IMPROVED CONFIDENCE SETS FOR THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

Nobuo Shinozaki

Tokyo Keizai University, Kokubunji, Tokyo 185, Japan

Abstract. A new class of confidence sets for the mean of a p-variate normal distribution ($p \geq 3$) is introduced. They are neither spheres nor ellipsoids. We show that we can construct our confidence sets so that their coverage probabilities are equal to the specified confidence coefficient. Some of them are shown to dominate the usual confidence set, a sphere centered at the observations. Numerical results are also given which show how small their volumes are.

Key words and phrases: Stein-type estimator, coverage probability, uniform improvement.

1. Introduction

Let $X = (X_1, \ldots, X_p)^t$ be a p-variate normal random variable with mean vector $\theta = (\theta_1, \ldots, \theta_p)^t$ and identity covariance matrix. Since Stein (1955) proved that the usual point estimator of θ, X, can be improved upon under the sum of squared error if $p \geq 3$, a great deal of research has been made to improve upon X. However, the problem of the confidence set has received comparatively little attention and its theory has seen advances only in recent years.

The usual confidence set for the mean is a sphere centered at X, i.e.,

$$(1.1) \qquad C^0(X) = \{\eta : \|\eta - X\| \leq c\},$$

where the radius c satisfies $P\{\chi_p^2 \leq c^2\} = 1 - \alpha$. This implies that the coverage probability of $C^0(X)$

$$P_\theta\{\theta \in C^0(X)\} = 1 - \alpha \quad \text{for all } \theta.$$

So $C^0(X)$ has the confidence coefficient $1 - \alpha$.

As in Casella and Hwang (1983) we consider a confidence set $C(X)$ to

be an improvement upon $C^0(X)$ if

(I) $P_\theta\{\theta \in C(X)\} \geq P_\theta\{\theta \in C^0(X)\}$ for all θ

and

(II)   volume of $C(X) \leq$ volume of $C^0(X)$ for all X,

with strict inequality either in (I) for some θ or in (II) for all X in some set with positive Lebesgue measure.

Stein (1962) developed the heuristic argument to indicate that $C^0(X)$ can be improved upon for large p. Brown (1966) and Joshi (1967) independently proved the existence of a dominating confidence set for $p \geq 3$. It was shown that $C^0(X)$ is dominated by the confidence sphere which has the same radius c and is centered at a Stein-type estimator. However, they did not give explicit improved confidence sets.

Attempts to construct usable improved confidence sets were made by several authors including Faith (1976), Berger (1980), Stein (1981), Hwang and Casella (1982, 1984) and Casella and Hwang (1983, 1986). Berger (1980) developed confidence ellipsoids associated with his robust generalized Bayes estimator. Although uniform dominance results were not obtained, he gave convincing analytical and numerical evidence that his confidence sets perform satisfactorily.

Hwang and Casella (1982) considered the confidence sphere which has radius c and is centered at a positive-part Stein estimator $\{1 - a/\|X\|^2\}^+ X$, and showed analytically that it is an improvement upon $C^0(X)$ for a specified range of values of a when $p \geq 4$. Hwang and Casella (1984) gave an alternative proof which yielded stronger results. Although their confidence set provides uniform improvement in coverage probability, it has the same volume and confidence coefficient as $C^0(X)$.

It would be desirable for the improving confidence set to have the same confidence coefficient as $C^0(X)$ but to have smaller volume than $C^0(X)$. Casella and Hwang (1983, 1986) constructed confidence spheres with variable radii through empirical Bayes considerations. They evaluated their coverage probabilities numerically and claimed that their confidence spheres dominate $C^0(X)$. However, no confidence set with variable volume seems to be available which has been analytically shown to be an improvement upon $C^0(X)$.

Here we introduce a class of confidence sets and analytically show that some of them improve upon $C^0(X)$. We construct the new confidence sets by shrinking the set $C^0(X)$ towards the origin. They are neither spheres nor ellipsoids, and there isn't a natural center for the new set, which might be the associating point estimator with it. In Section 2 we precisely define our confidence sets and obtain the necessary representations for their volumes

and coverage probabilities. We shall see that we may easily construct our confidence set so that it has the same coverage probability as $C^0(X)$ for all θ. In Section 3 we give a confidence set which has the same coverage probability and smaller volume. We cannot express its form explicitly and have to obtain it numerically. In Section 4 we show that we can obtain the numerical solution quite easily, and also give our numerical results which show how small the volume of our confidence set is.

2. A confidence set, its volume and coverage probability

In this section we introduce a class of confidence sets and show how to represent their volumes and coverage probabilities. Further we show how to construct our confidence set so that it has the same coverage probability as $C^0(X)$.

To begin with, we suppose that $\|X\| \geq c$ and let us define β_0 by $\sin \beta_0 = c/\|X\|$ (see Fig. 1(a)). As in Hwang and Casella (1982), we transform p-dimensional variate η to spherical coordinates and let $r = \|\eta\|$ and let β be the angle between X and η. Then η is on the sphere $\|\eta - X\|^2 = c^2$ if and only if

$$r^2 - 2r\|X\| \cos \beta + \|X\|^2 - c^2 = 0 \ .$$

This is a quadratic equation in r and has two distinct roots if $0 \leq \beta < \beta_0$. We denote them by r_+ and r_- ($r_- < r_+$), that is, $r_+ = \|X\| \cos \beta + D$ and $r_- = \|X\| \cos \beta - D$, where $D = \{c^2 - \|X\|^2 \sin^2 \beta\}^{1/2}$. We construct our confidence set by shrinking the boundary points η which are on the sphere $\|\eta - X\|^2 = c^2$ towards the origin, and by considering the resulting points to form the boundary of the new confidence set. Therefore, it is enough to specify two univariate functions $f(r_-; D)$ and $g(r_+; D)$ with parameter D
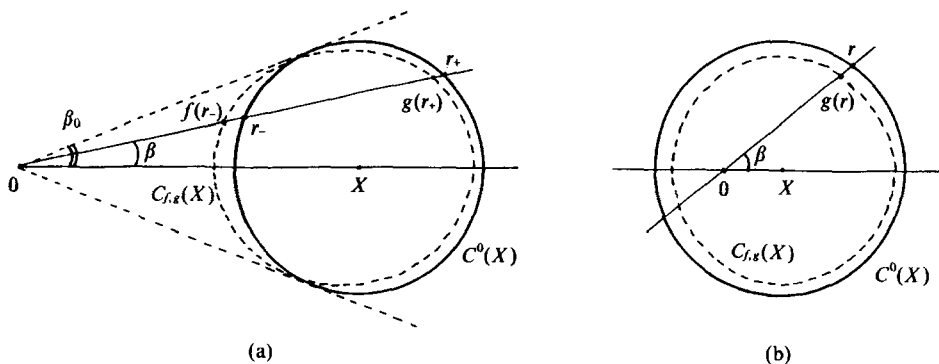

Fig. 1. Two-dimensional representation of $C^0(X)$ and $C_{f,g}(X)$. (a) Case of $\|X\| \geq c$: r_+, $g(r_+)$, r_- and $f(r_-)$ are the distances of the points from the origin. (b) Case of $\|X\| < c$: r and $g(r)$ are the distances of the points from the origin.

whose values are the distances of the resulting points from the origin. In order for this construction to make sense, it must hold true that $f(r_-; D) \leq g(r_+; D)$. This is exactly Condition (vi) given below. We occasionally drop D and write $f(r; D)(g(r; D))$ as $f(r)(g(r))$. It may be noticed that $2D = 2\{c^2 - \|X\|^2 \sin^2 \beta\}^{1/2}$ is the length of the line segment inside the sphere $C^0(X)$ which passes through the origin at the angle β with X and that D depends on both β and $\|X\|$.

For the case $\|X\| < c$ (see Fig. 1(b)), we always use the function $g(r; D)$ to shrink the points towards the origin. Therefore, if $f(r)$ and $g(r)$ are specified, we get a confidence set. Let it be denoted by $C_{f,g}(X)$. In the following we assume that $f(r; D)$ and $g(r; D)$ satisfy the following conditions.

CONDITIONS.
 (i) $f(r; D)$ and $g(r; D)$ are strictly increasing and continuous in r.
 (ii) $0 \leq f(r; D) \leq r$ and $0 \leq g(r; D) \leq r$.
 (iii) $f(0; D) = g(0; D) = 0$.
 (iv) $\lim_{r \to \infty} \{r - f(r; D)\} = \lim_{r \to \infty} \{r - g(r; D)\} = 0$.
 (v) $f(r; 0) = g(r; 0) = r$.
 (vi) $f(r; D) \leq g(r + 2D; D)$.

Now we discuss the volume of $C_{f,g}(X)$. Since $0 \leq g(r) \leq r$, it is clear that $C_{f,g}(X)$ has a smaller volume than $C^0(X)$ if $\|X\| < c$. Therefore, we give its representation only for the case $\|X\| \geq c$. By transforming to spherical coordinates we obtain

$$(2.1) \qquad \text{volume of } C_{f,g}(X) = K \int_0^{\beta_0} \int_{f(\|X\| \cos \beta - D; D)}^{g(\|X\| \cos \beta + D; D)} r^{p-1} \sin^{p-2} \beta \, dr \, d\beta \,,$$

where $\beta_0 = \arcsin(c/\|X\|)$, $D = \{c^2 - \|X\|^2 \sin^2 \beta\}^{1/2}$ and $K = 2 \cdot \prod_{i=0}^{p-3} \left\{ \int_0^\pi \sin^i t \, dt \right\}$. We give a sufficient condition for $C_{f,g}(X)$ to have smaller volume than $C^0(X)$ in the following lemma.

LEMMA 2.1. *If*

$$g(r + D) < r + D - \{r - D - f(r - D)\} \left(\frac{r - D}{r + D} \right)^{p-1} \quad \text{for all} \quad r > D \,,$$

then $C_{f,g}(X)$ has smaller volume than $C^0(X)$ for every X.

We give the proof of this lemma and those of Lemmas 3.1, 3.2 and 3.3 in the Appendix.

To give a representation of the coverage probability, we define

$$B^0(\theta) = \{X: \|X - \theta\| \le c\}$$

and

$$B_{f,g}(\theta) = \{X: \theta \in C_{f,g}(X)\} \, .$$

Since we use the equality $P_\theta\{\theta \in C_{f,g}(X)\} = P_\theta\{X \in B_{f,g}(\theta)\}$ as is often the case, we have to describe $B_{f,g}(\theta)$ explicitly. For that purpose we first note that $X \in B^0(\theta)(B_{f,g}(\theta))$ if and only if $PX \in B^0(P\theta)(B_{f,g}(P\theta))$, where P is an orthogonal matrix. Therefore, without loss of generality we may assume that θ is of the form $\theta = (\theta_1, 0, \ldots, 0)'$, where $\theta_1 \ge 0$. For $X = (X_1, \ldots, X_p)$, put $z = (X_2^2 + \cdots + X_p^2)^{1/2}$. If $z > c$, $X \notin B_{f,g}(\theta)$. Assuming that $z \le c$, let us define X_1^+ and X_1^- by

$$f(X_1^+ - (c^2 - z^2)^{1/2}; (c^2 - z^2)^{1/2}) = \theta_1$$

and

$$g(X_1^- + (c^2 - z^2)^{1/2}; (c^2 - z^2)^{1/2}) = \theta_1 \, .$$

Equivalently, if we denote the inverse function of $f(\cdot; D)(g(\cdot; D))$ (which really exist from Conditions (i), (iii) and (iv)) by $f^{-1}(\cdot; D) = f^{-1}(\cdot)(g^{-1}(\cdot; D) = g^{-1}(\cdot))$,

$$X_1^+ = f^{-1}(\theta_1; (c^2 - z^2)^{1/2}) + (c^2 - z^2)^{1/2}$$

and

$$X_1^- = g^{-1}(\theta_1; (c^2 - z^2)^{1/2}) - (c^2 - z^2)^{1/2} \, .$$

We can easily check that $X_1^- \le X_1^+$ because of Condition (vi). Thus, we see that for fixed $(X_2, \ldots, X_p)$ with the property $z \le c$, $(X_1, X_2, \ldots, X_p) \in B_{f,g}(X)$ if and only if $X_1 \in [X_1^-, X_1^+]$ because f and g are strictly increasing (see Fig. 2). We may notice that if $z = c$, $(\theta_1, X_2, \ldots, X_p)$ is the common boundary point of $B^0(\theta)$ and $B_{f,g}(\theta)$ because of Condition (v). Therefore, we have

$$B_{f,g}(\theta) = \{X: z \le c, X_1^- \le X_1 \le X_1^+\} \, .$$

We also notice that $B_{f,g}(0) = B^0(0)$.

As mentioned before, using the equality $P_\theta\{\theta \in C_{f,g}(X)\} = P_\theta\{X \in B_{f,g}(\theta)\}$, we have
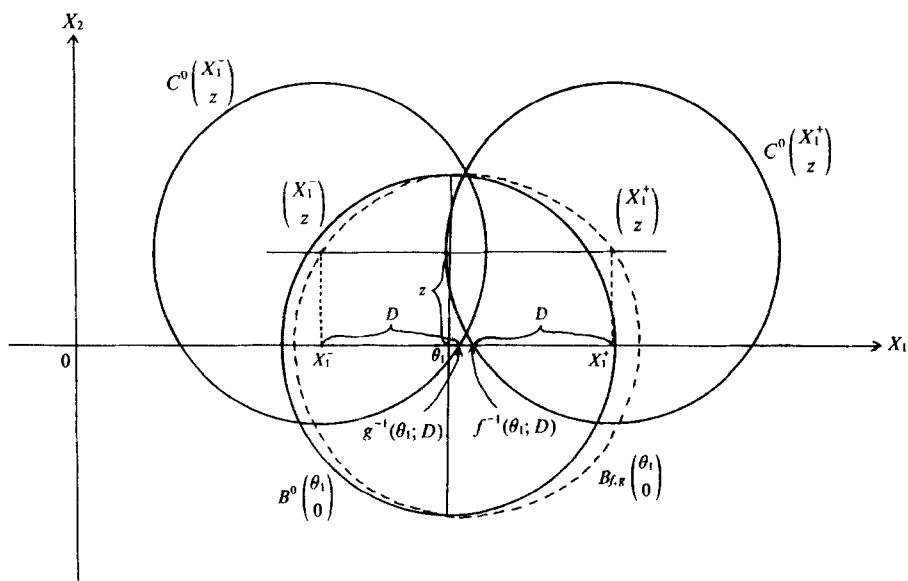
Fig. 2. Two-dimensional representation of $B^0(\theta)$ and $B_{f,g}(\theta)$.

(2.2) $\quad P_\theta\{\theta \in C_{f,g}(X)\}$

$$= \int \cdots \int_{z^2 \leq c^2} \prod_{i=2}^{p} \phi(x_i) \int_{g^{-1}(\theta_1;(c^2-z^2)^{1/2})-(c^2-z^2)^{1/2}}^{f^{-1}(\theta_1;(c^2-z^2)^{1/2})+(c^2-z^2)^{1/2}} \phi(x_1 - \theta_1) dx_1 dx_2 \cdots dx_p ,$$

where $z^2 = x_2^2 + \cdots + x_p^2$ and $\phi(\cdot)$ denotes the standard normal density. If $f(r; D)$ and $g(r; D)$ depend on D in a pathological manner, the inner integral of (2.2) may not even be a measurable function of z. However, we will choose f and g so that the integral is a well-behaved function of z and the expression (2.2) makes sense. From (2.2) we see that a sufficient condition for $P_\theta\{\theta \in C_{f,g}(X)\} \geq P_\theta\{\theta \in C^0(X)\}$ for all θ is that

(2.3) $\quad \int_{g^{-1}(\theta_1;D)-D}^{f^{-1}(\theta_1;D)+D} \phi(x_1 - \theta_1) dx_1 \geq \int_{\theta_1-D}^{\theta_1+D} \phi(x_1 - \theta_1) dx_1 ,$

for any $0 \leq D \leq c$ and for any $\theta_1 \geq 0$.

As a matter of fact, we may choose a function g for given f so that the equality holds in (2.3); that is,

(2.4) $\quad \Phi\{f^{-1}(\theta_1) - \theta_1 + D\} + \Phi\{\theta_1 - g^{-1}(\theta_1) + D\} - 2\Phi(D) = 0 ,$

where $\Phi(\cdot)$ is the standard normal distribution function. It should be noted that if f satisfies Conditions (i)–(v), then the function g defined by (2.4) also satisfies them and Condition (vi) as well. So we can always obtain a confidence set with the same coverage probability as $C^0(X)$ so far

as f satisfies Condition (i)–(v). Further, if f is appropriately chosen, we may have a confidence set which has the same coverage probability and also smaller volume. We discuss this in the next section.

It may be remarked that with some needed modification we may apply our method to confidence sets of the form (for example)

$$\|\theta - \delta(X)\| \le v(X) \,,$$

where $\delta(X)$ is some Stein-type estimator. When $v(X)$ is a constant our method does not work, and it is necessary that $\delta(X)$ and $v(X)$ are chosen so that the resulting confidence set is $C_{f,g}(X)$ for some choice of f and g. However, the corresponding f and g may not satisfy Conditions (i)–(vi) (especially Condition (v)). Even if this is the case, $C_{f,g}(X)$ may dominate $C^0(X)$. Here we do not pursue this possibility further.

3. A confidence set with the same coverage probability and smaller volume

In this section we show analytically that if we choose the function f appropriately, and if we define the function g by (2.4) for the given function f, then $C_{f,g}(X)$ improves upon $C^0(X)$. Since we define g by (2.4) for given f, $C_{f,g}(X)$ has the same coverage probability as $C^0(X)$ for every θ. Therefore, we need only to show that $C_{f,g}(X)$ has smaller volume than $C^0(X)$ for every X.

Now let a and b be some numbers which may depend on the parameter D and suppose that $a \ge b > 0$. Let

$$(3.1) \qquad\qquad f_0(r) = r - \frac{br}{a + r^2} \qquad r \ge 0 \,,$$

and let us define $g_0(r)$ by the equation

$$(3.2) \quad \Phi\{f_0^{-1}(\lambda) - \lambda + D\} + \Phi\{\lambda - g_0^{-1}(\lambda) + D\} - 2\Phi(D) = 0 \quad \lambda \ge 0 \,,$$

for the given function f_0. We note that the form of f_0 is the same as the Stein-type estimator which first appeared in Stein (1955) and was also considered in Shinozaki (1984). We can easily verify that f_0 and g_0 satisfy Conditions (i)–(iv) and (vi) given in the previous section. Since we will assume further that $b \le 4D^2$, they also satisfy Condition (v). We will give a sufficient condition on a and b for $C_{f_0,g_0}(X)$ to have smaller volume than $C^0(X)$ for every X.

Let us arbitrarily fix $r_0 > D$ and let us put $\lambda_0 = f_0(r_0 - D)$. From Lemma 2.1 we see that it suffices for us to show that

$$(3.3) \qquad g_0(r_0 + D) < r_0 + D - \frac{b(r_0 - D)}{a + (r_0 - D)^2} \left(\frac{r_0 - D}{r_0 + D} \right)^{p-1} ,$$

if a and b are appropriately chosen. To do this we give the following lemmas.

LEMMA 3.1. *If a and b satisfy the inequalities*

$$a \geq b \quad and \quad 0 < b \leq 4D^2 ,$$

and if f_0 is given by (3.1), then

$$f_0^{-1}(\lambda) > \lambda + \frac{b'}{\sqrt{\lambda^2 + 8D^2}} \quad for \ all \quad \lambda \geq \lambda_0 ,$$

where

$$(3.4) \qquad b' = b(r_0 - D)^2 / \{a + (r_0 - D)^2\} .$$

Using Lemma 3.1 we have the following lemma.

LEMMA 3.2. *If a and b satisfy the inequalities*

$$a \geq b \quad and \quad 0 < b \leq 4D^2 ,$$

and if we define g_0 by the equation (3.2) for f_0 which is given by (3.1), then

$$g_0^{-1}(\lambda) > \lambda + \frac{b'}{\sqrt{\lambda^2 + 8D^2}} \left(\frac{\lambda}{\lambda + 2D} \right)^s \quad for \ all \quad \lambda \geq \lambda_0 ,$$

where b' is given by (3.4) and $s = \max (b'/2, 1)$.

Finally from Lemma 3.2 we have the following lemma.

LEMMA 3.3. *If a and b satisfy the inequalities*

$$a \geq b \quad and \quad 0 < b \leq \min \{4D^2, 2(p - 2)\} ,$$

and if we define g_0 by the equation (3.2) for f_0 which is given by (3.1), then

$$(3.5) \qquad g_0(r + D) < r + D - \frac{b'}{r + D} \left(\frac{r - D}{r + D} \right)^{p-2} \quad for \ all \quad r \geq r_0 ,$$

where b' is given by (3.4).

If we put $r = r_0$ in (3.5), we have (3.3) which is the desired result. Therefore, we have shown the following theorem.

THEOREM 3.1. *Let a and b satisfy the inequalities*

$$a \geq b \quad and \quad 0 < b \leq \min \{4D^2, 2(p - 2)\} .$$

Let f_0 be defined by (3.1) *and for this f_0 let g_0 be defined by* (3.2). *Then $C_{f_0, g_0}(X)$ has the same coverage probability as $C^0(X)$ for every θ and has smaller volume than $C^0(X)$ for every X.*

We can easily see that if g is defined by (2.4), then $g^{-1}(\lambda) < f^{-1}(\lambda)$ for all $\lambda > 0$, which is equivalent to the condition $f(r) < g(r)$ for all $r > 0$. Therefore, we see that $g_0(\|X\| + c) > f_0(\|X\| + c) > \|X\|$, and this implies that X itself is in $C_{f_0, g_0}(X)$. We may also notice that $\{1 - b/(a + \|X\|^2)\}X$ is also in $C_{f_0, g_0}(X)$, where a and b are constants to be chosen so that they satisfy $a \geq b$ and $0 < b \leq \min \{4c^2, 2(p - 2)\}$.

To examine whether $B_{f_0, g_0}(\theta)$ is convex or not is of interest if we consider the associated hypothesis testing problem (see Casella and Hwang (1983)). However, even if we set $a = b = \min \{4D^2, 2(p - 2)\}$, for example, it seems quite difficult to settle the question, although $B_{f, g}(0) = B^0(0)$ is always convex. In general if $f(r; D)$ depends on D in a curious way and g is defined by (2.4), $B_{f, g}(\theta)$ will not be convex. Another convexity problem is that of $C_{f_0, g_0}(X)$ itself. This also seems difficult because g_0 is not explicitly defined.

4. Numerical results

In this section we discuss how to obtain the value of $g_0(r)$ defined by (3.2) and we also discuss how small the volume of $C_{f_0, g_0}(X)$ can be if we compare it with that of $C^0(X)$.

Suppose that we want to determine the value of $g_0(r)$ for a given value of r. If we put $\lambda = g_0(r)$ and if we define t by $\lambda = f_0(t)$, then the equation (3.2) can be rewritten as

$$(4.1) \qquad \Phi\{t - f_0(t) + D\} + \Phi\{f_0(t) - r + D\} - 2\Phi(D) = 0 .$$

If we can solve the equation (4.1) for t, then we can get the value of $g_0(r)$ by $g_0(r) = f_0(t)$. Since the left-hand side of (4.1) is negative for $t \leq r$, positive for $t \geq f^{-1}(r)$, and strictly increasing for $r < t < f^{-1}(r)$, we see that the solution t of (4.1) exists in the interval $(r, r + D)$ and is unique. Therefore, we can solve the equation (4.1) quite easily by some iterative method.

To get some idea about the behavior of $f_0(x; D)$ and $g_0(x; D)$, we give in Fig. 3 the two-dimensional loci of (r, β) for $C^0(X)$ and $C_{f_0, g_0}(X)$ when $p = 5$, $a = b = \min \{4D^2, 2(p - 2)\}$ and c satisfies $P\{\chi_5^2 \leq c^2\} = 0.95$. Figure 3(a) corresponds to the case $\|X\| = 0$ and $2c$, and (b) $\|X\| = c$ and $3c$. It may be seen that $g_0(r)$ differs from r by relatively little compared to $f_0(r)$, and that the value of $g_0(r)$ changes very slowly.

The ratio of volumes of $C_{f_0, g_0}(X)$ and $C^0(X)$ calculated numerically by using the expression (2.1) is given in Table 1. Again, in this case we have chosen $a = b = \min \{4D^2, 2(p - 2)\}$. We also calculated the ratio for the case $a = b = \min (2D^2, p - 2)$, but it was larger except when it was nearly equal to 1. From Table 1 it is seen that improvement is larger if p is larger, if $\|X\|$ is smaller, or if the confidence coefficient is smaller. We may say that improvement is substantial except in the case where p is small and
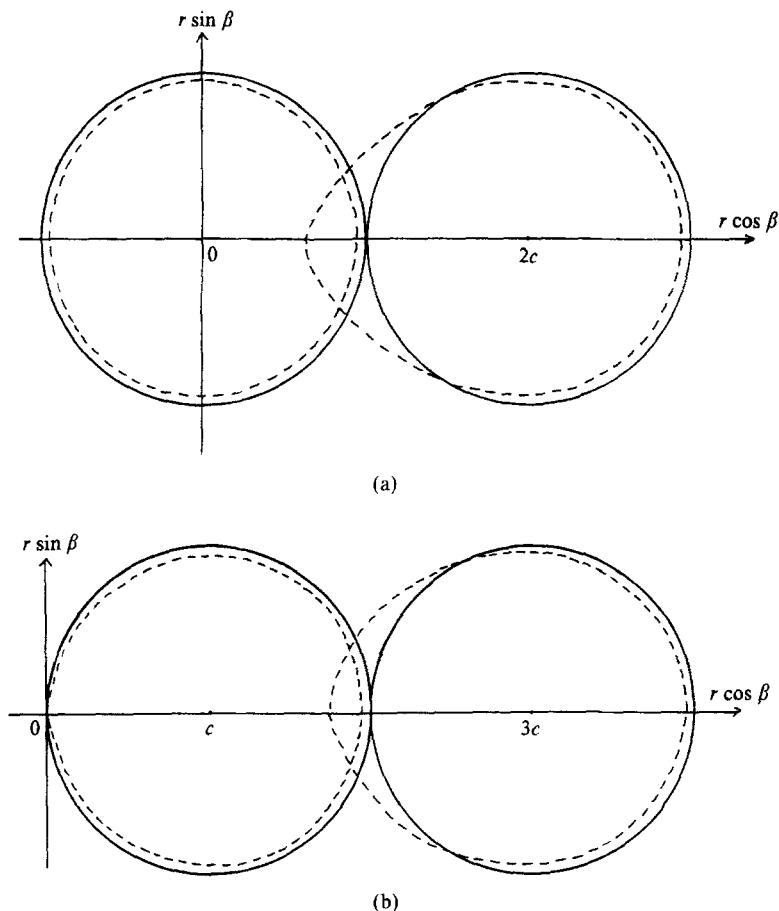

(a)

(b)

Fig. 3. Loci of (r, β) $(r = r_+$ or $r_-)$ for $C^0(X)$ (solid) and $C_{f_0, g_0}(X)$ (short dash) with $a = b = \min (4D^2, 2(p - 2))$ when $p = 5$. (a) $\|X\| = 0$ and $2c$, where c satisfies $P\{\chi_5^2 \leq c^2\} = 0.95$. (b) $\|X\| = c$ and $3c$, where c satisfies $P\{\chi_5^2 \leq c^2\} = 0.95$.

Table 1.   Volume ratio of $C_{f_0,g_0}(X)$ to $C^0(X)$ when $a = b = \min(4D^2, 2(p-2))$ and c is the constant such that $P\{\chi_p^2 \leq c^2\} = 1 - \alpha$.

$1-\alpha$	$\|X\|$	p				
		3	5	10	15	20
90%	0	.758	.695	.652	.631	.617
	c	.820	.729	.668	.642	.626
	$2c$	.955	.856	.751	.704	.673
	$3c$	.986	.933	.839	.791	.759
	$4c$	.994	.965	.895	.854	.826
	$5c$	.997	.979	.929	.895	.872
95%	0	.799	.738	.688	.663	.646
	c	.851	.767	.701	.673	.654
	$2c$	.965	.878	.775	.727	.696
	$3c$	.989	.944	.855	.807	.774
	$4c$	.995	.971	.907	.865	.837
	$5c$	.997	.983	.937	.903	.880
99%	0	.855	.799	.744	.714	.693
	c	.891	.818	.754	.721	.700
	$2c$	.978	.910	.813	.765	.734
	$3c$	.993	.960	.881	.833	.801
	$4c$	.997	.979	.924	.884	.855
	$5c$	.998	.988	.949	.917	.893

$\|X\|$ is large. Hwang and Casella (1982) have shown numerically that their confidence sphere has greatly increased coverage probability while maintaining the same volume. By trading coverage probability for volume, it should be possible to get a confidence set which has reduced volume and maintains the specified level of coverage probability. Our results seem to endorse this, although our confidence set is not a sphere.

To compare our confidence set with the ones given in Berger (1980) and Casella and Hwang (1983), we give in Table 2 the volume ratios of the three confidence sets to $C^0(X)$ when $1 - \alpha = 0.9$ and $p = 5$ and 10. Although the confidence sets of Berger and Casella and Hwang are not

Table 2.   Volume ratios of three confidence sets to $C^0(X)$ when $1 - \alpha = 0.9$ and c is the constant such that $P\{\chi_p^2 \leq c^2\} = 1 - \alpha$.

$\|X\|$	$p = 5$			$p = 10$		
	Berger	Casella-Hwang	C_{f_0,g_0}	Berger	Casella-Hwang	C_{f_0,g_0}
0	.410	.607	.695	.078	.189	.652
c	.593	.607	.729	.244	.189	.668
$2c$	.877	.905	.856	.581	.766	.751
$3c$	.946	.958	.933	.840	.896	.839
$4c$	.970	.977	.965	.891	.941	.893
$5c$	.981	.985	.979	.922	.962	.929

analytically proven to improve upon $C^0(X)$, they have smaller volume than $C_{f_0,g_0}(X)$ for $\|X\|$ near zero. This may be due to the fact that $f_0(r)$ (accordingly, also $g_0(r)$) does not shrink enough for r near zero. We may remark that f_0 given by (3.1) will not be the only choice of f which leads to the improvement upon $C^0(X)$, and another choice may lead to larger improvement. This may be a subject for further research.

Acknowledgements

Appendix

PROOF OF LEMMA 2.1.   To show that $C_{f,g}(X)$ has smaller volume than $C^0(X)$ for every X, we need only consider the case $\|X\| \geq c$. From the expression (2.1) we see that it is sufficient for us to show that

$$(A.1) \qquad \{(r+D)^p - (r-D)^p\} - \{g^p(r+D) - f^p(r-D)\} > 0,$$

for all $r > D$. We can easily verify that if $f(r-D)/(r-D) \geq g(r+D)/(r+D)$, then (A.1) holds. Therefore, we assume in the following that $f(r-D)/(r-D) < g(r+D)/(r+D)$, and show that (A.1) holds.

To evaluate $(r+D)^p - g^p(r+D)$, we use the following inequality which can be easily verified:

$$(A.2) \qquad \text{for} \quad q \geq 1, \quad 1 - (1-uv)^q \geq v\{1 - (1-u)^q\},$$

$$0 \leq u \leq 1, \quad 0 \leq v \leq 1.$$

We express $(r+D)^p - g^p(r+D)$ as $(r+D)^p\{1 - (1-uv)^p\}$ with $u = \{r - D - f(r-D)\}/(r-D)$ and $v = (r-D)\{r+D - g(r+D)\}/[(r+D)\{r-D - f(r-D)\}]$ and $(r-D)^p - f^p(r-D)$ as $(r-D)^p\{1 - (1-u)^p\}$. Noting that $0 \leq u \leq 1$ and $0 \leq v \leq 1$ and applying the inequality (A.2), we have

$$(r+D)^p - g^p(r+D) - (r-D)^p + f^p(r-D)$$

$$\geq \left[1 - \left\{ 1 - \frac{r - D - f(r-D)}{r-D} \right\}^p \right] (r+D)^{p-1}(r-D)$$

$$\cdot \left\{ \frac{r+D - g(r+D)}{r-D - f(r-D)} - \left(\frac{r-D}{r+D} \right)^{p-1} \right\},$$

which is positive from the condition of Lemma 2.1.

PROOF OF LEMMA 3.1.   Let us arbitrarily fix $\lambda(\geq \lambda_0)$ and put $r =$

$f_0^{-1}(\lambda)$. Then we have

(A.3)
$$\left(1 - \frac{b}{a + r^2}\right) r = \lambda,$$

or

(A.4)
$$r = \lambda + \frac{br^2}{a + r^2} \frac{1}{r}.$$

From (A.3) we have

$$r^2 - \lambda^2 = \frac{br^2}{a + r^2}\left(2 - \frac{b}{a + r^2}\right) < 2b \le 8D^2,$$

that is,

(A.5)
$$r < \sqrt{\lambda^2 + 8D^2}.$$

On the other hand,

(A.6)
$$\frac{br^2}{a + r^2} \ge \frac{b(r_0 - D)^2}{a + (r_0 - D)^2} = b',$$

since $r \ge r_0 - D$. Combining (A.4), (A.5) and (A.6), we have

$$r > \lambda + \frac{b'}{\sqrt{\lambda^2 + 8D^2}}.$$

PROOF OF LEMMA 3.2. From Lemma 3.1 and the equation (3.2), we see that it is sufficient for us to show that

$$\Delta(\lambda) \ge 0 \quad \text{for all} \quad \lambda \ge \lambda_0,$$

where

$$\Delta(\lambda) = \Phi\left\{\frac{b'}{\sqrt{\lambda^2 + 8D^2}} + D\right\}$$
$$+ \Phi\left\{D - \frac{b'}{\sqrt{\lambda^2 + 8D^2}}\left(\frac{\lambda}{\lambda + 2D}\right)^s\right\} - 2\Phi(D).$$

Since $\lim_{\lambda \to \infty} \Delta(\lambda) = 0$, we need only show that

$$\Delta'(\lambda) \leq 0 \quad \text{for all} \quad \lambda \geq \lambda_0 ,$$

which is equivalent to the condition

(A.7)
$$\exp\left[-\frac{1}{2}\left(D + \frac{b'}{\sqrt{\lambda^2 + 8D^2}} \right)^2 \right.$$
$$+ \frac{1}{2}\left\{ D - \frac{b'}{\sqrt{\lambda^2 + 8D^2}}\left(\frac{\lambda}{\lambda + 2D} \right)^s \right\}^2 \right]$$
$$\geq \left(\frac{\lambda}{\lambda + 2D} \right)^s \left\{ 1 - \frac{2Ds(\lambda^2 + 8D^2)}{\lambda^2(\lambda + 2D)} \right\},$$

for all $\lambda \geq \lambda_0$. (A.7) clearly holds if $1 - 2Ds(\lambda^2 + 8D^2)/\{\lambda^2(\lambda + 2D)\} \leq 0$. So, we may assume that it is positive and may take its logarithm. We can easily verify that the left-hand side of (A.7) is an increasing function of s. Therefore, if we set $s = 1$, it suffices for us to show that

(A.8)
$$-s \log\left(\frac{\lambda}{\lambda + 2D} \right) - \log\left\{ 1 - \frac{2Ds(\lambda^2 + 8D^2)}{\lambda^2(\lambda + 2D)} \right\}$$
$$- \frac{2b'D(\lambda + D)}{\sqrt{\lambda^2 + 8D^2}(\lambda + 2D)} - \frac{2b'^2 D(\lambda + D)}{(\lambda^2 + 8D^2)(\lambda + 2D)^2}$$

is nonnegative for all $\lambda \geq \lambda_0$. Since (A.8) approaches 0 as $\lambda \to \infty$, we need only show that (A.8) is a decreasing function of λ. Differentiating (A.8) with respect to λ and noting that $s \geq b'/2$, we see that (the derivative of (A.8)) $\times (\lambda + 2D)(b'D)^{-1}$ is not larger than

(A.9)
$$-\frac{1}{\lambda} - \left\{ 1 - \frac{2Ds(\lambda^2 + 8D^2)}{\lambda^2(\lambda + 2D)} \right\}^{-1} \left\{ \frac{\lambda^2 + 8D^2}{\lambda^2(\lambda + 2D)} + \frac{16D^2}{\lambda^3} \right\}$$
$$+ \frac{2\lambda(\lambda + D)}{(\lambda^2 + 8D^2)^{3/2}} - \frac{2D}{\sqrt{\lambda^2 + 8D^2}(\lambda + 2D)}$$
$$+ \frac{4b'\lambda(\lambda + D)}{(\lambda^2 + 8D^2)^2(\lambda + 2D)} + \frac{2b'\lambda}{(\lambda^2 + 8D^2)(\lambda + 2D)^2} .$$

Using the inequality

$$\left\{ 1 - \frac{2Ds(\lambda^2 + 8D^2)}{\lambda^2(\lambda + 2D)} \right\}^{-1} \geq 1 + \frac{2D(\lambda^2 + 8D^2)}{\lambda^3} ,$$

we can verify that (A.9) is not positive and this completes the proof.

PROOF OF LEMMA 3.3.   First we note that

$$g_0(r + D) < r + D - w \quad \text{for} \quad r \geq r_0 \,,$$

if and only if

$$r + D < g_0^{-1}(r + D - w) \quad \text{for} \quad r \geq r_0 \,,$$

where we put $w = b'(r - D)^{p-2}/(r + D)^{p-1}$. Noting that $0 < w < D$ by the condition $b \leq 4D^2$, we see that if $r \geq r_0$

$$r + D - w > r \geq r_0 > \lambda_0 \,.$$

So we can apply Lemma 3.2 with $\lambda = r + D - w$ and we have

$$g_0^{-1}(r + D - w) > r + D - w + \frac{b'}{\sqrt{(r + D - w)^2 + 8D^2}} \left(\frac{r + D - w}{r + 3D - w} \right)^{p-2}$$
$$r \geq r_0 \,.$$

Therefore, we need only show that

$$- w + \frac{b'}{\sqrt{(r + D - w)^2 + 8D^2}} \left(\frac{r + D - w}{r + 3D - w} \right)^{p-2} \geq 0 \quad \text{for} \quad r \geq r_0 \,,$$

which is equivalent to the condition

$$\left\{ \frac{r - D}{r + D} \frac{r + 3D - w}{r + D - w} \right\}^{p-2} \leq \frac{r + D}{\sqrt{(r + D - w)^2 + 8D^2}} \quad \text{for} \quad r > r_0 \,.$$

Since $(r - D)(r + 3D - w)/\{(r + D)(r + D - w)\} \leq 1$, it is sufficient for us to show that

$$h(w) \geq (r - D)^2/(r + D)^4 \,,$$

where

$$h(w) = \frac{1}{(r + D - w)^2 + 8D^2} \frac{(r + D - w)^2}{(r + 3D - w)^2} \,.$$

We can easily check that $\min_{0 \leq w \leq D} h(w) = \min \{h(0), h(D)\}$ and that both $h(0)$ and $h(D)$ are not less than $(r - D)^2/(r + D)^4$.

REFERENCES

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

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

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. A—Theory Methods*, **15**, 2043–2063.

Faith, R. E. (1976). Minimax Bayes set and point estimators of a multivariate normal mean, Tech. Report #66, University of Michigan.

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, (supplement issue), *Statist. Decisions*, **1**, 3–16.

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

Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss, *Ann. Statist.*, **12**, 322–335.

Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proc. Third Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 197–206, Univ. of 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.