The Annals of Statistics
1997, Vol. 25, No. 5, 2228–2258

GOOD EXACT CONFIDENCE SETS FOR A MULTIVARIATE NORMAL MEAN

BY YU-LING TSENG[1] AND LAWRENCE D. BROWN[2]

Academia Sinica, Taipei and University of Pennsylvania

A class of confidence sets with constant coverage probability for the mean of a p-variate normal distribution is proposed through a pseudo-empirical-Bayes construction. When the dimension is greater than 2, by combining analytical results with some exact numerical calculations the proposed sets are proved to have a uniformly smaller volume than the usual confidence region. Sufficient conditions for the connectedness of the proposed confidence sets are also derived. In addition, our confidence sets could be used to construct tests for point null hypotheses. The resultant tests have convex acceptance regions and hence are admissible by Birnbaum. Tabular results of the comparison between the proposed region and other confidence sets are also given.

1. Introduction. One of the most frequently used statistical techniques is the linear model, which includes both the analysis of variance and linear regression as special cases. In the usual formulation of a linear model, the estimation problem can be reduced to that of estimating a multivariate normal mean. [See, e.g., Hwang and Chen (1986).] In such a situation, the results of Stein (1956) and James and Stein (1961) lead immediately to a uniform, appreciable improvement in mean squared error over the least squares estimator when there are at least three parameters. Surprisingly, this phenomenon is not exceptional. Brown (1966) showed the same inadmissibility result for the best invariant estimator of location for a very wide variety of distributions and loss functions. In particular, his results implied that when the dimension is at least 3 the usual confidence region for the parameters in a linear model is inadmissible too. Since then considerable research has aimed at explicit constructions of dominating estimators and improved confidence sets for the mean vector of a multivariate normal distribution.

Let $X = (X_1, \dots, X_p)'$ have a p-variate normal distribution with mean vector $\theta = (\theta_1, \dots, \theta_p)'$ and identity covariance matrix I. There have been many breakthroughs in the theory of estimating θ. References to related

Received April 1995; revised January 1997.

[1]This is a revised version of Dr. Tseng's thesis presented to Cornell University in partial fulfillment of the requirements for a Ph.D. degree.

[2]Research supported in part by NSF Grant DMS-93-10228.

AMS 1991 *subject classifications*. Primary 62F25; secondary 62C20, 62C15.

Key words and phrases. Multivariate normal mean, volume, coverage probability, confidence sets, James–Stein estimator, Stein-type estimator, pseudo-empirical-Bayes construction.

works can be found in, for example, Faith (1976), Berger (1985) and Tseng (1994). In contrast to the rich development in point estimation, until recently there has been comparatively little research on finding improved confidence sets for θ.

The usual confidence set for θ is

$$C^0(X) = \{\theta : \|X - \theta\|^2 \le c\}, \tag{1}$$

where c is the $100(1-\alpha)$th percentile of the chi-squared distribution with p degrees of freedom and $\|\cdot\|$ is the Euclidean norm. It is an *exact* $1-\alpha$ confidence set for θ, that is,

$$P_\theta(\theta \in C^0(X)) = 1 - \alpha \qquad \forall\, \theta,$$

and it is a good confidence set estimator under several criteria. When $p = 1$ or 2, Brown (1966) and Joshi (1969) independently showed that under criterion (6) (see below) $C^0(X)$ is strongly admissible among all $1-\alpha$ confidence sets. Stein (1962) showed that for all dimensions $C^0(X)$ has the minimum volume among all $1-\alpha$ level confidence sets which are invariant under the group of translations in R^p, and it is minimax, that is, it satisfies

$$\sup_\theta E_\theta[\operatorname{Vol}(C^0(X))] = \inf_{C \in \mathscr{C}} \sup_\theta E_\theta[\operatorname{Vol}(C(X))],$$

where $\mathscr{C}$ is the class of all $1-\alpha$ confidence sets and $\operatorname{Vol}(C)$ for any set C in R^p is defined as

$$\operatorname{Vol}(C) = \int_{R^p} I(t \in C)\,\mu(dt)$$

with μ the Lebesgue measure in R^p. Stein (1962) also proved that the usual confidence set cannot be uniformly dominated in false coverage probability by any $1-\alpha$ confidence sets.

However, Stein (1962) also gave heuristic arguments claiming that the confidence sets associated with the now well known James–Stein estimators improve upon $C^0(X)$ for large dimensions and conjectured that the same result holds for all $p \ge 3$. It was independently proved in Brown (1966) and Joshi (1967) that, where the dimension is 3 or more, confidence spheres centered at a Stein-type estimator and having the same volume as $C^0(x)$ for all x can have higher coverage probability than $C^0(X)$ for all θ. They did not, however, provide the explicit form of better confidence regions.

For about a decade since the work of Brown (1966) and Joshi (1967), no significant progress was made in finding specific improved confidence sets for a multivariate normal mean. This is not because the problem is not statistically important, but rather because it involves great technical difficulty. Fortunately, better understanding of Stein phenomena and far greater computer facilities in the past two decades have made some significant breakthroughs in set estimation possible. The literature on this development includes Faith (1976), Stein (1981), Berger (1980), Hwang and Casella (1982, 1984), Casella and Hwang (1983, 1986) and Shinozaki (1989).

Faith (1976) developed Bayesian confidence regions and gave convincing numerical and theoretical arguments that they improve upon $C^0(X)$ when the dimension is 3 or 5. Unfortunately, his regions had complicated shapes due to their Bayesian derivation. Stein (1981) gave basic formulas for unbiased estimation of the risk of an arbitrary point estimator and utilized them to suggest some approximate confidence sets. However, the validity of the approximate confidence regions and their properties were not further studied. Berger (1980) considered confidence ellipsoids associated with an admissible generalized Bayes estimator for the mean vector. He derived necessary and sufficient conditions under which his sets have uniformly smaller volume than the usual confidence region. Although uniform dominance results in coverage probability were not obtained, he gave asymptotic theorems and convincing numerical evidence that his sets maintain satisfactory coverage probabilities. A major problem with this confidence set estimation is in its implementation. It involves complicated calculations for the generalized Bayes estimator and the inverse of a posterior covariance matrix.

Following in the spirit of Stein (1962), Hwang and Casella (1982, 1984) successfully showed that recentering the usual confidence set to the positive part of a Stein estimator results in uniform improvement in coverage probability. The recentered sets have the same volume as the usual one, though. Casella and Hwang (1983, 1986) considered recentered sets with variable radii by empirical Bayes arguments. Although they provided strong numerical results to support the superiority of their sets, analytical dominance results were not obtained. Shinozaki (1989) provided a class of confidence sets and showed that some of them have smaller volume than $C^0(X)$ and have the same confidence coefficient as $C^0(X)$. The sets were constructed by shrinking the boundary points of $C^0(X)$ toward the origin according to some functions which were not explicitly defined. The implicit functions used in the construction made the study concerning the geometry of the confidence sets and their associated acceptance regions extremely difficult and related questions remain unsettled. These are relevant because understanding the geometry of the associated acceptance regions of confidence sets is very important if the primary interest is in hypothesis testing.

Our principal goal here is to construct confidence regions which have smaller volume than $C^0(X)$ while retaining the constant coverage probability. Besides their theoretical importance, these kinds of improved confidence sets are more desirable in practice. Intuitively, smaller sets cover fewer points and, hence, are less likely to include false values. In other words, with the same preset acceptable confidence coefficient, sets with smaller volume provide higher precision for the practitioners. In fact, Ghosh (1961) and Pratt (1961) showed that the expected volume of a confidence region is equal to an integrated sum of its false coverage probabilities, now known as the Ghosh–Pratt identity. This identity was used in Cohen and Strawderman (1973) and Brown, Casella and Hwang (1995) in establishing volume optimality results for certain problems.

In Section 2 a class of exact confidence sets, denoted by $C^*(X)$, for θ is proposed through a pseudo-empirical-Bayes construction. The sets have an unfamiliar shape, somewhat like an egg (see Figure 1, in Section 4.) Nevertheless they give spherical, hence, convex acceptance regions. As a result, they are associated with admissible tests. By its construction, the volume of $C^*(X)$ depends on X only through its Euclidean norm $\|X\|$. In Section 3 we calculate the asymptotic volume difference, as $\|x\|$ tends to infinity, between $C^0(x)$ and $C^*(x)$ and give conditions such that $C^*(x)$ has a smaller asymptotic volume. In Section 4 we derive sufficient conditions under which $C^*(x)$ is connected for all x. Theorem 5.3 in Section 5 analytically establishes a value of $\|x\|$ after which the volume of $C^*(x)$ is always smaller than that of $C^0(x)$. With this theorem and with help from Section 4, we are able to provide, in Section 5, a computer-aided proof for the uniform dominance of $C^*(X)$ over $C^0(X)$. Finally, tabular results of the comparison between $C^*(X)$ and other confidence sets are given in Section 6.

2. Proposed confidence sets. If an improved confidence region is constructed by recentering $C^0(X)$ at a Stein-type estimator, hence keeping the same volume, one can view the improvement as due to the effect of moving the usual confidence set toward the origin. This results in a uniform improvement in coverage probability with largest improvement near the origin in the parameter space. In order to maintain the same constant coverage probability while decreasing the volume, the same idea of a shrinkage effect is introduced, however, in a different fashion.

For convenience, we first present the form of our confidence sets, $C^*(X)$. Some immediate properties of $C^*(X)$ and the motivation for its construction are given in Sections 2.1 and 2.2, respectively.

For any nonnegative number δ, let $c(\delta)$ be the $100(1-\alpha)$th percentile of a noncentral chi-squared distribution with p degrees of freedom and noncentrality parameter δ. The proposed class of confidence sets for θ is

$$C^*(X) = \left\{\theta: \|X - \theta(1+\gamma(t))\|^2 \le c(\lambda)\right\}, \tag{2}$$

where $\gamma(t) = (A+Bt)^{-1}$, $t = \|\theta\|^2$, $\lambda = t\gamma^2(t)$ and A, B are positive constants whose values will be specified later.

2.1. *Coverage probability and associated acceptance regions.* With this definition, automatically we have

$$P_\theta(\theta \in C^*(X)) = 1 - \alpha \qquad \forall\, \theta;$$

that is, like $C^0(X)$, $C^*(X)$ is also an exact $1-\alpha$ confidence set for θ.

Another important feature of $C^*(X)$ is the level-α acceptance regions, denoted by $A^*(\theta_0)$, for testing the hypotheses

$$H_0: \theta = \theta_0 \quad \text{versus} \quad H_1: \theta \ne \theta_0$$

as given through the usual correspondence

$$\theta_0 \in C^*(x) \quad \text{iff } x \in A^*(\theta_0).$$

For any fixed θ

$$A^*(\theta) = \{x: \|x - \theta(1 + \gamma(t))\|^2 \le c(\lambda)\},$$

which is a p-dimensional ball which is centered at $\theta(1 + \gamma(t))$ and with radius $c(\lambda)$, hence a convex set.

The results of Birnbaum (1955) showed that for this testing problem a necessary and sufficient condition for a test to be admissible is that it have a convex acceptance region. Hence we have the following theorem.

THEOREM 2.1. *Fix θ_0. For testing H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$, the level-α test with $A^*(\theta_0)$ as its acceptance region is admissible.*

2.2. *Motivation, the pseudo-empirical-Bayes construction.* The following theorem in Brown, Casella and Hwang (1995) motivates the construction of the improved confidence sets to be considered. It should be noticed that the proof of the Brown–Casella–Hwang theorem is essentially based on the Ghosh–Pratt identity [Ghosh (1961); Pratt (1961)].

THEOREM 2.2 [Brown, Casella and Hwang (1995)]. *Let $X \sim f(x|\theta)$, $\theta \sim \pi(\theta)$ and define the set $C_\pi(x)$ to be*

$$C_\pi(x) = \left\{\theta: \frac{m_\pi(x)}{f(x|\theta)} \le k(\theta)\right\},$$

where $m_\pi(x) = \int_\Theta f(x|\theta)\pi(\theta)\,d\theta$ and $k(\theta)$ is chosen such that

$$P_\theta(\theta \in C_\pi(X)) = 1 - \alpha.$$

Then $C_\pi(x)$ minimizes the expected Bayesian volume among all $1 - \alpha$ confidence sets.

Notice from their proof that the $1 - \alpha$ confidence set with smallest expected volume under the chosen prior π is found by inverting a class of most powerful tests of

$$H_0: X \sim f(x|\theta) \quad \text{versus} \quad H_1: X \sim m_\pi(x), \tag{3}$$

where $m_\pi(x)$ is the marginal distribution of X under the prior π.

For the normal case $X \sim N_p(\theta, I)$ consider $\theta \sim N_p(0, \tau^2 I)$ with some positive constant τ; that is,

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

and

$$\pi(\theta) = \left(\frac{1}{2\pi\tau^2}\right)^{p/2} \exp\left(-\frac{1}{2\tau^2}\|\theta\|^2\right).$$

The marginal distribution $m_\pi(x)$ of X is $N_p(0, (\tau^2 + 1)I)$. Most powerful tests of (3) accept H_0 iff

$$\frac{m_\pi(x)}{f(x|\theta)} \le \text{some constant}$$

$$\Leftrightarrow \quad \frac{\left(1/[2\pi(\tau^2+1)]\right)^{p/2} \exp\left[-\left(1/[2(\tau^2+1)]\right)\|x\|^2\right]}{\left(1/(2\pi)\right)^{p/2} \exp\left(-\frac{1}{2}\|x-\theta\|^2\right)} \le \text{some constant}$$

$$\Leftrightarrow \quad \left\| x - \theta\left(\frac{1+\tau^2}{\tau^2}\right)\right\|^2 \le c\left(\frac{\|\theta\|^2}{\tau^4}\right).$$

Hence the confidence set which minimizes the expected Bayesian volume under π is

$$C_\pi(x) = \left\{\theta : \left\| x - \theta\left(\frac{1+\tau^2}{\tau^2}\right)\right\|^2 \le c\left(\frac{\|\theta\|^2}{\tau^4}\right)\right\}. \tag{4}$$

We have been treating τ^2 as a constant so far, but to utilize (4) to produce confidence sets which have the potential to improve upon $C^0(X)$ in terms of (6), τ^2 needs to vary with either x or θ. Since we are aiming at constructing confidence sets which have coverage probability exactly $1 - \alpha$ and smaller volume than the usual confidence set, we need to replace τ^2 by some function of θ instead of a function of x.

In the usual empirical Bayes approach, hyperparameters are typically estimated from the observation x. As noted above we will instead use a function of the parameter θ to replace the hyperparameter τ. We call this kind of approach a pseudo-empirical-Bayes construction. An intuitive justification for this type of approach is that θ is more directly related to τ^2 than x. Earlier examples in which replacement of a nuisance parameter by a function of key parameters leads to better confidence sets than those using a function of x can be found in Hwang (1995) and Huwang (1991, 1995).

Since $C^0(X)$ is minimax in terms of expected volume, any $1 - \alpha$ confidence set having smaller volume than $C^0(x)$ for all x is minimax itself. A minimax procedure tries to do as well as possible in the worst case. One might expect that minimax procedures would be Bayes with least favorable prior distributions. Along this line, to find reasonable functions of θ to replace τ, let us revisit the previous testing problem (3). Note that, under H_0, $X \sim N_p(\theta, I)$ hence

$$\|x\|^2 \approx E(\|X\|^2) = p + \|\theta\|^2$$

and, under H_1, $X \sim N_p(0, (\tau^2 + 1)I)$ hence

$$\|x\|^2 \approx E(\|X\|^2) = p + pr^2.$$

For fixed θ, if we equate $p + \|\theta\|^2$ and $p + p\tau^2$ we make the hypotheses of (3) the most difficult to be distinguished, hence the resulting π close to being the least favorable distribution. This motivates the choice

$$p + \|\theta\|^2 = p + p\tau^2,$$

which is equivalent to

$$\tau^2 = \frac{\|\theta\|^2}{p}. \tag{5}$$

As an added motivation, note that $\|\theta\|^2/p$ is an unbiased "estimator" of τ^2 since $\theta \sim N_p(0, \tau^2 I)$.

Equation (5) motivates a particular linear function of $\|\theta\|^2$ as a reasonable replacement for τ^2. In this work, we obtain added flexibility by using $A + B\|\theta\|^2$ with some positive constants A and B to replace τ^2 in (4); note that (5) corresponds to $A = 0$, $B = 1/p$. This defines our new class of confidence sets for θ.

3. Asymptotic domination theorems. We consider one confidence set to be better than another if it has higher coverage probability at the same time it has no larger volume for every given observation. It was pointed out by Joshi (1969), however, that there is a technical difficulty to be avoided: by adding measure-zero sets to any confidence set C it is possible to increase its coverage probability without any increment in its volume. We should, therefore, compare confidence set estimators across different equivalence classes. Two confidence sets $C_1(X)$ and $C_2(X)$ are said to be equivalent if

$$\mathrm{Vol}(C_1(x) \,\Delta\, C_2(x)) = 0 \qquad \forall\, x,$$

where $C_1(x) \,\Delta\, C_2(x)$ is the symmetric difference set of $C_1(x)$ and $C_2(x)$. Let $\mathscr{C}^0$ be the equivalence class containing the usual confidence set $C^0(X)$. We say a confidence set $C(X)$ dominates $C^0(X)$ if $C(X) \notin \mathscr{C}^0$ and the following conditions are satisfied:

$$\begin{aligned} &\text{(i)} \quad P_\theta(\theta \in C(X)) \geq P_\theta(\theta \in C^0(X)) \qquad \forall \theta \\ &\text{(ii)} \quad \mathrm{Vol}(C(x)) \leq \mathrm{Vol}(C^0(x)) \qquad \forall\, x, \end{aligned} \tag{6}$$

with strict inequality either in (6)(i) for some θ or in (6)(ii) for all x in some set with positive Lebesgue measure.

Since $C^0(X)$ and $C^*(X)$ have the same constant coverage probability, to see if, for some choices of A and B, $C^*(X)$ dominates $C^0(X)$ under criteria (6), we need to calculate the volume of $C^*(x)$ for all x. In this section we calculate the asymptotic volume difference between $C^0(x)$ and $C^*(x)$, and we give the range for choosing A and B such that $C^*(x)$ has a smaller asymptotic volume.

To understand better the nature of $C^*(X)$ and the results of the asymptotic theorems to be presented, we first provide the following remark to clarify the relationship between $C^*(X)$ and the usual James–Stein estimator.

REMARK 3.1. For large $\|x\|$, $C^*(x)$ can be related to the recentered set

$$\{\theta: \|\delta^*(x) - \theta\|^2 \le c\},$$

where

$$\delta^*(x) = \left(1 - \frac{a}{Aa + \|x\|^2}\right)x \quad \text{and} \quad a = \frac{1}{B},$$

by the following approximating arguments. First note that

$$\begin{aligned} x - \theta(1+\gamma) &= x - \gamma(\theta - x + x) - \theta \\ &= (1-\gamma)x - \theta + \gamma(x - \theta) \\ &\approx \delta^*(x) - \theta \end{aligned}$$

since, as $\|x\| \to \infty$,

$$\gamma \approx \frac{1}{A + B\|x\|^2} \quad \text{and} \quad \gamma(x - \theta) \approx 0.$$

Also it is obvious that

$$\lim_{\|x\| \to \infty} c(\lambda) = c.$$

Hence

$$C^*(x) \approx \{\theta: \|\delta^*(x) - \theta\|^2 \le c\}$$

for large $\|x\|$.

It is interesting to note that

$$B \ge \frac{1}{2(p-2)} > 0 \quad \Leftrightarrow \quad 0 < a \le 2(p-2).$$

Since the condition on a is necessary and sufficient for the domination of $\delta^*(X)$ over X as point estimators, it is not surprising to see in the following theorems that this same condition on B implies asymptotic dominance of $C^*(X)$ over $C^0(X)$.

The following lemma is useful in proving the asymptotic theorems and several other proofs in this paper. For its proof readers are referred to Tseng (1994).

LEMMA 3.1. *For any given* $0 < \alpha < 1$, *let* $c(\lambda)$ *be the* $100(1-\alpha)$*th percentile of a noncentral chi-squared distribution with p degrees of freedom and noncentrality parameter* λ. *Then*

$$\frac{d}{d\lambda}c(\lambda) = \frac{f_{p+2,\lambda}(c(\lambda))}{f_{p,\lambda}(c(\lambda))} \tag{7}$$

$$\le \frac{c(\lambda)}{p}, \tag{8}$$

where $f_{m,\lambda}(\cdot)$ *is the p.d.f. of a noncentral chi-squared distribution with m degrees of freedom and noncentrality parameter* λ.

The following two theorems state ways to select the positive constants A and B in the proposed confidence set $C^*(X)$ such that it dominates the usual one in volume asymptotically. The proof for Theorem 3.1 is given in the Appendix and that for Theorem 3.2 is similar and hence omitted.

For simplicity, let $\nabla(x) = \mathrm{Vol}(C^0(x)) - \mathrm{Vol}(C^*(x))$ and define constants

$$k_0 = 2\pi \prod_{j=1}^{p-3} \int_0^{\pi} \sin^j(\beta)\, d\beta \quad \text{and} \quad k_p = \int_0^{\pi} \sin^{p-2}(\beta)\, d\beta$$

in our presentation hereafter.

THEOREM 3.1. *Suppose $A > 0$ and $B > 0$ in $C^*(X)$. Then*

$$\|x\|^2\, \nabla(x) = \frac{\mathrm{Vol}(C^0(x))}{2B^2}\{2B(p-2) - 1\} + o(1) \quad \text{as } \|x\| \to \infty,$$

where $\mathrm{Vol}(C^0(x)) = k_0 k_p c^{p/2}/p$, *which is independent of* x.

If in addition $p \geq 3$ and $B > 1/2(p-2)$, then $C^*(x)$ has an asymptotically smaller volume than that of $C^0(x)$ in the sense that

$$\lim_{\|x\| \to \infty} \left[\|x\|^2\, \nabla(x)\right] > 0.$$

THEOREM 3.2. *If $A > 0$ and $B > 0$ in $C^*(X)$, then*

$$(A + B\|x\|^2)\, \nabla(x) = \frac{\mathrm{Vol}(C^0(x))}{2B}\{2B(p-2) - 1\} + o(1) \quad \text{as } A + B\|x\|^2 \to \infty.$$

If in addition $p \geq 3$ and $B > 1/2(p-2)$, then

$$\lim_{A + B\|x\|^2 \to \infty} \left[(A + B\|x\|^2)\, \nabla(x)\right] > 0.$$

REMARK 3.2. The second theorem says that if B is fixed such that $B > 1/2(p-2)$, we can find a constant A_0 large enough that the proposed confidence set $C^*(X)$ dominates the set $C^0(X)$ in volume, *uniformly* in $\|x\|$, if $A > A_0$. In what follows we construct a moderate value of A under which $C^*(X)$ dominates $C^0(X)$. In this way $C^*(X)$ is significantly smaller than $C^0(X)$ when $\|X\|$ is not large. (See Section 6.)

4. The geometry of $C^*(x)$. In the normality setting, one property that should reasonably be required of a confidence region is that it be connected. In this section, a sufficient condition under which $C^*(x)$ is connected for all x is derived. We are aware of the possibility of improvement upon the sufficient conditions to be presented below. However, numerical results show that they are reasonably good in practice, so we do not pursue further in this direction.

The following theorem gives conditions under which $C^*(x)$ is a connected set for all x. The proof for this theorem is long and is given in the Appendix.

THEOREM 4.1. *The set $C^*(X)$ is connected if A and B are chosen such that* $A \geq 0.5$, $B > 0$ *and*

$$c(b) \leq p\sqrt{c}\,\min\left\{B(3A + B - 1), \frac{A^2(A + 3B + 1)}{B}\right\},$$

where $b = 1/(4AB)$.

To understand the geometry of $C^*(X)$ better, we present in the following some interesting properties concerning its shape. It is clear that the shape of $C^*(x)$ depends on x only through $\|x\|$. Thus, to graph the shape of $C^*(x)$ at a given x, we identify the x-axis with the observation x and vary β, the angle between x and $\theta - x$ from 0 to π; then (9) will give a sector of $C^*(x)$. The rest of the set is then generated by rotating the sector about the x-axis. Let $\mathscr{P}$ be the two-dimensional plane with x as one of its axes. The intersection of $\mathscr{P}$ and the boundary of $C^*(x)$, and also that of $C^0(x)$, is graphed in Figure 1 for various values of $\|x\|$ when $p = 3$ and $\alpha = 0.05$ with $A = 1$ and $B = 0.5$. Note that in this case $b = 0.5$, $c = 7.8147$ and $c(b) = 8.4509$. It is then easy to see that conditions for the connectedness of $C^*(x)$ for all x hold for this choice of A and B. In fact, $C^*(x)$ appears to also be convex in Figure 1.

Note that when $x = 0$, $C^*(x)$ is a smaller p-sphere contained in $C^0(x)$, and for small to moderate values of $\|x\|$ they are more like "egg-shaped" sets. As $\|x\|$ gets large $C^*(x)$ becomes more like a p-sphere again; in fact as $\|x\|$ tends to infinity $C^*(x)$ and $C^0(x)$ tend to coincide. It is interesting to see that, in the direction of x, $C^*(x)$ is wider than $C^0(x)$, while it is narrower than $C^0(x)$ in the direction which is perpendicular to x. This is a desirable property for a confidence region of multivariate normal means, as is explained in Berger [(1980), page 735].

Another important result we have on the geometry of $C^*(x)$ is the following lemma, which is used in proving the uniform dominance of $C^*(X)$ over $C^0(X)$ in Section 5. The lemma shows that the radius of the set $C^*(x)$ increases as the angle β between x and $\theta - x$ increases from 0 to π. For fixed $\beta \in [0, \pi]$, let $\varphi(\beta)$ denote this radius; that is, φ solves

$$\varphi^2 + \lambda + 2\gamma\varphi(\varphi + D\cos(\beta)) = c(\lambda). \tag{9}$$

LEMMA 4.1. *For $p \geq 3$, suppose A and B are selected such that $B \geq 1/2(p - 2)$, $A \geq 1$, $b \leq c$ and the following inequalities hold*:

$$A + 1 \geq Bc(b) + \frac{1}{4A} + 2B\sqrt{bc(b)},$$

$$A + 1 \geq Bc(b) + \frac{c(b)}{p},$$

$$A + 3 \geq Bc(b) + \frac{c(b)}{p} + \frac{1}{4A} + 2B\sqrt{bc(b)}$$

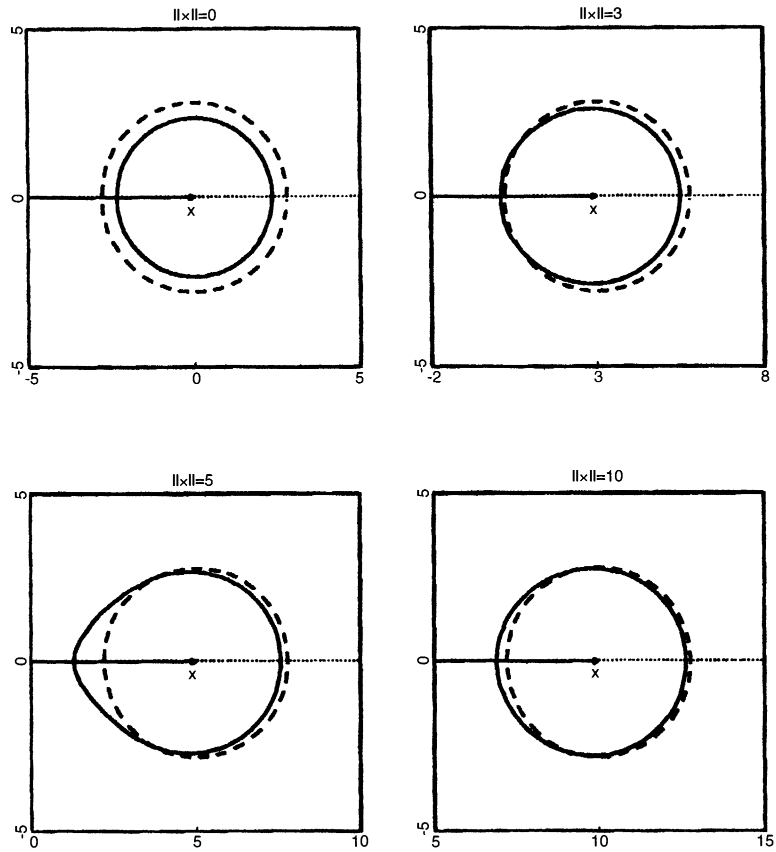

FIG. 1. *The boundaries of* $C^*(x)$ *(solid line) and* $C^0(x)$ *(dashed line) on the two-dimensional plane with x as one of its axes, when* $p = 3$ *and* $\alpha = 0.05$ *with* $A = 1$ *and* $B = 0.5$ *in* $C^*(X)$.

and

$$\left(A - 1 + \frac{c}{p}\right)\left[\frac{p(p-b)(p+2)}{bc + Bp(p-b)(p+2)}\right] \geq c(b).$$

Then $\varphi(B)$ *is a nondecreasing function of* β.

See Tseng and Brown (1995) for the proof, which also shows that $\varphi(\beta)$ is increasing when $A = 1$.

Table 1 gives the values of A and B satisfying the conditions of Theorem 4.1 and Lemma 4.1 for given p and α.

More remarks about the conditions are in order before we leave this section.

TABLE 1
Values of A satisfying conditions of Theorem 4.1 (∗) *and Lemma* 4.1 (•) *for* $B = 1/(p-2)$ *and* $B = 1/(2(p-2))$

	B = 1 / (p − 2)				**B = 1 / (2(p − 2))**			
	α = 0.05		**α = 0.1**		**α = 0.05**		**α = 0.1**	
p	*	•	*	•	*	•	*	•
3	0.50	9.52	0.50	7.43	0.93	5.71	0.86	4.42
4	0.78	6.27	0.73	4.99	1.49	4.06	1.39	3.22
5	1.01	5.09	0.95	4.20	1.87	3.45	1.75	2.80
6	1.17	4.46	1.11	3.79	2.16	3.12	2.03	2.57
7	1.31	4.05	1.24	3.54	2.40	2.90	2.27	2.43
8	1.42	3.77	1.35	3.36	2.61	2.75	2.47	2.33
9	1.51	3.56	1.44	3.24	2.79	2.63	2.64	2.27
10	1.60	3.44	1.52	3.14	2.96	2.54	2.81	2.22
11	1.68	3.33	1.60	3.06	3.11	2.46	2.95	2.19
12	1.75	3.25	1.67	3.00	3.25	2.40	3.09	2.16
13	1.82	3.18	1.74	2.95	3.37	2.35	3.22	2.13
14	1.88	3.12	1.80	2.90	3.49	2.30	3.33	2.11
15	1.94	3.07	1.86	2.86	3.61	2.26	3.45	2.09
16	1.99	3.03	1.91	2.83	3.72	2.22	3.55	2.07
17	2.05	2.99	1.97	2.80	3.82	2.19	3.66	2.06
18	2.10	2.96	2.02	2.77	3.92	2.16	3.76	2.04
19	2.15	2.93	2.06	2.75	4.01	2.14	3.85	2.03
20	2.19	2.90	2.11	2.73	4.11	2.11	3.94	2.02
21	2.24	2.87	2.15	2.71	4.19	2.10	4.03	2.01
22	2.28	2.85	2.20	2.69	4.28	2.09	4.11	2.00

REMARK 4.1. To understand the conditions of Theorem 4.1 and Lemma 4.1 better, note that for fixed p, α and B these conditions hold for all $A_2 \geq A_1$ if they hold when $A = A_1$.

To see why, note that when B is fixed $b = 1/(4AB)$ is a decreasing function of A. Hence $c(b)$ is also decreasing in A. However,

$$p\sqrt{c}\,\min\left\{B(3A + B - 1), \frac{A^2(A + 3B + 1)}{B}\right\}$$

is an increasing function of A. Also note that all the left-hand sides of the four inequalities in Lemma 4.1 increase with A, while the right-hand sides decrease with A.

REMARK 4.2. It is interesting to see how the volume of $C^*(x)$ changes with various choices for the constants A and B. In Figure 2 plots of the volume of $C^*(x)$ against that of $C^0(x)$ are given for $p = 3$ with several different values of A and B. The improvement in volume for small $\|x\|$ is substantial when both A and B are small. The largest improvement in

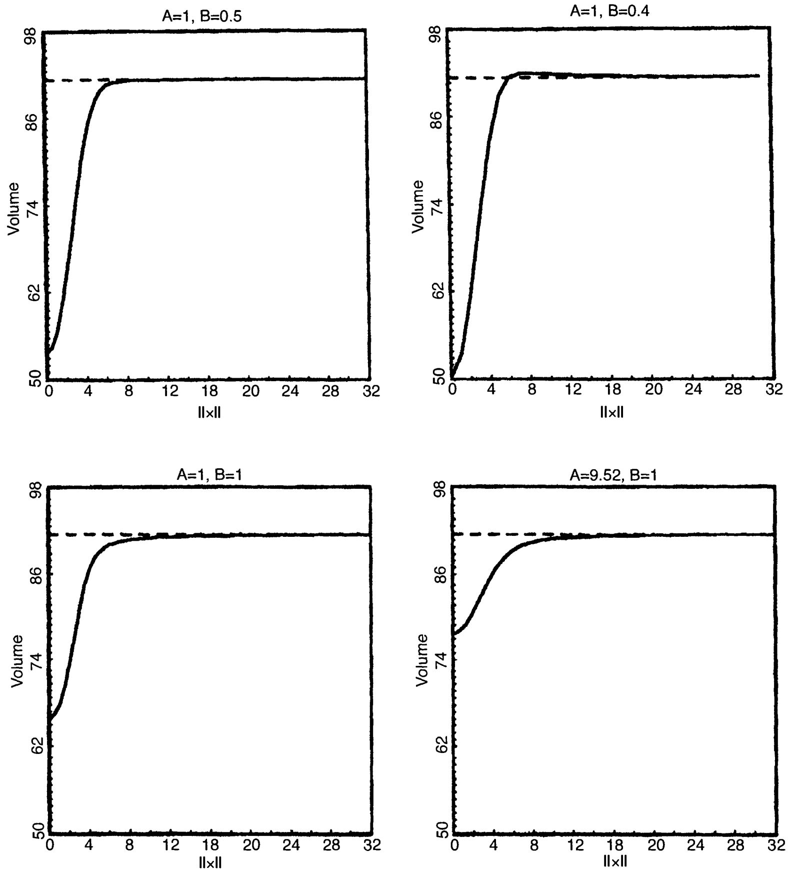

FIG. 2. *Plot of the volume of* $C^*(x)$ *(solid line) and that of* $C^0(x)$ *(dashed line) when* $\alpha = 0.05$ *and* $p = 3$, *with various choices of A and B in* $C^*(X)$.

volume is when $x = 0$ for all A and B. It is easy to see that, for a fixed A (B), smaller B (A) gives better performance for small $\|x\|$, at the minor price of slightly worse performance for large $\|x\|$. In fact, this is as expected if we recall that the variance parameter τ^2 in the normal prior we used was replaced by $A + B\|\theta\|^2$ in the construction of $C^*(X)$.

In the graph having $B = 0.4$, which is smaller than $1/2(p - 2) = 0.5$, it is interesting to see that, even though the improvement at $x = 0$ is larger than that when $B = 0.5$, $C^*(x)$ has larger volume than $C^0(x)$ at moderate values of $\|x\|$. This is not surprising in view of Theorem 3.1 or Corollary 5.1.

From the two remarks above, for fixed B we will choose the smallest value of A which satisfies the required conditions in Theorem 4.1 and Lemma 4.1 in later development of this paper.

REMARK 4.3. Easy calculations show that the conditions of Theorem 4.1 imply the set $C^*(x)$ is a p-sphere, a convex set, when $x = 0$.

REMARK 4.4. It is interesting to note that the condition $b \le c$ implies that $x \in C^*(x)$ for all x. Hence, for given p and α, if the constants A and B are chosen such that $b = 1/(4AB) \le c$, then for all x the set $C^*(x)$ is not an empty set.

5. Uniform domination. We have seen in Section 2 that $C^*(X)$ has a constant coverage probability $1 - \alpha$ at any given value θ. To show it dominates $C^0(X)$ under (6) one needs to prove that $C^*(x)$ has smaller volume than $C^0(x)$ for all x in some set with positive Lebesgue measure. In this section, we present an algorithm for a computer aided proof that $C^*(x)$ has uniformly smaller volume than $C^0(x)$.

First we describe the analytical results which are needed in the algorithm. To simplify the presentation hereafter, for given $\beta \in [0, \pi]$ and $D = \|x\| \ge 0$ we use the following notation in this section:

$$\begin{aligned}
V_*(x) &= \mathrm{Vol}(C^*(x)),\\
V_0 &= \mathrm{Vol}(C^0(x)),\\
J(D) &= \frac{d}{dD}\mathrm{Vol}(C^*(x)),\\
V_{*,c}(x) &= \text{the computer output for } \mathrm{Vol}(C^*(x))\\
&= V_*(x) + \varepsilon,\\
V_{0,c} &= \text{the computer output for } \mathrm{Vol}(C^0(x))\\
&= V_0 + \varepsilon,
\end{aligned}$$

where ε denotes the generic error term due to numerical integrations. Also let $\varphi(\beta|D)$ be the φ solving the equality

$$\varphi^2 + \lambda + 2\gamma\varphi(\varphi + D\cos\beta) = c(\lambda) \tag{10}$$

and

$$\psi(\beta|D) = \frac{d}{dD}\varphi(\beta|D).$$

We use the following results to prove the uniform dominance of $C^*(X)$ over $C^0(X)$:

(i) at $x = 0$, $\mathrm{Vol}(C^*(0)) < \mathrm{Vol}(C^0(0))$ (see Theorem 5.1);
(ii) a uniform upper bound U for $J(D)$ (see Theorem 5.2).

This number U gives, by the mean value theorem, the worst $V_*(x)$ in a neighborhood of any given x_0 one can possibly have; that is, we have

$$V_*(x) \le V_*(x_0) + U(D - D_0),$$

where $D = \|x\|$ and $D_0 = \|x_0\|$. Hence the worst value for $V_*(x)$ is $V_*(x_0) + U(D - D_0)$. So if we know that $V_*(x_0) < V_0$, we can be sure that $V_*(x) < V_0$ for those x's such that

$$D \le D_0 + \frac{V_{0,c} - V_{*,c}(x_0)}{U}$$

by controlling the numerical errors such that $V_{0,c} - V_{*,c}(x_0) < V_0 - V_*(x_0)$. (see Remark 5.1.) However, from (i) we know that $V_*(0) < V_0$; so we begin with $x_0 = 0$, that is, $D_0 = 0$, and we need to calculate $V_{*,c}(x)$ for the sequence of x's with Euclidean norms $0, D_1, D_2, \dots,$ where

$$D_i = D_{i-1} + \frac{V_{0,c} - V_{*,c}(x_{i-1})}{U} \quad \text{for } i = 1, 2, \dots,$$

and for any x with $\|x\|$ smaller than the maximum D value we have calculated for $V_{*,c}(x)$ we then know that $V_*(x) \le V_0$.

To complete the proof for the uniform dominance, however, we need to be able to stop calculating for $V_{*,c}(x)$ at some point:

(iii) a value D_* such that $V_0 - V_*(x) \ge 0$ for all x with $\|x\| \ge D_*$ (see Theorem 5.3.)

Hence the maximum value of D for which we have to calculate $V_{*,c}(x)$ is D_*.

REMARK 5.1. To make sure that $|V_{0,c} - V_{*,c}(x)| < |V_0 - V_*(x)|$, a GAUSS program is used to calculate $V_{*,c}(x)$ with slightly larger radii $\varphi(\beta|D)$ for many β's and to calculate $V_{0,c}$ with a slightly smaller c. That is,

$$V_*(x) = V_{*,c}(x) - \varepsilon_1,$$
$$V_0 = V_{0,c} + \varepsilon_2,$$

where both ε_1 and ε_2 are positive error terms due to numerical integrations. Hence if at some values of x our numerical results say $V_{*,c}(x) < V_{0,c}$, then

$$V_*(x) = V_{*,c}(x) - \varepsilon_1 < V_{0,c} - \varepsilon_1 = V_0 - \varepsilon_2 - \varepsilon_1 < V_0;$$

that is, at those x's, the volume of $C^*(x)$ is truly smaller than $C^0(x)$. A macro called INTSIMP in GAUSS is used to calculate

$$\mathrm{Vol}(C^*(x)) = \frac{k_0}{p}\int_0^{\pi} \sin^{p-2}\beta\varphi^p(\beta|D)\,d\beta,$$

using a slightly larger $\varphi(\beta|D)$ found by a dichotomy method using (10). The program is run on a 486DX personal computer and the error term in INTSIMP is set to be at the order of 10^{-7}.

Now, we present our analytical results for (i) and (ii). Notice that we assume the conditions of Lemma 4.1 in the theorem and lemma below; therefore the uniform dominance theorem holds only for $C^*(X)$ with properly selected constants A and B.

THEOREM 5.1. *Assume the conditions of Lemma* 4.1. *Then* $C^*(0)$ *is a subset of* $C^0(0)$, *that is*, $\mathrm{Vol}(C^*(0)) < \mathrm{Vol}(C^0(0))$.

PROOF. At $x = 0$, $C^0(0) = \{\theta\colon\ t \le c\}$ and $C^*(0) = \{\theta\colon\ t(1+\gamma)^2 \le c(\lambda)\}$; recall that $t = \|\theta\|^2$, $\gamma = (A + Bt)^{-1}$ and $\lambda = t\gamma^2$. It suffices to show, for $t > c$,

$$t(1+\gamma)^2 > c(\lambda). \tag{11}$$

Note that we need only the condition

$$A + 1 \ge Bc(b) + \frac{c(b)}{p}$$

since it implies

$$2A + 1 + 2Bc - c(b)/p > 0.$$

This proves (11) since $c(\lambda) \le c + c(b)\lambda/p$, $\gamma t = (A + Bt)\lambda$ and hence (11) holds if

$$t - c + \lambda\left\{2A + 2Bt + 1 - \frac{c(b)}{p}\right\} > 0$$

for $t > c$. □

THEOREM 5.2. *Let* $J(D) = (d/dD)(\mathrm{Vol}(C^*(x))$, *where* $D = \|x\|$. *Assume the conditions of Lemma* 4.1 *and* $A \ge 2(p-2)/c$. *Then*

$$J(D) \le \frac{p\pi^{p/2}M_1^{p-1}}{\Gamma(p/2+1)},$$

where $m_1 = \sqrt{c(b)} + \sqrt{b}$.

Two preliminary results are useful in proving this theorem. Their proofs are given in the Appendix.

LEMMA 5.1. *Assume the conditions of Lemma* 4.1. *For given* $\beta \in [0, \pi]$ *and* $D \ge 0$, *let* $\varphi(\beta|D)$ *be the* φ *solving the equality*

$$\varphi^2 + \lambda + 2\gamma\varphi(\varphi + D\cos\beta) = c(\lambda).$$

Then

$$M_0 \le \varphi(\beta|D) \le M_1 \qquad \forall\, \beta \text{ and } D,$$

where $M_0 = \sqrt{c} - \sqrt{b}$.

LEMMA 5.2. *Assume the conditions of Lemma* 4.1. *If in addition* $A \ge 2(p-2)/c$, *then*

$$\psi(\beta|D) \le 1 \qquad \forall\, \beta \in (0, \pi) \text{ and } D \ge 0,$$

where $\psi(\beta|D) = (d/dD)\varphi(\beta|D)$.

PROOF OF THEOREM 5.2. We have

$$\mathrm{Vol}(C^*(x)) = \frac{k_0}{p}\int_0^{\pi} \sin^{p-2} \beta \varphi^p(\beta|D)\, d\beta,$$

by the definition of $\varphi(\beta|D)$, and this implies

$$\begin{aligned} J(D) &= k_0 \int_0^{\pi} \sin^{p-2} \beta \varphi^{p-1}(\beta|D) \psi(\beta|D)\, d\beta \\ &\le k_0 M_1^{p-1} \int_0^{\pi} \sin^{p-2} \beta \psi(\beta|D)\, d\beta \quad \text{(by Lemma 5.1)} \\ &\le k_0 k_p M_1^{p-1} \quad \text{(by Lemma 5.2)} \\ &= \frac{p \pi^{p/2} M_1^{p-1}}{\Gamma(p/2+1)}, \end{aligned}$$

where the last equality is implied by an easy calculation for $k_0 k_p$. □

The next result we need is to find a D_* such that $V_0 - V_*(x) \ge 0$ for all x with $\|x\| \ge D_*$. For given p and α, we use the following theoretical approach to find a number for D_*:

If we correctly bound all of the remaining terms of all the approximation equalities we have in the proof of Theorem 3.1, we have the following inequality:

$$V_0 - V_*(x) \ge D^{-2} L_1 - D^{-4} L_2,$$

where

$$L_1 = k_0 k_p c^{p/2} \frac{2B(p-2) - 1}{2pB^2}$$

and L_2 is a positive constant resulting from the bound for the remaining terms. Therefore, to find D_* such that

$$V_0 - V_*(x) \ge 0 \qquad \forall\, x \text{ s.t. } \|x\| \ge D_*,$$

it is sufficient to find the smallest D such that

$$D^{-2} L_1 - D^{-4} L_2 \ge 0.$$

From this, a theoretical value for D_* is $\sqrt{L_2/L_1}$.

There are several untidy constants resulting from the bounds for all the D^{-4} or for smaller order terms when calculating $\mathrm{Vol}(C^0(x)) - \mathrm{Vol}(C^*(x))$. The calculations are very long and tedious. Hence, for convenience, we present here only the consequent theorem and for the details readers are referred to Tseng (1994).

THEOREM 5.3. *Assume the conditions of Lemma* 4.1. *Then*

$$\mathrm{Vol}(C^0(x)) - \mathrm{Vol}(C^*(x)) \ge 0 \qquad \forall\, x \text{ with } \|x\| \ge D_*,$$

where D_* *is a constant depending on* A, B, p *and* α.

Table 2 gives the values D_* with A and B satisfying the conditions of Lemma 4.1 for given p and α. It can be seen that Theorem 5.3 gives reasonably small values for D_*. This makes the described process for proving the uniform dominance of $C^*(X)$ upon $C^0(X)$ computationally feasible. For example, when $p = 3$ and $\alpha = 0.05$, D_* is 49.208. This means that, for all x with $\|x\| \geq 49.208$, we know that $C^*(x)$ has a smaller volume than $C^0(x)$. Therefore, to finish the whole proof we only need to follow (ii) and calculate $\text{Vol}(C^*(x))$ at various values of $\|x\|$ smaller than 49.208.

Notice that both the constant U in (ii) and D_* in (iii) depend on the dimension p and the confidence coefficient $1 - \alpha$. Therefore, for a given pair of p and α we can calculate a sequence of $V_{*,c}(x)$'s in the way described in (ii) and (iii) and complete the proof that $C^*(x)$ has uniformly smaller volume than $C^0(x)$. Thus we have the following corollary.

COROLLARY 5.1. *Assume the conditions in Lemma* 4.1 *and* $A \geq 2(p-2)/c$. *If the procedure described above is followed completely for a given pair of* $0 < \alpha < 1$ *and* $p \geq 3$. *Then* $C^*(x)$ *has uniformly smaller volume than* $C^0(x)$ *for those values of* α *and* p.

As a consequence, $C^*(X)$ is also minimax.

TABLE 2
Theoretical values of D_*

		$\alpha = 0.05$		$\alpha = 0.1$	
p	B	A	D_*	A	D_*
3	1	9.52	49.208	7.43	45.849
4	0.5	6.27	52.228	4.99	49.753
5	1/3	5.09	57.159	4.20	55.542
6	0.25	4.46	64.588	3.79	60.984
7	0.2	4.05	69.304	3.54	67.151
8	1/7	3.77	73.971	3.36	74.265
9	1/6	3.56	82.770	3.24	79.965
10	0.125	3.44	86.908	3.14	87.772
11	1/9	3.33	96.261	3.06	97.164
12	0.1	3.25	101.40	3.00	102.394
13	1/11	3.18	110.032	2.95	111.228
14	1/12	3.12	120.061	2.90	121.314
15	1/13	3.07	130,381	2.86	131.684
16	1/14	3.03	140.997	2.83	142.316
17	1/15	2.99	147.280	2.80	153.273
18	1/16	2.96	163.102	2.77	164.560
19	1/17	2.93	174.618	2.75	176.065
20	1/18	2.90	186.431	2.73	189.317
21	1/19	2.87	198.597	2.71	207.157
22	0.05	2.85	210.930	2.69	219.788

Note: B is taken to $1/(p-2)$ here.

TABLE 3
Ratio of the effective radii (RER) of $C^*(x)$ and $C^0(x)$ at $x = 0$

		$\alpha = 0.05$		$\alpha = 0.1$	
p	B	A	RER	A	RER
3	1	9.52	0.946	7.43	0.932
4	0.5	6.27	0.919	4.99	0.901
5	1/3	5.09	0.90	4.20	0.882
6	0.25	4.46	0.888	3.79	0.869
7	0.2	4.05	0.878	3.54	0.860
8	1/6	3.77	0.869	3.36	0.852
9	1/7	3.56	0.862	3.24	0.846
10	0.125	3.44	0.857	3.14	0.841
11	1/9	3.33	0.852	3.06	0.837
12	0.1	3.25	0.848	3.00	0.834
13	1/11	3.18	0.845	2.95	0.831
14	1/12	3.12	0.842	2.90	0.828
15	1/13	3.07	0.839	2.86	0.825
16	1/14	3.03	0.837	2.83	0.823
17	1/15	2.99	0.834	2.80	0.821
18	1/16	2.96	0.832	2.77	0.819
19	1/17	2.93	0.830	2.75	0.817
20	1/18	2.90	0.828	2.73	0.816

Note: B is taken to $1/(p-2)$ here.

For an illustration of using (i)–(iii) to prove the uniform dominance of $C^*(X)$ over $C^0(X)$ for various values of p and α, the reader is referred to Tseng (1994). In short, the described algorithm is completely followed in Tseng (1994) for $\alpha = 0.05$ and $p = 3, 4, 5, 9$ or 10. Table 3 gives the ratio of the effective radii (this is defined in Section 6) of $C^*(x)$, with A and B satisfying the conditions of Corollary 5.1, and $C^0(x)$ at $x = 0$ for $\alpha = 0.05$ or 0.1 and $p = 3, \dots, 20$.

6. Comparison with other confidence regions. As mentioned in the Introduction, several other confidence regions have been proposed since the works of Brown (1966) and Joshi (1967). Unfortunately, it is impossible to make the comparison of $C^*(X)$ with all of them. Therefore, here we compare our confidence sets with the usual one, the ones given in Berger (1980) and those considered by Casella and Hwang (1983).

Comparisons between the two confidence sets are made in terms of the ratio of the effective radii of these sets. More precisely, the ratio of the effective radii of $C_1(x)$ and $C_2(x)$ is defined as

$$\left[\frac{\mathrm{Vol}(C_1(x))}{\mathrm{Vol}(C_2(x))}\right]^{1/p}.$$

This is proposed by Faith (1976) to make the comparison between confidence sets more nearly independent of the dimension p.

TABLE 4
Ratio of the effective radii of $C^*(X)$ and $C^0(X)$

	Norm of x							
p	0	1	2	4	6	8	10	20
3	0.837	0.853	0.892	0.971	0.997	0.999	0.999	0.999
6	0.745	0.761	0.798	0.885	0.947	0.979	0.991	0.999
12	0.699	0.709	0.734	0.805	0.872	0.923	0.956	0.996

Note: $\alpha = 0.05$, $A = 1$ and $B = 1/2(p - 2))$ for each p.

In Table 4 the ratio of the effective radii of $C^*(x)$, with $A = 1$ and $B = 1/2(p - 2)$, and $C^0(x)$ is given for various values of $\|x\|$ with $p = 3$, 6 and 12 and $\alpha = 0.05$. It can be seen that $C^*(x)$ is clearly smaller than $C^0(x)$. In particular, the improvement can be substantial for small $\|x\|$ and large p. This suggests that $C^*(X)$ is superior to $C^0(X)$ for small area estimation problems.

Table 5 gives the comparison of our confidence sets with the ones in Berger (1980) and Casella and Hwang (1983) for $p = 6$ and 12 with $\alpha = 0.1$. The constant A in $C^*(X)$ is taken to be 1 and two choices of B are considered; $B = 1/2(p - 2)$ and $B = 1/(p - 2)$. Since their confidence sets have higher coverage probability, by trading coverage probability for volume it is not surprising that $C^*(x)$ has smaller volume. Table 5 shows that, for both $B = 1/2(p - 2)$ and $B = 1/(p - 2)$, this is the case. In these comparisons we observe the following: for a fixed p, the improvement is greater for

TABLE 5
Comparison of $C^*(X)$ with those in Berger (1980) *and Casella and Hwang* (1983): *ratio of the effective radii with $C^0(x)$*

	Norm of x								
	0	1	2	4	6	8	10	20	50
	$p = 6$								
Casella and Hwang	0.881	0.881	0.881	0.943	0.978	0.989	0.993	0.998	0.999
Berger	0.816	0.822	0.840	0.908	0.960	0.978	0.986	0.997	0.999
$C^*(x)$, $B = 1/8$	0.717	0.736	0.780	0.879	0.947	0.980	0.992	0.999	0.999
$C^*(x)$, $B = 1/4$	0.775	0.793	0.833	0.912	0.958	0.977	0.996	0.999	0.999
	$p = 12$								
Casella and Hwang	0.821	0.821	0.821	0.821	0.941	0.971	0.983	0.996	0.999
Berger	0.764	0.766	0.772	0.806	0.875	0.931	0.957	0.990	0.998
$C^*(x)$, $B = 1/20$	0.679	0.690	0.717	0.794	0.868	0.922	0.956	0.996	0.999
$C^*(x)$, $B = 1/10$	0.729	0.741	0.769	0.840	0.898	0.935	0.957	0.989	0.998

Note: $\alpha = 0.1$ and $A = 1$.

TABLE 6
Ratio of the effective radii of $C^*(X)$ and $C^0(X)$ when $p = 6$ and $\alpha = 0.1$ with A and B satisfying the conditions of Theorem 4.1 and Lemma 4.1

	Norm of x								
	0	**1**	**2**	**4**	**6**	**8**	**10**	**20**	**50**
Casella and Hwang	0.881	0.881	0.881	0.943	0.978	0.989	0.993	0.998	0.999
Berger	0.816	0.822	0.840	0.908	0.960	0.978	0.986	0.997	0.999
$C^*(x)$, $B = 0.125$									
$A = 2.03$	0.788	0.797	0.823	0.889	0.943	0.973	0.987	0.999	0.999
$A = 2.57$	0.811	0.819	0.839	0.895	0.943	0.971	0.986	0.998	0.999
$C^*(x)$, $B = 0.25$									
$A = 1.11$	0.782	0.799	0.836	0.912	0.957	0.977	0.985	0.996	0.999
$A = 3.79$	0.869	0.875	0.890	0.929	0.959	0.976	0.985	0.996	0.999

smaller $\|x\|$. When $\|x\|$ is large, these three sets are not very different from each other; neither are they significantly different from the usual one.

Note that, even though the numerical results show $C^*(X)$ is better than $C^0(X)$ for the A and B used in Tables 4 and 5, we have not proved it dominates $C^0(X)$ for those values since they do not satisfy the conditions of Lemma 4.1. Table 6 gives the ratio of the effective radii of $C^*(x)$ and $C^0(x)$ when $p = 6$ and $\alpha = 0.1$ with A and B satisfying the conditions of Theorem 4.1 and Lemma 4.1 (see Table 1). With these choices of A and B, $C^*(x)$ still has satisfactory performance except for the case where $A = 3.79$ and $B = 0.25$. In that case, Berger's confidence region, which has, however, a probability of coverage smaller than 0.9 for a range of θ, has smaller volume when $\|x\|$ is small. Nevertheless, this indicates that the sufficient conditions we have for the dominance of $C^*(X)$ over $C^0(X)$ are not entirely satisfactory. There appears to be room for improvement and we defer this problem for future study.

APPENDIX

PROOF OF THEOREM 3.1. First note that both $\gamma = \gamma(t)$ and λ depend on θ only through $\|\theta\|$. Then

$$\theta \in C^0(x) \quad (C^*(x)) \quad \text{iff } P\theta \in C^0(Px)\ (C^*(Px)),$$

where P is any $p \times p$ orthogonal matrix. Therefore, without loss of generality we can assume that $x = (x_1, 0, \dots, 0)'$, where $x_1 \geq 0$. Let $\varphi = \|\theta - x\|$, $D = \|x\| = x_1$, $\beta \in [0, \pi]$ s.t. $(\theta - x)'x = \varphi D \cos\beta$, and $\eta = \varphi + D\cos\beta$. Then

$$\begin{aligned}
&\|x - \theta(1 + \gamma(t))\|^2 \\
&\quad = \|x - \theta\|^2 + \gamma^2(t)\|\theta\|^2 - 2\gamma(t)\theta'(x - \theta) \\
&\quad = \varphi^2 + \gamma^2(t)t - 2\gamma(t)(\theta - x)'(x - \theta) - 2\gamma(t)x'(x - \theta) \\
&\quad = \varphi^2 + \lambda + 2\gamma\varphi\eta.
\end{aligned}$$

Thus, for any fixed x,

$$C^*(x) = \{\theta: \|x - \theta(1 + \gamma(t))\|^2 \le c(\lambda)\}$$
$$= \{\theta: \varphi, \beta \text{ s.t. } \varphi^2 \le c(\lambda) - \lambda - 2\gamma\varphi\eta\}.$$

The idea is for any given x to approximate the term $c(\lambda) - \lambda - 2\gamma\varphi\eta$, for θ on the boundary of $C^*(x)$, by a function of D and β, say $R(D, \beta)$. Then the difference in volume between $C^0(x)$ and $C^*(x)$ can be approximated as

$$\frac{k_0}{p}\int_0^{\pi} \sin^{p-2}(\beta)\left[c^{p/2} - R^{p/2}(D, \beta)\right] d\beta,$$

which under the described condition is then shown to be a $O(D^{-2})$ term with positive coefficient if $p \ge 3$.

Define

$$W = \frac{A + B(\varphi^2 + 2\varphi D\cos\beta)}{BD^2 + A + B(\varphi^2 + 2\varphi D\cos\beta)}.$$

Then it is easy to see that $W = O(D^{-1})$ as $D \to \infty$, since A and B are constants and φ is bounded.

Now, by the definition of γ,

$$\begin{aligned}
\gamma &= \frac{1}{A + B(D^2 + \varphi^2 + 2\varphi D\cos\beta)}\\
&= \frac{1}{BD^2}(1 - W)\\
&= \frac{1}{BD^2}\left[1 - \left(\frac{2\varphi\cos\beta}{D} + \frac{A + B\varphi^2}{BD^2} - \frac{(A + B(\varphi^2 + 2\varphi D\cos\beta))W}{BD^2}\right)\right]\\
&= \frac{1}{BD^2}\left[1 - \left(\frac{2\varphi\cos\beta}{D}\right)\right] + o(D^{-3}) \quad \text{since } W = O(D^{-1}).
\end{aligned} \tag{12}$$

This implies that

$$\begin{aligned}
\lambda = t\gamma^2(t) &= (D^2 + \varphi^2 + 2\varphi D\cos\beta)\gamma^2(t)\\
&= \frac{1}{B^2D^2} + \frac{1}{B^2D^2}(W^2 - 2W) + \frac{\varphi^2 + 2\varphi D\cos\beta}{B^2D^4}(1 - W)^2.
\end{aligned}$$

Hence

$$\lambda = \frac{1}{B^2D^2} + o(D^{-2}). \tag{13}$$

Now, by (12) and easy calculations, we have

$$
\begin{aligned}
2\varphi\gamma\eta &= \frac{2\varphi\eta}{BD^2}(1-W) \\
&= 2\varphi\eta\frac{1}{BD^2}\left[1-\frac{2\varphi\cos\beta}{D}\right]+o(D^{-3}) \quad \text{[by (12)]} \\
&= \frac{2\varphi^2-4\varphi^2\cos^2\beta}{BD^2}+\frac{2\varphi\cos\beta}{BD}+o(D^{-3}).
\end{aligned} \tag{14}
$$

For the term $c(\lambda)$, the Taylor series approximation gives

$$
\begin{aligned}
c(\lambda) &= c+\frac{d}{d\lambda}c(\lambda)\bigg|_{\lambda=0}\lambda+o(\lambda), \quad (\text{for } \lambda\approx 0) \\
&= c+\frac{f_{p+2,\lambda}(c(\lambda))}{f_{p,\lambda}(c(\lambda))}\bigg|_{\lambda=0}\lambda+o(\lambda) \quad (\text{by Lemma 3.1}) \\
&= c+\frac{c}{pB^2D^2}+o(D^{-2}),
\end{aligned} \tag{15}
$$

where the last equality is implied by (13).

Therefore by (13), (14) and (15), on the boundary of $C^*(x)$ we have the following equalities:

$$
\begin{aligned}
\varphi^2 &= c(\lambda)-\lambda-2\gamma\varphi\eta \\
&= c+\frac{c}{pB^2D^2}-\frac{1}{B^2D^2}-\frac{2\varphi^2-4\varphi^2\cos^2\beta}{BD^2}-\frac{2\varphi\cos\beta}{BD}+o(D^{-2}) \\
&= c-\frac{2\varphi\cos\beta}{BD}+o(D^{-1}),
\end{aligned}
$$

which implies

$$
\begin{aligned}
\varphi &= \sqrt{c-\frac{2\varphi\cos\beta}{BD}+o(D^{-1})} \\
&= \sqrt{c}-\frac{\varphi\cos\beta}{BD\sqrt{c}}+o(D^{-1}) \\
&= \sqrt{c}-\frac{\cos\beta}{BD}+o(D^{-1}).
\end{aligned}
$$

Applying this to (14) we have

$$
\begin{aligned}
2\varphi\gamma\eta &= \frac{2\varphi^2-4\varphi^2\cos^2\beta}{BD^2}+\frac{2\varphi\cos\beta}{BD}+o(D^{-2}) \\
&= \frac{2}{BD^2}\left(c+\sqrt{c}\,D\cos\beta-\frac{\cos^2\beta}{B}-2c\cos^2\beta\right)+o(D^{-2}).
\end{aligned} \tag{16}
$$

Define

$$\Delta(\beta) = \frac{c - p - 2pB(c + \sqrt{c}D\cos\beta) + 2p\cos^2\beta + 4pBc\cos^2\beta}{pB^2D^2}.$$

Then $\Delta(\beta) = O(D^{-1})$.

Hence by applying (13), (15) and (16) we have, for large D,

$$c(\lambda) - \lambda - 2\gamma\varphi\eta = R(D, \beta),$$

where

$$R(D, \beta) = c + \Delta(\beta) + o(D^{-2}).$$

Therefore when D is large

$$\begin{aligned} C^*(x) &= \{\theta\colon \varphi, \beta \text{ s.t. } \varphi^2 \le R(D, \beta)\} \\ &= \{\theta\colon \varphi, \beta \text{ s.t. } \varphi^2 \le c + \Delta(\beta) + o(D^{-2})\}. \end{aligned}$$

Note that a Taylor series approximation gives

$$\begin{aligned} &\left(1 + \frac{\Delta(\beta) + o(D^{-2})}{c}\right)^{p/2} \\ &= 1 + \frac{p}{2}\frac{\Delta(\beta) + o(D^{-2})}{c} + \frac{1}{2}\frac{p}{2}\left(\frac{p}{2} - 1\right)\left(\frac{\Delta(\beta) + o(D^{-2})}{C}\right)^2 \qquad (17) \\ &\quad + o(D^{-2}) \\ &= 1 + \frac{p}{2}\frac{\Delta(\beta)}{c} + \frac{p(p-2)}{8}\left(\frac{\Delta(\beta)}{c}\right)^2 + o(D^{-2}). \end{aligned}$$

Now we can calculate the volume difference for large D as

$$\begin{aligned} \nabla(x) &= \int_{\{\theta \in C^0(x)\}} d\theta - \int_{\{\theta \in C^*(x)\}} d\theta \\ &= \frac{k_0}{p}\int_0^{\pi} \sin^{p-2}(\beta)\left[c^{p/2} - \left(c + \Delta(\beta) + o(D^{-2})\right)^{p/2}\right] d\beta \\ &= \frac{k_0}{p}c^{p/2}\int_0^{\pi} \sin^{p-2}(\beta)\left[1 - \left(1 + \frac{\Delta(\beta) + o(D^{-2})}{c}\right)^{p/2}\right] d\beta \\ &= k_0 c^{p/2}\int_0^{\pi} \sin^{p-2}(\beta)\left[-\frac{\Delta(\beta)}{2c} - \frac{(p-2)\,\Delta^2(\beta)}{8c^2}\right] d\beta + o(D^{-2}) \\ &= k_0 c^{p/2}\int_0^{\pi} \sin^{p-2}(\beta)\left[-\frac{\Delta(\beta)}{2c} - \frac{p-2}{8c^2}\frac{4c\cos^2\beta}{B^2D^2}\right] d\beta + o(D^{-2}) \\ &= \frac{k_0 k_p c^{p/2}}{2pB^2D^2}\{2B(p-2) - 1\} + o(D^{-2}), \end{aligned}$$

where the fourth equality followed from (17) and the last equality followed from the facts that

$$\int_0^{\pi} \sin^{p-2}(\beta)\cos\beta\, d\beta = 0 \quad \text{and} \quad \int_0^{\pi} \sin^{p-2}(\beta)\cos^2\beta\, d\beta = \frac{k_p}{p}.$$

That is,

$$D^2\,\nabla(x) = \frac{\mathrm{Vol}(C^0(x))}{2B^2}\{2B(p-2)-1\} + o(1),$$

since it is easy to see that

$$\mathrm{Vol}(C^0(x)) = \int_{R^p} I[\theta \in C^0(x)]\, d\theta = \frac{k_0 k_p c^{p/2}}{p}.$$

But for $p \ge 3$ and $B > 0$

$$\{2B(p-2)-1\} > 0 \quad \text{iff } B > \frac{1}{2(p-2)}.$$

This completes the proof. □

Theorem 3.1 of Casella and Hwang (1983) is used in the proof of Theorem 4.1. The translated form of the Casella–Hwang theorem in the present context is the following.

THEOREM A.1 [Casella and Hwang (1983)]. *The set $C^*(x)$ is connected for any given x iff the set $C(D)$ is an interval for any given $D \ge 0$, where*

$$C(D) = \{\tau \ge 0\colon H(\tau|D) \le c(\lambda)\},$$

$$H(\tau|D) = \left(\tau + \frac{\tau}{A + B\tau^2} - D\right)^2 \quad \text{and} \quad \lambda = \frac{\tau^2}{(A+B\tau^2)^2}.$$

PROOF OF THEOREM 4.1. Let $\gamma(\tau) = 1/(A + B\tau^2)$ and $g(\tau) = \tau + \tau\gamma(\tau)$. Then

$$H(\tau|D) = (g(\tau) - D)^2,$$

and

$$\frac{d}{d\tau}H(\tau|D) = 2[g(\tau) - D][A^2 + A + B^2\tau^4 + B\tau^2(2A-1)]\gamma^2(\tau).$$

An easy calculation shows that $g(\tau)$ is strictly increasing from 0 to ∞ in $[0,\infty)$ provided that $A \ge 0.5$; hence there is a unique $\tau_0 = \tau_0(D)$ such that $H(\tau_0|D) = 0$.

Since $\lambda = \tau^2/(A + B\tau^2)^2$, we have

$$\frac{d}{d\tau}\lambda = 2\tau(A - B\tau^2)(A + B\tau^2)^{-3},$$

which has zeros at 0 and τ^*, where $\tau^* = \sqrt{A/B}$.

Also it is easy to check that $(d^2/d\tau^2)\lambda|_{\tau=\tau^*} < 0$. Thus we have the following facts about the noncentrality λ as a function of τ: (1) it is an increasing function on $[0, \tau^*]$; (2) it is a decreasing function on $[\tau^*, \infty)$; and (3) it has its maximal value b, which is defined to be $\lambda|_{\tau=\tau^*} = 1/(4AB)$, at $\tau^* = \sqrt{A/B}$. As a result, and applying Lemma 3.1, we have

$$\frac{d}{d\tau}c(\lambda) = \left[\frac{d}{d\tau}\lambda\right]\frac{d}{d\lambda}c(\lambda)$$

$$= \left[\frac{d}{d\tau}\lambda\right]\frac{f_{p+2,\lambda}(c(\lambda))}{f_{p,\lambda}(c(\lambda))}\begin{cases} > 0, & \text{on } [0, \tau^*), \\ < 0, & \text{on } (\tau^*, \infty), \\ = 0, & \text{at } \tau^*, \end{cases}$$

since $f_{p+2,\lambda}(c(\lambda))/f_{p,\lambda}(c(\lambda))$ is a positive number.

Hence the function $c(\lambda)$ behaves similarly to the function λ; that is, as a function of of τ, $c(\lambda)$ increases on $[0, \tau^*]$, decreases on $[\tau^*, \infty)$, has its maximal value $c(b)$ at $\tau^* = \sqrt{A/B}$ and also tends to c as τ tends to 0 or ∞ since λ tends to 0 in both cases.

The goal is to prove the set $C(D)$ is an interval for any given $D \geq 0$, where

$$C(D) = \{\tau \geq 0 : H(\tau|D) \leq c(\lambda)\}.$$

Let the two sets $C_1(D)$ and $C_2(D)$ be defined as

$$C_1(D) = C(D) \cap \{\tau : \tau \leq \tau_0\},$$
$$C_2(D) = C(D) \cap \{\tau : \tau \geq \tau_0\}.$$

Then $C(D) = C_1(D) \cup C_2(D)$.

Note that

$$H(\tau_0|D) = 0 \quad \text{by the definition of } \tau_0, \text{ and } c(\lambda) \geq 0 \text{ for all } \tau \geq 0,$$

which implies

$$\tau_0 \in C_1(D) \cap C_2(D) \quad \text{that is, } C_1(D) \cap C_2(D) \neq \varnothing,$$

and we have the following fact:

FACT A.1. $C(D)$ is an interval iff $C_1(D)$ and $C_2(D)$ are intervals.

CLAIM A.1. $C_1(D)$ is an interval.

PROOF OF THE CLAIM. On $C_1(D)$, $\tau \leq \tau_0$.

There are two possibilities for the relationship between τ^* and τ_0:

Case 1. $\tau_0 \leq \tau^*$. In this case, the function $H(\tau|D) - c(\lambda)$ is decreasing on $[0, \tau_0]$, thus $C_1(D)$ is an interval.

Case 2. $\tau_0 > \tau^*$. We have the following subcases.

Case 2.1. $H(\tau|D) \leq c$ *for all* $\tau \in [0, \tau_0]$. Obviously $C_1(D)$ is an interval.

Case 2.2. $H(\tau|D) > c$ *for some* $\tau \in [0, \tau_0]$. Then there is a unique number

$\tau_1 = \tau_1(D)$ such that $H(\tau_1|D) = c$ since $H(\tau|D)$ is strictly decreasing on $[0, \tau_0)$ and $H(\tau_0|D) = 0$.

Case 2.2.1. $\tau_1 \le \tau^*$. Obviously $C_1(D)$ is an interval.

Hence, we are left to consider the case that $H(\tau|D) > c$ for some τ in $[0, \tau_0]$ and $\tau_0 \ge \tau_1 > \tau^*$.

To prove that $C_1(D)$ is an interval for the case that $H(\tau|D) > c$ for some $\tau \in [0, \tau_0]$ and $\tau_0 \ge \tau_1 > \tau^*$, it suffices to prove that the function $H(\tau|D) - c(\lambda)$ is decreasing on $[0, \tau_1]$; in fact we only need to prove that on $[\tau^*, \tau_1]$, since we can write $C_1(D) = \{C_1(D) \cap [0, \tau_1]\} \cup [\tau_1, \tau_0]$ and $H(\tau|D) - c(\lambda)$ is decreasing on $[0, \tau^*)$.

Define $W(\tau) = (d/d\tau)H(\tau|D) - (d/d\tau)c(\lambda)$, for $\tau \ge 0$. Then we want to show that $W(\tau) < 0$ on $[\tau^*, \tau_1]$ if A and B are selected as described in the theorem.

Note that on $(\tau^*, \tau_1]$, $B\tau^2 - A > 0$,

$$
\begin{aligned}
\frac{1}{2}\gamma^{-2}(\tau)W(\tau)
&= [g(\tau) - D][A^2 + A + B^2\tau^4 + B\tau^2(2A - 1)] \\
&\quad + \frac{\tau(B\tau^2 - A)}{A + B\tau^2}\frac{f_{p+2,\lambda}(c(\lambda))}{f_{p,\lambda}(c(\lambda))} \\
&< [g(\tau) - D][A^2 + A + B^2\tau^4 + B\tau^2(2A - 1)] + \frac{\tau(B\tau^2 - A)}{A + B\tau^2}\frac{c(b)}{p} \\
&< -\sqrt{c}[A^2 + A + B^2\tau^4 + B\tau^2(2A - 1)] + \frac{\tau(B\tau^2 - A)}{A + B\tau^2}\frac{c(b)}{p} \\
&= \frac{1}{p(A + B\tau^2)}\mathscr{M},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathscr{M} = {} & -A\tau c(b) + B\tau^3 c(b) \\
& - p\sqrt{c}(AB\tau^2)[A^2 + A + B^2\tau^4 + B\tau^2(2A - 1)].
\end{aligned}
$$

We have applied Lemma 3.1 for the first inequality, and used the fact that

$$g(\tau) - D \le -\sqrt{c} \quad \text{on } [\tau^*, \tau_1],$$

by the definition of τ_1, for the second inequality.

We want to show that $\mathscr{M} \le 0$, hence $W(\tau) < 0$, on $[\tau^*, \tau_1]$ if A and B are selected as described in the theorem. Two cases are to be considered:

Case 1. $\tau \ge 1$. Hence $\tau^3 \le \tau^n$, $\forall\, n \ge 3$. Thus,

$$
\begin{aligned}
\mathscr{M} &\le \tau^3\{Bc(b) - p\sqrt{c}[AB^2 + B^3 + B^2(2A - 1)]\} \\
&\le 0 \quad \text{if } c(b) \le p\sqrt{c}B(3A + B - 1).
\end{aligned}
$$

Case 2. $\tau < 1$. Hence $\tau^3 \le \tau^n$, $\forall\ n \le 3$. Thus

$$\mathscr{M} \le \tau^3\{Bc(b) - p\sqrt{c}\,[A(A^2 + A + 2AB - B) + B(A^2 + A)]\}$$
$$\le 0 \quad \text{if } c(b) \le p\sqrt{c}\,\frac{A^2(A + 3B + 1)}{B}.$$

This proves that $C_1(D)$ is an interval, for all $D \ge 0$ provided that A and B satisfy the condition described in the theorem. By similar arguments, we can prove that $C_2(D)$ is an interval, for all $D \ge 0$, under the same condition, by proving that

$$W(\tau) > 0 \quad \text{on } [\tau_2, \tau^*],$$

where

$$\tau_2 \text{ uniquely solves } H(\tau|D) = c \text{ on } (\tau_0, \tau^*).$$

Therefore, $C(D)$ is an interval, for all $D \ge 0$ provided that A and B satisfy the condition described in the theorem. Finally, applying Theorem A.1 we have proved that the confidence set $C^*(x)$ is connected for any $x \in R^p$ if A and B satisfy the condition described in the theorem. □

PROOF OF LEMMA 5.1. From the proof of Lemma 4.1, we have, for all $D \ge 0$,

$$\varphi_0 \le \varphi(\beta|D) \le \varphi_1 \ \forall\, \beta \in (0, \pi),$$

where $\varphi_0 = \varphi(0|D)$ and $\varphi_1 = \varphi(\pi|D)$.

Hence if we show that

$$\varphi_0 \ge M_0 \quad \text{and} \quad \varphi_1 \le M_1,$$

then we prove this lemma.

First note that, when $\beta = \pi$, $t = \|\theta\|^2 = D^2 + \varphi_1^2 + 2\varphi_1 D\cos\beta = (\varphi_1 - D)^2$, $\eta = \varphi_1 - D$ and $\gamma = (A + Bt)^{-1} = (A + B(\varphi_1 - D)^2)^{-1}$. By the definition of φ_1 we have

$$\varphi_1^2 + t\gamma^2 + 2\gamma\varphi_1(\varphi_1 - D) = c(t\gamma^2).$$

However, $c \le c(t\gamma^2) \le c(b)$ for all $D \ge 0$, we then have

$$\varphi_1^2 + t\gamma^2 + 2\gamma\varphi_1(\varphi_1 - D) \le c(b),$$

which implies

$$\sqrt{c(b)} \ge \varphi_1 + \gamma(\varphi_1 - D)$$
$$\ge \varphi_1 + \inf_{s \in \mathscr{R}} \frac{s}{A + Bs^2}$$
$$= \varphi_1 - \sqrt{b}$$

since it is easy to show that

$$\inf_{s \in \mathscr{R}} \frac{s}{A + Bs^2} = -\sqrt{b}$$

Hence we have proved the inequality that

$$\varphi_1 \le \sqrt{c(b)} = \sqrt{b} = M_1.$$

Now, when $\beta = 0$, $t = \|\theta\|^2 = D^2 + \varphi_0^2 + 2\varphi_0 D \cos\beta = (\varphi_0 + D)^2$, $\eta = \varphi_0 + D$ and $\gamma = (A + Bt)^{-1} = (A + B(\varphi_0 + D)^2)^{-1}$. Then similarly we have

$$c \le \varphi_0^2 + t\gamma^2 + 2\gamma\varphi_0(\varphi_0 + D),$$

which implies

$$\begin{aligned}\sqrt{c} &\le \varphi_0 + \gamma(\varphi_0 + D)\\ &\le \varphi_0 + \sup_{s\ge 0} \frac{s}{A + Bs^2}\\ &= \varphi_0 + \sqrt{b},\end{aligned}$$

since it is equally easy to show that

$$\sup_{s\ge 0} \frac{s}{A + Bs^2} = \sqrt{b}.$$

Thus we have proved the other inequality that

$$\varphi_0 \ge \sqrt{c} - \sqrt{b} = M_0.$$

This completes the proof for Lemma 5.1. □

Proof of Lemma 5.2. We will use the same notation as in Lemma 5.1. Straightforward calculations show that

$$\psi(\beta|D) = \frac{-\mathscr{K}D - (\gamma + \mathscr{K})\varphi\cos\beta}{\varphi + \gamma\varphi + \eta(\gamma + \mathscr{K})},$$

where $\mathscr{K} = \gamma^3(1 - (d/d\lambda)c(\lambda))(A - Bt) - 2B\varphi\gamma^2\eta$ and $\varphi = \varphi(\beta|D)$.

The conditions of Lemma 4.1 imply that

$$\varphi + \gamma\varphi + \eta(\gamma + \mathscr{K}) > 0 \quad \text{and} \quad \gamma + \mathscr{K} > 0.$$

Then we have

$$\begin{aligned}&\psi(\beta|D) \le 1\\ (18)\qquad &\Leftrightarrow \quad \varphi + \gamma\varphi + \eta(\gamma + \mathscr{K}) + \mathscr{K}D + (\gamma + \mathscr{K})\varphi\cos\beta \ge 0\\ &\Leftrightarrow \quad \varphi + \gamma\varphi + (\gamma + \mathscr{K})\varphi(1 + \cos\beta) + \mathscr{K}D + (\gamma + \mathscr{K})D\cos\beta \ge 0.\end{aligned}$$

Note that since $1 + \cos\beta \ge 0$ and $\gamma + \mathscr{K} > 0$, we have

$$\begin{aligned}(19)\qquad &\varphi + \gamma\varphi + (\gamma + \mathscr{K})\varphi(1 + \cos\beta) + \mathscr{K}D + (\gamma + \mathscr{K})D\cos\beta\\ &\ge \varphi + \gamma\varphi + \mathscr{K}D - (\gamma + \mathscr{K})D = \varphi + \gamma(\varphi - D).\end{aligned}$$

Note that, when $\varphi < D$, we have

$$\begin{aligned}\gamma = (A + Bt)^{-1} &\le \frac{1}{A + B(\varphi - D)^2}\\ \Rightarrow \quad \gamma(\varphi - D) &\ge \frac{\varphi - D}{A + B(\varphi - D)^2} \ge -\sqrt{b},\end{aligned}$$

where the last inequality was proved in Lemma 5.1. Also $\varphi \geq M_0$ by Lemma 5.1; hence

$$\varphi + \gamma(\varphi - D) \geq \begin{cases} M_0 \geq 0, & \text{if } \varphi \geq D, \\ M_0 - \sqrt{b} \geq 0, & \text{if } \varphi < D, \end{cases}$$

where

$$M_0 - \sqrt{b} = \sqrt{c} - 2\sqrt{b} \geq 0 \quad \text{since } A \geq \frac{2(p-2)}{c} \text{ by the assumption.}$$

Then, by (18) and (19), we have proved Lemma 5.2. □

REFERENCES

BERGER, J. O. (1980). A robust generalized Bayes estimate and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, Berlin.

BIRNBAUM, A. (1955). Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Ann. Math. Statist.* **26** 21–36.

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

BROWN, L. D., CASELLA, G. and HWANG, J. T. G. (1995). Optimal confidence sets, bioequivalence and the limacon of pascal. *J. Amer. Statist. Assoc.* **90** 880–889.

CASELLA, G. and HWANG, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78** 688–698.

CASELLA, G. and HWANG, J. T. (1986). Confidence sets and the Stein effect. *Comm. Statist. Theory Methods* **15** 2043–2063.

COHEN, A. and STRAWDERMAN, W. E. (1973). Admissibility implications for different criteria in confidence estimation. *Ann. Statist.* **1** 363–366.

FAITH, R. E. (1976). Minimax Bayes set and point estimators of a multivariate normal mean. Technical Report 66, Univ. Michigan.

GHOSH, J. K. (1961). On the relation among shortest confidence intervals of different types. *Calcutta Statist. Assoc. Bull.* 147–152.

HUWANG, L. (1991). Good coverage probability confidence sets in linear errors-in-variables models. Ph.D. dissertation, Statistics Center, Cornell Univ.

HUWANG, L. (1995). Interval estimation in structural error-in-variables model with partial replication. *J. Multivariate Anal.* **55** 230–245.

HWANG, J. T. G. (1995). Fieller's problems and resampling techniques. *Statist. Sinica* **5** 161–172.

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

HWANG, J. T. and CASELLA, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Decisions (Suppl.)* **1** 3–16.

HWANG, J. T. and CHEN, J. (1986). Improved confidence sets for the coefficients of a linear model with spherically symmetric errors. *Ann. Statist.* **14** 444–460.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss function. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

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

JOSHI, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40** 1042–1067.

PRATT, J. W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 541–567.

SHINOZAKI, N. (1989). Improved confidence sets for the mean of a multivariate normal distribution. *Ann. Inst. Statist. Math.* **41** 331–346.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

STEIN, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–296.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

TSENG, Y. L. (1994). Good exact confidence sets and minimax estimators for the mean vector of a multivariate normal distribution, Ph.D. dissertation, Statistics Center, Cornell Univ.

TSENG, Y. L. and BROWN, L. D. (1995). Good exact confidence sets for the mean vector of a multivariate normal distribution. Technical report, Inst. Statistical Science, Academia Sinica, Taipei.

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI 11529, TAIWAN
E-MAIL: yltseng@stat.sinica.edu.tw

DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302
E-MAIL: lbrown@compstat.wharton.upenn.edu