DIAMETER AND VOLUME MINIMIZING CONFIDENCE SETS IN BAYES AND CLASSICAL PROBLEMS

by

Anirban DasGupta*
Purdue University

Technical Report #88-63C

Department of Statistics
Purdue University

December 1988
Revised May 1990

---

*Research supported by NSF Grant DMS–87–02620

△

Abstract

If $X \sim P_\theta$, $\theta \varepsilon \Omega$ and $\theta \sim G << \mu$ where $\frac{dG}{d\mu}$ belongs to the convex family $\Gamma_{L,U} = \{g: L \leq cg \leq U$, for some $c > 0\}$, then the sets minimizing $\lambda(S)$ subject to $\inf_{G \varepsilon \Gamma_{L,U}} P_G(S|X) \geq p$ are derived, where $P_G(S|X)$ is the posterior probability of S under the prior G, and λ is any nonnegative measure on Ω such that $\mu << \lambda << \mu$. Applications are shown to several multiparameter problems and connectedness (or disconnectedness) of these sets is considered. The problem of minimizing the diameter is also considered in a general probabilistic framework. It is proved that if $\mathcal{X}$ is any finite dimensional Banach space with a convex norm, and $\{P_\alpha\}$ is a tight family of probability measures on the Borel σ–algebra of $\mathcal{X}$, then there always exists a closed connected set minimizing the diameter under the restriction $\inf_\alpha P_\alpha(S) \geq p$. It is also proved that if P is a spherical unimodal measure on R^m, then volume (Lebesgue measure) and diameter minimizing sets are the same. A result of Borell (1975) is then used to conclude that diameter minimizing sets are spheres whenever the underlying distribution P is symmetric absolutely continuous and the density f is such that $f^{-1/m}$ is convex. All standard symmetric multivariate densities satisfy this condition. Applications are made to several Bayes and classical problems and admissibility implications of these results are discussed.

Key words: Lebesgue measure, diameter, posteriors, equivariant, confidence set, spheres, Steiner symmetrization, Banach space, convex, connected, Minkowski sum, unimodal

1. Introduction. One of the most widely studied problems in statistical theory is the construction of a good confidence set for a (vector) parameter θ. In the classical framework where one has X distributed as P_θ, an often adopted criterion is to minimize the probability of covering false values subject to having a large enough probability of covering the true value. Formally, one wants to find a family of sets $\{S(x)\}$ minimizing $P_\theta(S(X) \ni \theta')$ for $\theta \neq \theta'$ subject to a restriction that $P_\theta(S(X) \ni \theta) \geq p$ for every θ. Such sets have been called uniformly most accurate (UMA) confidence sets in the literature. It is well known, however, that even in problems with some kind of invariance, a UMA confidence set often does not exist even in the class of invariant confidence sets. See Lehmann (1986) for examples. Instead, one often considers the problem of finding the smallest set subject to a restriction of the form $P_\theta(S(X) \ni \theta) \geq p$ for every θ. Typically, the size of a confidence set S is measured by considering $\lambda(S)$ where λ is a nonnegative measure on the parameter space; in problems with a continuous parameter θ, λ is usually taken as the Lebesgue measure while in the case of a discrete parameter, λ may be taken as the counting measure. For specific results and general exposition, see Brown (1966), Cohen and Strawderman (1973), Hwang and Casella (1982), Joshi (1967), Naiman (1986), Wijsman (1979), Wynn (1984) and Lehmann (1986). Analogous problems are of interest in the Bayesian domain also. Thus if $X \sim P_\theta$ and θ has a prior G belonging to a family Γ, one may consider the problem of minimizing $\lambda(S)$ among all S such that $\inf_G P_G(\theta \varepsilon S|X) \geq p$, where $P_G(\theta \varepsilon S|X)$ denotes the posterior probability of the set S when the prior is G. Note that in the case of a single prior, Γ is a singleton set. Various authors have considered Bayesian set estimation problems: see Berger (1985), Box and Tiao (1973), Lehmann (1986), Diaconis and Freedman (1986), Pratt (1961), etc. In this article we mostly focus attention on this latter robust Bayes problem when the priors G are assumed to be absolutely continuous with respect to some measure μ on $\Omega \subseteq \mathrm{R}^m$ and the density $g = \frac{dG}{d\mu}$ belongs to the convex family

$$\Gamma_{L,U} = \{g : \text{ for some } c > 0,\ L(\theta) \leq cg(\theta) \leq U(\theta) \ \ \forall \theta\}, \tag{1.1}$$

where L and U are fixed nonnegative functions. The density $g(\theta)(\int g d\mu = 1)$ is thus chosen such that it is proportional to some function between $L(\theta)$ and $U(\theta)$. The family (1.1) was first considered in DeRobertis (1978). In the case when $U = kL$ for some

$k > 1$, $\Gamma_{L,U}$ becomes a metric neighborhood of L; in this case one may interpret $\Gamma_{L,kL}$ as a neighborhood of a subjectively elicited (proper or improper) prior density L. For a comprehensive discussion of $\Gamma_{L,U}$ and a variety of results, also see Berger (1987), DasGupta and Studden (1988a,b), DeRobertis and Hartigan (1981) etc. We will merely mention here that the family $\Gamma_{L,U}$ has many attractive properties and is one of the few mathematically tractable classes of priors considered in the literature.

The approach of minimizing the measure of a set subject to a restriction on its probability content can in general lead to disconnected sets. Indeed we will see a very natural example in the scenario of the above mentioned robust Bayes problem where the Lebesgue measure minimizing confidence set is disconnected. While we are not suggesting that a disconnected confidence set is necessarily undesirable, we believe that a connected set is more attractive from the viewpoint of effectively communicating where the unknown parameter lies. See Example 8 in section 3 and section 4 for further discussion on this issue. An alternative possibility, apparently not studied in the statistical literature, is instead to minimize the diameter of the confidence set in a suitable metric subject to a restriction on the set's probability content. Since the Lebesgue measure of an arbitrary Borel set in $\mathbb{R}^m$ with diameter d cannot exceed that of a sphere with the same diameter (see section 3), controlling the diameter should keep the Lebesgue measure in control although the converse assertion is badly false. In addition, it will be seen that the potential disconnectedness problem disappears if one minimizes the diameter as opposed to some nonnegative measure.

In section 2, we state and prove a general theorem on the existence and the form of the set S that minimizes $\lambda(S)$ subject to $\inf_{G \epsilon \Gamma_{L,U}} P_G(S|X) \geq p$, where λ is a nonnegative measure and $\Gamma_{L,U}$ is defined in (1.1). Several applications of this theorem are then shown to standard multiparameter estimation problems and also to a problem on construction of smallest width (average or maximum) linear Bayesian bands for the true regression line in the case of simple linear regression. The important practical issue of the convexity (or connectedness) of these sets is considered and in important specific examples we point out under what conditions these sets are connected. Posterior confidence sets for a normal mean θ when θ has a t prior are considered in good detail.

In section 3, we address the problem of minimizing the diameter. Without explicitly differentiating between Bayesian and non-Bayesian problems, we consider the diameter minimization problem in a general probabilistic framework and then show that our results apply to Bayes as well as classical problems. It is proved that if $\mathcal{X}$ is a finite dimensional Banach space with a convex norm $\|\cdot\|$, and if $\{P_\alpha, \alpha\varepsilon I\}$ is a tight family of probability measures, then there always exists a closed convex set S minimizing the diameter subject to the restriction $\inf_{\alpha\varepsilon I} P_\alpha(S) \geq p$. We next address the question whether the two approaches of minimizing a measure and minimizing the diameter ever lead to a common answer. We prove in section 3 that if $\{P_\alpha\}$ consists of a single spherically symmetric unimodal measure P on R^m, then the volume and diameter minimizing sets are both spheres. This result, although intuitive, is not immediate because an arbitrary bounded set A in R^m cannot be enclosed in a sphere of the same diameter as that of A. Indeed one has to use the Steiner symmetrizations of Borel sets to achieve this result for a spherical P. We give applications of these results to the robust Bayes problem of section 2 and also point out some applications to classical set estimation problems. We also give a one dimensional example where the two approaches do not lead to the same answer; the optimal sets for both approaches are derived and some general comparison is made.

Intuitively, one would expect diameter minimizing sets to be spherical under more general conditions. For example, it is natural to ask what would be the shape of diameter minimizing confidence sets when the underlying distributions are elliptical normal. We have a general theorem geared to handling location problems. We have shown that if P is an absolutely continuous probability measure on R^m and if the density f is symmetric about some vector $\underset{\sim}{\mu}$ and if moreover $f^{-\frac{1}{m}}$ is convex, then sets minimizing the diameter subject to a restriction on the probability content are spheres. Some results of Borell (1975) and Prekopa (1971) are useful in proving this theorem. This result, for example, implies that for all standard location problems (including elliptical normal, t, and the double exponential) spheres are diameter minimizing among equivariant sets. Similarly, if X has a density $f_\theta(x)$ satisfying the above properties and θ has a prior G belonging to

$\Gamma_{L,kL}$ where L is uniform, then diameter minimizing robust Bayesian sets are also spheres. Finally, in section 4 we briefly discuss the admissibility implications of the results of section 3 in classical set estimation problems. We hope that the results of our present article will find applications to other Bayesian as well as classical problems.

2. Smallest measure robust Bayes sets. In this section we consider the problem of finding size minimizing robust Bayes confidence sets when size is measured in terms of a nonnegative measure λ and the family of priors considered is $\Gamma_{L,U}$ defined in (1.1). To our knowledge this problem has not been explicitly treated for any other classes of priors in the literature although related point estimation results have been obtained by many authors.

Theorem 2.1. Let $X \sim P_\theta << \nu$ for some σ-finite measure ν on a sample space $\mathcal{X}$. Let $f_\theta(x) = \frac{dP_\theta}{d\nu}$. Assume $\theta \varepsilon \Omega$ has a prior $G << \mu$ where $\frac{dG}{d\mu}$ belongs to $\Gamma_{L,U}$. Let $M_2 = M_2(x) = \int_\Omega fU d\mu$ and assume $M_2 < \infty$. Let λ be any nonnegative measure on Ω such that λ and μ are mutually absolutely continuous. Define a measure Q on Ω as $dQ = \{(1-p)fL + pfU\}d\mu$ where $0 < p < 1$ is a fixed number. Let $S = \{\theta \varepsilon \Omega : \frac{dQ}{d\lambda} \geq c\}$ where c is such that $Q(S) = pM_2$ (we assume such a c exists). Then S minimizes $\lambda(A)$ among all A such that $\inf_{G\varepsilon\Gamma_{L,U}} P_G(A|x) \geq p$.

Proof: It is easy to prove that for any A,

$$\begin{aligned}
&\inf_{G\varepsilon\Gamma_{L,U}} P_G(A|x) = \frac{\int_A fL d\mu}{\int_A fL d\mu + \int_{A^c} fU d\mu}. \\
&\therefore \inf_{G\varepsilon\Gamma_{L,U}} P_G(A|x) \geq p \\
&\iff \int_A \{(1-p)fL + pfU\} d\mu \geq pM_2 \\
&\iff Q(A) \geq pM_2.
\end{aligned} \tag{2.1}$$

The theorem now follows from the Neyman–Pearson Lemma on noting that $Q << \mu << \lambda$ and hence $\frac{dQ}{d\lambda}$ exists.

Remark. A case of special interest is when $U = kL$ for some $k > 1$. In this case, it follows from Theorem 2.1 that the set S minimizing $\lambda(A)$ (where λ denotes Lebesgue measure) among A such that $P_G(A|x) \geq p$ for all $G\varepsilon\Gamma_{L,kL}$ satisfies $p^* = P_L(S|x) = \frac{kp}{kp+(1-p)} > p$.

Roughly speaking, this means that if one starts with a subjectively elicited L and considers a neighborhood $\Gamma_{L,kL}$, then to find the smallest volume posterior set with a minimum posterior probability of p for all priors in $\Gamma_{L,kL}$, one merely has to find the smallest volume set S with respect to L, but under L this S must have a posterior probability somewhat larger than p. We find this connection between robust Bayes sets and sets optimal with respect to the initial prior L quite surprising. The value of p^* for various p and k are given below. Note that p^* does not depend on L.

Table 1

p \ k	2	3	4	5	6	8	10
.5	.67	.75	.8	.83	.86	.89	.91
.75	.86	.9	.92	.94	.95	.96	.97
.9	.95	.96	.97	.98	.98	.99	.99
.95	.97	.98	.99	.99	.99	.99	.99

Thus, for example, to construct a robust 90% set with $k = 4$, one has to construct a 97% set with respect to L. Note the marginal diminishing effect of increasing k.

We now give several applications of Theorem 2.1 to illustrate how it works in practice and also to make the point that the result will apply to a wide variety of situations.

Example 1. Consider the canonical linear model where $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, I)$ and $\underset{\sim}{\theta}$ has a prior in $\Gamma_{L,kL}$ where L is the density of a $N_m(\underset{\sim}{\mu}, \Sigma)$ distribution where $\underset{\sim}{\mu}, \Sigma$ are assumed known. Notice that with such a choice of L, $\Gamma_{L,kL}$ does not include other multivariate normal priors. The results of this example, therefore, will not apply to situations where simultaneous inclusion of many normal priors is considered desirable. Suppose we want to find the set with the smallest Lebesgue measure such that for all priors in $\Gamma_{L,kL}$, the posterior probability exceeds p. Then from Theorem 2.1 we at once have that S is the ellipsoid

$$S = \left\{\underset{\sim}{\theta} \colon (\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}})' D^{-1} (\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}) \leq \chi^2_{p^*}(m)\right\},$$

where $\hat{\underset{\sim}{\theta}} = (X'X + \Sigma^{-1})^{-1}(X'\underset{\sim}{y} + \Sigma^{-1}\underset{\sim}{\mu})$, $D = (X'X + \Sigma^{-1})^{-1}$, $p^* = \frac{kp}{kp+(1-p)}$, and $\chi^2_p(m)$ is the pth percentile of a central chi–square distribution with m degrees of freedom. Thus, in this case, the volume minimizing sets are convex and hence connected. The volume of S equals $c_m(\chi^2_{p^*}(m))^{\frac{m}{2}}$ where $c_m = \frac{\pi^{\frac{m}{2}}}{\Gamma(1+\frac{m}{2})}|D|^{\frac{1}{2}}$. Table 2 below gives the values of $\chi^2_{p^*}(m)$ when $m = 2$ for various combinations of p and k. Again note the marginal diminishing effect of increasing k. For example, there is virtually no increase in the volume of S if one increases k from 8 to 10 and $p = .9$.

Table 2

	k							
p	1	2	3	4	5	6	8	10
.5	1.39	2.22	2.77	3.22	3.54	3.93	4.41	4.82
.75	2.77	3.93	4.61	5.05	5.63	5.99	6.44	7.01
.9	4.61	5.99	6.44	7.01	7.82	7.82	9.21	9.21
.95	6.44	7.82	9.21	9.21	9.21	9.21	9.21	9.21

Example 2. Suppose $X_i \overset{\text{indep}}{\sim}$ Gamma (α_i, θ_i) where $\{\alpha_i\}$ are the known shape parameters and $\{\theta_i\}$ are the scale parameters, $1 \le i \le m$. Let $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_m)'$ have a prior belonging to $\Gamma_{L,kL}$ where L is a product Gamma prior (i.e., θ_i are independent with density $e^{-r_i\theta_i}\theta_i^{\beta_i}$ under L). Using an easy concavity argument it follows that S is convex in this case if $\alpha_i + \beta_i \ge 1$. Since $\{\theta_i\}$ are scale parameters, one may also take λ as the measure with density (with respect to Lebesgue) $p(\underset{\sim}{v}) = \frac{1}{\prod_{i=1}^m v_i}$. The general shape of the required set S remains unchanged and it is convex for all values of α_i and $\beta_i > 0$.

Example 3. We now give an application of Theorem 2.1 to the construction of linear confidence bands for the true regression line in simple linear regression. Suppose $y \sim N(\theta_0 + \theta_1 x,\ I)$ where x is within the bounds $a \le x \le b$; without loss of generality, one may assume $a = 0$ and $b = 1$. As usual, we assume there are n pairs of observations (x_i, y_i), $1 \le i \le n$, and suppose $\underset{\sim}{\theta} = (\theta_0, \theta_1)'$ has a prior belonging to $\Gamma_{L,kL}$ where L is a $N(\underset{\sim}{\mu}, \Sigma)$ density. We want to find linear bands $f(x) = a_1 + b_1 x$ and $g(x) = a_2 + b_2 x$ such that $P_G[f(x) \le \theta_0 + \theta_1 x \le g(x) \forall x \varepsilon [0,1] | y] \ge p$ for all G in $\Gamma_{L,kL}$ and the width of the band is minimized in some sense; for example, one may minimize either the average width $= \int \{g(x) - f(x)\} dW(x)$ for some measure W or the maximum width $\max_{0 \le x \le 1}\{g(x) - f(x)\}$.

Observe now that by the remark following Theorem 2.1,

$$\begin{aligned} &P_G\left[\bigcap_{0\le x\le 1}\{f(x)\le\theta_0+\theta_1 x\le g(x)\}|y\right]\ge p \text{ for all } G,\\ &\Longleftrightarrow P_L\left[a_1\le\theta_0\le a_2,\ a_1+b_1\le\theta_0+\theta_1\le a_2+b_2|y\right]\\ &\ge p^*=\frac{kp}{kp+(1-p)}. \end{aligned}\tag{2.2}$$

Now under L, $\underset{\sim}{Z}=(Z_1,Z_2)'$ where $Z_1=\theta_0$ and $Z_2=\theta_0+\theta_1$ has a $N(\underset{\sim}{\nu},R)$ distribution where

$$\underset{\sim}{\nu}=A(X'X+\Sigma^{-1})^{-1}(X'y+\Sigma^{-1}\mu)$$
$$\text{and } R=A(X'X+\Sigma^{-1})^{-1}A'$$

where $A=\begin{pmatrix}1&0\\1&1\end{pmatrix}$.

Furthermore, $\int\{g(x)-f(x)\}dW(x)$

$$=(a_2-a_1)+\lambda(b_2-b_1)\tag{2.3}$$

where $\lambda=\int x dW(x)$ and W is a probability measure. From (2.2) and (2.3) it is then clear that we want to find a rectangle $S=[c_1,c_2]\times[d_1,d_2]$ such that under the $N(\underset{\sim}{\nu},R)$ distribution this rectangle has a probability p^* and such that among all rectangles with this property it has the smallest weighted perimeter $(1-\lambda)(c_2-c_1)+\lambda(d_2-d_1)$. From Anderson's theorem (see Anderson (1955)) it follows that S must be centered at $\underset{\sim}{\nu}$, i.e., if $\underset{\sim}{\nu}=(\nu_1,\nu_2)'$, then $|c_1-\nu_1|=|c_2-\nu_1|$ and $|d_1-\nu_2|=|d_2-\nu_2|$. Once this simplification is made, minimizing the weighted perimeter subject to a restriction on its probability becomes a routine calculus problem. See DasGupta (1988) for these details. Figure 1 gives a plot of the optimal 60% band for $k=1.7, 2.5$, and 3.25 when W is the uniform distribution, $n=9$, 3 observations are taken at each of $x=0, .5$, and 1, $\underset{\sim}{\mu}=\begin{pmatrix}0\\1\end{pmatrix}$, $\Sigma=\begin{pmatrix}\frac{1}{12}&0\\0&4\end{pmatrix}$ and the least squares estimate equals $(.5,1.5)'$. Note that the same bands would have been obtained for any other probability measure W with a mean of $\frac{1}{2}$. For the case of minimizing the maximum width, there is the additional simplification that S must be a square. Figure 2 gives a plot of the optimal band in this case. Note that the bands minimizing the maximum width will always be parallel. Before closing this example, we will like to point out that our techniques given here will not apply to the problem of finding more general nonlinear bands with small widths.

Example 4. We now give a common example where the volume minimizing set is not necessarily connected. Suppose $X \sim N(\theta, \sigma^2)$ and θ has a prior belonging to $\Gamma_{L,kL}$ where L is the density of a central t distribution with m degrees of freedom, median μ, and scale parameter τ^2. From theorem 2.1 it follows that for each k and p, the required volume (i.e., Lebesgue measure) minimizing sets are of the form $S = \{\theta : \pi_L(\theta|x) \geq c\}$ where $\pi_L(\theta|x)$ denotes the posterior density of θ under L.

From definition, we have

$$\pi_L(\theta|x) = \text{ constant } \frac{e^{-\frac{1}{2\sigma^2}(\theta-x)^2}}{\left(1+\frac{\theta^2}{m\tau^2}\right)^{\frac{m+1}{2}}}. \tag{2.4}$$

We will prove that for suitable $x, \pi_L(\theta|x)$ is bimodal and hence the Lebesgue measure minimizing set would obviously be disconnected for appropriate choices of p and k. It is easy to see that as $\theta \to \pm\infty, \pi_L(\theta|x)$ goes to zero in a monotone way. Thus $\pi_L(\theta|x)$ will be bimodal if and only if its derivative (or equivalently the derivative of $-\log \pi_L(\theta|x)$) has three real zeros. Now,

$$g(\theta) = -\log \pi_L(\theta|x) = \text{ Constant } + \frac{1}{2\sigma^2}(\theta - x)^2 + \frac{m+1}{2}\log\left(1+\frac{\theta^2}{m\tau^2}\right).$$

$$\therefore g'(\theta) = \frac{\theta - x}{\sigma^2} + \frac{(m+1)\theta}{m\tau^2(1+\frac{\theta^2}{m\tau^2})} = 0$$

$$\iff \omega^3 - z\omega^2 + (\alpha + c)\omega - cz = 0, \tag{2.5}$$

where $\omega = \frac{\theta}{\sigma}, z = \frac{x}{\sigma}, c = \frac{m\tau^2}{\sigma^2}$, and $\alpha = m + 1$.

$$\text{Defining } Q = \frac{z^2 - 3(\alpha + c)}{9},$$

$$\text{and } R = \frac{-2z^3 + 9(\alpha + c)z - 27cz}{54},$$

it is well known that (2.5) has three real roots if and only if $Q^3 - R^2 \geq 0$, which on algebra reduces to

$$4cz^4 - [\alpha^2 + 20\alpha c - 8c^2]z^2 + 4(c + \alpha)^3 \leq 0. \tag{2.6}$$

We want to know if there exists z such that the inequality in (2.6) holds. It is easy to see that there do not exist any z giving the required inequality in (2.6) if

$$(\alpha^2 + 20\alpha c - 8c^2)^2 < 64c(c+\alpha)^3, \tag{2.7}$$

which is equivalent to $\alpha < 8c \Leftrightarrow \frac{m\tau^2}{\sigma^2} > \frac{m+1}{8}$. We have thus proved that $\pi_L(\theta|x)$ is unimodal for all x and hence the Lebesgue measure minimizing sets are connected for all x if $\frac{m\tau^2}{\sigma^2} > \frac{m+1}{8}$. This, for example, implies that if $X \sim N(\theta, 1)$ and θ has a standard Cauchy prior, then the posterior is in fact unimodal for every x. If, however, $\frac{m\tau^2}{\sigma^2} \leq \frac{m+1}{8}$, then the required inequality in (2.6) holds for all values of z^2 in the range

$$\frac{\alpha^2 + 20\alpha c - 8c^2 - \sqrt{\alpha(\alpha - 8c)^3}}{8c} \leq z^2 \leq \frac{\alpha^2 + 20\alpha c - 8c^2 + \sqrt{\alpha(\alpha - 8c)^3}}{8c}. \tag{2.8}$$

Hence, for $|x|$ in an interval, the Lebesgue measure minimizing sets will be disconnected for suitable k and p. We mention here without proof that analogous characterizations for the estimation of a multivariate normal mean with general spherically symmetric priors have recently been obtained in DasGupta, Ghosh and Zen (1990). Figures 3 through 6 give plots of the posterior density of $\underset{\sim}{\theta}_{2\times 1}$ in some cases where the posterior has more than peak. Clearly, volume minimizing sets are not necessarily connected in such cases.

3. Diameter minimizing sets. In this section we focus attention on the problem of minimizing the diameter subject to a restriction on the probability content of the set. It will be seen that the problem of disconnectedness disappears in very general setups by minimizing the diameter. Without constraining ourselves to specific statistical problems, we first prove an existence theorem in a general probabilistic framework that will apply simultaneously to many Bayesian as well as classical problems of interest.

Theorem 3.1. Let $(\mathcal{X}, \mathcal{B})$ be a finite dimensional Banach space with a convex norm $||\cdot||$, where $\mathcal{B}$ is the usual Borel σ-field on $\mathcal{X}$. For $A \subseteq \mathcal{X}$, let $d(A) = \sup_{x,y \epsilon A} ||x - y||$. Let $\{P_\alpha, \alpha \epsilon I\}$ be a tight family of probability measures on $\mathcal{X}$. Let $\mathcal{F} = \{A \epsilon \mathcal{B} : \inf_{\alpha \epsilon I} P_\alpha(A) \geq p\}$ where $0 < p < 1$ is a fixed number. Then there exists a closed convex set $S \epsilon \mathcal{F}$ such that $d(S) = \inf_{A \epsilon \mathcal{F}} d(A)$.

Proof: Observe that we only need consider closed sets A such that $\inf_{\alpha\epsilon I} P_\alpha(A) \geq p$. We claim that we can moreover assume without loss of generality that there is a common compact set K_1 such that $A \subseteq K_1$. Towards this end, by using the tightness of $\{P_\alpha\}$ first find a compact set F such that $\inf_\alpha P_\alpha(F) \geq p$; let d denote the diameter of F. Next let x_0 be any point in $\mathcal{X}$. Let $K_0 = S(x_0, r)$ be a closed sphere with center at x_0 and radius r such that $\inf_\alpha P_\alpha(K_0) > 1-p$. Also let K_1 denote the closed sphere S $(x_0,\ r+2d)$ with center at x_0 and radius $r+2d$. We claim that if A is any set such that $\inf P_\alpha(A) \geq p$ and $d(A) \leq d$, then A must be contained in K_1. For if $A \cap K_1^c$ is not empty, then because of the fact that $d(A) \leq d$, we must have $A \cap K_0 = \emptyset$, implying that $P_\alpha(A) \leq P_\alpha(K_0^c) < p$ for every α. This is a contradiction to the hypothesis that $\inf P_\alpha(A) \geq p$.

Define now $\mathcal{F}_0 = \{A : A \text{ is closed, } A \subseteq K_1\}$. Since $\mathcal{X}$ is a finite dimensional Banach space, the open mapping theorem implies that the closed sphere K_1 is compact. The proof of the theorem will now consist of exhibiting a topology on $\mathcal{F}_0$ such that $d(A)$ is continuous and

$$\mathcal{F}_1 = \{A\epsilon\mathcal{F}_0 : \inf_\alpha P_\alpha(A) \geq p\}$$

is compact with respect to this topology. For this purpose, we metrize $\mathcal{F}_0$ by the Hausdorff metric

$$\rho(A,B) = \max\{\sup_{x\epsilon A} d(x,B), \sup_{y\epsilon B} d(y,A)\}, \text{ where } d(x,B) = \inf_{y\epsilon B} ||x-y||.$$

Notice that $\rho(A,B) < \epsilon$ implies that for each x in A, there exists $y\epsilon B$ such that $||x-y|| < \epsilon$ and for each y in B there exists $x\epsilon A$ such that $||x-y|| < \epsilon$. It is clear therefore that $d(A)$ is continuous with respect to the metric ρ.

It remains to show $\mathcal{F}_1$ is compact. Define $Q(A) = \inf_\alpha P_\alpha(A)$. Let $A\epsilon\mathcal{F}_0$ be such that $P_\alpha(A) < p_0$, where $\alpha\epsilon I$is fixed. If now $\rho(A,B) < \epsilon$, then B is contained in $A^\epsilon = \{x\epsilon\mathcal{X} : \inf_{y\epsilon A} ||x-y|| \leq \epsilon\}$. Hence $P_\alpha(B) \leq P_\alpha(A^\epsilon)$. Since $P_\alpha(A) < p_0$, there exists $\delta > 0$ such that $P_\alpha(A) < p_0 - \delta$. Choose now an $\epsilon > 0$ suitably so that $P_\alpha(A^\epsilon) \leq P_\alpha(A) + \delta$. Then

$$\begin{aligned} &\rho(A,B) < \epsilon \\ \Longrightarrow &P_\alpha(B) \leq P_\alpha(A^\epsilon) < p_0. \end{aligned}$$

This proves that $P_\alpha(A)$ is upper semicontinuous with respect to the metric ρ for each α and hence $Q(A) = \inf_\alpha P_\alpha(A)$ must also be upper semicontinuous with respect to ρ. It follows that $\mathcal{F}_1$ is closed. Also, $\mathcal{F}_0$ is compact with respect to ρ because K_1 is compact in $\mathcal{X}$ with respect to the norm $||\cdot||$ (see Dieudonné (1960), page 58). Hence $\mathcal{F}_1$ also must be compact. Finally, since the norm on $\mathcal{X}$ is convex, the convex hull of a bounded set A has the same diameter as A. This proves the theorem. We now give a simple illustration of this theorem.

Example 5. Let $\underset{\sim}{X} \sim N_m(\underset{\sim}{\theta}, \Sigma)$ where $\underset{\sim}{\theta} \epsilon \mathrm{R}^m$ is unknown and Σ is assumed to be known. The usual Hotelling confidence set $S(\underset{\sim}{x}) = \{\underset{\sim}{\theta} : (\underset{\sim}{\theta} - \underset{\sim}{x})'\Sigma^{-1}(\underset{\sim}{\theta} - \underset{\sim}{x}) \leq c^2\}$ has the property that it minimizes the Lebesgue measure among all equivariant sets. Now equivariant sets here are of the form $S = S_0 + \underset{\sim}{x}$ where S_0 is any fixed set in R^m. The restriction that $P_{\underset{\sim}{\theta}}(S \ni \underset{\sim}{\theta}) \geq p \ \forall \ \theta$ amounts to the restriction that $P_{\underset{\sim}{\theta}=\underset{\sim}{0}}(S_0) \geq p$. It follows from Theorem 3.1 that there is a closed bounded convex diameter minimizing set among the equivariant sets. Notice, however, the Hotelling ellipsoid is not a diameter minimizing set here because the smallest sphere containing it has the same diameter and a strictly larger probability (under $N(\underset{\sim}{0}, \Sigma)$). Consequently one can find a smaller diameter sphere C such that $P_{\underset{\sim}{\theta}=\underset{\sim}{0}}(C)$ is still as large as p. It will shortly be seen that diameter minimizing sets are spheres again in this situation and therefore volume and diameter minimizing sets are not the same.

The example given above naturally leads to the following question: do volume minimizing sets ever have the diameter minimization property as well? This is the content of the following theorem.

Theorem 3.2. On R^m, let P be an absolutely continuous (with respect to Lebesgue) spherically symmetric probability measure with Radon-Nikodym derivative f. Assume $f(\underset{\sim}{x})$ is of the form $g((\underset{\sim}{x} - \underset{\sim}{\mu})'(\underset{\sim}{x} - \underset{\sim}{\mu}))$ for some fixed $\underset{\sim}{\mu}$ and suppose $g(\cdot)$ is decreasing. Let C be a sphere centered at $\underset{\sim}{\mu}$ such that $P(C) = p$. Then C solves each of the following problems:

Problem 1: Minimize $\lambda(A)$ subject to $P(A) \geq p$, where $\lambda(\cdot)$ denotes Lebesgue measure and $0 < p < 1$ is a fixed number.

Problem 2: Minimize $d(A)$ subject to $P(A) \geq p$, where $d(\cdot)$ denotes diameter in L_2 norm.

Before giving a proof of this theorem, we will give an example to show that diameter minimizing sets (subject to a restriction on the probability content) are in general not necessarily spheres if the measure P is not spherical.

Example 6. In R^2, consider a bounded equilateral triangle T and let P be a probability measure with support T. The smallest radius circle containing T has its center at the centroid of T; since it has a larger diameter than that of T, circles cannot in general be diameter minimizing sets with respect to this measure P. Indeed, the reason diameter minimization is general not trivial is that arbitrary Borel sets cannot be enclosed in spheres of the same diameter (although an arbitrary bounded set in R^m with diameter d can be enclosed in a sphere with diameter not exceeding $\sqrt{\frac{2m}{m+1}} \cdot d$. See Federer (1969)). Of course, if one restricts attention to only symmetric convex sets (a set S is symmetric about $\underset{\sim}{0}$ if $\underset{\sim}{x} \epsilon S \Longrightarrow -\underset{\sim}{x} \epsilon S$; if S is symmetric about $\underset{\sim}{0}$, then $S + \underset{\sim}{\mu}$ is symmetric about $\underset{\sim}{\mu}$), then spheres will be diameter minimizing sets. In the example above, the triangle T is not symmetric about any $\underset{\sim}{\mu}$.

In order to prove Theorem 3.2, we need the following result.

Lemma 3.3. (Isodiametric inequality). Let A be any bounded Borel set in R^m with diameter d. Then the Lebesgue measure $\lambda(A)$ of A satisfies the inequality $\lambda(A) \leq \alpha(m)2^{-m} \cdot d^m$, where $\alpha(m) = \frac{\pi^{m/2}}{\Gamma(1+\frac{m}{2})}$.

Proof: A detailed proof can be found in Federer (1969). We will sketch the idea of the proof for the sake of completeness. Given a bounded Borel set A in R^m, it is possible to find a Borel set B such that B is symmetric about $\underset{\sim}{0}$, has the same Lebesgue measure as A, and such that $d(B) \leq d(A) = d$, where $d(\cdot)$ denotes diameter of a set in L_2 norm (the set B is called the Central Steiner symmetrization of A; for a proof of the existence of B, see Federer (1969)). Since B is symmetric (i.e., $\underset{\sim}{x} \epsilon B \Longrightarrow -\underset{\sim}{x} \epsilon B$) and since $d(B) \leq d$, it follows that $B \subseteq S(\underset{\sim}{0}, d)$ where $S(\underset{\sim}{0}, d)$ is the closed sphere of diameter d with center at $\underset{\sim}{0}$. Lemma 3.3 now follows immediately because $\alpha(m)2^{-m}d^m$ equals the Lebesgue measure of $S(\underset{\sim}{0}, d)$.

Lemma 3.4. Among all Borel sets in $\mathbb{R}^m$ with Lebesgue measure $\geq c (c > 0$ fixed), the spheres with Lebesgue measure c have the smallest diameter.

Proof: The proof is immediate from Lemma 3.3 if one considers the dual problem of maximizing the Lebesgue measure subject to an upper bound on the diameter.

Proof of Theorem 3.2: Define $\mathcal{F}_0 = \{A : A \text{ Borel}, P(A) \geq p\}$. Since P is spherically symmetric unimodal, it follows from the Neyman-Pearson Lemma that there exists a sphere C (with center at $\underset{\sim}{\mu}$) such that C minimizes the Lebesgue measure among sets in $\mathcal{F}_0$. Let c denote $\lambda(C)$, the Lebesgue measure of C. Now define $\mathcal{F}_1 = \{A : A \text{ Borel}, \lambda(A) \geq c\}$. Evidently, $\mathcal{F}_0 \subseteq \mathcal{F}_1$. By Lemma 3.4, C minimizes the diameter among sets in $\mathcal{F}_1$ and hence also minimizes the diameter among sets in $\mathcal{F}_0$ because $C \epsilon \mathcal{F}_0$.

It immediately follows from the above theorem that, for example, if $\underset{\sim}{X} \sim N_m(\underset{\sim}{\theta}, I)$ then the usual Hotelling confidence set minimizes the diameter among all equivariant sets with a coverage probability at least p. In fact, an analogous result is true for all spherically symmetric unimodal distributions.

We now briefly indicate applications of these results to the robust Bayes framework of section 2.

Example 7. Consider the framework of Theorem 2.1. If the likelihood function f and the envelopes L and U are such that the measure Q is spherically symmetric unimodal, then it follows from Theorem 3.2 that the volume minimizing set S of Theorem 2.1 (which is a sphere) also minimizes the diameter among all sets with the property $\inf\limits_{G \epsilon \Gamma_{L,U}} P_G(S|x) \geq p$. Examples of problems where Q is spherically symmetric unimodal are when f is a spherically symmetric normal, so is L, and $U = kL$, or when f is any spherically symmetric unimodal distribution and one takes a neighborhood of the uniform prior by letting $L = 1$ and $U = k$. Note that even though situations where Q is exactly spherical for all sample sizes are relatively rare, as the sample size becomes large Q will be nearly spherical for practically any L and U as long as f is spherical (see page 224 in Berger (1985)). Also notice that even if the original class of priors Γ (not necessarily $\Gamma_{L,U}$) is not tight, the restriction $\inf\limits_{G \epsilon \Gamma} P_G(S|x) \geq p$ may be reducible to a restriction $\inf\limits_{\alpha \epsilon I} P_\alpha(S) \geq p$ where $\{P_\alpha\}$

are tight, thereby making Theorem 3.2 applicable.

Although volume and diameter minimizing sets may not be the same for nonspherical distributions, diameter minimizing sets continue to be spheres in more general situations. The result proved below does not apply to common scale parameter problems but resolves the location problem in very satisfactory generality. We need the following definitions; see Borell (1975) and Prekopa (1971) for details.

Definition. Let P be a probability measure on R^m. For $-\infty \leq s \leq \infty$, we say that $P \epsilon m_s(\mathsf{R}^m)$ if $P(\frac{A+B}{2}) \geq \left[\frac{P^s(A)}{2} + \frac{P^s(B)}{2}\right]^{\frac{1}{s}}$ for all nonempty Borel sets A and B where $\frac{A+B}{2}$ denotes the Minkowski sum $\{z : z = \frac{1}{2}x + \frac{1}{2}y, x \epsilon A, y \epsilon B\}$. In the above, the cases $s = 0$ and $s = -\infty$ are interpreted by continuity; for example, we say that $P \epsilon m_{-\infty}(\mathsf{R}^m)$ if $P(\frac{A+B}{2}) \geq \min\{P(A), P(B)\}$ and $P \epsilon m_o(\mathsf{R}^m)$ if $P(\frac{A+B}{2}) \geq \sqrt{P(A)P(B)}$.

It is well-known that $m_s(\mathsf{R}^m)$ is empty for $s > \frac{1}{m}$ and that $s_1 \leq s_2$ implies $m_{s_2}(\mathsf{R}^m) \subseteq m_{s_1}(\mathsf{R}^m)$. Thus the family $m_{-\infty}(\mathsf{R}^m)$ contains the maximum number of probability measures. Prekopa (1971) shows that $P \epsilon m_o(\mathsf{R}^m)$ if P is absolutely continuous with respect to the Lebesgue measure λ and if $f = \frac{dP}{d\lambda}$ is log concave (i.e., m_o is the class of strongly unimodal absolutely continuous distributions). Thus the measures corresponding to the multivariate normal distributions are in m_o but those corresponding to the multivariate t are not. However, the multivariate t measures are in m_s for $s = -\frac{1}{\alpha}$ where α is the number of degrees of freedom. In particular, the multivariate t measures are in $m_{-\infty}$. We refer the reader to the excellent article of Borell (1975) for a comprehensive introduction to the families m_s.

Theorem 3.5 On R^m, let P be an absolutely continuous probability measure with density f. Suppose f is symmetric about some point $\underset{\sim}{\mu}$ and $f^{-\frac{1}{m}}$ is convex. Then sets minimizing $d(A)$ among A such that $P(A) \geq p (0 < p < 1)$ are spheres.

Proof: Without loss the point $\underset{\sim}{\mu}$ may be assumed to be zero. Let A be any bounded (Borel) set. Define S as the Minkowski sum $S = \frac{A+A_-}{2}$, where $A_- = \{-x : x \epsilon A\}$. Thus $P(A_-) = P(A)$. Since P is absolutely continuous and the density f is such that $f^{-\frac{1}{m}}$ is convex, it follows from Theorem 3.2 in Borell (1975) that $P \epsilon m_{-\infty}$. Therefore,

$P(S) \geq P(A)$. Notice now that S is symmetric about zero and $d(S) \leq d(A)$. Hence S is contained in the closed sphere $C = S(\underset{\sim}{o}, d(A))$, implying that for any bounded set A with diameter $d(A)$, there exists a sphere C of the same diameter such that $P(C) \geq P(A)$. This proves the theorem.

Corollary 3.6. Let $\underset{\sim}{x} \sim f(\underset{\sim}{x} - \underset{\sim}{\theta})$, where $f(\underset{\sim}{x})$ is symmetric and strongly unimodal. Then among all equivariant sets A with $\inf_{\theta} P_\theta(A \ni \theta) \geq p$, the sphere $C = C_o + \underset{\sim}{X}$ where $P_{\underset{\sim}{\theta}=\underset{\sim}{o}}(C_o) = p$ has the smallest diameter.

Proof: Use the fact that $m_o \subseteq m_{-\infty}$.

Corollary 3.6 applies in particular to the important case when $\underset{\sim}{X} \sim N(\underset{\sim}{\theta}, \Sigma), \Sigma$ known. For other symmetric unimodal distributions (which are not strongly unimodal) such as the elliptical t, the assertion of corollary 3.6 will still hold as long as $f^{-\frac{1}{m}}$ is convex.

Corollary 3.7. Consider the setup of Theorem 2.1. Assume P_θ is absolutely continuous with respect to the Lebesgue measure λ (on R^m) for every θ. Let $f(x - \theta) = \frac{dP_\theta}{d\lambda}$ and suppose $f(\cdot)$ satisfies the two assumptions of Theorem 3.5. Let $L \equiv 1, U \equiv k$. Then sets minimizing $d(A)$ among A such that $\inf_G P_G(A|x) \geq p$ are spheres.

Proof: Use the fact that the posterior under L must also satisfy the two assumptions of Theorem 3.5.

Again, Corollary 3.7 will apply to all common location parameter problems if one takes a neighborhood of the uniform prior as formulated via (1.1) with $L \equiv 1, U \equiv k$.

Finally, in order to make things complete, we give a simple one dimensional example where diameter and Lebesgue measure minimizing sets are different and we explicitly show what these sets are.

Example 8. Consider a random variable $X \sim g(x - \theta)$ with g given by the following linear spline:

$$\begin{aligned} \frac{11}{16} g(x) &= x \text{ if } 0 \leq x \leq \frac{1}{4} \\ &= \frac{1}{2} - x \text{ if } \frac{1}{4} \leq x \leq \frac{1}{2} \end{aligned}$$

$$= 10(x - \frac{1}{2}) \text{ if } \frac{1}{2} \leq x \leq \frac{3}{4}$$

$$= 10(1 - x) \text{ if } \frac{3}{4} \leq x \leq 1. \tag{3.1}$$

Direct algebra gives that under $g(x)$ the set S_0 with smallest Lebesgue measure among sets with probability at least $\frac{10}{11}$ is $S_0 = [.2384, .2616] \cup [.5238, .9762]$ with a Lebesgue measure of .4756 and a diameter of .7378. On the other hand, the set S_1 with the smallest diameter with the same probability is the interval $[.5, 1]$, which has a slightly larger Lebesgue measure but a substantially smaller diameter. Thus the smallest Lebesgue measure equivariant set is $[X - .2616, X - .2384] \cup [X - .9762, X - .5238]$ while the smallest diameter equivariant set is $[X - 1, X - .5]$. Of course, this example should be taken just as an artifact; the point here is that if the underlying distribution is bimodal (or multimodal) but one of the peaks is considerably higher than the other(s), it may be more desirable to use the smallest diameter connected set rather than a disconnected set simply to gain a small amount in Lebesgue measure. Examples of exactly this kind can also be easily found in common inference problems.

4. Final remarks. In the final analysis, it is true that diameter minimizing sets will in general be much harder to find than volume minimizing ones. It is reassuring that in spherical problems both approaches lead to the same answer and that we know the form of these sets for most location problems. We are not suggesting that the criterion of minimum volume should be abandoned in favor of minimum diameter. We simply want to make the point that in some problems it may be preferable to do so, especially if connectedness is considered desirable. In the problems where volume and diameter minimizing sets are different, further study and comparison of their frequentist properties are needed before the diameter minimizing sets can be seriously recommended for practice. A possible criterion for assessing confidence sets in a framework of decision theory is to consider the loss $L_1(\theta, C) = (1 - \epsilon)P_\theta(C \not\ni \theta) + \epsilon d(C)$; clearly all symmetric convex sets other than spheres will be inadmissible with respect to L_1. A second possibility is to consider a loss $L_2(\theta, C) = \epsilon_1 P_\theta(C \not\ni \theta) + \epsilon_2 d(C) + \epsilon_3 \lambda(C)$. In the problem of estimating the mean of a $N(\underset{\sim}{\theta}, I)$ distribution, Hwang and Casella (1982) showed that the Hotelling confidence sets can be improved in terms of coverage probability by recentering them at

a positive-part James-Stein estimator. We have been able to show that the improved sets of Hwang-Casella (1982) are inadmissible under L_2 if $\frac{\epsilon_2}{\epsilon_1} \geq \epsilon_0$, where ϵ_0 is a fixed positive number. Whether this holds for arbitrary ϵ_i may be of some interest.

Acknowledgement. I am deeply indebted to Steve Lalley, Herman Rubin, and Bill Studden for valuable advice and suggestions. Jim Berger and Leon Gleser gave active encouragement. The Associate Editor and two referees made valuable suggestions for a clearer and more compact presentation of the material. Mr. Sanjib Basu and Mr. F.C. Chang helped with the numerical work. They all have my sincerest appreciation.

References

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.

Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, 2nd Ed., Springer–Verlag, New York.

Berger, J.O. (1987). Robust Bayesian Analysis: Sensitivity to the Prior, Technical Report, Department of Statistics, Purdue University.

Borell, C. (1975). Convex set functions in *d*-space. *Period. Math. Hung.*, **6**, 111–136.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Massachusetts.

Brown, L.D. (1966). On the Admissibility of Invariant Estimators of One or More Location Parameters. *Ann. Math. Statist.*, **37**, 1087–1136.

Cohen, Arthur and Strawderman, W.E. (1973). Admissibility Implications for Different Criteria in Confidence Estimation. *Ann. Statist.*, **1**, 363–366.

DasGupta, A. (1988). Volume and Diameter Minimizing Confidence Sets in Bayes and Classical Problems. Technical Report, Department of Statistics, Purdue University.

DasGupta, A. and Studden, W.J. (1988a). *Robust Bayesian Analysis and Optimal Experimental Designs in Normal Linear Models with many Parameters – I,* Technical Report, Department of Statistics, Purdue University.

DasGupta, A. and Studden, W.J. (1988b). *Robust Bayesian Analysis and Optimal Experimental Designs in Normal Linear Models with many parameters – II,* Technical Report, Department of Statistics, Purdue University.

DasGupta, A., Ghosh, J.K. and Zen, M.M. (1990). Bayesian analysis of a multivariate normal mean under flat tailed priors. Technical Report, Department of Statistics, Purdue University.

DeRobertis, L. (1978). *The Use of Partial Prior Knowledge in Bayesian Inference.* Ph.D.

Thesis, Yale University, New Haven.

DeRobertis, L. and Hartigan, J.A. (1981). Bayesian Inference Using Intervals of Measures. *Ann. Statist.*, **9**, 235–244.

Diaconis, P. and Freedman, D. (1986). On the Consistency of Bayes Estimates. *Ann. Statist.*, **14**, 1–67.

Dieudonne, J. (1960). *Foundations of Modern Analysis*, Academic Press, New York.

Federer, H. (1969). *Geometric Measure Theory*, Springer-Verlag, New York.

Hotelling, H. (1939). Tubes and Spheres in n-spaces, and a Class of Statistical Problems. *Amer. J. Math*, **61**, 440–460.

Hwang, J.T. and Casella, G. (1982). Minimax Confidence Sets for the Mean of a Multivariate Normal Distribution. *Ann. Statist.*, **10**, 868–881.

Joshi, V.M. (1967). Inadmissibility of the Usual Confidence Sets for the Mean of a Multivariate Normal Population. *Ann. Math. Statist.*, **38**, 1868–1875.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2nd Ed., John Wiley and Sons, New York.

Naiman, Daniel (1986). Conservative Confidence Bands in Curvilinear Regression. *Ann. Statist.*, **14**, 896–906.

Pratt, J.W. (1961). Length of Confidence Intervals. *Jour. Amer. Statist. Assoc.*, **56**, 549–567.

Prekopa, A. (1971). Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math.*, **32**, 301–316.

Wijsman, R.A. (1979). Constructing All Smallest Simultaneous Confidence Sets in a Given Class, with Applications to MANOVA. *Ann. Statist.*, **7**, 1003–1018.

Wynn, H.P. (1984). An Exact Confidence Band for Polynomial Regression. *Biometrika*, **71**, 375–379.

FIGURE 1 : SMALLEST AVERAGE WIDTH BANDS

(Example 3)

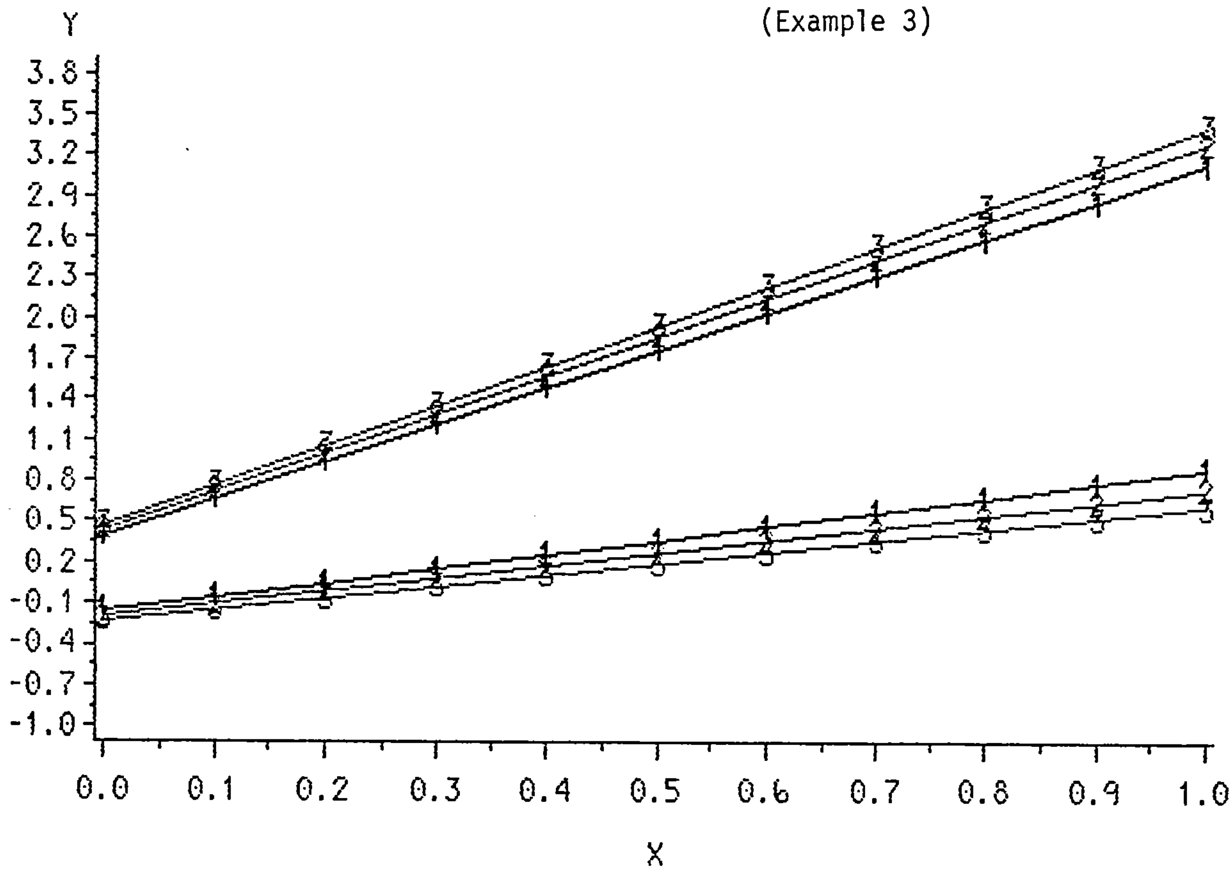

FIGURE 2 : SMALLEST MAXIMUM WIDTH BANDS

(Example 3)

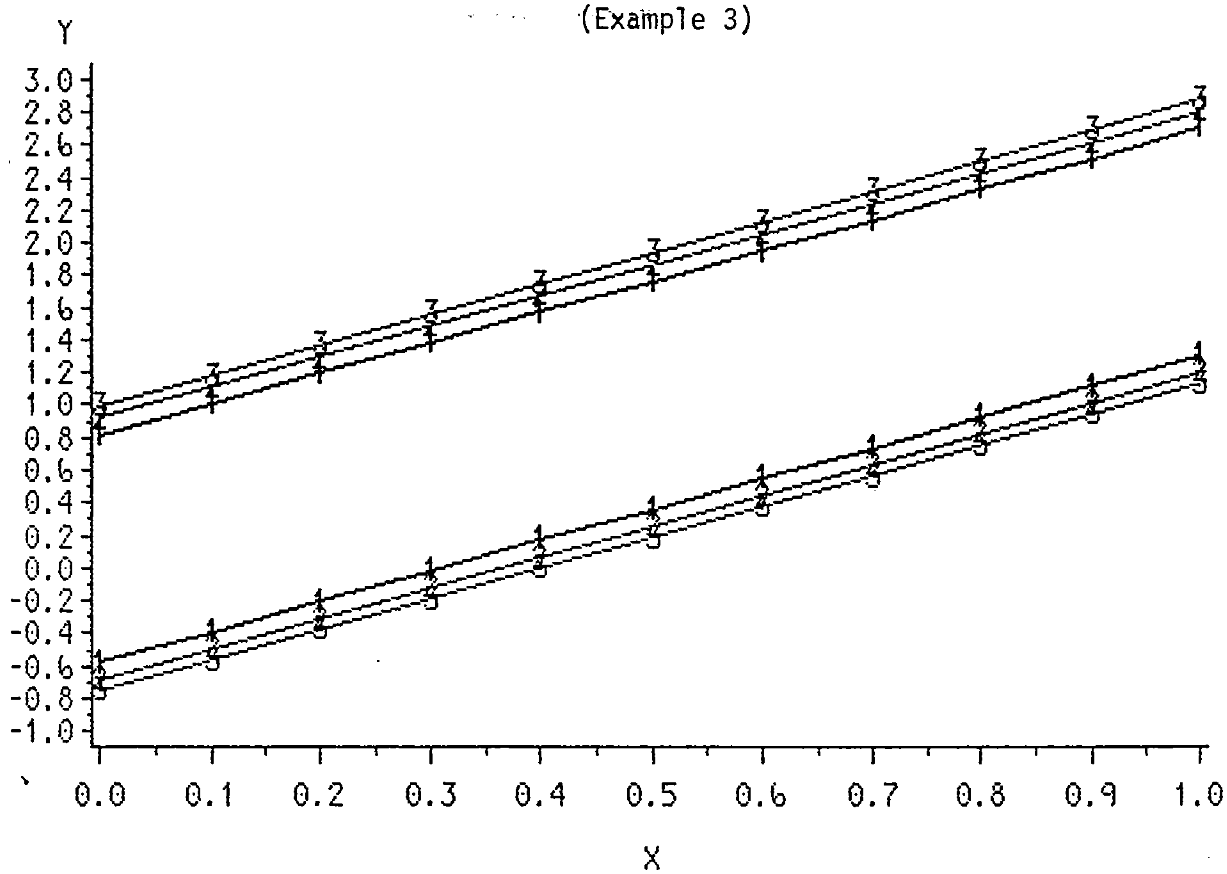

FIGURE 3: POSTERIOR FOR NORMAL MEANS WITH CAUCHY PRIOR, $||X|| = 2.6$

(Example 4)

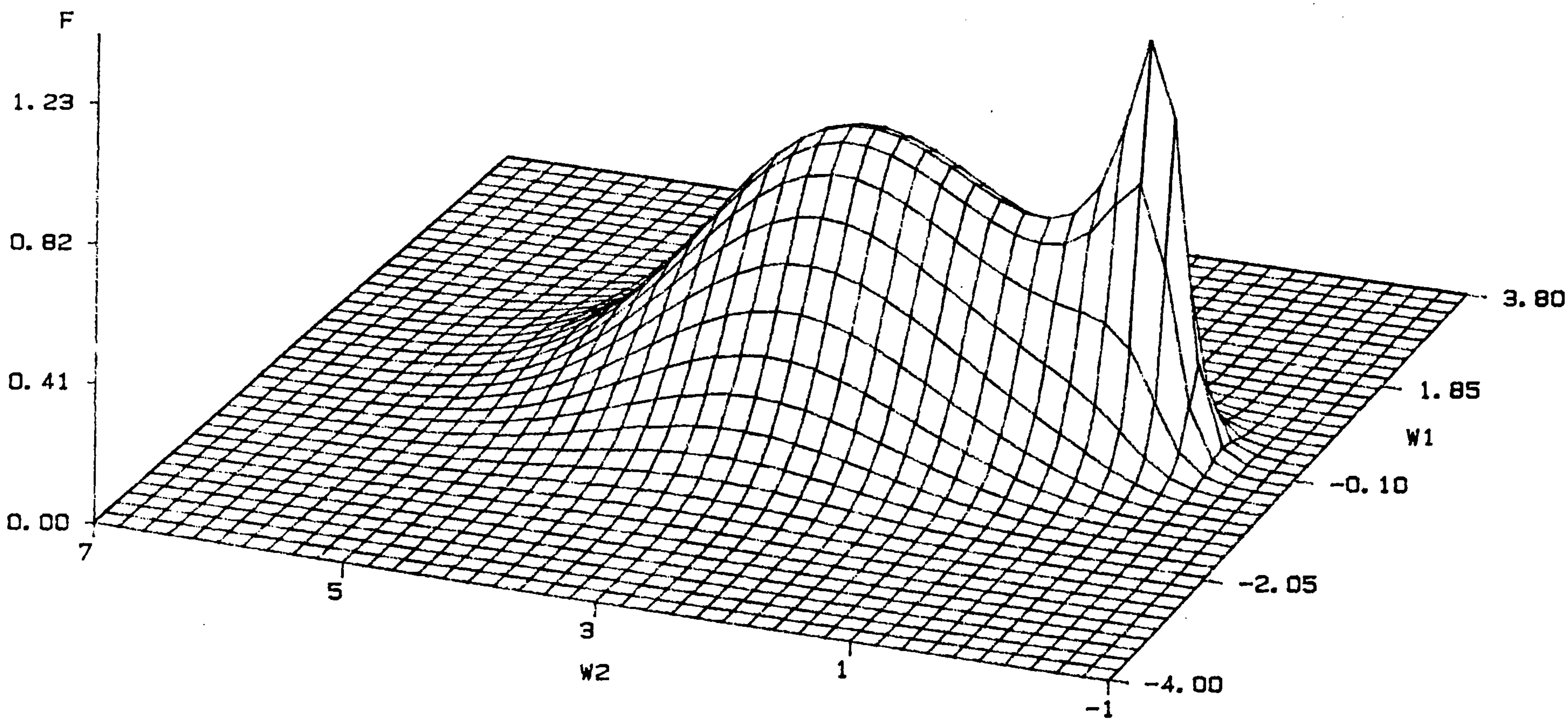

FIGURE 4: POSTERIOR FOR NORMAL MEANS WITH CAUCHY PRIOR, $||X|| = 2.8$

(Example 4)

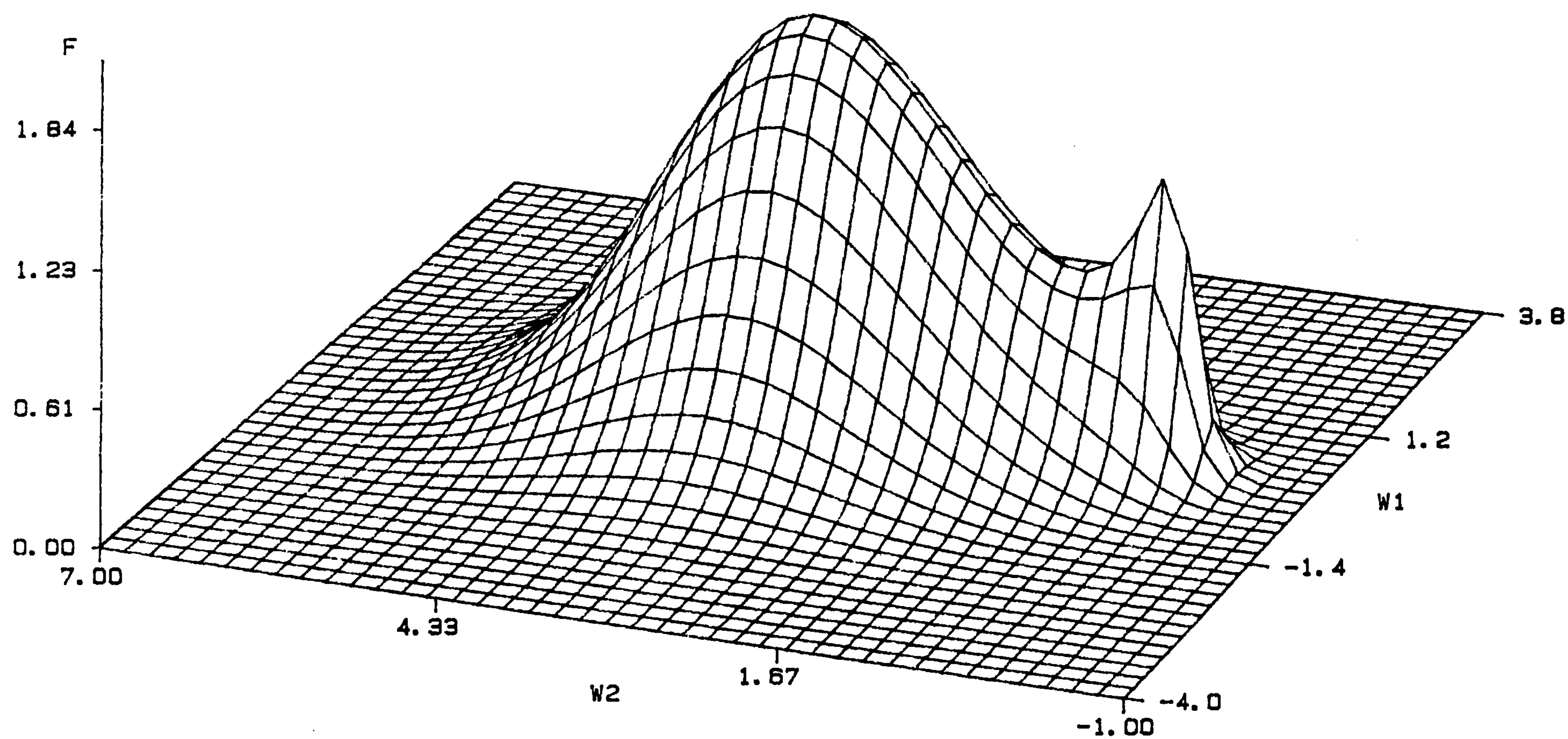

FIGURE 5: POSTERIOR FOR NORMAL MEANS WITH CAUCHY PRIOR, $||X|| = 3$

(Example 4)

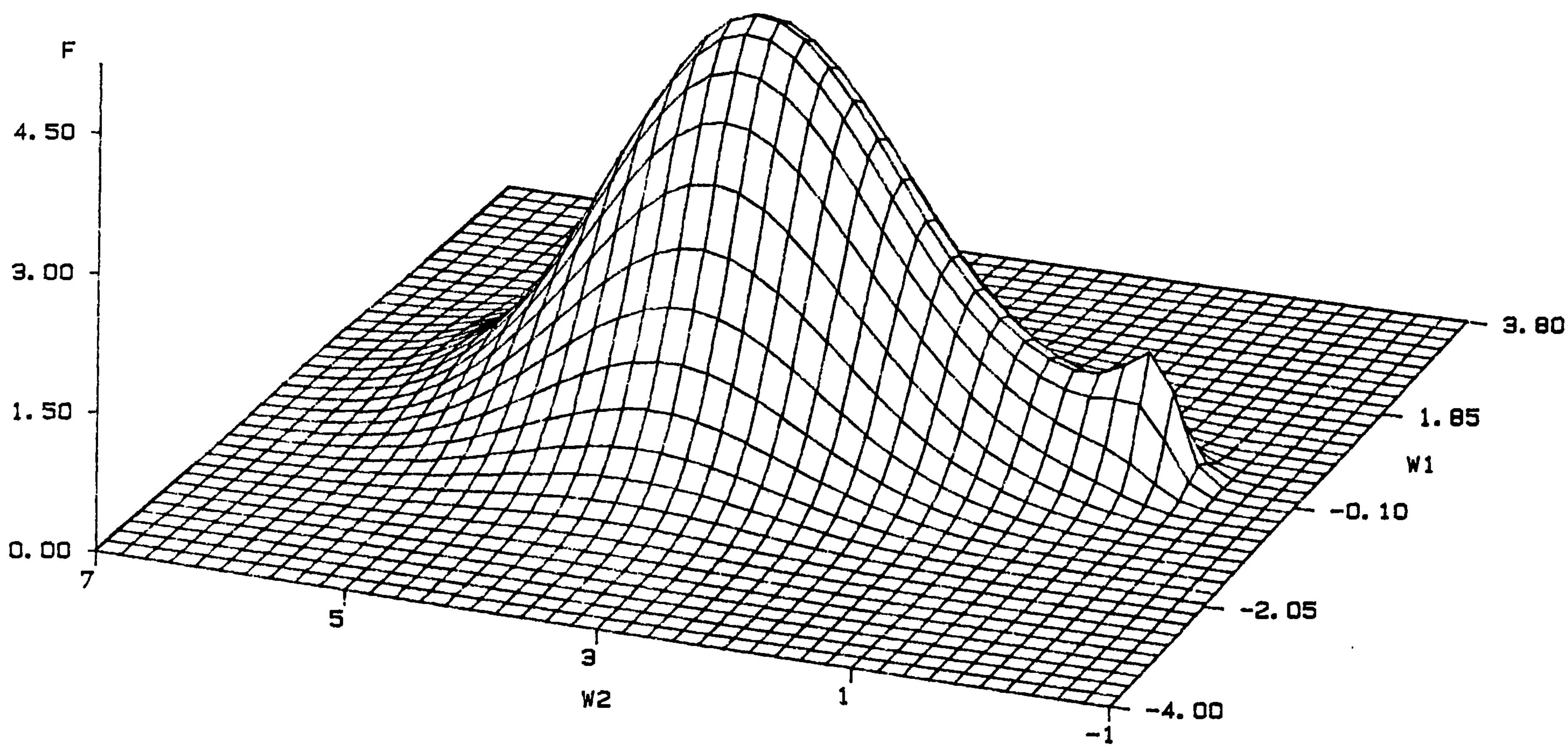

FIGURE 6: POSTERIOR FOR NORMAL MEANS WITH CAUCHY PRIOR, $||X|| = 3.2$

(Example 4)