A NEW GENERAL METHOD FOR CONSTRUCTING CONFIDENCE SETS IN ARBITRARY DIMENSIONS: WITH APPLICATIONS

by

A. DasGupta
Purdue University

J. K. Ghosh
Indian Statistical Institute
and Purdue University

and

M. M. Zen
National Cheng Kung University

Technical Report #91-66C

Department of Statistics
Purdue University

November 1991
Revised - June 1993

◇

A NEW GENERAL METHOD FOR CONSTRUCTING CONFIDENCE SETS IN ARBITRARY DIMENSIONS: WITH APPLICATIONS

by

A. DasGupta*
Purdue University

J. K. Ghosh
Indian Statistical Institute
and Purdue University

M. M. Zen*
National Cheng Kung University

ABSTRACT

Let $\underset{\sim}{X}$ have a star unimodal distribution P_0 on $\mathbf{R}^p$. We describe a general method for constructing a star-shaped set S with the property $P_0(\underset{\sim}{X} \in S) \geq 1-\alpha$, where $0 < \alpha < 1$ is fixed. This is done by using the Camp-Meidell inequality on the Minkowski functional of an arbitrary star-shaped set S and then minimizing Lebesgue measure in order to obtain size-efficient sets. Conditions are obtained under which this method reproduces a level (high density) set. The general theory is then applied to two specific examples: set estimation of a multivariate normal mean using a multivariate t prior and classical invariant estimation of a location vector $\underset{\sim}{\theta}$ for a mixture model. In the Bayesian example, a number of shape properties of the posterior distribution are established in the process. These results are of independent interest as well. A computer code is available from the authors for automated application. The methods presented here permit construction of explicit confidence sets under very limited assumptions when the underlying distributions are calculationally too complex to obtain level sets.

Key words: confidence set, star unimodal, star-shaped sets, Minkowski functional, invariant sets, level sets, prior, posterior, HPD sets.

AMS Classifications: Primary 62F25, Secondary 60E15, 62C10

---

*Research supported by NSF grant DMS 89-23071. This work was part of the Ph.D. dissertation of the third author at Purdue University.

1. INTRODUCTION

1.1. Previews

Set estimation of unknown parameters is a problem of major statistical importance and mathematical interest. The common approach is to minimize a reasonable measure of size of the set subject to a lower bound on the set's probability content. This is generally accepted as a good formulation of the problem and in common situations results in many standard and time-tested estimation procedures. The z-interval for an unknown normal mean and the Hotelling confidence ellipsoid for an unknown multivariate normal mean vector are two prime examples.

The success of the method depends crucially on the ability to identify the high probability sets of a relevant distribution. This is usually not difficult in many standard problems, because the underlying distribution often has a structure, like a spherical or an elliptical structure. But the identification of the high probability sets becomes very difficult if such structure is not present. Indeed, it may even be argued that realistic models will not result in such nice structures in the underlying distributions. A simple example is the case

$$\underset{\sim}{X} = \underset{\sim}{\theta} + \underset{\sim}{Z}, \tag{1.1}$$

where $\underset{\sim}{\theta}$ is an unknown parameter vector in $\mathbf{R}^p$ and the error $\underset{\sim}{Z}$ has a distribution with density

$$f(\underset{\sim}{z}) = (1-\lambda)\frac{1}{(2\pi)^{p/2}}e^{-\frac{1}{2}\underset{\sim}{z}'\underset{\sim}{z}} + \lambda \cdot \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}\underset{\sim}{z}'\Sigma^{-1}\underset{\sim}{z}}, \tag{1.2}$$

where Σ is a known positive definite matrix, and $0 < \lambda < 1$ is also taken to be known. Even in such a simple mixture problem, determination of, for example, the best invariant confidence set is rather difficult. The problem stems from the fact that sets of the form

$$\{\underset{\sim}{z}: f(\underset{\sim}{z}) \geq c\}$$

are no longer multiples of one another; the shapes vary with c and the correct threshold c for a specified confidence level $1-\alpha$ can only be found by repeated trial and error integration. The same problem arises in practically any Bayesian problem whenever the prior distribution for the unknown parameter is such that the high probability sets of the posterior have no fixed shapes or are difficult to determine. A simple example is the case when $\underset{\sim}{X}$ is $N(\underset{\sim}{\theta}, I)$ and $\underset{\sim}{\theta}$ has a $t(m, \underset{\sim}{\mu}, I)$ prior, i.e., a t prior with m degrees of freedom, location $\underset{\sim}{\mu}$ and scale matrix I. Even for such a reasonable prior and a standard problem, the Bayesian HPD sets are difficult to nearly impossible to determine. Numerical methods, such as simulation from the posterior, do not work in general because it is not even known that the HPD sets are convex. An attempt to identify the high posterior density points by simulation from the posterior followed by a numerical construction of its convex hull cannot be mathematically justified and in any case is a formidable project in high dimensions.

In this article, we provide a new method for constructing confidence sets in any (finite) dimension. The method is explicit: it requires only an appropriate integration involving the underlying distribution. The sets are constructed to have a specified $1-\alpha$ probability content, where $0 < \alpha < 1$ is fixed but arbitrary. The construction is specifically directed towards finding the smallest set possible using this technique. We give ample evidence that sensible sets with reasonable to excellent size properties come out of this method. The sets are in general not high probability sets. But rather surprisingly, the contour of the sets produced by our method are in fact high probability contours under a particular assumption. At a mathematical level, the results we present here bring out a novel connection between unimodality and high probability sets and demonstrate that the classical Camp-Meidell inequalities are much more useful than are generally thought to be. A Fortran program for use in practical cases is available from the authors.

1.2. The Key Idea: An Illustrative Example

Suppose $X \sim U[\theta-1, \theta+1]$, where $-\infty < \theta < \infty$ is an unknown location parameter. Trivially, the interval $X \pm (1-\alpha)$ is the best (location) invariant $100(1-\alpha)\%$ confidence interval for θ. Since the 'optimal' interval is known and is so simple, there would be no reason to construct frequentist confidence intervals by other methods. However, other methods can be used, if one wants. Indeed, by Chebyshev's inequality, for any K, $r > 0$,

$$P[|X-\theta| > K] \leq \frac{E|X-\theta|^r}{K^r} = \frac{1}{K^r(r+1)}. \tag{1.3}$$

(1.3) immediately implies that the interval $X \pm \sqrt[r]{\frac{1}{\alpha(r+1)}}$ is also a $100(1-\alpha)\%$ confidence interval. Since Chebyshev's inequality is usually not very sharp, use of this interval results in loss of efficiency in the sense of size. For example, if $\alpha = .1$ and one takes $r = 2$, then the length of the Chebyshev interval is twice more than that of the best invariant interval $X \pm .9$.

However, the following observation is interesting. Since $X - \theta \sim U[-1, 1]$ and the $U[-1, 1]$ distribution is unimodal about zero, by the Camp-Meidell inequality (see Camp (1922), Meidell (1922), Dharmadhikari and Joag-dev (1988)), for every K, $r > 0$,

$$P[|X - \theta| > K] \leq \left(\frac{r}{r+1}\right)^r \cdot \frac{E|X - \theta|^r}{K^r} = \left(\frac{r}{r+1}\right)^r \cdot \frac{1}{K^r(r+1)}. \tag{1.4}$$

(1.4) gives the shorter interval $X \pm K$ with

$$K = K(r) = \frac{r}{r+1} \cdot \sqrt[r]{\frac{1}{\alpha(r+1)}} \tag{1.5}$$

as a $100(1-\alpha)\%$ interval, for any $r > 0$. Rather surprisingly, on minimizing (1.5) over $r > 0$ by using elementary calculus, one returns exactly to the interval $X \pm (1 - \alpha)$, corresponding to $r = \frac{1-\alpha}{\alpha}$! This simple example indicates that optimal use of the Camp-Meidell inequality may be useful in set estimation problems where the underlying distribution is unimodal.

In this article, we demonstrate that the above example is not an accidental coincidence and we show ways in which this technique can be used with reward in any number of dimensions, subject to an appropriate kind of multivariate unimodality. The theory is then illustrated with two concrete examples.

1.3. Outline, Overview

In section 2, we present the general theory extending the technique of the previous example to many dimensions. As indicated in that example, some form of unimodality is required for arriving at useful sets by using this method. The appropriate form of unimodality precisely suited for our analysis is the so called star unimodality. In this definition, the high probability sets of the underlying distribution are assumed to be star-shaped about a mode $\underset{\sim}{\nu}$. The very appealing feature is that unlike some other notions of multivariate unimodality, there are usable and verifiable characterizations of star unimodality. It is also one of the weaker notions of multivariate unimodality, making it more likely that the underlying distribution is star unimodal. This in turn makes the methods presented in this article more widely applicable.

It is proved in section 2 that if a random vector $\underset{\sim}{X}$ is star unimodal about some $\underset{\sim}{\nu}$, then for any set S with $\underset{\sim}{0}$ in its interior and also star-shaped about $\underset{\sim}{0}$ (zero), and for any given $0 < \alpha < 1$,

$$P(\underset{\sim}{X} - \underset{\sim}{\nu} \in kS) \geq 1 - \alpha \tag{1.6}$$

for suitable k, depending on a generic constant $r > 0$, the value of α, and the set S. This essentially says that a prespecified probability content of $1 - \alpha$ can be achieved by starting with any set S star-shaped about $\underset{\sim}{0}$, inflating it sufficiently, and then recentering it at the mode $\underset{\sim}{\nu}$.

Next, we address the problem of minimizing the volume (Lebesgue measure) of the set kS. Since each set kS in (1.6) guarantees a probability content of $1 - \alpha$, this minimization is simply for deriving the smallest set obtainable by using our method. For any given $r > 0$, we give an explicit analytic description of the star-shaped set $S^* = S^*(r)$ that solves this minimization problem over all possible star-shaped sets S. It is proved that the optimal family of sets $S^*(r)$ have the following invariance property: for any given α_1, α_2, $0 < \alpha_1, \alpha_2 < 1$, the corresponding optimal star-shaped sets $S^*(r, \alpha_1)$, $S^*(r, \alpha_2)$ are multiples of each other, i.e., the solutions we present are mutually homothetic. This property is attractive from a communication and interpretation point of view and is not necessarily shared by the family of high probability sets. We then give an indirect stochastic majorization argument to show that if the high probability sets do have this property, then for any $r > 0$ the contour of our set $S^*(r)$ is exactly a high probability (density) contour. Examples are given where this is indeed the situation.

For best results to be obtained, it is clearly necessary to optimize not only over all possible star-shaped sets S, but also over $r > 0$. This is evident from the illustrative example presented in subsection 1.2. This two stage optimization is then carried out in two concrete examples. The theory we present is completely general, subject to star unimodality. It therefore applies in general, subject to star unimodality. The two examples to which the theory is applied are the ones mentioned in subsection 1.1, namely estimation of a multivariate normal mean with multivariate t priors, and invariant classical set estimation for the mixture model (1.2). Notice use of t priors in the first problem is very reasonable, and calculational complexities aside, is preferred by many over conjugate normal priors because of the thicker tails of t distributions.

As stated before, the theory presented here applies only when the underlying distribution is star unimodal (although a direct extension of our theory will also cover the cases of α-unimodal distributions, as in Olshen and Savage (1970); this will be useful for the cases when star unimodality is lacking). In the Bayesian example, therefore, it is necessary that one has the posterior of θ given the observed data $\underset{\sim}{X}$ to be star unimodal. In section 3, we describe explicit results in this direction. These results are of independent interest, regardless of the present context in which they are needed. We prove that if $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I)$ and $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \tau^2 I)$, where $m, \underset{\sim}{\mu}, \sigma^2, \tau^2$ are given, then the posterior is star unimodal for all $\underset{\sim}{X}$ if and only if $\frac{\tau^2}{\sigma^2} \geq \frac{m+p}{8m}$. Under this condition, therefore, our set estimation methods can be applied regardless of which $\underset{\sim}{X}$ is observed. If $\frac{\tau^2}{\sigma^2} < \frac{m+p}{8m}$, we give a complete analytic description of the set of all $\underset{\sim}{X}$ for which the posterior is star unimodal. Indeed, we prove that the posterior is star unimodal if and only if $\underset{\sim}{X}$ lies outside a spherical band

$$\{\underset{\sim}{X}\colon\ a < ||\underset{\sim}{X} - \underset{\sim}{\mu}|| < b\}, \tag{1.7}$$

with explicit formulas for a and b. This will enable direct immediate verification of whether or not the set estimation methods we present are applicable. We also state without proof a necessary and sufficient condition for the posterior to be star unimodal for all $\underset{\sim}{X}$ in the more general case $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \Sigma_1)$ and $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \Sigma_2)$; again $m, \underset{\sim}{\mu}, \Sigma_1, \Sigma_2$ are assumed given. Under this condition, our methods can again be used without worrying about which $\underset{\sim}{X}$ was observed.

To get a fair picture of the situation, it is necessary to find out how efficient our sets are in the sense of size. In the Bayesian example, having obtained a specific confidence set after the two stage optimization on $r > 0$ and star-shaped sets S, we evaluate their efficiency by the following method: we squeeze the obtained set until the desired probability of $1 - \alpha$ is exactly attained and then measure efficiency by taking the pth root of the ratio of the volumes of the two sets. Taking the pth root is reasonable because in high dimensions, a negligible increase in (say) the diameter of a set can result in a very significant increase in the volume. For example, if S_1 is the unit sphere in $\mathbf{R}^{20}$ and S_2 is the sphere with radius

1.02, the volume of S_2 is 48.6% more than that of S_1! These efficiencies are evaluated for various p, m, σ^2, τ^2 and data $\underset{\sim}{X}$ and are reported and discussed in section 4. Section 4 also deals with the case when the underlying distribution is exactly a normal. This is not a case where our methods need or should be used in practice. But applying our methods to the normal case is useful as a benchmark.

Section 5 treats the invariant estimation example stated before.

The principal results of this article are then the following:

a. demonstrating that Camp-Meidell inequalities can be very useful;

b. presenting a general method of set estimation in arbitrary dimensions subject to star unimodality, when determination of, say, high density sets can be at least formidable;

c. establishing a new connection between unimodality and high density sets; and

d. description of the shape behavior of posteriors under t priors for normal means. These results are needed for our results to be applicable in the first place; but we hope they are of their own interest as well.

Notice that particularly exciting is the potential application of our methods to prediction problems; construction of prediction regions is very valuable for planning purposes.

1.4. History

There is a truly vast literature on confidence sets. As such, it is impossible to give a complete account of the work in this area. For results on invariant and decision-theoretic set estimation, see Brown (1966), Cohen and Strawderman (1973), DasGupta (1991), Hooper (1982), Hwang and Casella (1982), Joshi (1967) and Naiman (1984); for general exposition and discussion on Bayesian confidence sets, see Box and Tiao (1973), Ferguson (1973) and Lehmann (1986); for various notions and implications of multivariate unimodality, see Anderson (1955), Dharmadhikari and Joag-dev (1988), DasGupta (1980), Eaton (1982), Eaton and Perlman (1977), Kanter (1977), Marshall and Olkin (1974), Mudholkar (1966), Olshen and Savage (1970), Tong (1980) and Wells (1978); for a recent use of the Camp-Meidell inequality in a different context, see Bickel and Krieger (1989).

2. THE GENERAL THEORY

2.1. Basic Definitions

For a lucid treatment of unimodality in high dimensions, see Dharmadhikari and Joag-dev (1988). We will merely cite some definitions and theorems which will be employed throughout this section.

Definition 2.1. A real random variable X with distribution function F is called *unimodal* about a mode ν if F is convex on $(-\infty, \nu)$ and concave on (ν, ∞).

Definition 2.2. A set $S \subset \Re^p$ is said to be *star-shaped* about $\underset{\sim}{\xi} \in S$ if, for every $\underset{\sim}{x} \in S$, the line segment joining $\underset{\sim}{\xi}$ and $\underset{\sim}{x}$ is completely contained in S.

Evidently the star-shaped property is weaker than convexity, except in one dimension.

Definition 2.3. Let $\underset{\sim}{X} \sim F$ be an absolutely continuous random variable on $\Re^p$ with density $f(\underset{\sim}{x})$. We say that $\underset{\sim}{X}$ (or equivalently f) is *star unimodal about* $\underset{\sim}{0}$ if, and only if $f(t\underset{\sim}{x}) \geq f(s\underset{\sim}{x})$ for all $0 < t < s < \infty$ and all $\underset{\sim}{x}$ or, equivalently, if, and only if, for every $s > 0$, the level set

$$C_s = \{\underset{\sim}{x} \in \Re^p \colon\ f(\underset{\sim}{x}) \geq s\} \tag{2.1}$$

is star-shaped about $\underset{\sim}{0}$.

Remark 1. Notice that if f is differentiable, then star unimodality is equivalent to $\frac{d}{dt} f(t\underset{\sim}{x}) \leq 0$ for all $t > 0$ and all $\underset{\sim}{x}$.

Definition 2.4. $\underset{\sim}{X}$ is said to be *star unimodal about* $\underset{\sim}{\nu}$ if $\underset{\sim}{X} - \underset{\sim}{\nu}$ is star unimodal about $\underset{\sim}{0}$.

$\underset{\sim}{\nu}$ will be called the *mode* of $\underset{\sim}{X}$. Notice a particular $\underset{\sim}{X}$ may be star unimodal about several $\underset{\sim}{\nu}$.

Example: Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \Sigma)$, where Σ is a given nonnegative definite $p \times p$ matrix. Then $\underset{\sim}{X}$ is star unimodal about $\underset{\sim}{\theta}$.

Remark 2. Anderson (1955) defined unimodality by the property that the level set C_s in (2.1) is convex for every s. Clearly, thus, star unimodality is a weaker notion than the one suggested by Anderson. While the definition proposed by Anderson is really quite natural, unfortunately there seems to be no verifiable or usable characterizations of random vectors satisfying it. In that sense, the following Theorem is very useful. So is Remark 1.

Theorem 2.1. The p dimensional random vector $\underset{\sim}{X}$ is star unimodal about $\underset{\sim}{0}$ if and only if $\underset{\sim}{X}$ is distributed as $U^{\frac{1}{p}}\underset{\sim}{Z}$, where U and $\underset{\sim}{Z}$ are independent and U is uniformly distributed on (0,1).

Proof: See page 40, Dharmadhikari and Joag-dev (1988).

Remark 3. Theorem 2.1 immediately gives that for $p = 1$, star unimodality is equivalent to usual unimodality.

From Theorem 2.1, one also gets the assertion of the following example immediately.

Example 1. Let $\underset{\sim}{X}$ be star unimodal about $\underset{\sim}{0}$. Then

$$||\underset{\sim}{X}||_2^p = (\sum_{i=1}^{p} X_i^2)^{p/2} \tag{2.2}$$

is real-valued unimodal about 0.

2.2. A Basic Result

Example 1 above leads to the following simple result.

Proposition 2.2. Let $\underset{\sim}{X}$ be star unimodal about $\underset{\sim}{\nu}$. Then for every $0 < \alpha < 1$, and every $r > 0$, the L_2 ball

$$S_2(\underset{\sim}{\nu}) = \{\underset{\sim}{X}:\ ||\underset{\sim}{X} - \nu||_2 \leq k\} \tag{2.3}$$

where

$$k = (\frac{r}{(r+1)\cdot\sqrt[r]{\alpha}})^{1/p}(E||\underset{\sim}{X} - \underset{\sim}{\nu}||_2^{pr})^{1/pr} \tag{2.4}$$

has a probability content of at least $1-\alpha$; i.e.,

$$P(\underset{\sim}{X} \in S_2(\underset{\sim}{\nu})) \geq 1-\alpha. \tag{2.5}$$

Remark 4. Proposition 2.2 is formally valid even if $E||\underset{\sim}{X}-\underset{\sim}{\nu}||_2^{pr}=\infty$, but obviously is not useful in that case.

Remark 5. From Proposition 2.2, one sees that if the distribution of $\underset{\sim}{X}$ is star unimodal about some $\underset{\sim}{\nu}$, then one can construct appropriate L_2 balls as confidence sets with a guaranteed probability content $1-\alpha$.

Proof of Proposition 2.2: By Example 1, since $\underset{\sim}{X}$ is star unimodal about $\underset{\sim}{\nu}$, $||\underset{\sim}{X}-\underset{\sim}{\nu}||_2^p$ is real-valued unimodal about 0. The proposition now follows on using the Camp-Meidell inequality for real valued unimodal (about 0) random variables Z, namely,

$$P(|Z|>a) \leq (\frac{r}{r+1})^r \cdot \frac{E|Z|^r}{a^r} \tag{2.6}$$

for any r, $a>0$.

Remark 6. Proposition 2.2 generalizes to L_k balls for any $k>0$; i.e., one can construct appropriate L_k balls as confidence sets with a guaranteed probability content. This is because the only property of the L_2 norm $||\underset{\sim}{X}||_2$ that is needed for the unimodality of $||\underset{\sim}{X}-\underset{\sim}{\nu}||_2^p$ in one dimension is that L_2 norm is homogeneous, i.e., for $c\geq 0$, $||c\underset{\sim}{X}||_2 = c\cdot||\underset{\sim}{X}||_2$. However, this is true of the L_k norm for any $k>0$ and thus Proposition 2.2 generalizes in an obvious way to L_k balls. Indeed, since the above homogeneity property is valid for functions much more general than L_k norms, it is possible to strongly generalize Proposition 2.2. The following definition is needed for this purpose.

Definition 2.5. Let $S \subseteq \Re^p$ be any star-shaped set, star-shaped about $\underset{\sim}{0}$. The *Minkowski functional* π_S of the set S is defined as

$$\pi_S(\underset{\sim}{x}) = \inf\{a>0: \underset{\sim}{x} \in aS\},\ \underset{\sim}{x} \in \Re^p. \tag{2.7}$$

Example 2.

(a) Let $S = \{\underset{\sim}{X} \in \mathrm{R}^p : ||\underset{\sim}{X}||_2 \leq 1\}$.

Then $\pi_S(\underset{\sim}{X}) = ||\underset{\sim}{X}||_2$.

(b) Let $S = \{\underset{\sim}{X} \in \mathrm{R}^p : ||\underset{\sim}{X}||_k \leq 1\}, k > 0$.

Then $\pi_S(\underset{\sim}{X}) = ||\underset{\sim}{X}||_k$.

(c) Let $S = \{\underset{\sim}{X} \in \mathrm{R}^p : \underset{\sim}{X}'\Sigma^{-1}\underset{\sim}{X} \leq 1\}$, Σ positive definite

Then $\pi_S(\underset{\sim}{X}) = \sqrt{\underset{\sim}{X}'\Sigma^{-1}\underset{\sim}{X}}$.

In each of the above examples, $\pi_S(\underset{\sim}{x})$ has the homogeneity property

$$\pi_S(c\underset{\sim}{x}) = c\pi_S(\underset{\sim}{x}) \qquad \forall\, c \geq 0,\ \forall\, \underset{\sim}{x} \in \mathrm{R}^p. \tag{2.8}$$

This is true in general.

Proposition 2.3. Let $S \subseteq \mathrm{R}^p$ be star-shaped about $\underset{\sim}{0}$. Then $\pi_S(\underset{\sim}{x})$, the Minkowski functional of the set S, is homogeneous of degree 1, i.e., $\pi_S(\underset{\sim}{x})$ satisfies (2.8).

Proof: Obvious.

We are now ready to prove the following generalization of Proposition 2.2.

Proposition 2.4. Let $\underset{\sim}{X}$ be star unimodal about $\underset{\sim}{\nu}$ and absolutely continuous and let $S \subseteq \mathrm{R}^p$ be star-shaped about $\underset{\sim}{0}$, with $\underset{\sim}{0}$ in its interior. Then for every $0 < \alpha < 1$, and every $r > 0$,

$$P(\underset{\sim}{X} - \underset{\sim}{\nu} \in kS) \geq 1 - \alpha,$$

where

$$k = \left(\frac{r}{(r+1)\cdot \sqrt[r]{\alpha}}\right)^{1/p} \cdot (E[\pi_S(\underset{\sim}{X} - \underset{\sim}{\nu})]^{pr})^{1/pr}. \tag{2.9}$$

Remark 7. Proposition 2.4 shows how an arbitrary S star-shaped about $\underset{\sim}{0}$ can be sufficiently blown up and then recentered in order to guarantee a probability content of $1 - \alpha$.

Proof of Proposition: It follows from (2.8) and Theorem 2.1 that $(\pi_S(\underset{\sim}{X} - \underset{\sim}{\nu}))^p$ is real valued unimodal about 0. (2.9) now follows exactly in the lines of Proposition 2.2 on using the Camp-Meidell inequality (2.6).

2.3. Construction of Optimal Star Shaped Sets and Their Efficiency Properties

From Proposition 2.4, it follows that for any star-shaped set S, kS (recentered) is a confidence set of guaranteed probability of $1-\alpha$, where k is as in (2.9). Construction of an 'optimal' star-shaped set S^0 is the main goal of this section. Let $\lambda(S)$ denote the Lebesgue measure of S; for a fixed $r > 0$, we find a star-shaped set S^* which minimizes (2.9), subject to the restriction that $\lambda(S) = 1$. Notice that the restriction $\lambda(S) = 1$ results in no loss of generality, since $\lambda(kS) = k^p\lambda(S)$, and there is a scale-invariance in the particular problem we now have.

The following definition and notation will be subsequently used.

Definition 2.6. For given $r > 0$, let $S^*(r)$ (if it exists) minimize $E[\pi_S(\underset{\sim}{X} - \underset{\sim}{\nu})]^{pr}$ among all star-shaped (about $\underset{\sim}{0}$) sets S such that $\lambda(S) = 1$. Then we will call $S^*(r)$ the optimal star-shaped set of order r.

We now present a general and explicit result describing the contour of $S^*(r)$ for any $r > 0$. We will give the details for the case $p = 2$ for ease of understanding. The technique is just the same for any $p \geq 2$. The general result will be stated precisely.

Let then $S \subseteq \mathbf{R}^2$ be star-shaped about $\underset{\sim}{0}$. Recall that $\underset{\sim}{X} - \underset{\sim}{\nu}$ is assumed to have a star unimodal distribution. We can and will assume here that $\underset{\sim}{\nu} = \underset{\sim}{0}$. We thus want to minimize $E[\pi_S(\underset{\sim}{X})]^m$ subject to $\lambda(S) = 1$, where $m = pr = 2r$, and $\pi_S(\underset{\sim}{X}) = \inf\{c > 0 : \underset{\sim}{X} \in cS\}$.

Transforming to polar coordinates

$$X_1 = \rho\cos\phi \quad \text{and} \quad X_2 = \rho\sin\phi,$$

where $\rho > 0$ and $0 < \phi \leq 2\pi$, and denoting $\psi(\phi)$ as the radius of S along angle ϕ, i.e., $\psi(\phi) = \sup\{\rho : (\rho\cos\phi,\ \rho\sin\phi) \in S\}$, one has $\pi_S(\underset{\sim}{X}) = \frac{\rho}{\psi(\phi)}$. The restriction $\lambda(S) = 1$ is equivalent to

$$\int_S d\underset{\sim}{\theta} = 1$$

$$\iff \int_0^{2\pi} \int_0^{\psi(\phi)} \rho d\rho d\phi = 1$$
$$\iff \int_0^{2\pi} \frac{\psi^2(\phi)}{2} d\phi = 1. \tag{2.10}$$

Observe that the star property of S is being used to deduce (2.10). Let now $P(\rho, \phi)$ denote the density of $\underset{\sim}{X}$ written as a function of ρ and ϕ. The problem is then to minimize

$$\int_0^{2\pi} \int_0^{\infty} (\frac{\rho}{\psi(\phi)})^m \rho P(\rho, \phi) d\rho d\phi \tag{2.11}$$

subject to $\int_0^{2\pi} \frac{\psi^2(\phi)}{2} d\rho = 1$. Note that $\psi \geq 0$ is arbitrary measurable.

Denoting

$$P_m(\phi) = \int_0^{\infty} \rho^{m+1} P(\rho, \phi) d\rho \text{ and } d\nu(\phi) = \frac{\psi^2(\phi)}{2} d\phi, \tag{2.12}$$

(2.11) reduces to $\int_0^{2\pi} \frac{P_m(\phi)}{\psi^m(\phi)} d\phi$.

Now, the problem is equivalent to minimizing

$$\begin{aligned} &\int_0^{2\pi} \frac{P_m(\phi)}{\psi^m(\phi)} d\phi \\ =&2 \int_0^{2\pi} \frac{P_m(\phi)}{\psi^{m+2}(\phi)} \frac{\psi^2(\phi)}{2} d\phi \\ =&2 \int_0^{2\pi} \frac{P_m(\phi)}{\psi^{m+2}(\phi)} d\nu(\phi) \\ \text{subject to } &\int_0^{2\pi} d\nu(\phi) = 1. \end{aligned} \tag{2.13}$$

By Holder's inequality,

$$2 \int_0^{2\pi} \frac{P_m(\phi)}{\psi^{m+2}(\phi)} d\nu(\phi) \geq 2 (\int_0^{2\pi} (\frac{P_m(\phi)}{\psi^{m+2}(\phi)})^s d\nu(\phi))^{\frac{1}{s}}, \quad \text{if } s \leq 1. \tag{2.14}$$

Choosing $s = \frac{2}{m+2} = \frac{1}{r+1} < 1$ and $\nu(\cdot)$ as in (2.12), direct computation gives that

$$\text{R. H. S. of (2.14) } = 2 (\int_0^{2\pi} \frac{P_m^s(\phi)}{2} d\phi)^{\frac{1}{s}}. \tag{2.15}$$

Hence, (2.14) reduces to

$$\int_0^{2\pi} \frac{P_m(\phi)}{\psi^m(\phi)} d\phi \geq (\int_0^{2\pi} P_m^s(\phi) d\phi)^{\frac{1}{s}} \cdot 2^{1-\frac{1}{s}}. \tag{2.16}$$

If we now choose $\psi_0(\cdot)$ such that

$$\frac{P_m(\phi)}{\psi_0^{m+2}(\phi)} = c \Longleftrightarrow \psi_0(\phi) = (\frac{P_m(\phi)}{c})^{\frac{1}{m+2}}, \tag{2.17}$$

where c is a constant, then the restriction (2.10) forces

$$\begin{aligned} c &= (\frac{1}{2}\int_0^{2\pi} P_m^{\frac{2}{m+2}}(\phi)d\phi)^{\frac{m+2}{2}} \\ &= (\frac{1}{2})^{\frac{1}{s}}(\int_0^{2\pi} P_m^s(\phi)d\phi)^{\frac{1}{s}}. \end{aligned} \tag{2.18}$$

On simplification, (2.13) now reduces to

$$\begin{aligned} &\int_0^{2\pi} \frac{P_m(\phi)}{\psi_0^m(\phi)}d\phi \\ =&\int_0^{2\pi} \frac{P_m(\phi)}{(\frac{P_m(\phi)}{c})^{\frac{m}{m+2}}}d\phi \\ =&c^{1-s}\int_0^{2\pi} P_m^s(\phi)d\phi \\ \overset{(2.18)}{=}&2^{1-\frac{1}{s}}[\int_0^{2\pi} P_m^s(\phi)d\phi]^{\frac{1}{s}}, \\ =&2^{-r}[\int_0^{2\pi} P_m^{\frac{1}{r+1}}(\phi)d\phi]^{r+1}, \end{aligned} \tag{2.19}$$

which attains the R.H.S. of (2.16).

Since $s = \frac{1}{r+1}$, combining (2.18) and (2.19), one has that

$$\psi_0(\phi) = \left[\frac{2P_m^{\frac{1}{r+1}}(\phi)}{\int_0^{2\pi} P_m^{\frac{1}{r+1}}(\phi)d\phi}\right]^{\frac{1}{2}} \tag{2.20}$$

produces the contour of the required set $S^*(r)$. Thus we have the following result:

Theorem 2.5. For $p = 2$ and any fixed $r > 0$, the contour of $S^*(r)$ is given by (2.20) and the minimum value of $E[\pi_S(\underset{\sim}{X})]^{pr}$ is given by (2.19).

For $p \geq 3$, similar arguments are employed again. We omit the unnecessary details but give the corresponding result. Transforming to polar coordinates,

$$\underset{\sim}{X} = (X_1, X_2, \ldots, X_p) \longrightarrow (\rho, \phi_1, \ldots, \phi_{p-1}),$$

where $p > 0$, $0 < \phi_1, \ldots, \phi_{p-2} \leq \pi$ and $0 < \phi_{p-1} \leq 2\pi$,

$$\begin{aligned}
\text{and} \qquad X_1 &= \rho \sin\phi_1 \sin\phi_2 \cdots \sin\phi_{p-2} \sin\phi_{p-1} \\
X_2 &= \rho \sin\phi_1 \sin\phi_2 \cdots \sin\phi_{p-2} \cos\phi_{p-1} \\
&\vdots \\
X_{p-1} &= \rho \sin\phi_1 \cos\phi_2 \\
X_p &= \rho \cos\phi_1, \\
\text{one has} \qquad \rho^2 &= X_1^2 + X_2^2 + \cdots + X_p^2 \\
\text{and} \qquad |J| &= \rho^{p-1} \sin^{p-2}\phi_1 \cdots \sin\phi_{p-2},
\end{aligned}$$

where $|J|$ denotes the Jacobian of the transformation. Let $P(\rho, \phi_1 \ldots \phi_{p-1})$ denote the density of $\underset{\sim}{X}$ in terms of ρ, $\phi_1, \ldots, \phi_{p-1}$.

With $m = pr$, define

$$P_m(\phi_1, \ldots, \phi_{p-1}) = \int_0^\infty \rho^{m+p-1} P(\rho, \phi_1, \ldots, \phi_{p-1}) d\rho, \tag{2.21}$$

and

$$\psi_0(\phi_1, \ldots, \phi_{p-1}) \tag{2.22}$$
$$= \left[\frac{p P_m^{\frac{1}{r+1}}(\phi_1, \ldots, \phi_{p-1})}{\int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi P_m^{\frac{1}{r+1}}(\phi_1, \ldots, \phi_{p-1}) \sin^{p-2}\phi_1 \ldots \sin\phi_{p-2} d\phi_1 \ldots d\phi_{p-1}} \right]^{\frac{1}{p}}.$$

Therefore we have an extension of Theorem 2.5.

Theorem 2.6. For general p and for any fixed $r > 0$, the contour of $S^*(r)$ is given by (2.22), and the minimum value of $E[\pi_S(\underset{\sim}{X})]^{pr}$ equals $p^{-r} \left[\int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi P_m^{\frac{1}{r+1}} (\phi_1, \ldots, \phi_{p-1}) \right.$ $\left. \sin^{p-2}\phi_1 \ldots \sin\phi_{p-2}\ d\phi_1 \ldots d\phi_{p-1}\right]^{r+1}$.

The following example gives an illustration of these results.

Example 3. Let $\underset{\sim}{X} \sim N(\underset{\sim}{0}, tI)$, $t > 0$. Then, by definition,

$$P(\rho, \phi_1, \ldots, \phi_{p-1}) \equiv \frac{1}{(2\pi t)^{p/2}} e^{-\frac{\rho^2}{2t}}. \tag{2.23}$$

Therefore, on straightforward integration,

$$P_m(\phi_1,\ldots,\phi_{p-1}) = \frac{2^{\frac{m}{2}-1}t^{\frac{m+1}{2}}}{\pi^{p/2}}\Gamma(\frac{m+p}{2}), \tag{2.24}$$

a constant (actually, as long as it is a constant, the exact value is not important).

Using now the fact that

$$\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} \sin^{p-2}\phi_1\ldots\sin\phi_{p-2}d\phi_1\ldots d\phi_{p-1} = \frac{2\pi^{p/2}}{\Gamma(\frac{p}{2})}, \tag{2.25}$$

one gets from (2.22) that

$$\psi_0(\phi_1,\ldots,\phi_{p-1}) = \left(\frac{\Gamma(\frac{1+p}{2})}{\pi^{p/2}}\right)^{1/p}. \tag{2.26}$$

The very interesting fact is that (2.26) is a constant independent of $\phi_1,\ldots,\phi_{p-1}$, implying that for any $r > 0$, the set $S^*(r)$ is a sphere. Thus the contours of the sets we propose are high density contours in this example.

The phenomenon of the above example actually generalizes to a much broader situation. An attractive result shows that whenever the distribution of the underlying variable is such that its high density sets are mutually homothetic, the method of Theorem 2.6 reproduces a high density contour. We must caution the reader, however, that even though our methods will produce a high density set, it is not necessarily the $100(1-\alpha)\%$ high density set exactly. Separate efficiency calculations will therefore be necessary.

Theorem 2.7. Let $\underset{\sim}{X}$ be distributed as P, where P is absolutely continuous star unimodal. Suppose the high density sets of P are mutually homothetic, i.e., if f denotes the density of P, then the sets

$$\{\underset{\sim}{x}: f(\underset{\sim}{x}) \geq c\}$$

are mutual multiples of each other for different c. Then, for any fixed $r > 0$, the contour of the star-shaped set $S^*(r)$ is a high density contour.

Proof: Let S_0 denote the high density set and let S be any other star-shaped set with $\lambda(S_0) = \lambda(S) = 1$. It suffices to prove that

$$E[\pi_{S_o}(\underset{\sim}{X})]^m \leq E[\pi_S(\underset{\sim}{X})]^m, \ \forall\, m > 0.$$

Since S_0 is a high density set and $\lambda(S_0) = \lambda(S)$, we have

$$P(\underset{\sim}{X} \in S_0) \geq P(\underset{\sim}{X} \in S).$$

$$\text{Hence,} \quad P(\underset{\sim}{X} \in cS_0) \geq P(\underset{\sim}{X} \in cS); \tag{2.27}$$

this is because the high density sets of P are mutually homothetic, i.e., cS_0 is another high density set. By definition of $\pi(\cdot)$, since P is absolutely continuous,

$$P(\pi_S(\underset{\sim}{X}) \leq c) = P(\underset{\sim}{X} \in cS) \tag{2.28}$$

$$\text{and} \qquad P(\pi_{S_0}(\underset{\sim}{X}) \leq c) = P(\underset{\sim}{X} \in cS_0). \tag{2.29}$$

Combining (2.27), (2.28) and (2.29), we have

$$P(\pi_{S_0}(\underset{\sim}{X}) \leq c) \geq P(\pi_S(\underset{\sim}{X}) \leq c).$$

Since $c > 0$ is arbitrary, this means

$$\pi_{S_0}(\underset{\sim}{X}) \prec \pi_S(\underset{\sim}{X})$$

where '$\prec$' means '*stochastically less than*'. Therefore, as is well known,

$$E[\pi_{S_0}(\underset{\sim}{X})]^m \leq E[\pi_S(\underset{\sim}{X})]^m, \ \forall\, m > 0,$$

which completes the proof.

Remark 8. The common examples where high density sets are mutually homothetic are general spherically or elliptically symmetric distributions or uniform distributions on hyperrectangles etc. There are other situations in which this will be the case as well. For example, if an observable $\underset{\sim}{X}$ and a parameter $\underset{\sim}{\theta}$ have a joint elliptical distribution, the conditional (posterior) distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is elliptical too and the mutual homothetic nature of Bayes confidence sets is apparent. A case of particular interest is when $(\underset{\sim}{X}, \underset{\sim}{\theta})$ is jointly distributed as a $2p$ dimensional elliptical t; in this case, given $\underset{\sim}{\theta}$, $\underset{\sim}{X}$ is a p dimensional elliptical t with $\underset{\sim}{\theta}$ as a location parameter. Thus the interpretation of $\underset{\sim}{\theta}$ is maintained; the conditional covariance matrix of $\underset{\sim}{X}$ given $\underset{\sim}{\theta}$ is a function of $\underset{\sim}{\theta}$, however. See Muirhead (1982) for more details.

Remark 9. After the optimal contour of Theorem 2.6 is derived the actual optimal confidence set is obtained by using

$$C^*(r) \overset{def.}{=} k \cdot S^*(r),$$

where k is as in (2.9) with $S = S^*(r)$.

Finally, in this section, we point out the following rather nice property of the family of sets $C^*(r)$.

Proposition 2.8. For given α_1, α_2, $0 < \alpha_1, \alpha_2 < 1$, and any given $r > 0$, let $C_1^*(r)$ and $C_2^*(r)$ denote the optimal confidence sets for $\alpha = \alpha_1, \alpha_2$ respectively in the sense of Remark 9. Then, for a suitable constant $b > 0$, $C_2^*(r) = bC_1^*(r)$.

Remark 10. The mutual homotheticity property of the above proposition is attractive from an interpretation and communication viewpoint. The user can immediately visualize the effect of increasing or decreasing the confidence level.

Proof of Proposition 2.8: The proof is transparent on noting that $S^*(r)$ does not depend on α and on using the definition of k in (2.9).

3. AN APPLICATION TO BAYESIAN DECISION THEORY

3.1 Introductory Remarks

The theory of the preceding section makes no reference to any specific problem. Consequently, in principle, it applies in general, subject to having a star unimodal distribution.

In this section, we give an example on Bayesian set estimation of a multivariate normal mean when it has a multivariate t prior. Formally, then, consider the model

$$\left.\begin{array}{r} \underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I) \\ \underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \tau^2 I) \end{array}\right\} \tag{3.1}$$

where $\underset{\sim}{\theta}$ is the only unknown quantity and all others have the same meaning as in subsection 1.1. Extensive previous research exists on Bayesian inference about a normal mean with respect to t priors. For discussions on the general appeal of t priors and other related references, see Berger (1985).

3.2. Star Unimodality of Posteriors

The goal in this subsection is to apply the theory of section 2 for constructing Bayesian confidence sets for $\underset{\sim}{\theta}$ under model (3.1). In order that the theory be applicable, we need star unimodality of the posterior. The next few results give a complete picture for this problem. The posterior is star unimodal if it has the stronger property of logconcavity. The converse is known to be not necessarily true. The first theorem below gives the rather surprising result that for t priors as in (3.1), the posterior is logconcave for all $\underset{\sim}{X}$ if it is starunimodal for all $\underset{\sim}{X}$. In fact, we prove it for more general priors. This result is useful because checking whether the posterior is logconcave for all $\underset{\sim}{X}$ is easier than directly checking if it is star unimodal for all $\underset{\sim}{X}$. Also note that the theory of section 2 can be applied without regard of which particular $\underset{\sim}{X}$ was obtained if it is known that the posterior is star unimodal for all $\underset{\sim}{X}$.

Theorem 3.1. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I)$ and let $\underset{\sim}{\theta} \sim \pi(\frac{||\underset{\sim}{\theta} - \underset{\sim}{\mu}||^2}{\tau^2})$. Assume $\underset{\sim}{\mu}, \sigma^2, \tau^2$ are known. Suppose $\pi(\cdot)$ is twice differentiable and decreasing. Then the posterior distribution of $\underset{\sim}{\theta}$ is logconcave for all $\underset{\sim}{X}$ if and only if it is star unimodal for all $\underset{\sim}{X}$.

Proof: Clearly we only need prove the if part. We will, without loss of generality, assume $\underset{\sim}{\mu} = \underset{\sim}{0}$ and further let $\sigma^2 = \tau^2 = 1$. The case of general σ^2 and τ^2 is exactly similar. Let $\pi(\underset{\sim}{\theta}|\underset{\sim}{X})$ denote the posterior density of $\underset{\sim}{\theta}$. Clearly, $-\log \pi(\underset{\sim}{\theta}|\underset{\sim}{X})$ is proportional to $\phi_1(||\underset{\sim}{\theta}||^2) + \frac{1}{2}||\underset{\sim}{\theta} - \underset{\sim}{X}||^2$, where $\phi_1 = -\log \pi$. Suppose now $\pi(\underset{\sim}{\theta}|\underset{\sim}{X})$ is star unimodal for all

$\underset{\sim}{X}$ with mode at $\underset{\sim}{\nu} = \underset{\sim}{\nu}(\underset{\sim}{X})$. Therefore, $\underset{\sim}{Z} = \underset{\sim}{\theta} - \underset{\sim}{\nu}$ is star unimodal for all $\underset{\sim}{X}$ with mode at $\underset{\sim}{0}$. It then follows that $\forall\, t > 0,\ \forall\, \underset{\sim}{Z},\ \forall\, \underset{\sim}{X}$,

$$\begin{aligned} &\frac{d}{dt}\{\phi_1(\|t\underset{\sim}{Z} + \underset{\sim}{\nu}\|^2) + \frac{1}{2}\|t\underset{\sim}{Z} + \underset{\sim}{\nu} - \underset{\sim}{X}\|^2\} \geq 0 \\ \Longleftrightarrow\ &2\phi_1'(\|t\underset{\sim}{Z} + \underset{\sim}{\nu}\|^2)(t\|\underset{\sim}{Z}\|^2 + \underset{\sim}{\nu}'\underset{\sim}{Z}) + (t\|\underset{\sim}{Z}\|^2 + \underset{\sim}{Z}'(\underset{\sim}{\nu} - \underset{\sim}{X})) \geq 0, \\ &\forall\, t > 0\ \forall\, \underset{\sim}{Z},\ \forall\, \underset{\sim}{X}. \end{aligned} \tag{3.1}$$

Letting $t \to 0$ in (3.1), we have

$$\begin{aligned} &2\phi_1'(\|\underset{\sim}{\nu}\|^2) \cdot \underset{\sim}{\nu}'\underset{\sim}{Z} + \underset{\sim}{Z}'(\underset{\sim}{\nu} - \underset{\sim}{X}) \geq 0\ \forall\, \underset{\sim}{Z},\ \forall\, \underset{\sim}{X} \\ \Longleftrightarrow\ &(2\phi_1'(\|\underset{\sim}{\nu}\|^2) \cdot \underset{\sim}{\nu} + \underset{\sim}{\nu} - \underset{\sim}{X})'\underset{\sim}{Z} \geq 0\ \forall\, \underset{\sim}{Z},\ \forall\, \underset{\sim}{X}, \end{aligned} \tag{3.2}$$

from which it immediately follows

$$(2\phi_1'(\|\underset{\sim}{\nu}\|^2) + 1) \cdot \underset{\sim}{\nu} = \underset{\sim}{X}, \tag{3.3}$$

i.e., $\underset{\sim}{\nu} = a\underset{\sim}{X}$ for suitable a. Notice 'a' may (and will, usually) depend on $\underset{\sim}{X}$. Given $\underset{\sim}{X} \neq \underset{\sim}{0}$, '$a$' can be found from the equation

$$a(2\phi_1'(a^2\|\underset{\sim}{X}\|^2) + 1) - 1 = 0; \tag{3.4}$$

it is easy to check that $0 < a < 1$ and that $a = a(\|\underset{\sim}{X}\|)$ is continuous in $\|\underset{\sim}{X}\|$. Furthermore, $\inf_{\underset{\sim}{X}} a\|\underset{\sim}{X}\| = 0$ and $\sup_{\underset{\sim}{X}} a\|\underset{\sim}{X}\| = 1$. Substituting $a\underset{\sim}{X}$ for $\underset{\sim}{\nu}$ and $\underset{\sim}{Z} = \underset{\sim}{X}$ in (3.1), one then has

$$2\phi_1'((t + a)^2\underset{\sim}{X}'\underset{\sim}{X})(t + a) + t + a - 1 \geq 0\ \forall\, t,\ \forall\, \underset{\sim}{X}. \tag{3.5}$$

Multiplying both sides of (3.5) by $\|\underset{\sim}{X}\|$, writing ω for $(t+a)\|\underset{\sim}{X}\|$, ω_0 for $a\|\underset{\sim}{X}\|$ and letting $f(\omega) = 2\phi_1'(\omega^2)\omega + \omega$, one then obtains that given $\|\underset{\sim}{X}\| > 0$, there exists $\omega_0 > 0$ such that

$$\begin{array}{rrcll} & f(\omega) \geq f(\omega_0) & = & \|\underset{\sim}{X}\| & \text{if } \omega \geq \omega_0 \\ \text{and} & f(\omega) \leq f(\omega_0) & = & \|\underset{\sim}{X}\| & \text{if } \omega \leq \omega_0\ . \end{array}$$

For this one uses (2.1) for both $\underset{\sim}{Z} = \underset{\sim}{X}$ and $\underset{\sim}{Z} = -\underset{\sim}{X}$. Since f must then be nondecreasing, it follows that $\pi(\underset{\sim}{\theta}|\underset{\sim}{X})$ is logconcave for all $\underset{\sim}{X}$. This is because direct computations give that

$$\text{the Hessian matrix of } -\log \pi(\underset{\sim}{\theta}|\underset{\sim}{X}) \text{ is equal to } (1 + 2\phi_1'(\|\underset{\sim}{\theta}\|^2)) \cdot I + 4\phi_1''(\|\underset{\sim}{\theta}\|^2)\underset{\sim}{\theta}\underset{\sim}{\theta}',$$

which is nonnegative definite if

$$1 + 2\phi_1'(||\underset{\sim}{\theta}||^2) + 4||\underset{\sim}{\theta}||^2\phi_1''(||\underset{\sim}{\theta}||^2) \geq 0.$$

This last inequality, however, follows if $f(\omega)$ is a nondecreasing function.

Corollary 3.2. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I)$ and $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \tau^2 I)$; then the posterior of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is star unimodal for all $\underset{\sim}{X}$ if and only if $\frac{m\tau^2}{\sigma^2} \geq \frac{m+p}{8}$.

Proof: We give the proof here for the case $\sigma^2 = \tau^2 = 1$ (the proof for the general case is essentially the same). Clearly, we can assume $\underset{\sim}{\mu} = \underset{\sim}{0}$. From Theorem 3.1, the posterior is star unimodal for all $\underset{\sim}{X}$ if and only if it is logconcave for all $\underset{\sim}{X}$. The condition for logconcavity is $1 + 2\phi_1'(||\underset{\sim}{\theta}||^2) + 4||\underset{\sim}{\theta}||^2\phi_1''(||\underset{\sim}{\theta}||^2) \geq 0$ for all $\underset{\sim}{\theta}$ which reduces to $m > \frac{p}{7}$ on computation, as required.

The above corollary implies that if $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \tau^2 I)$ and if $\frac{m\tau^2}{\sigma^2} < \frac{m+p}{8}$, then there exist appropriate $\underset{\sim}{X}$ such that the posterior of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is not star unimodal. We have a complete description of this set for any given m, $\underset{\sim}{\mu}$, σ^2 and τ^2. The full proof is rather lengthy and can be obtained from the authors; however, at least an indication is necessary for completeness.

We assume $\underset{\sim}{\mu} = 0$; if $\underset{\sim}{\mu} \neq \underset{\sim}{0}$, all assertions are valid with $\underset{\sim}{X}$ replaced by $\underset{\sim}{X} - \underset{\sim}{\mu}$. We will use the following notation:

$$\frac{\underset{\sim}{\theta}}{\sigma} = \underset{\sim}{Z},\ \frac{\underset{\sim}{X}}{\sigma} = \underset{\sim}{X}_0,\ \frac{\sigma^2}{m\tau^2} = \alpha,\ m + p = \beta,\ \gamma = \frac{1}{\alpha},\ y = \frac{1}{||\underset{\sim}{X}_0||^2}.$$

Theorem 3.3. For given m, σ^2 and τ^2, the set of $\underset{\sim}{X}$ for which the posterior of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is not star unimodal is given by

$$S_m = \{\underset{\sim}{X}: a_0 < ||\underset{\sim}{X}|| < b_0\},$$

where

$$a_0 = \frac{\beta^2 - 8\gamma^2 + 20\beta\gamma - \sqrt{\beta(\beta - 8\gamma)^3}}{8\gamma}$$

and

$$b_0 = \frac{\beta^2 - 8\gamma^2 + 20\beta\gamma + \sqrt{\beta(\beta - 8\gamma)^3}}{8\gamma}; \tag{3.6}$$

whenever the quantities a, b are not well defined (i.e., if $\beta < 8\gamma$), the set S_m equals the empty set.

Discussion: Notice the very interesting aspect of Theorem 3.3 that for large values of $||\underset{\sim}{X}||$ (i.e., when the prior and the data are totally incompatible), the prior *is* star unimodal. This is essentially because of the difference in tails of normal and t distributions; if prior and data are compatible, then the posterior is star unimodal as intuition would suggest. If they are very incompatible, then only the dominant tail matters and the posterior is again star unimodal. The general assertion of Theorem 3.3 is false if the likelihood and the prior were each a t distribution; in that case, the posterior is star unimodal if and only if $||\underset{\sim}{X} - \underset{\sim}{\mu}||$ is sufficiently small.

Indication of Proof of Theorem 3.3:

A detailed proof is available in Zen (1991). The main steps are the following:

Step 1. Showing that if the posterior is star unimodal for a particular $\underset{\sim}{X}$, then it is necessarily star unimodal about

$$\underset{\sim}{\nu} = a\underset{\sim}{X}_0,$$

where a solves the cubic equation

$$h(a) = a^3 - a^2 + (\beta + \gamma)ay - \gamma y = 0 \tag{3.7}$$

Step 2. If the posterior is star unimodal, then there is a unique root in (0,1) of (3.7) and the $\underset{\sim}{X}$ for which $h(a)$ has this unique root property form the set

$$S_{1,m} = \{\underset{\sim}{X}: 4(\beta + \gamma)^3 y^2 - (\beta^2 - 8\gamma^2 + 20\beta\gamma)y + 4\gamma \geq 0\}. \tag{3.8}$$

Step 3. By definition, star unimodality also implies that the posterior density is nonincreasing as one moves away from the mode along any ray. More formally, suppose $h(a)$

has a unique root, say, a^*. Then $\underset{\sim}{\theta}$ is star unimodal if and only if $\underset{\sim}{V} = \underset{\sim}{\theta} - a^* \underset{\sim}{X}_0$ is star unimodal about $\underset{\sim}{0}$, which implies

$$\frac{d}{dt} f(t\underset{\sim}{V}|\underset{\sim}{X}) \leq 0 \; \forall\, t > 0, \; \forall\, \underset{\sim}{V}, \tag{3.9}$$

where $f(\underset{\sim}{V}|\underset{\sim}{X})$ denotes the posterior density of $\underset{\sim}{V}$. Then, on lengthy algebra, (3.9) is equivalent to

$$a^{*2}(1+\alpha\beta) + a^*(2-\alpha\beta) + 1 > 0. \tag{3.10}$$

Step 4. Consequently, the posterior of $\underset{\sim}{\theta}$ is star unimodal if and only if $\underset{\sim}{X}$ is in the set

$$S_{2,m} = \{\underset{\sim}{X} \in S_{1,m}\colon (3.10) \text{ also holds } \} \tag{3.11}$$

Step 5. $S_{1,m}$ is a set of the form

$$S_{1,m} = \{\underset{\sim}{X}\colon ||\underset{\sim}{X}|| \geq b_1 \text{ or } ||\underset{\sim}{X}|| \leq a_1\}. \tag{3.12}$$

This is trivial on using the definition of y.

Step 6. The set $S_{2,m}$ is a set of the form

$$S_{2,m} = \{\underset{\sim}{X} \in S_{1,m}\colon ||\underset{\sim}{X}|| \geq b_2 \text{ or } ||\underset{\sim}{X}|| \leq a_2\}. \tag{3.13}$$

This is not entirely trivial, but requires proving that if $\underset{\sim}{X}_i$, $i = 1, 2$ are such that for each $\underset{\sim}{X}_i$, (3.7) has a unique root $a^*(||\underset{\sim}{X}_i||)$, then $a^*(||\underset{\sim}{X}_1||) \leq a^*(||\underset{\sim}{X}_2||)$ whenever $||\underset{\sim}{X}_1|| \leq ||\underset{\sim}{X}_2||$. After that, step 6 is immediate.

Step 7. Explicit formulas for a_i, b_i, $i = 1, 2$ can be found, which on patient calculations give $a_1 = a_2 = a_0$, $b_1 = b_2 = b_0$, where a_0, b_0 are as in (3.6).

Corollary 3.4. For any given p, σ^2 and τ^2, the set S_m in (3.6) is either empty for all large m or converges to the empty set as $m \to \infty$.

Proof: If $\frac{8\tau^2}{\sigma^2} > 1$, then the first assertion holds. If $\frac{8\tau^2}{\sigma^2} \leq 1$, then the second assertion holds as can be seen on noting that the quantity a_0 defined in (3.6) goes to ∞ as $m \to \infty$.

Remark 11. This result is in a sense intuitive because the t prior for very large m essentially looks like a normal prior.

Finally, in this section, we state the following generalization of Theorem 3.1.

Theorem 3.5. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \Sigma_1)$ and let $\underset{\sim}{\theta} \sim \pi((\underset{\sim}{\theta} - \underset{\sim}{\mu})'\Sigma_2^{-1}(\underset{\sim}{\theta} - \underset{\sim}{\mu}))$. Assume π is twice differentiable and nonincreasing and $\underset{\sim}{\mu}$, Σ_1, Σ_2 are known. Then the posterior of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is logconcave for all $\underset{\sim}{X}$ if and only if it is star unimodal for all $\underset{\sim}{X}$.

Proof: See Zen (1991).

The following corollary is of particular interest.

Corollary 3.6. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \Sigma_1)$ and let $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \Sigma_2)$. Then the posterior of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is star unimodal for all $\underset{\sim}{X}$ if and only if $\lambda \min(\Sigma_1^{-1}\Sigma_2) \geq \frac{m+p}{8m}$, where $\lambda \min(\cdot)$ denotes the minimum eigenvalue.

Proof: Use of Theorem 3.5 and direct verification of the negative definiteness of the Hessian matrix for the log posterior results in the corollary.

4. Efficiency Calculations

4.1. Description of Calculations

In this section, we first briefly describe the application of the method outlined in section 2 to the Bayesian set estimation problem for the model given in (3.1) and then report and discuss the efficiencies of the obtained confidence sets; the precise definition of efficiency is given below. Without loss of generality, we can assume $\underset{\sim}{\mu} = \underset{\sim}{0}$; furthermore, since only the ratio $\frac{\tau^2}{\sigma^2}$ is important, we can also set $\sigma^2 = 1$. Thus, in the following, we will use $\underset{\sim}{\mu} = \underset{\sim}{0}$ and $\sigma^2 = 1$.

We first describe very briefly the procedure used to obtain the confidence sets. For ease of understanding, let us consider the specific case $p = 2$. Since the posterior depends on $\underset{\sim}{X}$ only through $||\underset{\sim}{X}||$, we can take all but one coordinate of $\underset{\sim}{X}$ to be 0 and the remaining one as $||\underset{\sim}{X}||$. For $p = 1$, in the notation of section 2, this results in

$$P(\rho, \phi) \propto \frac{e^{-\frac{1}{2}\rho^2+\rho||\underset{\sim}{X}||\sin\phi}}{(1+\frac{\rho^2}{n\tau^2})^{\frac{n}{2}+1}}. \tag{4.1}$$

Consequently, the function $P_m(\phi)$ (see (2.12)) is proportional to

$$\int_0^\infty \frac{\rho^{m+1}}{(1+\frac{\rho^2}{n\tau^2})^{\frac{n}{2}+1}} e^{-\frac{1}{2}\rho^2+\rho||\underset{\sim}{X}||\sin\phi} d\rho \tag{4.2}$$

For given $r > 0$, the optimal contour $\psi_0(\phi)$ (see (2.20)) is then proportional to the $\frac{1}{2(r+1)}$th power of the expression in (4.2). (4.2) was evaluated by a straightforward one dimensional numerical integration; we are not aware of any special functions that (4.2) clearly relates to. The numerical integration was easy. This was done for a grid of r values and (2.19) was evaluated by another numerical integration (on ϕ). Finally, on substituting (2.19) into (2.9), the penultimate minimization over $r > 0$ was done by a numerical search. We would like to specifically point out here that simulation from the posterior was not necessary or done in order to obtain the confidence sets.

However, simulation was necessary to evaluate the efficiencies of the sets. As stated in subsection 1.3, in order to calculate efficiencies, we shrunk the initial confidence set until it had a posterior probability of (exactly) $1 - \alpha$ and then took the pth root of the ratio of the volumes. The shrinking was done by gradually reducing the constant k in (2.9) from its initial value. The stage at which the desired $1 - \alpha$ posterior probability was reached was decided on the basis of a simulated sample from the posterior. As a matter of fact, this method of shrinking the initial confidence set until the desired content was attained, may be a good idea whenever such sequential shrinkage and simulation are not difficult.

The calculations were done over a wide range of p, m, τ^2 and various α, and $||\underset{\sim}{X}||$. Table 1 provides efficiencies in some selected cases. In the tables, $\underset{\sim}{X}'\underset{\sim}{X}$ corresponds to the three quartiles of the marginal distribution of $\underset{\sim}{X}'\underset{\sim}{X}$ for the particular given combination of p, m and τ^2. The HPD set was obtained in a (mathematically) ad-hoc way by simulating from the posterior and forming a set of sufficiently many high density points, starting from the one with the highest posterior density, and then by constructing the convex hull of these points in a visual manner. Thus, even this ad-hoc method of finding a HPD set will not work for $p > 2$. All simulation was done by using an acceptance-rejection scheme.

Table 1: Table of Efficiency

95%				$\underline{x'x}$	
p	m	τ^2	25%	50%	75%
2	1	10.00	0.929	0.910	0.895
		1.00	1.000	1.000	0.953
	3	10.00	0.910	0.922	0.914
		1.00	0.984	1.000	1.000
	5	10.00	0.902	0.910	0.910
		1.00	0.964	0.988	1.000
	30	10.00	0.884	0.887	0.891
		1.00	0.910	0.925	0.922
	70	10.00	0.872	0.869	0.865
		1.00	0.839	0.839	0.835
3	1	10.00	0.729	0.729	0.744
		1.00	0.710	0.749	0.715
	3	10.00	0.729	0.729	0.734
		1.00	0.729	0.691	0.700
	5	10.00	0.739	0.734	0.739
		1.00	0.719	0.710	0.700
	30	10.00	0.759	0.754	0.744
		1.00	0.729	0.724	0.719
	70	10.00	0.754	0.754	0.759
		1.00	0.749	0.739	0.734

4.2. Discussion of Efficiencies

We did not see any clear trend or pattern in the efficiencies. But generally speaking, the efficiencies seemed to be better for smaller p. Notice that for $p = 2$ in particular, near 100% efficiency was reported for many combinations of m, τ^2, $||\underset{\sim}{X}||$ and α.

Even when the efficiency is not very good, the information about the contour is useful knowledge. As a matter of practice, it is probably a good idea to work with a somewhat smaller value of $1 - \alpha$ than the one actually desired. The conservatism of the method will hopefully automatically take us near the desired level, while also automatically producing smaller sets. We believe this is a useful general recommendation in practice. It may also be a good idea to compute such a region and report its estimated content on the basis of simulation.

4.3. The Normal Case

It follows from Theorem 2.7 that if our method was applied to a multivariate normal distribution with (known) covariance matrix Σ, then we will always obtain an ellipsoid, oriented as high density ellipsoids are. As a general benchmark, it is interesting to evaluate the efficiencies of these ellipsoids, where efficiency is defined the same way as in subsection 4.1. In this case there is a more or less closed form formula for the efficiency. For arbitrary positive definite Σ, the efficiency can be easily proved to be equal to

$$e \doteq e(\alpha, p) = \frac{\chi_\alpha(p)}{c_\alpha(p)},$$

where $\chi^2_\alpha(p)$ denotes the 100 (1-α)th percentile of the central chi-square distribution with p degrees of freedom and

$$c_\alpha(p) = \frac{1}{\sqrt{2}} \cdot \left[\inf_{r>0} \left\{ \frac{r}{(r+1)\cdot \sqrt[r]{\alpha}} \cdot \left[\frac{\Gamma\left(\frac{p(r+1)}{2}\right)}{\Gamma(\frac{p}{2})} \right]^{\frac{1}{r}} \right\} \right]^{\frac{1}{p}} . \tag{4.3}$$

(4.3) applies to classical or Bayesian problems alike, but obviously the sets have different interpretation. An attractive feature of formula (4.3) is its nondependence on Σ. Table 2 gives some values of e for various α and p. Comparison with Table 1 shows that for all $\underset{\sim}{X}$ considered, the efficiencies there are close to the corresponding normal distribution efficiencies when the degree of freedom m is large. This is reassuring.

Table 2. Efficiency in the Normal Case

	α					
p	0.001	0.010	0.025	0.050	0.100	0.200
2	0.93403	0.91895	0.91020	0.90172	0.89098	0.87666
3	0.93777	0.92394	0.91582	0.90800	0.89778	0.88382
4	0.94053	0.92563	0.91995	0.91256	0.90285	0.88938
5	0.94273	0.93044	0.92321	0.91615	0.90690	0.89387
10	0.94985	0.93963	0.93360	0.92768	0.91985	0.90868
15	0.95414	0.94509	0.93972	0.93447	0.92750	0.91752
20	0.95721	0.94892	0.94404	0.93922	0.93286	0.92371

5. An Application to Classical Invariant Estimation

5.1 Derivation of the Confidence Set

In this section, we consider the example of constructing a classical invariant confidence set for $\underset{\sim}{\theta}$ when $\underset{\sim}{Z} = \underset{\sim}{X} - \underset{\sim}{\theta}$ has the density given in (1.2). Motivation was discussed at length in subsection 1.1. So we will only describe the work involved in the construction of the confidence set. Since the work, in spirit, is the same as that in the Bayesian example of section 4, we will keep details to a minimum; only the case $p = 2$ will be illustrated and we will assume (actually with no loss of generality) that

$$\Sigma^{-1} = \text{ diag } (d_1^2, d_2^2).$$

Easy algebra and integration gives that for any $r > 0$, the optimal contour corresponding to (2.20) is proportional to

$$\left(\lambda + \frac{(1-\lambda)d_1 d_2}{(d_1^2 \cos^2\phi + d_2^2 \sin^2\phi)^{1+r}}\right)^{\frac{1}{2(1+r)}} \tag{5.1}$$

It is easily seen that optimizing over r corresponds to minimizing

$$(\Gamma(1+r))^{\frac{1}{2r}} \cdot \left[\int_0^{2\pi} \left\{\lambda + \frac{(1-\lambda)d_1 d_2}{(d_1^2 \cos^2\phi + d_2^2 \sin^2\phi)^{1+r}}\right\}^{\frac{1}{1+r}} d\phi\right]^{\frac{1+r}{2r}}$$
$$\times \left(\frac{r}{(r+1)\cdot \sqrt[r]{\alpha}}\right)^{\frac{1}{2}}, \tag{5.2}$$

over $r > 0$.

For any given λ, d_1, d_2 and α, this can be done by a numerical integration followed by a search. Usually, the minimum is at some $r_0 > 0$ (as opposed to $r_0 = 0$). Henceforth, when we refer to r, it will be understood that we mean $r = r_0$. Once the optimal set corresponding to (2.9) is found, $\underset{\sim}{X}$ is added to it, thereby giving invariant (family of) sets.

5.2. Properties of the Set

1. Formula (5.1) in a sense says that the eventual confidence set has a spherical component and also an elliptical component. This is nice from a visual as well as communication point of view. The actual high density sets do not have an obvious interpretation such as this.

2. Suppose, without loss of generality, $d_2^2 > d_1^2$ (if $d_2^2 = d_1^2$, the problem is somewhat uninteresting). This means Var $X_1 >$ Var X_2. One may, therefore, like to see a confidence set stretched more in the direction of θ_1 than θ_2. Examination of (5.1) shows that this is indeed the case, on forcing $\phi = 0$ and π respectively.

3. Examination of (5.1) also gives it is periodic with a period of π. This implies that the confidence set is symmetric in each direction. Again, this is probably a desirable property.

4. In fact, examination of (5.1) gives the following stronger property: the set is most stretched along the direction θ_1 and the amount of stretch decreases monotonically as one approaches the θ_2 direction; and in this direction, the set is the least stretched. This is because (5.1) is monotone in ϕ.

5.3. A Specific Case

Figures 1 through 4 give plots of the actual 95% optimal confidence set in the above example, corresponding to $d_1 = 2.5$, $d_2 = 7.5$, and $\lambda = 0$, .25, .5, and 1 respectively. Note the very interesting transition of the confidence set from a circle to an ellipse. In each picture, the value of $r = r_0$ for which (5.2) is minimized is given.

Acknowledgement

We are deeply grateful to Jim Berger, Larry Brown, Morris Eaton, Jiunn Hwang, Glen Meeden, Friedrich Pukelsheim, Bill Studden and Jim Zidek for many interesting comments. S. Mukhopadhyay helped very kindly with the computing. We are glad to thank all of them. Joop Kemperman very graciously pointed to an error in the example of section 5.3.

References

1. Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6**, 170–176.

2. Berger, J. (1985). Statistical decision theory and Bayesian analysis. Springer Verlag, New York.

3. Bickel, P. J. and Krieger, A. M. (1989). Extensions of Chebychev's inequality with applications. Preprint.

4. Box, G. E. P. and Tiao, G. C. (1973). Bayesian inference in statistical analysis. Addison Wesley, Reading, Mass.

5. Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37**, 1087–1136.

6. Camp, B. H. (1922). A new generalization of Tchebycheff's statistical inequality. *Bull. Amer. Math. Soc.* **28**, 427–432.

7. Cohen, A. and Strawderman, W. (1973). Admissible confidence interval and point estimation for translation or scale parameters. *Ann. Statist.* **1**, 545–550.

8. DasGupta, A. (1991). Diameter and volume minimizing confidence sets in Bayes and classical problems. *Ann. Statist.* **19**, 1225–1243.

9. DasGupta, S. (1980). Brunn-Minkowski inequality and its aftermath. *J. Mult. Analysis* **10**, 296–318.

10. Dharmadhikari, S. and Joag-Dev, K. Unimodality, convexity and applications. Academic Press, New York.

11. Eaton, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.*, 11–43.

12. Ferguson, T. S. (1973). Mathematical Statistics: a decision theoretic approach. Academic Press, New York.

13. Hooper, P. M. (1982). Invariant confidence sets with smallest expected measure. *Ann. Statist.* **10**, 1283–1294.

14. Hwang, J. T. and Casella, G. (1982). Minimax confidence sets for the mean of a normal distribution. *Ann. Statist.* **10**, 868–881.

15. Joshi, V. M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38**, 1868–1875.

16. Kantner, M. (1977). Unimodality and dominance for symmetric random vectors. *Trans. Amer. Math. Soc.* **229**, 65–85.

17. Lehmann, E. L. (1983). Theory of point estimation. John Wiley, New York.

18. Marshall, A. W. and Olkin, I. (1979). Inequalities: the theory of majorization with applications to combinatorics, probability, statistics and matrix theory. Academic Press, New York.

19. Meidell, B. (1922). Sur un probleme du calcul des probabilities et les statistiques mathematiques. *C. R. Acad. Sci. Paris*, t, **137**, 806–808.

20. Mudholkar, G. S. (1966). The integral of an invariant unimodal function over an invariant convex set — an inequality and applications. *Proc. Amer. Math. Soc.* **17**, 299–306.

21. Muirhead, R. (1982). Aspects of multivariate statistical theory. John Wiley, New York.

22. Naiman, D. Q. (1984). Optimal simultaneous confidence bounds. *Ann. Statist.* **12**, 702–715.

23. Olshen, R. A. and Savage, L. J. (1970). A generalized unimodality. *J. Appl. Prob.* **7**, 21–34.

24. Tong, Y. L. (1980). Probability inequalities in multivariate distributions. Academic Press, New York.

25. Wells, D. R. (1978). A result in multivariate unimodality. *Ann. Statist.* **6**, 926-931.

26. Zen, M. M. (1991). Bayesian inference with nonconjugate priors. Ph.D. thesis, Dept. of Statistics, Purdue University.

Figure 1

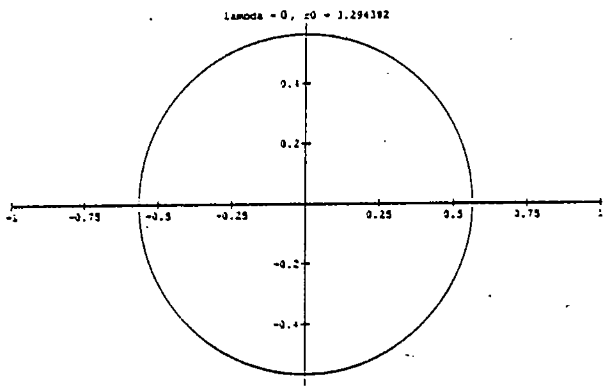

Figure 2

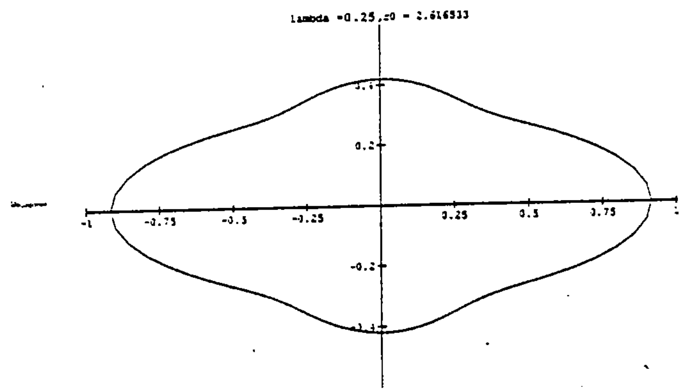

Figure 3

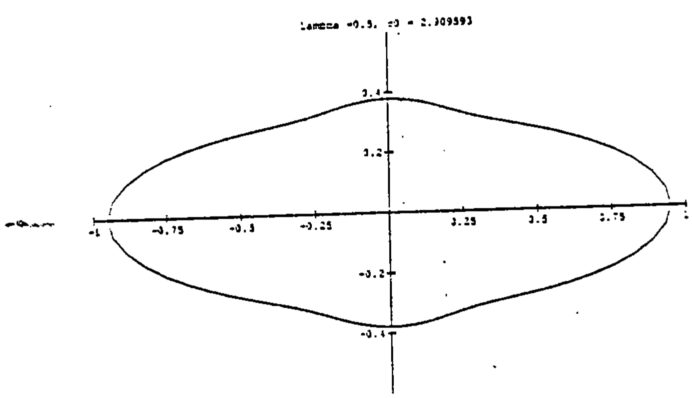

Figure 4