined as required.

The author will illustrate a particular HPD region with a numerical example. Suppose that both μ and λ are unknown and a $1 - \alpha$ HPD region is desired for μ. From (2.8) the marginal of μ is given as:

$$\pi(\mu) \propto \frac{\Gamma\left(\frac{n+1}{2}\right)}{\mu^{3/2}(a\mu^{-2} - n\mu^{-1} + b)^{\frac{n+1}{2}}} \qquad o < \mu < \infty \qquad (4.11)$$

A sample of size 50 is drawn. The data represented in Table I were generated by a program developed by the author. A listing of INGAUS is included in the appendix. The data in the table come from an inverse Gaussian distribution with parameters $\mu = 2$ and $\lambda = 8$. The data points were calculated to seven decimal places. The data in the Table have been rounded to two decimal places for ease of presentation.

TABLE I

A RANDOM SAMPLE OF 50 RANDOM DEVIATES FROM AN INVERSE GAUSSIAN DISTRIBUTION WITH PARAMETERS MU=2 AND LAMBDA=8

1.20	1.24	4.13	5.21	1.63
1.81	0.71	1.74	2.36	1.26
3.10	1.63	0.83	4.88	1.02
0.92	3.78	1.77	1.92	4.40
0.87	0.64	1.74	2.37	2.32
1.75	1.71	1.24	1.06	3.77
1.46	0.99	1.73	1.04	1.89
3.54	2.27	0.95	2.16	3.76
1.47	1.88	1.56	2.29	1.95
1.62	2.23	1.57	4.40	1.07

Recall that $a = \frac{\Sigma X_i}{2}$, n is the sample size, and $b = \frac{1}{2} \Sigma X_i^{-1}$. For the data in the above Table, the constants in (4.11) are a = 51.4109826, n = 50, and b = 15.8444061. From Table I the following statistics were also calculated:

$$\bar{x} = 2.0567627$$

$$\bar{x}_h = 1.5778437$$

$$s_x = 1.1548328$$

$$\hat{\lambda} = 6.7761965$$

The modal estimators of μ were calculated to be

$$\hat{\mu}_1 = 2.038683552 \quad \text{(average over } \lambda)$$

$$\hat{\mu}_2 = 2.03799392 \quad \text{(simultaneous)}$$

The author notes that in this case both Bayesian estimators are nearer the true value of $\mu = 2$ than the classical estimator $\bar{x}$. The normalizing constant for (4.11) is $(1.477 \times 10^9)^{-1}$. This constant has an absolute error of less than $1. \times 10^5$ or a relative error of about 10^{-3}.

The density that was used in this example is:

$$\pi(\mu) = \frac{\Gamma(25.5)}{(1.477325 \times 10^9)\mu^{3/2}(a\mu^{-2} - n\mu^{-1} + b)^{25.5}}$$

A 95% and 99% HPD for μ will be displayed. Approximately 30 to 35 iterations were required to determine each of the regions.

The 95% and 99% HPD regions computed by HPDU for μ were (1.76, 2.42) and (1.68, 2.59), respectively. If one fits a confidence interval based on $\bar{x}$ and s^2 one obtains the corresponding intervals (1.78, 2.33) and (1.63, 2.48). The major discrepancy is due to the fact that $\pi(\mu)$ is positively skewed causing larger departures at the right endpoints of the actual and approximated intervals.

The author regrets that he is not able to display a multivariate HPD region for μ and λ. This numerical problem took up approximately one month of the author's time with no results. The reader should not infer that the problem cannot be solved. It was merely felt that the expenditure of time and money for a single numerical example were becoming unrealistic.

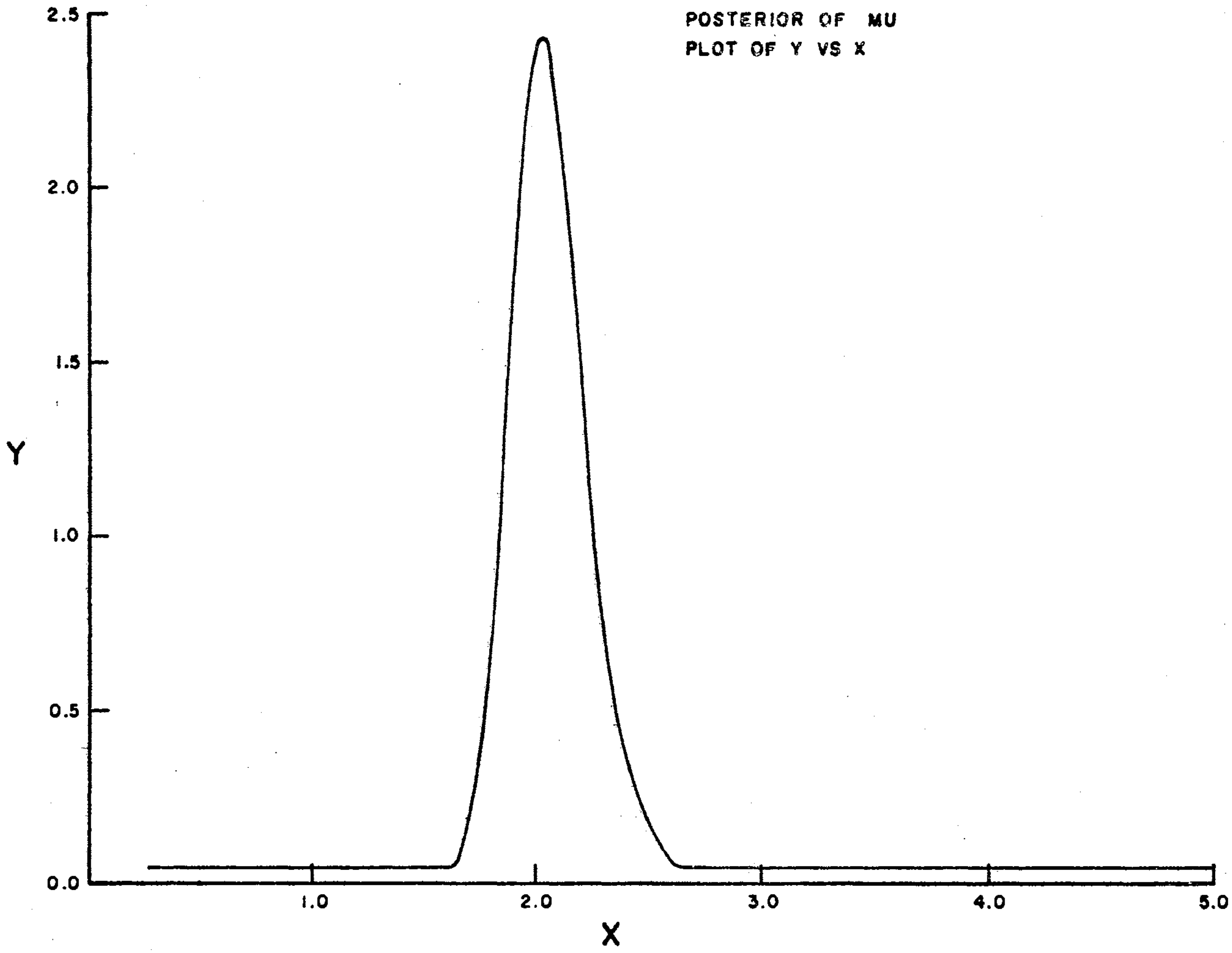

Figure 1. Graph of Posterior Distribution of mu

The reader may find the following contour graph of the posterior of $\pi(\mu,\lambda)$ to be informative. One may note that for some values of α the HPD region may not be a convex set.

In concluding this chapter the author will briefly explain the basis for INGAUS, the generator used for selecting the fifty data points employed in the numerical examples involving HPD regions.

If $X \sim$ Inverse Gaussian (μ,λ) then from Chhikara (4) it is known that $Y = \frac{\lambda(X-\mu)^2}{\mu^2 X} \sim \chi^2(1)$. If one solves for Y in terms of X one obtains:

$$X_1 = \tfrac{1}{2}(A - B) \qquad \text{and} \qquad X_2 = \tfrac{1}{2}(A + B)$$

$$\text{where } A = \mu\left(2 + \frac{\mu Y}{\lambda}\right) \qquad \text{and} \qquad B = \sqrt{\frac{\mu^3 Y}{\lambda}\left(4 + \frac{\mu Y}{\lambda}\right)}.$$

The fact that this relationship between the inverse Gaussian distribution and the chi-square distribution is not one-to-one does not create insurmountable problems. The reason is that if one examines the graph of $Y = \frac{\lambda(X-\mu)}{\mu^2 X}$ it is apparent that $X_1 < \mu$ and $X_2 > \mu$.

Since μ and λ are input variables to INGAUS it is easy to determine $\Pr\{X \leq \mu\} = p$ to be:

$$p = .5 + \exp(2\lambda/\mu)[1 - \Phi(2\sqrt{\lambda/\mu})]$$

where $\Phi(\cdot)$ is the cdf of a standard normal deviate. The procedure followed is then:

(1) generate a standard normal deviate Z by GAUSF;

(2) generate Y, a $\chi^2(1)$ deviate by using $[\text{GAUSF}(0)]^2$;

(3) solve for X_1, X_2;

(4) compute p;

(5) generate a uniform (0,1) deviate R by RANF;

(6) select X_1 if $R < p$ and X_2 otherwise.

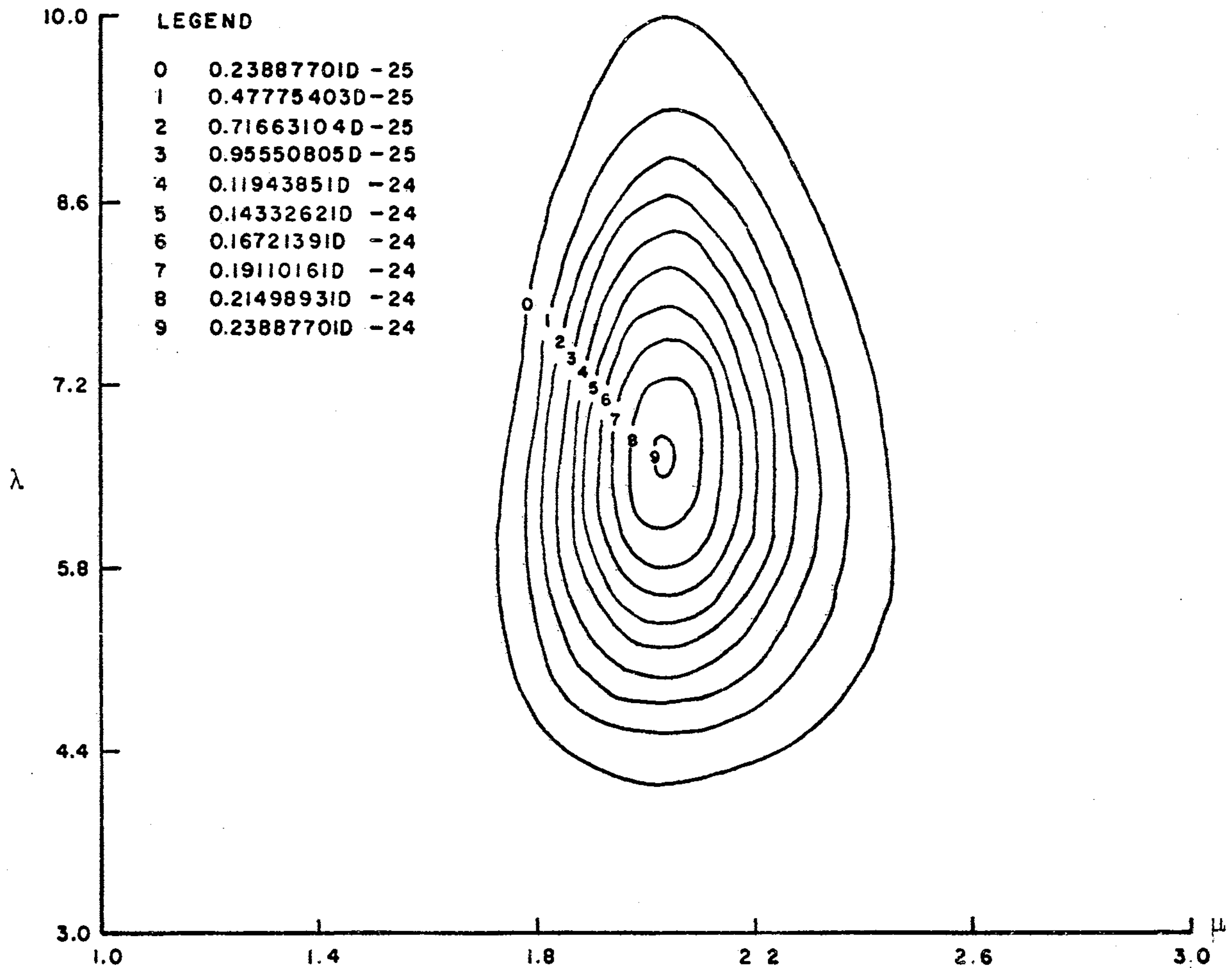

Figure 2. Contours of the Joint Posterior of mu and lambda

CHAPTER V

SOME PRINCIPLES OF EMPIRICAL BAYESIAN ANALYSIS

Empirical Bayesian (EB) procedures are applicable in the following context. An experimenter has available m samples of distinct r.var. These samples may be labeled $\{X_{1i}\}_{i=1}^{n}, \{X_{2i}\}_{i=1}^{n} \ldots \{X_{mi}\}_{i=1}^{n}$. It is not necessary or even very probable that each of the samples will be of the same size. However, using equal sample sizes simplifies the notation and in practice is no hindrance. These distinct samples may have been collected by the experimenter at different times or in different places. In any case the experimenter believes that although the r.var. $X_1, X_2, \ldots X_n$ are distinct they originate from the same conditional distribution defined by the cdf $F(x|\theta)$. It is assumed that $X_1, X_2, \ldots X_n$ are independent and are distinct r.var. only through random fluctuation in the r.var. Θ. The distribution $G(\theta)$ is assumed to remain constant throughout all experimentation. The experimenter may know the functional form of $G(\theta)$ or he may not. It is the purpose of EB techniques to develop inferential procedures regarding the realizations of Θ in this experimental situation. In the resolution of questions regarding the unknown variate Θ one makes a decision d dependent upon the outcome of the experiment resulting in a realization of X. Let D denote the space of all possible decisions d regarding θ. Further suppose that upon observing $X = x$ and in making the decision $d(x)$ one incurs a loss denoted by $L(d(x),\theta)$, where θ is the true value of Θ when X is observed.

Definition 5.1: The risk associated with the decision rule d is the expected loss when $\Theta = \theta$. The risk is denoted $R(d;\theta)$ and is therefore defined by

$$R(d;\theta) = \int_X L(d(x);\theta)f(x|\theta)dx$$

assuming $X|\theta$ is abs. continuous.

Definition 5.2: The Bayes Risk $W(d)$ associated with the decision rule d is the expected value over Θ of the risk $R(d;\theta)$ i.e.

$$W(d) = \iint_X L(d(x);\theta)f(x|\theta)dxg(\theta)d\theta,$$

Definition 5.3: The Bayes estimator of θ is that function $d(x)$ for which $W(d)$ is minimized so that d is a Bayes estimate with respect to $g(\theta)$ and $L(d,\theta)$ iff for each d^* it follows that $W(d) \leq W(d^*)$.

The most commonly encountered loss functions are squared-error loss $(L(d(x);\theta)) = (d(x)-\theta)^2)$ and absolute error loss. These are frequently used when $d(x)$ represents a point estimate of θ. When $d(x)$ is used in the context of testing then a 0 - 1 or linear loss function is most frequently used. These loss functions will be discussed in greater detail in Chapter VII.

The author will now present some general results of Bayesian estimation which are incorporated in EB procedures.

Theorem 5.1: The Bayes estimator of θ with respect to the distribution $g(\theta)$ and loss function $L(d(x);\theta) = (d(x)-\theta)^2$ is the mean of the posterior distribution of θ.

Proof:

The Bayes estimator is that estimator which minimizes W(d) where

$$W(d) = \iint_X (d(x)-\theta)^2 f(x|\theta)g(\theta)dxd\theta.$$

By Tonelli's theorem the order of integration may be interchanged so that W(d) may be rewritten as

$$W(d) = \int_X [\ (d(x)-\theta)^2 f(x|\theta)g(\theta)d\theta]dx \qquad (5.1)$$

The function W(d) will be minimized if for each fixed x the integrand in brackets above is minimized. Expanding the integrand in (5.1) one obtains:

$$[d(x)]^2 f(x) - 2d(x)\int \theta\ f(x;\theta)d\theta + \int \theta^2 f(x;\theta)dx.$$

The above is a quadratic function of d and for each fixed x attain its minimum at

$$d(x) = \frac{\int \theta f(x;\theta)d\theta}{f(x)} = \int \theta f(\theta|x)dx.$$

Therefore, the decision rule minimizing w(d) is the posterior mean of θ.

As was previously discussed the assumption in EB situations is that the experimenter has a sequence of past observation from which have been derived a sequence of estimators $d_i(x)$ $i=1,2,\ldots,m$. The form of $d_k(\cdot)$ is dependent upon the previous observations $\vec{x}_1,\vec{x}_2,\ldots\vec{x}_k$. Because all EB procedures require estimation of $g(\theta)$ or of $\int f(x|\theta)g(\theta)d\theta$ and because there are various approaches available in these estimation procedures there may be several distinct sequences $\{d_i(x)\}_{i=1}^{m}$. It is therefore necessary to consider what would be desirable properties of an EB estimator. Since the EB estimate $d_i(x)$ is a function of the r.var. X the EB estimator is a r.var. The loss incurred in this manner is taken into account by the Bayes risk. However, the form of the EB estimator $d_n(x)$ depends upon the n previous realizations of the r.var. X so that the Bayes risk is itself a random variable. The author will denote the Bayes risk associated with $d_n(x)$ by $W(d_n)$. From Maritz (16) one observes that one measure of the effectiveness of an EB estimator $d_n(x)$ is $E_n W(d_n)$ where the expectation is with respect to all past and present samples of size n. That is the expectation is taken with respect to the unknown density $\prod_{i=1}^{n} (\int f(x_i|\theta)g(\theta)d\theta)$.

Definition 5.4: An EB estimator d is strongly asymptotically optimal (a.o.) iff

$$\lim_{n\to\infty} E_n W(d_n) = W(d^*)$$

where d* is the Bayes estimator of θ with respect to g(θ).

Definition 5.5: An EB estimator d is weakly asymptotically optimal iff

$$\lim_{n\to\infty} W(d_n) \overset{p}{=} W(d^*)$$

Proposition 5.1: If d is strongly a.o. then d is weakly a.o.

Proof:

Let d* be the Bayes estimator of θ with respect to g(θ). Then by definition it must be the case that $W(d_n) \geq W(d^*)$ so that $E_n W(d_n) \geq E_n W(d^*) = W(d^*)$. Now suppose that d_n is strongly a.o., i.e.

$\lim\limits_{n\to\infty} E_n(W(d_n)) = W(d^*)$. Therefore, $\lim\limits_{n\to\infty} E_n(W(d_n) = \lim\limits_{n\to\infty} [E_n(W(d_n) - E_n(W(d^*)] = \lim\limits_{n\to\infty} E_n(W(d_n) - W(d^*)) = \lim\limits_{n\to\infty} E_n \; W(d_n) - W(d^*) \; = 0.$

Let $A = \left\{ \vec{x}: \dfrac{W(d_n) - W(d^*)}{\epsilon} \leq 1 \right\}$ and let $B = \sim A$.

Then given $\epsilon > o$ there exists an N such that for $n > N$

$E_n \left| \dfrac{W(d_n) - W(d^*)}{\epsilon} \right| < \epsilon$ because of the pointwise convergence.

One may then write:

$$E_n \left| \frac{W(d_n) - W(d^*)}{\epsilon} \right| = \int_A \left| \frac{W(d_n) - W(d^*)}{\epsilon} \right| dF_n(\vec{x})$$

$$+ \int_B \left| \frac{W(d_n) - W(d^*)}{\epsilon} \right| dF_n(\vec{x})$$

$$\geq \int_B \left| \frac{W(d_n) - W(d^*)}{\epsilon} \right| dF_n(\vec{x})$$

By the definition of B the last integral is not less than $\int_B dF_n(\vec{x}) = \Pr \left\{ |W(d_n) - W(d^*)| > \epsilon \right\}$. Therefore, given $\epsilon > o$ there exists an N such that for $n > N$ it follows that $\Pr \left\{ |W(d_n) - W(d^*)| > \epsilon \right\} < \epsilon$. Hence, d_n is weakly a.o. for θ.

A few theorems regarding asymptotic optimality that are relevant to this study will be presented. The following is due to Robbins (20).

<u>Theorem 5.2</u>: Given the EB estimator $d_n(x)$ of θ with respect to $g(\theta)$ and loss function $L(d(x);\theta)$ then $d_n(x)$ is strongly a.o. if

(i) $\int \left\{ \sup\limits_{d} L(d(x);\theta) \right\} g(\theta) d\theta < \infty$

and (ii) The Bayes risks for $d_n(x)$ converges in probability to the Bayes risks of the Bayes estimator for almost all x. That is $d_n(x)$ is weakly a.o.

If the parameter space is not bounded then (i) above cannot hold for the squared error loss function.

The following two results are due to Rutherford and Krutchkoff (22).

Theorem 5.3: Given the EB estimator $d_n(x)$ of θ with respect to quadratic loss and density $q(\theta)$ then $d_n(x)$ is weakly a.o. if

(i) $g(\theta)$ possesses a bounded moment of order greater than 2,

and (ii) $d_n(x)$ is a consistent estimator of the Bayes estimator $d^*(x)$

Theorem 5.4: If X is an absolutely continuous r.var. for all values of Θ such that the density satisfies

$$D_x \log f(x|\theta) = a(x) + b(x)\theta$$

where $a(x)$ and $b(x)$ are any functions of x such that $b(x) \neq o$ and if $\int D_x f(x|\theta)g(\theta)d\theta = D_x \int f(x|\theta)g(\theta)d\theta$ then the Bayes estimator of θ is:

$$d^*(x) = \frac{1}{b(x)f(x)} D_x f(x) - \frac{a(x)}{b(x)}$$

where $f(x) = \int f(x|\theta)g(\theta)d\theta$.

From the work of Parzen (17) one may obtain a consistent sequence of estimators for $d^*(x)$ by using consistent estimators of $f(x)$ and its derivatives. Some of the techniques suggested by Parzen are discussed below. The utility of the last result is that it provides a method for obtaining EB point estimate in a manner which circumvents explicit estimation of $g(\theta)$.

It should also be remarked that apart from comparing EB procedures, the determination of $E_n W(d_n)$ is useful for comparing non-Bayesian procedures with those derived by EB techniques.

CHAPTER VI

EMPIRICAL BAYESIAN POINT ESTIMATION

1. Estimation When λ is Known

The standard parametrization of the inverse Gaussian density is given in (1.1). In the formulation of EB problems the density $g(\theta)$ of Chapter V is frequently called a "mixing distribution" instead of the previously used "prior distribution." At this point the author desires to avoid estimation of the mixing distribution. To this point the author has consistently used the (μ,λ) parametrization of the inverse Gaussian distribution as many classical results are formulated in these terms. It is now necessary to reparametrize the density to obtain estimates and simultaneously avoid the mixing distribution. To this end the author will apply the transformation $v = \mu^{-2}$. The parameter v so defined is, in terms of the Brownian motion described in Chapter I, equal to $(\text{velocity/distance})^2$ so that the parameter v is connected with the kinetic energy of the particles under Brownian motion making the transformation $\mu = v^{-\frac{1}{2}}$ gives:

$$f(x;v,\lambda) = \frac{1}{\sqrt{2\pi}}\,\lambda^{1/2}x^{-3/2}\exp\left[\frac{-\lambda v}{2x}\,(x - v^{-1/2})^2\right]$$

where $0 < v < \infty$.

If a random sample of size n is obtained then the distribution $\bar{x}$ is:

$$h(\bar{x};v,\lambda) = (n\lambda/2\pi)^{1/2}\bar{x}^{-3/2}\exp\left[-n\lambda(\bar{x} - v^{-1/2})^2/2\bar{x}\right]$$

for $o < x < \infty$ and zero elsewhere.

Following Rutherford and Krutchkoff technique presented in theorem 5.4 one obtains:

$$\log h(\bar{x};v,\lambda) = \frac{1}{2}\log(n\lambda/2\pi) - \frac{3}{2}\log\bar{x} - n\lambda v\bar{x}/2$$

$$+ n\lambda v^{\frac{1}{2}} - n\lambda/2\bar{x}$$

$$D_{\bar{x}}\log h(\bar{x};v,\lambda) = [h(\bar{x};v,\lambda)]^{-1}D_{\bar{x}}\,h(\bar{x};v,\lambda)$$

$$= -3/2\bar{x} - n\lambda v/2 + n\lambda/2\bar{x}^2$$

$$= a(\bar{x}) + b(\bar{x})v$$

where: $a(\bar{x}) = -3/2\bar{x} + n\lambda/2\bar{x}^2$

and $b(\bar{x}) = -n\lambda/2$

From theorem 5.4 the Bayes estimator with respect to any mixing distribution $G(v)$ and quadratic loss is:

$$d_g(\bar{x}) = D_{\bar{x}}\,h^*(\bar{x})/(b(\bar{x})h^*(\bar{x})) - a(\bar{x})/b(\bar{x})$$

or in this case:

$$d_g(\bar{x}) = 2D_{\bar{x}}\,h^*(\bar{x})/(n\lambda h^*(\bar{x})) + (\bar{x}^{-2} - 3/n\lambda\bar{x}) \qquad (6.1)$$

where $h^*(\bar{x}) = \int h(\bar{x};v,\lambda)G(dv)$ is unknown and must be estimated.

It is the author's purpose to obtain the classical competitor of $d_g(\bar{x})$ since it was not possible by this procedure to obtain the EB competitor of $\bar{x}$ under the parametrization of (1.1). To this end note that since $\bar{x}$ is sufficient for μ it is sufficient for v. Also the family of inverse Gaussian distributions is complete so that $\bar{x}$ is a complete sufficient statistic for v. From Johnson and Kotz (12) one obtains:

$$E(1/\bar{x}) = 1/\mu + 1/n\lambda = v^{\frac{1}{2}} + 1/n\lambda \tag{6.2}$$

and

$$E(1/\bar{x}^2) = 1/\mu^2 + 3/n\lambda\mu + 3/n^2\lambda^2 \tag{6.3}$$

$$= v + 3v^{\frac{1}{2}}/n\lambda + 3/n^2\lambda^2$$

From (6.2) and (6.3) one may then conclude that

$$\hat{v}_c = \frac{1}{\bar{x}}\left(\frac{1}{\bar{x}} - \frac{3}{n\lambda}\right) \tag{6.4}$$

is unbiased estimator of v and since it is based on the complete sufficient statistic it is therefore possible to conclude from an extension of the Rao-Blackwell theorem as found in Hogg and Craig (9) that $\hat{v}_c$ is the MVUE of v.

If one compares (6.1) with (6.4) it is apparent that the Bayes estimate is biased by $\frac{2}{n\lambda h^*(\bar{x})} D_{\bar{x}} h^*(\bar{x})$.

A difficulty encountered in using $d_g(\bar{x})$ is that its estimation requires the estimation of $h^*(\bar{x})$ and its derivative $D_{\bar{x}}h^*(\bar{x})$. Rutherford and Krutchkoff (22) refer to Parzen (17) who develops consistent estimates $h_n^*(\bar{x})$ of $h^*(\bar{x})$ of the form

$$h_n(\bar{x}) = \int_{-\infty}^{\infty} \frac{1}{p(n)} K\left(\frac{\bar{x} - y}{p(n)}\right) dH_n^*(y)$$

where $H_n^*(y)$ is the empirical cdf of $\bar{x}$. The function K is thought of as a weighting function and p(n) is a constant dependent upon the sample size.

Parzen shows that if the following hold then $h_n^*(\bar{x})$ is asymptotically unbiased at the continuity points of $h^*(\bar{x})$. The aforementioned conditions are:

(i) $\lim_{n\to\infty} p(n) = 0$

(ii) $\sup_{y\in R} |K(y)| < \infty$

(iii) $\int_{-\infty}^{\infty} |K(y)|\, dy < \infty$

(iv) $\lim_{y\to\infty} |yK(y)| = 0$

(v) $\int_{-\infty}^{\infty} K(y)dy = 1$

If one replaces (i) by (i'): $\lim_{n\to\infty} np(n) = \infty$ then one obtains a mean square consistent estimator of $h^*(n)$. Parzen also additionally limits $K(y)$ to be an even function of y and establishes that $\lim_{n\to\infty} np(n)\ \mathrm{Var}[h_n^*(n)] = h^*(\bar{x})\int_{-\infty}^{\infty} K^2(y)dy$.

Some of the $K(y)$ suggested by Parzen are given below:

$K(y)$	$\int_{-\infty}^{\infty}[K(y)]^2 dy$
$(2\pi)^{-\frac{1}{2}} \exp(-y^2/2)$	$1/2\sqrt{\pi}$
$\frac{1}{2} \exp - \lvert y \rvert$	$1/2$
$[\pi(1 + y^2)]^{-1}$	$1/\pi$
$(1/2\pi)(\sin(y/2)/(y/2))$	$1/3\pi$

Given Parzen's consistent estimate of a density it is possible to develop consistent estimates of the derivative of the density. There are likely more powerful assertions than the following but the author believes this to be sufficient for most purposes.

<u>Proposition 6.1</u>: Suppose an unknown density $f(x)$ is estimated by a

sequence $\{f_n\}$ defined by:

$$f_n(x) = \int_{-\infty}^{\infty} \frac{1}{p(n)} K\left(\frac{x-y}{p(n)}\right) dF_n(y)$$

where $F_n(y)$ is the empirical c.d.f. of x and where:

(i) $\lim_{n\to\infty} n[p(n)]^2 = \infty$,

(ii) $\sup_y |K(y)| < \infty$,

(iii) $\int_{-\infty}^{\infty} |K(y)|\, dy < \infty$,

(iv) $\lim_{y\to\infty} |yK(y)| = 0$,

(v) $\int_{-\infty}^{\infty} K(y)dy = 1$,

(vi) $K(y) = K(-y)$,

(vii) $f(x)$ is uniformly continuous and differentiable,

and (viii) $K(y)$ is differentiable.

Then $f_n'(x) \xrightarrow{p} f'(x)$.

Proof:

Under conditions (i) - (vii) Parzen (17) establishes that given $\epsilon > 0$ there exist an N such that for each $n > N$ then $\Pr\left\{\sup_{-\infty < x < \infty} |f_n(x) - f(x)| > \epsilon\right\} < \epsilon$. Let $h > 0$ and consider $|f_n(x+h) - f(x+h)|$. Let

$$|f_n(x^*) - f(x^*)| = \sup_{x \le t \le x+h} |f_n(t) - f(t)| .$$

Then from Parzen's results one concludes that since $|f_n(x+h)| + |f_n(x) - f(x)| < 2\,|f_n(x^*) - f(x^*)|$ then:

$$\Pr\{|f_n(x+h) - f(x+h)| + |f_n(x) - f(x)| > \epsilon\} <$$

$$\Pr\{2\,|f_n(x^*) - f(x^*)| > \epsilon\} < \Pr\{\sup|f_n(x) - f(x)| > \epsilon/2 < \epsilon/2 < \epsilon.$$

This gives:

$$\Pr\{|f_n(x+h) - f(x+h)| + |f(x) - f_n(x)| > \epsilon\} < \epsilon \qquad (6.5)$$

From the triangle inequality one obtains:

$$|f_n(x+h) - f(x+h)| + |f(x) - f_n(x)| \geq |f_n(x+h) - f_n(x) - (f(x+h) - f(x))|.$$

Using the above in (6.5) gives:

$$\Pr\{|f_n(x+h) - f_n(x) - (f(x+h) - f(x))| > \epsilon\} < \epsilon.$$

Now put $\epsilon = h\delta$ to obtain:

$$\Pr\left\{\left|\frac{f_n(x+h) - f_n(x)}{h} - \frac{f(x+h) - f(x)}{h}\right| > \delta\right\} < h\delta < \delta$$

for small h. Therefore since $f_n(x)$ and $f(x)$ are both differentiable it follows that:

$$\Pr\left\{\lim_{h\to o}\left|\frac{f_n(x+h) - f_n(x)}{h} - \frac{f(x+h) - f(x)}{h}\right| > \delta\right\} < \delta$$

or $\Pr\{|f_n'(x) - f'(x)| > \delta\} < \delta$. Hence $f_n'(x)$ is a consistent estimator of $f'(x)$.

2. Parametric Estimation of μ

In this case the author makes the accumption that the mixing distribution is known to be a member of a particular family. It is the author's opinion that this is an unusual situation but possible. For example, the mixing distribution of μ might actually be believed to be a modified conjugate prior or perhaps μ is thought to be normally distributed. This is dependent upon the investigator's opinion or on real prior knowledge.

In this context assume μ has a density $g(\mu;\vec{\alpha})$. If m samples of size n are obtained then the appropriate likelihood is

$$L(\vec{\alpha};\vec{x}) = \prod_{i=1}^{m} \int \left(\prod_{j=1}^{n} f(x_{ij};\mu,\lambda)g(\mu;\vec{\alpha})d\mu \right) \tag{6.6}$$

where $\vec{x}$ represents all present and prior observations and λ is assumed known.

One of the most acceptable solutions to the problem would be to maximize (6.6) with respect to $\vec{\alpha}$ by numerical procedures. In some cases there will be less sophisticated but computationally simpler procedures. If we assume the ML estimate of $\vec{\alpha}$ obtained from (6.6) is $\vec{\alpha}^*$ then the EB estimate of μ at the mth stage is

$$\widehat{dg}(\vec{x}) = \frac{\int \mu f(\vec{x};\lambda,\mu)\ \widehat{g}(\mu)d\mu}{\int f(\vec{x};\lambda,\mu)\ \widehat{g}(\mu)d\mu}$$

The ML estimate $\vec{\alpha}^*$ is generally a consistent estimate of $\vec{\alpha}$. If one can conclude that $\widehat{dg}(x) \xrightarrow{p} dg(x)$ then $\widehat{dg}(x)$ is weakly a.o. provided g possesses a bounded moment of order exceeding two.

Other methods may be employed to estimate $\vec{\alpha}$ as will be seen later. However, it is more plausible to believe that even the functional form of g is unknown. One then must choose some procedure for estimating g.

Since any continuous function can be approximated within any desired degree of accuracy by a step function [Royden (21)], one method of estimating the mixing distribution in point estimation problems is to take

$$G_k(\theta) = \sum_{i=1}^{j} p_i \quad \theta_j \leq \theta < \theta_{j+1} \text{ where } j = 1,2\ldots k-1 \tag{6.7}$$

Now if $G(\theta)$ is unknown then both the p_i and θ_j must be determined in some manner. If one assumes m samples of size n are available then the joint likelihood of (p,θ) based on all samples is:

$$L(\vec{x};p_1,p_2,\dots p_{kj}\theta_1,\dots\theta_k) = \prod_{i=1}^{m} [\sum_{j=1}^{k} p_j f(\vec{x}_i \mid \theta_j)] \tag{6.8}$$

One may determine the step function approximation either by:

(i) Use a numerical optimization procedure to maximize $L(x;\vec{p},\vec{\theta})$ with respect to $\vec{p}$ and $\vec{\theta}$.

or (ii) Obtain from each $\vec{x}_i$ the MLE $\hat{\theta}_i$ and then maximize $L^*(\vec{x};\vec{p},\vec{\theta})$ with respect to $\vec{p}_{m\times 1}$

or (iii) Set $p_i = \frac{1}{m}$ and use the MLE $\hat{\theta}_i$

Any of these procedures results in a step function which approximates to $G(\theta)$ and results in an EB estimator of θ of the form:

$$d_m(\vec{x}_m) = \frac{\sum_{j=1}^{m} \tilde{p}_j \tilde{\theta}_j f(\vec{x}_m;\tilde{\theta}_j)}{\sum_{j=1}^{m} \tilde{p}_j f(\vec{x}_m;\tilde{\theta}_j)}.$$

There are other procedures for obtaining "smooth estimates" of the mixing distribution and not all of these procedures will be considered in this chapter. An alternative to (iii) above is to estimate G by a finite mixture of densities. In the case of estimation of μ with m samples of size n one could do the following. The ML estimate of μ is $\bar{x}$ and $\bar{x}$ follows an inverse Gaussian probability law with parameters μ and $n\lambda$. If the i^{th} sample results in $\hat{\mu}_i = \bar{x}_i$ then a smooth estimate of dG is obtained by taking

$$g_m(\mu) = \sum_{i=1}^{m} \frac{1}{m} \sqrt{\frac{n\lambda}{2\pi\bar{x}_i}} \exp\left[\frac{-n\lambda}{2\bar{x}_i\mu^2}(\bar{x}_i - \mu)^2\right]$$

The practical utility of any EB procedure is difficult to determine. This investigation takes on the aspect of a case study and inferences to other situations could be misleading.

What is involved in such a study? First, one must guess as to what type of mixing distribution is likely to be encountered and then this distribution must be specified exactly for simulation purposes. The conditional distribution of the observable r.var. is known. One may then have one or more EB estimation and classical procedures that one may wish to compare.

Since the mixing distribution is known the Bayes risk for the classical estimator can be computed and is a constant W_c. The reader will recall that the Bayes risk for any EB procedure is a random variable. The question is whether or not to use an EB estimator in preference to a classical estimator. The answer is a function of the individual sample size, the loss function, the number of samples, and the mixing distribution. If all of these are fixed then the investigator can approximate the distribution of the Bayes risk W_d for the EB estimator d. One then attempts to determine the answer to the following questions: Is $EW_d < W_c$? What is $\Pr\{W_d < W_c\}$? How many past samples m of size n are required so that $\Pr\{W_d < W_c\} > 1 - \epsilon$ for a specified ϵ? These questions are of paramount importance to determine the utility of d. Unfortunately they may also be quite expensive to calculate. This is due to the fact that the EB estimator is expensive to calculate and one must determine the value of the EB estimate repeatedly to determine a single realization of the Bayes risk (definition 5.2).

As a compromise between doing a case study on point estimates based on a step function and the economically unfeasible case study involving smooth EB estimates, the author will present an evolutionary example illustrated in Table II. The example is not based on real data but does depart from using a "natural conjugate prior." The between sample distribution of μ is N ($\mu = 1$, $\sigma = .2$) and the hypothetical experimenter has specified a normality constraint on the mixing distribution. For each value of μ a sample of size ten is taken which is used for estimating the parameters of the normal mixing distribution. At preselected points the total number of samples taken at that point is noted and the other quantities identified are printed. This example cannot be presented as representative, since no replication at any stage is presented and because the stopping points are weighted toward those points with a large number of previous samples. The parametric Bayes estimate at stage i is:

$$\int_0^{\infty} \mu T(\mu) d\mu \Big/ \int_0^{\infty} T(\mu) d\mu$$

where n = 10 and $\lambda = 1$, and $T(\mu)$ is given by:

$$T(\mu) = \sqrt{5/(\pi x^{-3})}\ \exp\left[\frac{-5(x_i - \mu)^2}{x_i \mu^2}\right] \sqrt{1/(2\pi\hat{\sigma}_i^2)}\ \exp\left[-\frac{(\mu - \hat{\mu}_i)^2}{2\hat{\sigma}_i^2}\right]$$

3. Estimation When λ is Known

As before the author will initially derive the Bayes estimate of λ which bypasses estimation of the mixing distribution of λ. When μ is known $T(\vec{x}) = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{x_i}$ is sufficient for λ and from Chhikara (4)

TABLE II

AN EXAMPLE OF CLASSICAL AND PARAMETRIC EB ESTIMATION

NUMBER OF SAMPLES	ESTIMATED PARAMETERS MU-HAT	ESTIMATED PARAMETERS SIGMA-HAT	ACTUAL VALUE OF MU	CLASSICAL ESTIMATE	PARAMETRIC BAYES ESTIMATE
5	0.9522E 00	0.1455E 00	0.9149E 00	0.1024E 01	0.9765E 00
7	0.1108E 01	0.4737E 00	0.9070E 00	0.8508E 00	0.1059E 01
9	0.1106E 01	0.4609E 00	0.1109E 01	0.6781E 00	0.9010E 00
12	0.1293E 01	0.6802E 00	0.1295E 01	0.2697E 01	0.1958E 01
15	0.1221E 01	0.6433E 00	0.1114E 01	0.1423E 01	0.1498E 01
20	0.1174E 01	0.5816E 00	0.6645E 00	0.8247E 00	0.1057E 00
27	0.1103E 01	0.5418E 00	0.7242E 00	0.3714E 00	0.4588E 00
35	0.1174E 01	0.5143E 00	0.1049E 01	0.1488E 01	0.1481E 01
47	0.1189E 01	0.4957E 00	0.9975E 00	0.1230E 01	0.1340E 01
62	0.1234E 01	0.5565E 00	0.9867E 00	0.1677E 01	0.1570E 01
81	0.1241E 01	0.5277E 00	0.1325E 00	0.9273E 00	0.1122E 01
107	0.1256E 01	0.5403E 00	0.8143E 00	0.8022E 00	0.1107E 01
142	0.1240E 01	0.5306E 00	0.1040E 01	0.1849E 01	0.1454E 01
188	0.1212E 01	0.5575E 00	0.1090E 01	0.1604E 01	0.1591E 01
248	0.1184E 01	0.5337E 00	0.9631E 00	0.1194E 01	0.1289E 01
328	0.1176E 01	0.5352E 00	0.1019E 01	0.6320E 00	0.8654E 00
433	0.1171E 01	0.5154E 00	0.7877E 00	0.6354E 00	0.8516E 00
573	0.1181E 01	0.5237E 00	0.9843E 00	0.7728E 00	0.1017E 01
757	0.1165E 01	0.5146E 00	0.1159E 01	0.5343E 00	0.7198E 00

$T(\tilde{x})$ is distributed as a $\mu^2\chi^2/\lambda$ variate. From this one can conclude that

$$h(t;\lambda) = \begin{cases} \left(\dfrac{\lambda}{2\mu^2}\right)^{n/2} \dfrac{t^{n/2-1} e^{-\frac{\lambda}{2\mu^2}t}}{\Gamma\left(\frac{n}{2}\right)} & t > o \\ o & \text{elsewhere} \end{cases}$$

Now employing the aforementioned procedure of Rutherford and Krutchkoff gives:

$$\log h = \frac{n}{2}\log(\lambda/2\mu^2) - \log\Gamma(n/2) + (n/2-1)t - \lambda t/(2\mu^2)$$

$$D_t \log h = (n/2 - 1) - \lambda/(2\mu^2) \qquad (6.9)$$

Equation (6.9) satisfies the conditions of Rutherford and Krutchkoff with:

$$a(t) = \frac{n}{2} - 1 \quad b(t) = -1/(2\mu^2).$$

Therefore the EB estimator of λ when μ is known is:

$$d_m(t_m) = 2\mu^2 \left[n/2 - 1 - D_t h^*(t)/h^*(t)\Big|_{t=t_m}\right]$$

where $h^*(t)$ is the estimated density of t.

In the previous section some general estimation techniques were discussed. The estimates so obtained cannot in general be presented unless the mixing distribution is known or data are available. For example, the mean of the truncated conjugate prior is not known explicitly so the EB estimator cannot be given in closed form. This is a partial motivation for the preceding numerical example. Then the case when μ is known the situation is artificially improved in that the natural conjugate prior distribution of λ is the well-known gamma probability law. From Maritz (16) one notes that even in EB problems there appears to be a magnetic attraction to using the natural conjugate prior as the mixing distribution. This is again a case of parametric EB estimation, in that the mixing distribution is taken as

$$g(\lambda) = \begin{cases} \dfrac{a^p \lambda^{p-1} \exp(-a\lambda)}{\Gamma(p)} & \lambda \geq 0 \\ 0 & \lambda < 0 \end{cases}$$

Chhikara previously provided the ML estimators of μ and λ which were modified to produce $\hat{\mu}$ and $\hat{\sigma}^2$ of the previous numerical example. In this case the author will demonstrate that direct ML estimation of the parameters a and p is not possible.

The likelihood for a simple sample is:

$$L(a,p;x) = \int_0^{\infty} f(x;\mu,\lambda)g(\mu)d\lambda$$

where $$f(\vec{x};\mu,\lambda) = (\lambda/2\pi)^{n/2}\left(\prod_{i=1}^{n} x_i^{-3/2}\right) \exp\left[-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}\right] \quad (6.10)$$

Upon integrating (6.10) one obtains:

$$L(a,p;\vec{x}) = (1/2\pi)^{n/2}\left(\prod_{i=1}^{n} x_i^{-3/2}\right)\frac{a^p\,\Gamma(\frac{n}{2}+p)}{(\Sigma + a)^{n/2+p}\,\Gamma(p)}$$

where $\Sigma = (1/2\mu^2)\displaystyle\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}$

so that

$$\log L = c + p\ln a - (\tfrac{n}{2}+p)\ln(\Sigma + a) + \ln\Gamma(\tfrac{n}{2}+p) - \ln\Gamma(p)$$

Taking $D_a \log L$, $D_p \log L$, and equating these to zero gives:

$$(D_p \log L) \qquad \ln\left|\frac{a}{a + \Sigma}\right| - \psi(p) + \psi(\tfrac{n}{2} + p) = 0 \qquad (6.11)$$

where $\psi(x) = D_x \ln \Gamma(x)$ is the digamma function.

$$(D_a \log L) \qquad \frac{p}{a} - (\tfrac{n}{2} + p)/(\Sigma + a) = 0$$

$$\text{or } a = \frac{2p\Sigma}{n} \qquad (6.12)$$

Since $\frac{a}{a + \Sigma} = \frac{2p}{2p + n}$ substitution of a from (6.12) in (6.11) gives:

$$\ln[2p/(2p + n)] - \psi(p) + \psi(n/2 + p) = 0 \qquad (6.13)$$

If one notes that $\ln\left(\frac{2p}{2p + n}\right) = -\ln\left(\frac{2p + n}{2}\right) + \ln p$ then equation (6.13) can be written as

$$\psi(p) - \ln(p) = \psi(p + n/2) - \ln(p + n/2) \qquad (6.14)$$

If one defines $z(p) = \psi(p) - \ln(p)$ then (6.14) indicates z(p) is cyclic with period equal to one-half.

Consequently, since the likelihood equation is continuous the derived equation either has an infinite number of solutions or no solutions.

From Abramowitz and Segun (1) one finds that

$$\psi(p) = -2e\int_0^{\infty} \frac{t \cdot \exp(-2\pi t)}{t^2 + p^2}\,dt$$

$$-\frac{1}{2p} + \ln p$$

hence $z(p) = -2e\int_0^{\infty} \frac{t}{t^2 + p^2} \exp(-2\pi t)dt - \frac{1}{2p} < 0$

Since the derivative of the likelihood is bounded and nonzero for p in $(0,\infty)$ one must conclude that L has no interior extrema. In view of this it is not possible that the joint likelihood of m sample will possess interior extrema.

A much simpler technique may be developed by using ML estimation in conjunction with the method of moments. Each sample yields a ML estimate of λ_i given by

$$\tilde{\lambda}_i = n\mu^2 \Big/ \left[\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{x_i} \right]$$

By the method of moments one obtains the equations:

$$E(\lambda) = \frac{p}{a} \doteq \bar{\tilde{\lambda}}$$

$$\mathrm{Var}(\lambda) = \frac{p}{a^2} \doteq s^2_{\tilde{\lambda}}.$$

From the above, estimates of a and p can be derived. These estimates are:

$$\tilde{a} = \bar{\tilde{\lambda}}/s^2_{\tilde{\lambda}} \qquad \tilde{p} = (\bar{\tilde{\lambda}})^2/s^2_{\tilde{\lambda}}.$$

The EB estimator corresponding to this estimation scheme is then:

$$d(\vec{x}) = (\tilde{P} + n/2) \Big/ \left[\tilde{a} + \frac{1}{2\mu^2} \sum_{j=1}^{n} \frac{(x_j - \mu)^2}{x_j} \right]$$

4. Smooth EB Estimation When μ is Known

It is now assumed that the mixing distribution of λ is unknown. The author will present no material regarding the step function approximation of the mixing distribution as these procedures are inherently numerical and the author intends to undertake no additional numerical analyses.

If one reparametrizes the inverse Gaussian by $\theta = \frac{1}{\lambda}$ then one can obtain a smooth approximation to the mixing distribution based on MVUE of θ. From Chhikara (4) one can conclude that

$$\text{if } \hat{\theta} = (1/n\mu^2) \sum_{j=1}^{n} (x_i - \mu)^2/x_i$$

then

$$\hat{\theta} \sim \text{Gamma}(a = \frac{n}{2\theta} \text{ , } p = \frac{n}{2})$$

Given m samples and the corresponding estimates $\hat{\theta}_1 \ldots \hat{\theta}_m$ then a smooth estimate of $g(\theta)$ is:

$$g^*(\theta) = \sum_{i=1}^{m} \frac{(1/m)(n/2\hat{\theta}_i)^{n/2} \, \theta^{\frac{n}{2}-1} \exp(-n\theta/2\hat{\theta}_i)}{\Gamma(\frac{n}{2})}$$

with the resultant EB estimator of θ being

$$d(\vec{x}_m) = \int \theta v(\theta) d\theta / \int v(\theta) d\theta \qquad (6.15)$$

where $v(\theta) = f(\vec{x}_m | \theta) g^*(\theta)$.

The above parametrization is not necessary but may be of practical value since θ is proportional to the diffusion constant in the context of Brownian motion mentioned in Chapter I.

If one is interested in smooth estimation of λ then noting that $t(\vec{x}) = \frac{\Sigma(x_i - \mu)^2}{x_i}$ is sufficient for λ and that

$$t \sim \text{Gamma } (a = \frac{\lambda}{2\mu^2} \text{ , } p = n/2)$$

then $\frac{\lambda}{2\mu^2} t \sim Ga(1, \frac{n}{2})$ from which one concludes that $W = \frac{n}{2} \frac{\lambda}{\mu^2} \frac{2\mu^2}{\lambda t} = \frac{n}{t}$ follows an inverted gamma distribution with parameters $v = n$ and $\psi = \lambda/\mu^2$.

Since $E(W) = \frac{\psi v}{v - 2} = \frac{n\lambda}{\mu^2(n - 2)}$ it is apparent that $\hat{\lambda} = \frac{(n - 2)\mu^2}{n} W$ is unbiased for λ and since $\hat{\lambda}$ is a one to one transformation of the complete sufficient statistic t then $\hat{\lambda}$ is MVUE for λ.

The point is, however, that the density of $\hat{\lambda}$ can be determined to be:

$$g(\hat{\lambda}) = \frac{[(n - 2)\lambda/2\hat{\lambda}]^{\frac{n}{2}+1} \exp[-(n - 2)\lambda/2\hat{\lambda}]}{(n - 2)\lambda\,\Gamma(\frac{n}{2})}$$

so that from m samples one first obtains the estimates $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m$ and take

$$g^*(\lambda) = \sum_{i=1}^{m} \frac{(1/m)[(n - 2)\hat{\lambda}_i/2\lambda]^{\frac{n}{2}=1} \exp[-(n - 2)\hat{\lambda}_i/2\lambda]}{(n - 2)\,\hat{\lambda}_i\,\Gamma(\frac{n}{2})}$$

The resulting smooth EB estimation is then of the same form as (6.15), with the obvious substitutions.

5. Simultaneous Estimation of μ and λ

As far as the author is aware multiparameter estimation problems have not received much attention in EB literature. The work put forth involves location and scale parameters which is not appropriate for the inverse Gaussian under the μ, λ parametrization. In this situation the author is not aware of estimation procedures which bypass explicit estimation of the mixing distribution $g(\mu,\lambda)$. Of course $g(\mu,\lambda)$ could be approximated by a bivariate step function. Since no real evaluation of any EB procedure is presently possible the author will present a smooth EB procedure. In determining a smooth approximation to the mixing density the reader will note that the ML estimates of μ and λ^{-1} are:

$$\tilde{\mu} = \bar{x} \text{ and } \tilde{\lambda}^{-1} = \frac{1}{n}\Sigma(x_i^{-1} - \bar{x}^{-1})$$

It has been established as referenced in Johnson and Kotz (12) that $\tilde{\mu}$ and $\tilde{\lambda}^{-1}$ are independent. Therefore, $\tilde{\mu}$ and $g(\tilde{\lambda}^{-1})$ are independent. One also notes that

$$\tilde{\mu} \sim \text{Inverse Gaussian } (\mu, n\lambda)$$

and that $\tilde{\mu}$ is MVUE for μ. For this case the author has deduced from the fact that $\lambda\Sigma(x_i^{-1} - \bar{x}^{-1}) \sim \chi^2(n - 1)$ that

$$\hat{\lambda} = (n - 3)/[\Sigma(x_i^{-1} - \bar{x}^{-1})]$$

follows an inverted gamma distribution with parameters $v = n - 1$ and $\psi = (n - 3)\lambda/(n - 1)$. This assumes that $n > 3$. In this case $\hat{\lambda}$ is MVUE for λ and being a function of $\tilde{\lambda}^{-1}$ alone is independent of $\tilde{\mu}$. This independence permits the formation of the joint density of $(\hat{\mu},\hat{\lambda})$ which, as before, is the basis of the author's smooth estimation procedure. Following the previous developments, the author's estimate of the density $g(\mu,\lambda)$ based on m samples is:

$$g^*(\mu,\lambda) = \sum_{i=1}^{m} (1/m)h_i(\mu,\lambda)$$

where $h_i(\mu,\lambda) = \sqrt{n\lambda_i/(2\pi\mu^3)}\ \exp[-n\hat{\lambda}_i(\mu - \hat{\mu}_i)^2/(2\hat{\mu}_i^2\mu)]$

$$\times \frac{[(n-3)\hat{\lambda}_i/2\lambda]^{\frac{n+1}{2}} \exp[-(n-3)\hat{\lambda}_i/(2\lambda)]}{[(n-3)\hat{\lambda}_i/2]\,\Gamma(\frac{n-1}{2})}$$

for $\mu > o$, $\lambda > o$, zero elsewhere.

From this the EB estimate of μ based on the m^{th} sample is:

$$d(\vec{x}_m) = \iint \mu w(\vec{x}_m,\mu,\lambda)d\mu d\lambda / \iint w(\vec{x}_m,\mu,\lambda)d\mu d\lambda$$

where

$$w(\vec{x}_m,\mu,\lambda) = f(\vec{x}_m;\mu,\lambda)g^*(\mu,\lambda)$$

The EB estimate for λ is similarly defined.

CHAPTER VII

TESTS OF HYPOTHESES

1. Introduction

First, a general introduction to some of the Bayesian procedures used in hypothesis testing will be discussed. These techniques are those that are used in Empirical Bayes procedures.

One will be interested in testing H_o: $\theta \epsilon P_o$ vs. H_1: $\theta \epsilon P_1$. It will be assumed that P_o and P_1 may be simple or composite and that $P_o \cup P_1$ covers the parameter space. The Bayesian approach to testing H_o vs. H_1 is usually of one of two forms. One procedure entails computing the $(1 - \alpha)$ HPD region Q for θ. Recall that this is the Bayesian analogue of a classical confidence interval. The procedure is to then accept H_o and reject H_1 if $\Pr\{P_o \cap Q\} > \Pr\{P_1 \cap Q\}$ or, in the case of a simple vs. composite accept $H_o = \theta = \theta_o$ if $\theta_o \epsilon Q$. If both H_o and H_1 are simple one accepts H_o if $d(P_o \cap Q) > d(P_1 \cap Q)$ where d is the posterior density of θ and all aforementioned probabilities are with respect to this posterior distribution.

Another Bayesian approach that sometimes leads to the same procedure involves selecting a rejection region from the sample space of the observed r.var. with the criterion that the rejection region should result in the minimization of the Bayes risk (for a specified loss function and mixing distribution) with respect to all such rejection regions. This is a more versatile approach. In classical testing procedures one may encounter situations where no UMP test exists. Likewise one may encounter situations where no rejection region exists which minimizes the Bayes risk.

In EB procedures there are two forms of loss functions that are most commonly encountered in tests of hypotheses. These are the 0 - 1 loss and linear loss. The latter loss function will be treated below. By 0 - 1 loss it is meant that the experimenter incurs 0 loss if a correct decision is made and a unit loss if an incorrect decision is rendered. Let $\bar{X}$ denote the sample space of the observed deviate X and let R and A denote respectively the rejection and acceptance regions for the test H_o: $\theta \epsilon P_o$ vs. H_1: $\theta \epsilon P_1$. Let $f(x \mid \theta)$ and $g(\theta)$ represent the conditional of X and marginal of θ. The Bayes risk is then:

$$W(R) = \Pr \left\{ \text{An incorrect decision is made} \right\}$$

$$= \int_R \int_{P_o} f(x \mid \theta) dG(\theta) dx + \int_A \int_{P_1} f(x \mid \theta) dG(\theta) dx$$

Letting $f_i(x) = \int_{P_i} f(x \mid \theta) dG(\theta)$, $i = 0,1$ then

$$W(R) = \int_R f_o(x) dx + \int_A f_1(x) dx > o. \qquad (7.1)$$

It will be assumed that $X = R \cup A$ so that no sequential or randomized decision rule is encountered. Under this constraint:

$$W(R) = \int_R f_o(x) dx + \int_{X-R} f_1(x) dx$$

$$= K + \int_R (f_o - f_1)(x) dx \qquad (7.2)$$

where $K = \int_X f_1(x) dx > o.$

In view of (7.1) W(R) will be minimized by allowing the integration in (7.2) to be over regions where $f_o - f_1 < o$.

Since $f_o(x) > 0$ this is equivalent to defining

$$R = \left\{ x: \frac{f_o(x)}{f_1(x)} < 1 \right\}$$

The 0 - 1 loss function then leads to the test criteria of "highest posterior odds" since if $q(x)$ is the marginal of X then $f_o/f_1 = (f_o/q)/(f_1/q)$ is a ratio of posterior probabilities.

Frazer (7) has shown that the above analysis can be extended to multiple hypotheses and that the procedure which minimizes the Bayes risk is that which accepts the hypothesis with highest posterior probability. The above tests can also be considered from the classical viewpoint. That is, one may inquire as to their size, power, and expected power. These comparisons are interesting but are usually of little value since the two approaches have fundamentally different goals.

It may be well at this point to introduce some notation and simultaneously review the notion of asymptotic optimality. For a given EB procedure concerned with discriminating between H_o: $\theta \epsilon A_o$ and H_1: $\theta \epsilon A_1$ there will exist a sequence of rejection regions $R_m, m = 1,2\dots$. The Bayes risk is

$$W(R_m) = \int_\theta \int_X L(R_m, \theta) dF(x \mid \theta) dG(\theta) \tag{7.3}$$

As before, the EB procedure is strongly asymptotically optimal iff $E_m W(R_m) \longrightarrow W(R_g)$ pointwise, where R_g is the Bayes region which minimizes equation (7.3). The procedure is weakly a.o. iff $W(R_m) \xrightarrow{P} W(R_g)$.

2. EB Test Procedures for Simple Hypotheses for the Inverse Gaussian Distribution

It will be assumed throughout that there are available a sequence of samples $\{\vec{x}_i\}_{i=1}^{m}$ each of size n.

Consider first the problem of discriminating between the simple hypotheses:

$$H_o: \ \mu = \mu_o$$
$$H_1: \ \mu = \mu_1$$

where $\mu_o < \mu_1$ and λ may or may not be known. A usual EB procedure is used and it is assumed that the 0 - 1 loss function is acceptable. The distribution $G(\mu)$ is assumed to be of the form:

$$dG(\mu_o) = p_o$$
$$dG(\mu_1) = p_1$$

The mean of the r.var. x is then given by

$$EX = \int_o^{\infty} xf(x \mid \mu)dG(\mu)dx = \int_o^{\infty} x[p_o f(x \mid \mu_o) + p_1 f(x \mid \mu_1)]dx$$

so that

$$EX = p_o\mu_o + p_1\mu_1$$

At the m^{th} stage of the decision process one may estimate EX by $\Sigma\bar{X}_i/m$. Using the method of moments one obtains:

$$p_o\mu_o + p_1\mu_1 = \Sigma\bar{X}_i/m \qquad (7.4)$$

which gives $\tilde{p}_{om} = (m\mu_1 - \Sigma\bar{X}_i)/(\mu_1 - \mu_o)$.

One may then take as an estimate of p_o

$$p^*_{om} = \max\,[o, \tilde{p}_{om}]$$
$$p^*_{1m} = 1 - p^*_{om}$$

and in this manner avoid negative estimates of p_o.

Consider now the behavior of $\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i$. If one bears in mind that $\bar{X}_i$ is the sample mean of realizations of $x_j \mid M_i$, $j = 1,2,\ldots n$ then one may write

$$\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i - \frac{1}{m}\sum_{i=1}^{m} M_i = \frac{1}{m}\sum_{i=1}^{m} [(\bar{X}_i \mid M_i) - M_i] \tag{7.5}$$

and employing Chebyshev's first inequality as referenced in Fisz (6) one obtains from (7.4) that

$$\Pr\left\{ \frac{1}{m^2}\left(\sum_{i=1}^{m} ((\bar{X}_i \mid M_i) - M_i)\right)^2 \geq \epsilon \right\}$$

is less than or equal to $\frac{1}{\epsilon m^2} E\left[\sum_{i=1}^{m} (X_i \mid M_i) - M_i)\right]^2$. Since the M_i are independent r.var. and $X_i \mid M_i$ and $X_j \mid M_j$ are independent for $i \neq j$ then the above expectation is $\sum_{i=1}^{m} E[(\bar{X}_i \mid M_i) - M_i]^2$ which is just the conditional variance of $X_i \mid M_i$ or $\frac{M_i^3}{n\lambda}$. Hence

$$\Pr\left\{ \left| \frac{1}{m}\Sigma[(\bar{X}_i \mid M_i) - M_i] \right| \geq \epsilon \right\}$$

$$= \Pr\left\{ \frac{1}{m^2}[\Sigma(\bar{X}_i \mid M_i) - M_i]^2 \geq \epsilon^2 \right\} \leq \frac{1}{n\lambda\epsilon^2 m^2}\sum_{i=1}^{m} M_i^3 \tag{7.6}$$

Now note that the random variables M_i^3 are independent identically distributed. If $E(M_i^3) < \infty$ then by Kolmogorov's Law of large numbers the

$\frac{1}{m}\sum_{i=1}^{m} M_i^3 \longrightarrow EM_i^3$ a.s. so that if one takes the limit of (7.6) as $m \rightarrow \infty$ one obtains that the $\frac{1}{m}\sum_{i=1}^{m} (\bar{X}_i \mid M_i) \xrightarrow{p} \frac{1}{m}\Sigma M_i$ a.s.$(G(\mu))$. By assuming M possesses a bounded third moment one insures that $\frac{1}{m}\Sigma M_i \xrightarrow{p(\mu)} E(M)$.

Hence, $\frac{1}{m}\sum_{i=1}^{m} \bar{X}_i$ almost surely converges in probability to the mean of the random variable M. The only constraint the author imposed on the mixing r.var. M is that it have a bounded third moment exist. There are perhaps weaker conditions which can be imposed.

Referring again to equation (7.4) in conjunction with Slutsky's theorem, one can conclude that for almost all μ, $\tilde{p}_{om} \xrightarrow{p} p_O$. In this case $G_m(\mu) \longrightarrow G(\mu)$.

Consider $W(R_m) - W(R_g) =$

$$\iint_{L(R_m) \neq L(R_g)} (\pm 1) dF(x \mid \mu) d[Gm(\mu) - G(\mu)]$$

$$+ \int \int_{L(R_m) = L(R_g)} L(R_g, \mu) dF(x \mid \mu) d[Gm(\mu) - G(\mu)] \qquad (7.7)$$

The quantity defined by Equation (7.7) is less than

$$\int \left\{ \int_X \mid dF(x \mid \mu) \right\} d\left\{ Gm(\mu) - G(\mu) \right\} = p_{om} - p_o + p_{1m} - p_1$$

Therefore $o \leq \lim_{m \to \infty} \left| W(R_m) - W(R_g) \right| < \lim_{m \to \infty} \left| p_{om} - p_o + p_{1m} - p_1 \right.$

$$\leq \lim_{m \to \infty} \left| p_{om} - p_o \right| + \lim_{m \to \infty} \left| p_{1m} - p_1 \right| = o \text{ in probability}$$

so that $W(R_m) \xrightarrow{p} W(R_g)$ and R_m is weakly a.o.

One may derive in a similar manner procedures for testing H_o: $\lambda = \lambda_o$ vs. H_i: $\lambda = \lambda_1$. In this case it is assumed μ is known and

$\lambda_o < \lambda_1$.

Using the same formulation as before one obtains

$$\text{Var } x = \mu^3(p_o/\lambda + p_1/\lambda_1)$$

Using the estimate of the variance from all samples and the method of moments, one obtains as an estimate of $p_o = \Pr\{\lambda = \lambda_o\}$:

$$\tilde{p}_{om} = \frac{\lambda_o\lambda_1 \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \mu)^2 - mn\lambda_o}{mn(\lambda_1 - \lambda_o)}$$

The resultant estimator is $\hat{p}_{om} = \max[o, \tilde{p}_{om}]$. If $p_{om} \xrightarrow{P} p_o\ (\mu)$ then as before the sequence R_m of EB decision rules is weakly a.o.

3. One Sided Tests With One Parameter Known

Consider first the hypotheses H_o: $\mu \leq \mu_o$ vs. H_1: $\mu > \mu_o$. It is a consequence of the Bayes test procedure that the hypotheses with the highest posterior probability is accepted when a 0 - 1 loss function is used. For the one sided hypotheses the judgement may be made either by computing the probabilities or by computing the posterior median $\mu_{.5}$ of the parameter. The decision procedure is then to accept H_o iff $\mu_{.5} \leq \mu_o$. As far as the author is aware these procedures require explicit estimation of the mixing distribution of μ. The author will discuss methods that have been successful in other applications.

The author has previously discussed some methods for obtaining smooth estimates of $g(\mu)$ in the context of estimation. Maritz (16) suggests using the following approximation $G^*(\mu)$ of $G(\mu)$ where

$$dG^*(\mu) = \frac{d\mu}{(k-1)(\hat{\mu}_{j+1} - \hat{\mu}_j)} \quad \text{for } \hat{\mu}_j < \mu < \hat{\mu}_{j+1}$$

and $j = 1,2,\ldots,k-1$

and where $\hat{\mu}_1, \hat{\mu}_2, \ldots \hat{\mu}_k$ are obtained by maximizing:

$$\sum_{i=1}^{m} \log \left\{ \sum_{j=1}^{k-1} \frac{1}{(k-1)(\mu_{j+1} - \mu_j)} \int_{\mu_j}^{\mu_{j+1}} \sqrt{\frac{n\lambda}{2\pi \bar{x}_i^3}} \exp\left[\frac{-n\lambda(\bar{x}_i - \mu)^2}{2\mu^2 \bar{x}_i} \right] d\mu \right\}$$

with respect to $\mu_1, \mu_2, \ldots \mu_k$. The EB decision rule is to accept H_o iff the solution $\mu_{.5}$ to $G^*_m(\mu) = .5$ is less than or equal to μ_o.

The motivation for this technique is the minimization of the Kulbeck-Leibler distance between the actual and approximate marginal distribution of the r.var. $\bar{X}$. Maritz indicates that this procedure is useful in that it is weakly a.o. or almost weakly a.o. in the sense that the Bayes risk for the EB procedure converges in probability to the $W(d_g) + \delta$ where d_g is the Bayes rule. The question as to which is the case cannot be resolved without knowing $G(\mu)$ exactly. The utility of this decision process as applied to the inverse Gaussian is unknown and could be determined only through Monte Carlo studies.

The author previously remarked that an alternate loss structure is sometimes of value in EB tests of hypotheses. This loss structure is introduced in the context of testing:

$$H_o: \ \mu \leq \mu_o$$

$$\text{vs. } H_1: \ \mu > \mu_o$$

Partition the sample space of X into A_o and A_1 so that $A_o = \{ \bar{x} \mid H_o \text{ is accepted} \}$ and $A_1 = \sim A_o$. The loss function is defined in the following manner:

$$\text{If } H_o \text{ is true then } L(A_o,\mu) = \begin{cases} o \text{ if } \bar{x}\epsilon A_o \\ \mu_o - \mu \text{ if } \bar{x}\epsilon A_1 \end{cases}$$

$$\text{If } A_1 \text{ is true then } L(A_o,\mu) = \begin{cases} o \text{ if } \bar{x}\epsilon A_1 \\ \mu - \mu_o \text{ if } \bar{x}\epsilon A_o \end{cases}$$

The corresponding Bayes risk is:

$$W(A_o) = \int_{A_o} \int_{\mu_o}^{\infty} (\mu - \mu_o) f(\bar{x} \mid \mu) dG(\mu) d\bar{x}$$

$$+ \int_{A_1} \int_{-\infty}^{\mu_o} (\mu_o - \mu) f(\bar{x} \mid \mu) dG(\mu) d\bar{x} \,. \qquad (7.8)$$

The Bayes decision rule is by definition the rule which minimizes $W(a_o)$ with respect to A_o.

Rewriting equation (7.8) gives:

$$W(A_o) = \int_{A_o} \int_{\mu_o}^{\infty} (\mu - \mu_o) f(\bar{x} \mid \mu) dG(\mu) d\bar{x}$$

$$+ \int_{X-A_o} \int_{-\infty}^{\mu_o} (\mu_o - \mu_o) f(\bar{x} \mid \mu) dG(\mu) d\bar{x}$$

$$= C(\mu_o) + \int_{A_o} \left\{ \int_{-\infty}^{\infty} (\mu - \mu_o) f(\bar{x} \mid \mu) dG(\mu) \right\} d\bar{x}$$

$$\text{where } C(\mu_o) = \int_X \int_{-\infty}^{\mu_o} (\mu_o - \mu) f(\bar{x} \mid \mu) dG(\mu) d\bar{x} > o.$$

Since $W(A_o) > o$ it is apparent that $W(A_o)$ is minimized by choosing A_o so that

$$\int_{A_o} \left\{ \int_{-\infty}^{\infty} (\mu - \mu_o) f(\bar{x} \mid \mu) dG(\mu) \right\} d\bar{x} < o \,.$$

This is accomplished by assigning to A_0 each $\bar{x}$ such that

$$\int_{-\infty}^{\infty}(\mu - \mu_0)f(\bar{x} \mid \mu)dG(\mu) < 0 \tag{7.9}$$

Rewrite (7.9) to obtain:

$$\frac{\int_{-\infty}^{\infty}\mu f(\bar{x} \mid \mu)dG(\mu)}{\int_{-\infty}^{\infty}f(\bar{x} \mid \mu)dG(\mu)} < \mu_0 . \tag{7.10}$$

Equation (7.10) represents the basic concept in the Bayes and EB tests involving the so-called linear loss. From (7.10) one concludes that the Bayes rule is to accept H_0 iff $E(\mu \mid \bar{x}) \leq \mu_0$.

The EB procedure mimics the above through estimation of $G(\mu)$. The reason that the linear loss function is popular is perhaps two-fold. First, it considers the magnitude of the error and second, it leads to (7.10). As was mentioned in Chapter VI, the estimation of the posterior mean in some cases can be accomplished without estimating the mixing distribution. In this case the hypotheses cannot be tested in this manner but must be formulated in terms of $v = \frac{1}{\mu^2}$. Under this parametrization one-sided hypotheses about v may be tested. Because of the parameter space of μ the above reparametrization does not preclude the utilization of this procedure for one-sided test of μ. The same procedure may be used to test one-sided hypotheses concerning λ without reparametrizing the density.

Not all hypotheses that could be formulated have been discussed in this chapter. It is hoped that some of the general approaches used have been informative. Some further remarks concerning the content of this chapter will be presented in the next and final chapter.

CHAPTER VIII

SUMMARY AND EXTENSIONS

1. Summary

This study was directed toward developing Bayesian and EB procedures which were useful in making inferences regarding the parameters of the inverse Gaussian distribution.

The Bayesian study was limited to considering the results obtained in the standard analysis involving a Jeffrey's diffuse prior or modified natural conjugate prior. Chapter II was devoted to developing the posterior distribution corresponding to the various cases in which one or both of the parameters were unknown. These densities were the basis for the following two chapters. In Chapter III the results of Chapter II were extended by determining the normalizing constants for the posterior distributions. The existence and derivation of mean, median, and modal estimates were also presented. In addition, some of the sampling properties of the modal estimators were presented. In Chapter IV the HPD region, or Bayesian analogue of the classical confidence interval, was introduced. Closed expressions were presented where possible and a numerical procedure was devised for computing the HPD region for the other univariate posterior densities. A numerical example illustrating this anslysis was presented along with the classical results.

Chapter V provides an introductory framework for the Empirical Bayesian analysis including theorems of general utility in EB procedures.

Chapter VI and Chapter VII are concerned with EB point estimation and testing procedures applied to the inverse Gaussian family of distributions. Although some relatively simple point estimators are developed in Chapter VI, this chapter is more of an adaptation of suitable EB technique and algorithms. A numerical example is presented, wherein, a parametric EB estimate is computed along with its classical competitor. Chapter VII is a rather brief investigation of EB testing procedures. The author considered the more commonly encountered test of hypotheses for which there exist applicable EB procedures.

2. Extensions

With regard to the pure Bayesian analysis the following extensions would be useful.

1. Investigate the variety of families of prior distributions derived in this study.
2. Consider alternative parametrizations of the inverse Gaussian that lead to prior and posterior distributions which are more mathematically tractable.
3. Develop other families of distributions which can represent plausible prior knowledge of the parameter(s) of the inverse Gaussian and which are mathematically tractable.
4. Determine the predictive densities associated with the posterior distributions derived in this study.
5. Develop numerical algorithms for constructing multivariate HPD regions for the inverse Gaussian.

The empirical Bayesian analysis could be extended in several ways.

1. Consider alternative EB procedures. There are many techniques

for density estimation not considered in this paper.

2. Consider the subject of identifiability. Can the mixing distribution be uniquely determined?

3. Consider the utility of a particular EB technique by varying:

   (a) the loss function

   (b) the mixing distribution

   (c) the sample size

   (d) the number of samples

   (e) the tolerable computational cost of the procedure

4. Develop EB techniques appropriate for testing composite hypotheses concerning both μ and λ.

Finally, one might consider the mixing distribution as a stochastic process and consider the possibility of estimating this process.

A SELECTED BIBLIOGRAPHY

(1) Abramowitz, M. and I. A. Segun. Handbook of Mathematical Functions. New York: Dover Publications, 1968.

(2) Apostol, T. M. Mathematical Analysis. Reading, Massachusetts: Addison Wesley, 1957.

(3) Box, G. E. P. and G. C. Tiao. Bayesian Inference in Statistical Analysis. Reading, Massachusetts: Addison Wesley, 1973.

(4) Chhikara, R. S. "Statistical Inference Related to the Inverse Gaussian Distribution." (Unpub. Ph.D. dissertation, Oklahoma State University, 1972.)

(5) Feller, W. An Introduction to Probability Theory and Its Applications 2. New York: John Wiley and Sons, 1972.

(6) Fisz, M. Probability Theory and Mathematical Statistics. New York: John Wiley and Sons, 1963.

(7) Frazer, D. A. S. Nonparametric Methods in Statistics. New York: John Wiley and Sons, 1959.

(8) Hildebrand, F. B. Advanced Calculus for Application. Englewood Cliffs, New Jersey: Prentice Hall, 1962.

(9) Hogg, R. V. and A. T. Craig. Introduction to Mathematical Statistics. New York: Macmillan Company, 1965.

(10) Hunt, H. H. "A Prior Distribution for the Mean of the Inverse Gaussian Distribution." (Unpub. article, Texas A & M University, 1971.)

(11) Jeffreys, H. Theory of Probability. London: Oxford University Press, 1966.

(12) Johnson, N. L. and S. Kotz. Continuous Distributions 1. Boston: Houghton Mifflin Company, 1970.

(13) Kendall, M. G. and A. Stuart. The Advanced Theory of Statistics I. New York: Hafner Publishing Co., 1969.

(14) LaValle, I. H. An Introduction to Probability, Decision, and Inference. New York: Holt, Rinehart, and Winston, 1970.

(15) Lyness, N. N. "Algorithm 379 SQUANK (Simpson Quadrature Used Adaptively—Noise Killed)." Communication of the ACM, 13 (1970), 270-273.

(16) Maritz, J. S. Empirical Bayes Methods. II New Fetter Lane, London: Meuthen and Co., 1970.

(17) Parzen, E. "On Estimation of a Probability Density Function and Its Mode." Ann. Math. Stat., 33 (1962), 1065-1076.

(18) Pennisi, L. L. Elements of Complex Variables. New York: Holt, Rinehart, and Winston, 1966.

(19) Raiffa, H. and R. Schlaifer. Applied Statistical Decision Theory. Clinton, Mass.: Colonial Press Inc., 1961.

(20) Robbins, H. "The Empirical Bayes Approach to Statistical Decision Problems." Ann. Math. Stat., 35 (1964), 1-20.

(21) Royden, H. L. Real Analysis. New York: The Macmillan Company, 1963.

(22) Rutherford, J. R. and R. G. Krutchkoff. "Some Empirical Bayes Techniques in Point Estimation." Biometrika, 56 (1969).

(23) Schrodinger, E. "Zur Theorie der Fall und Steigversuche an Teilchenn mit Browscher Bewegung." Physikalische Zeitschrift, 16 (1915), 189-195.

(24) Shuster, J. "On the Inverse Gaussian Distribution Function." Journal of the American Statistical Association, 63 (1968), 1514-1516.

(25) Tweedie, M. C. K. "Statistical Properties of the Inverse Gaussian Distributions." Ann. Math. Stat., 28 (1957), 362-377.

(26) Wald, A. Sequential Analysis. New York: John Wiley and Sons, 1947.

(27) Wasan, M. T. and L. K. Roy. "Tables of Inverse Gaussian Probabilities." Ann. Math. Stat., 38 (1967), 299.

(28) Zellner, A. An Introduction to Bayesian Inference in Econometrics. New York: John Wiley and Sons, 1971.

APPENDIX

```
      SUBROUTINE RINT(A,B,F,ERROR,TER,OP,MESS,AREA,COV)
C     A...LEFT ENDPT
C     B...RIGHT ENDPOINT
C     F...FUNCTION TO INTEGRATED...REQUIRES EXTERNAL DECLARATION IN CALLING PROG
C     TER...MAX # OF FUNCTION EVALUATIONS PERMITTED...REAL
C     OP...OPTIONS...OP=1, INTEGRATION FROM A TO B...OP=2 INTEGRATION FROM
*     A TO INFINITY...OP=3INTEGRATION FORM - INFINITY TO A...OP=4 INTEGRATION
*     FROM -INFINITY TO INFINITY
C     MESS=1...ERROR MESSAGE PRINTED
C     MESS=2...ERROR MESSAGE SUPPRESSED
C     AREA...ESTIMATE OF THE DESIRED INTEGRAL, THE INTEGRATION FROM LEFT TO RIGH
C     COV...ESTIMATE OF THE COEFFICIENT OF VARIATION OF AREA HEN ROUTINE
C     TERMINATED. EXCEPTIONS:WHEN AREA=0 OR OVER(UNDER)FLOW CONDITIONS
C     ARE ENCOUNTERED COV IS SET TO 0.
C     ERROR...DESIRED UPPER BOUND ON COV, NORMALLY RINT TERMINATES WHEN COV.LE.
C     ERROR.
      INTEGER OP
      ASSIGN 40 TO M
      ASSIGN 61 TO N
      ASSIGN 13 TO K
      S=0.
      SS=0.
      CNT=0.
      GO TO (10,20,30,35,),OP
   10 C=B-A
   11 X=C*RANF(0)+A
      X=F(X)/C
   12 S=S+X
      CNT=CNT+1.
      IF(CNT.EQ.TER)GO TO 60
      GO TO K,(13,41)
   13 Y=ABS(X)
      IF(Y.EQ.0..OR.(Y.GT.1.E-37.AND.Y.LT.1.E37))GO TO 14
      ASSIGN 71 TO N
      ASSIGN 41 TO K
      GO TO (15,41),MESS
   15 WRITE(6,1)
    1 FORMAT(1H ,'$$$$$DUE TO THE SIZE OF THE NUMBERS ENCOUNTERED CONTINUED
     2UED CALCULATION OF COV COULD BE MEANINGLESS AND WILL RESULT IN A')
      WRITE(6,2)
    2 FORMAT(1H ,'PROGRAM INTERUPT. THEREFORE THIS PORTION OF THE ROUTINE
     2E IS DISCONTINUED AND COV IS SET TO 0 AND ALL ITERATIONS ARE USED.')
     3')
      GO TO 41
   14 SS=SS+X*X
      GO TO M,(40,50)
   20 X=A-ALOG(RANF(0))
      X=F(X)/(EXP(A-X))
      GO TO 12
   30 X=A+ALOG(RANF(0))
      X=F(X)/(EXP(X-A))
      GO TO 12
   35 X=10.*GAUSF(0)
      X=25.0662827*F(X)/(EXP((-X*X)/200.))
      GO TO 12
   40 ASSIGN 50 TO M
   41 GO TO (11,20,30,35),OP
   50 AREA=S/CNT
```

```
...RINT CONTINUED
      IF(AREA.EQ.0.)GO TO 41
      E=SQRT((SS-S*S/CNT)/(CNT*(CNT-1.)))
      COV=E/AREA
      IF(COV.GT.ERROR)GO TO 41
      RETURN
   60 GO TO N,(61,71)
   61 AREA=S/CNT
      IF(AREA.EQ.0.)GO TO 71
      E=SQRT((SS-S*S/CNT)/(CNT*(CNT-1.)))
      COV=E/AREA
      RETURN
   71 AREA=S/CNT
      COV=0.
      RETURN
      END
```

```
      REAL FUNCTION INGAUS(AMU,ALAM)
C     INGAUS MUST BE DECLARED REAL IN ANY ROUTINE WHICH CALLS INGAUS
C     PETURNS AN INVERSE GAUSSIAN VARIATE WITH MEAN AMU AND PARAMETER ALAM
C     INGAUS REQUIRES     GAUSF TO GENERATE STD NORMAL DEVIATES
C     INGAUS USES RANF TO GENERATE UNIFORM (0,1) DEVIATES
C     INGAUS USES CNORM TO EVALUATE THE STD NORMAL DISTRIBUTION FUNTION
      P=0.5+EXP(2.0*ALAM/AMU)*(1.0-CNORM(2.0*SQRT(ALAM/AMU)))
      Y=(GAUSF(0))**2
      B=SQRT((4.0+AMU*Y/ALAM)*Y*(AMU**3)/ALAM)
      A=AMU*(2.0+AMU*Y/ALAM)
      IF(RANF(0)-P)1,1,2
    1 INGAUS=(A-B)/2.0
      RETURN
    2 INGAUS=(A+B)/2.0
      RETURN
      END
```

```
      SUBROUTINE AINVRT(A,B,C,F,ERROX,ERROY,ITER,BIN,M)
C    AINVRT ATTEMPTS TO FIND SOLUTION(S) TO F(X)=C WHERE A<SOLUTION<B
C    PARAMETERSD    A...LEFT  ENDPT
C                   B...RIGHT ENDPT
C                   C...VALUE  OF F TO BE INVERTED
C                   ERROX... ERROR TOLERANCE ON X...ABS(X1-X2)>ERROX CAUSES CON-
C                   TINUATION
C                   ERROY...ERROR TOLERANCE ON F(X)-C...ABS(F(X)-C)-ERROY>0
C                   CAUSES CONTINUATION
C                   ITER...NUMBER OF ITERATIONS
C                   AIN...F-INVERSE(C)
C    M...CONTROL VARIABLE...
C                   M<0 SINGLE VALUE RETURNED
C                   M=0 ALL ROOTS IN (A,B) DISCOVERED ARE PRINTED
C                   M=1... SMALLEST ROOT RETURNED
C                   M=2...SECOND SMALLEST ROOT IS RETURNED, ETC
      EXTERNAL F
      IF(B-A)1,1,2
    1 WRITE(6,3)A,B
    3 FORMAT(1H ,'LEFT ENDPT',E10.3,' > RIGHT ENDPT:',E10.3)
      RETURN
    2 IF(M)5,6,6
    5 BIN=TECBIN(A,B,C,F,ERROX,ERROY,ITER)
      RETURN
    6 CALL BMTPLR(A,B,C,F,ERROX,ITER,BIN,M)
      RETURN
      END
      SUBROUTINE BMTPLR(A,B,C,F,ERROX,ITER,BIN,M)
      EXTERNAL F
      DIMENSION R(100),RD(100)
      K=0
      HI=(B-A)/FLOAT(ITER)
      X1=A
      DO 20 I=1,ITER
      X2=X1+HI
      Y1=F(X1)-C
      Y2=F(X2)-C
      U=Y1*Y2
      IF(U)1,10,13
    1 K=K+1
      R(K)=TINRT(X1,X2,C,F,ERROX,ITER)
      X1=X2
      GO TO 20
   10 IF(Y1)11,12,11
   12 K=K+1
      R(K)=X1
      X1=X2
      GO TO 20
   11 IF(Y2)13,14,13
   14 K=K+1
      R(K)=X2
      X1=X2
      GO TO 20
   13 X1=X2
   20 CONTINUE
      IF(K)21,21,22
   21 WRITE(6,3)
    3 FORMAT(1H ,'*****NO ROOTS FOUND*****INCREASE ITER')
```

```
...AINVRT CONTINUED
      RETURN
   22 IF(M)24,24,27
   24 WRITE(6,5)R(1)
    5 FORMAT(1H ,'*****ROOT=',E14.7)
      I=1
   37 I=I+1
   36 IF(I-K)38,38,27
   38 J=I-1
      IF(R(I)-R(J))25,37,25
   25 WRITE(6,5)R(I)
      GO TO 37
   27 RD(1)=R(1)
      L=1
      DO 30 I=2,K
      J=I-1
      IF(R(I)-R(J))28,30,28
   28 L=L+1
      RD(L)=R(I)
   30 CONTINUE
      DO 35 I=1,L
   35 R(I)=RD(I)
      IF(M-L)65,65,70
   65 BIN=R(M)
      RETURN
   70 WRITE(6,100)L
  100 FORMAT(1H ,'*****AINVRT...ONLY ',I2,' ROOTS FOUND, LARGEST RETURNE
     2D')
      BIN=R(L)
      RETURN
      END
      FUNCTION TINRT(A,B,C,F,ERROX,ITER)
      IF(B-A)1,1,2
    1 WRITE(6,3)A,B
    3 FORMAT(1H '*****TINRT WAS CALLED WITH LEFT ENDPT:',E14.7,'> THAN RIGHT
     2IGHT ENDPT:',E14.7,' NO VALUE RETURNED')
      RETURN
    2 X1=A
      X2=B
      DO 60 I=1,ITER
      Y2=F(X2)-C
      Y1=F(X1)-C
      IF(Y2-Y1)10,15,10
   15 IF(Y2)16,17,16
   16 IF(Y1)18,19,18
   17 TINRT=X2
      RETURN
   19 TINRT=X1
      RETURN
   18 WRITE(6,5)X1,X2,Y2
    5 FORMAT(1H ,'*****TINRT ERROR*****X1=',E14.7,' X2=',E14.7,' F(X1)-C=
     2=F(X2)-C=',E14.7,'X2 RETURNED')
      TINRT=X2
      RETURN
   10 X3=X2-Y2*(X2-X1)/(Y2-Y1)
      IF(ABS(X3-X2)-ERROX)20,20,25
   20 TINRT=X3
      RETURN
```

```
...AINVRT CONTINUED
   25 IF(ABS(X3-X1)-ERROX)20,20,30
   30 Y3=F(X3)-C
      IF(Y1*Y3)35,50,40
   35 X2=X3
      GO TO 60
   40 X1=X3
      GO TO 60
   50 IF(Y3)60,55,60
   55 TINRT=X3
      RETURN
   60 CONTINUE
      X3=X2-Y2*(X2-X1)/(Y2-Y1)
      WRITE(6,65)X2,X3
   65 FORMAT(1H ,'*****TINVRT*****ERROR BOUND NOT ATTAINED, LAST TWO VALUES
     2UES:',2(E14.7,2X))
      RETURN
      END
      FUNCTION TECBIN(A,B,C,F,ERROX,ERROY,ITER)
      EXTERNAL F
      X1=A
      X2=B
      DO 50 I=1,ITER
      Y1=F(X1)-C
      Y2=F(X2)-C
      IF(Y1-Y2)3,11,3
    3 X3=X2-Y2*(X2-X1)/(Y2-Y1)
      X1=X2
      X2=X3
      IF(ABS(X1-X2)-ERROX)16,16,5
   16 IF(ABS(Y2)-ERROY)45,45,5
    5 U=(F(X1)-C)*(F(X2)-C)
      IF(U)6,7,50
    6 U=AMIN1(X1,X2)
      V=AMAX1(X1,X2)
      TECBIN=TINRT(U,V,C,F,ERROX,ITER)
      RETURN
    7 IF(F(X1)-C)8,9,8
    8 IF(F(X2)-C)50,10,50
   10 TECBIN=X2
      RETURN
    9 TECBIN=X1
      RETURN
   11 IF(ABS(X1-X2)-ERROX)45,45,14
   14 X1=AMIN1(X1,X2)
      X2=AMAX1(X1,X2)
      X4=X1+(X2-X1)/3.
      X5=X4+(X2-X1)/3.
      X1=X4
      X2=X5
      X3=(X1+X2)/2.
      Y1=F(X1)-C
      Y2=F(X2)-C
      Y3=F(X3)-C
      S1=ABS(Y3-Y1)
      S2=ABS(Y2-Y3)
      IF(S1-S2)17,17,18
   18 IF(ABS(Y1)-ABS(Y3))20,20,21
   20 X2=X1
      X1=X3
```

```
...AINVRT CONTINUED
      GO TO 50
   21 X2=X3
      GO TO 50
   17 IF(ABS(Y2)-ABS(Y3))22,22,23
   22 X1=X3
      GO TO 50
   23 X1=X2
      X2=X3
      GO TO 50
   45 IF(A-X2)47,47,46
   46 WRITE(6,100)X2
  100 FORMAT(1H ,'*****AINVRT RETURNS THE VALUE',E10.3,'< LEFT ENDPT SPE
     2CIFIED')
      TECBIN=X2
      RETURN
   47 IF(X2-B)48,48,49
   49 WRITE(6,101)X2
  101 FORMAT(1H ,'*****AINVRT RETURNS THE VALUE',E10.3,'> RIGHT ENDPT SP
     2ECIFIED')
      TECBIN=X2
      RETURN
   48 TECBIN=X2
      RETURN
   50 CONTINUE
      TECBIN=X2
      WRITE(6,102)X1,X2
  102 FORMAT(1H ,'*****AINVRT....ABS(X1-X2)>ERROR...LAST TWO VALUES: X1=
     2',E14.7,2X,'X2=',E14.7)
      RETURN
      END
```

```
C..........PROGRAM USED TO GENERATE UNIVARIATE HPD REGION FOR MU...............
      EXTERNAL AUC,F
      COMMON XMAX,ALPHA,ERROR
      XMAX=2.0386836
      ALPHA=.01
      XP=XMAX/ALPHA
      ERROR=.001
      A=-ALPHA
      B=1.-ALPHA
      Y=AINVRT(A,B,0.,AUC,ERROR)
      XL=AINVRT(0,XMAX,Y,F,ERROR)
      XR=AINVRT(XMAX,XP,Y,F,ERROR)
      WRITE(6,1)ALPHA,XL,XR
    1 FORMAT(1H ,'...THE 1-',F4.3,' HPD REGION IS (',E14.7,',',E14.7,')'
     2)
      CALL EXIT
      END
      FUNCTION F(X)
      A=51.41908268D0
      B=15.8444061D0
      IF(X)1,1,2
    1 F=0.
      RETURN
    2 C=GAMMA(25.5)
      CL=ALOG(C)
      DL=ALOG(X)*1.5
      E=A/X**2-50./X+B
      EL=ALOG(E)*25.5
      GL=CL-DL-EL
      F=EXP(GL)/.1477325E10
      RETURN
      END
      FUNCTION AUC(C)
      EXTERNAL F
      COMMON XMAX,ALPHA,ERROR
      ERR=.001
      XL=BINVRT(0.,XMAX,C,F,ERROR)
      XP=XMAX/ALPHA
      XR=BINVRT(XMAX,XP,C,F,ERROR)
      CALL SOANK(0.,XL,ERR,FIFTH,RUM,NO,F,AREA)
      AREA1=AREA
      XTRUN=10.*XR
      CALL SOANK(XR,XTRUN,ERR,FIFTH,RUM,NO,F,AREA)
      AREA2=AREA
      SUM=AREA1+AREA2
      AUC=SUM-ALPHA
      WRITE(6,1)XL,AREA1,XR,XTRUN,AREA2,SUM
    1 FORMAT(1H ,'I(0,',E10.3,') = ',E10.3,'  I(',E10.3,',',E10.3,') = ',
     2,E10.3,'  TOTAL TAIL AREA=',E10.3)
      RETURN
      END
```

VITA

John Mark Palmer

Candidate for the Degree of

Doctor of Philosophy

Thesis: CERTAIN NON-CLASSICAL INFERENCE PROCEDURES APPLIED TO THE INVERSE GAUSSIAN DISTRIBUTION

Major Field: Statistics

Biographical:

Personal Data: Born in Wichita, Kansas, December 1, 1945, the son of Paul and Kara Palmer.

Education: Attended elementary school in Independence, Kansas; graduated from Belle Fourche High School, Bell Fourche, South Dakota in 1963; received the Bachelor of Science degree in mathematics from Oklahoma State University in May, 1967; received the Master of Science degree in mathematics from Oklahoma State University in May, 1970; completed requirements for the Doctor of Philosophy degree in July, 1973.

Professional Experience: Served as Graduate Assistant in the Department of Mathematics, Oklahoma State University, 1967-1973; served as computer programmer for Cities Service Oil Company, Summer 1967, 1968, and 1969.