J. R. Statist. Soc. B (2005)
67, *Part* 3, *pp.* 343–361

Small confidence sets for the mean of a spherically symmetric distribution

Richard Samworth

University of Cambridge, UK

[Received July 2002. Revised November 2004]

Summary. Suppose that X has a k-variate spherically symmetric distribution with mean vector θ and identity covariance matrix. We present two spherical confidence sets for θ, both centred at a positive part Stein estimator $T_S^+(X)$. In the first, we obtain the radius by approximating the upper α-point of the sampling distribution of $\|T_S^+(X)-\theta\|^2$ by the first two non-zero terms of its Taylor series about the origin. We can analyse some of the properties of this confidence set and see that it performs well in terms of coverage probability, volume and conditional behaviour. In the second method, we find the radius by using a parametric bootstrap procedure. Here, even greater improvement in terms of volume over the usual confidence set is possible, at the expense of having a less explicit radius function. A real data example is provided, and extensions to the unknown covariance matrix and elliptically symmetric cases are discussed.

Keywords: Conditional properties; Confidence sets; Coverage probability; Location parameter; Multivariate normal distribution; Parametric bootstrap; Spherically symmetric distribution; Stein estimator; Volume

1. Introduction

Let X have a k-dimensional spherically symmetric distribution about θ, with density $f(\|x-\theta\|^2)$. The usual $(1-\alpha)$-level confidence set for θ is

$$C^0(X)=\{\theta\in\mathbb{R}^k : \|X-\theta\|^2\leqslant c^2\},$$

where c^2 satisfies

$$\int_{\mathbb{R}^k} f(\|x\|^2)\,\mathbb{1}_{\{\|x\|^2\leqslant c^2\}}\,\mathrm{d}x=1-\alpha.$$

This paper is concerned with the construction of confidence sets for θ which improve—in a sense which is made clear later—on $C^0(X)$ when $k\geqslant 3$. Specifically, we consider sets of the form

$$\{\theta\in\mathbb{R}^k : \|T_S^+(X)-\theta\|^2\leqslant v^2(\|X\|)\},$$

where

$$T_S^+(X)=\left(1-\frac{a}{\|X\|^2}\right)_+X$$

is a positive part Stein estimator, $h_+=\max(h,0)$ and $a>0$. This version of the positive part Stein estimator shrinks the observations towards the origin, with greater shrinkage as a increases. We investigate two methods of construction of the radius function $v(\cdot)$, both involving direct

Address for correspondence: Richard Samworth, Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK.
E-mail: r.j.samworth@statslab.cam.ac.uk

 1369–7412/05/67343

approximation of the upper α-point of the sampling distribution of $\|T_S^+(X)-\theta\|^2$. The first is an analytic procedure, giving an explicit expression for $v(\cdot)$ which is never larger than c and which can be considerably smaller. Despite this, subject to minor conditions on the underlying density, we can show that the resulting confidence set dominates $C^0(X)$ in terms of coverage probability, provided that $\|\theta\|$ is either less than a given bound, or sufficiently large. The result for large $\|\theta\|$ is an immediate corollary of theorem 5.1 of Hwang and Chen (1986), whereas our theorem 2 finds the optimal range of values of $\|\theta\|$ such that the θ-section corresponding to the analytic confidence set (which is defined immediately before lemma 1) contains the corresponding θ-section of $C^0(X)$. This technique of proof cannot therefore be extended to cover intermediate values of $\|\theta\|$. Simulations suggest that dominance may be attained for all values of $\|\theta\|$, at least for moderate or large k. An alternative to the analytic procedure is to apply the parametric bootstrap. Here, an even greater improvement in volume over the original confidence set is possible, without the coverage probability dropping below the nominal level, but at the expense of a less explicit radius function.

Structurally, the confidence sets are of the same form as those of Casella and Hwang (1983), who considered only the multivariate normal case and who obtained their radius by modifying the solution to an empirical Bayes problem. However, the sets that are constructed in this paper, as well as having a more natural motivation, compare favourably in the region of the parameter space which is of most interest when applying the positive part Stein estimator (see the discussion in the final paragraph of Section 5).

Interest in the problem of point estimation of θ when X has a multivariate normal distribution was sparked by the celebrated discovery of Stein (1956), who proved the existence of estimators which strictly dominate X with respect to the squared error loss function when $k \geqslant 3$. Brandwein and Strawderman (1978) and Brandwein (1979) extended these results to cover spherically symmetric distributions. It is now known that the Stein phenomenon applies to a very wide class of distributions and loss functions—see, for example, Brandwein and Strawderman (1990) or Evans and Stark (1996). By contrast, progress on the confidence set problem has been much slower, to the extent that results for confidence sets which strictly dominate the obvious confidence set in terms of volume are still restricted to the multivariate normal distribution. As several researchers testify, this is not to do with the lesser importance of the confidence set problem, but rather because of its technical difficulty.

As an application of our techniques, consider for $n \geqslant k$ the linear model

$$\underset{n\times 1}{X} = \underset{n\times k}{A}\,\underset{k\times 1}{\theta} + \sigma \underset{n\times 1}{\varepsilon}, \tag{1.1}$$

where the design matrix A is assumed to be of full rank k, and where the error vector ε has a density which is spherically symmetric about the origin. Of course, this model includes the standard linear model with normally distributed errors as an important special case. Zellner (1976) cited several references in which the linear model with spherically symmetric errors as a model for practical situations was considered and proposed other scenarios himself. Properties of the usual least squares estimator, $\hat{\theta}=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}X$, in model (1.1), have been studied by Thomas (1970), Zellner (1976) and Box (1953), among others.

Hwang and Chen (1986) showed how the problem of finding a confidence set for θ in model (1.1) can be reduced to the simpler form that is studied here, provided that the error variance σ^2 is known. The need to assume knowledge of σ^2 may be regarded as a weakness of our method. Indeed, the analytic theory is greatly complicated by replacing σ^2 with an estimate from the data, though it is straightforward to extend the parametric bootstrap procedure to this situation; some simulations and discussion are presented in Section 4.1. However, the known σ^2 model is a common assumption in nonparametric function estimation problems (Brown and

Low, 1996; Brown *et al.*, 1997), for instance when wavelet methods are used in nonparametric regression problems with Gaussian noise. There the vector of wavelet coefficients has a multivariate normal distribution, and σ^2 can typically be estimated accurately from the wavelet coefficients at fine scales (which are discarded in the signal reconstruction) and is therefore treated as known. Donoho and Johnstone (1994) and Donoho *et al.* (1995) have given further details. We give a different application to some baseball data in Section 6, in which we may take $\sigma^2 = 1$ after a variance stabilizing transformation of binomial data.

A loss function is rarely stated explicitly in the confidence set problem, though Casella and Hwang (1983) and Beran (1995) are exceptions in this regard. Instead, different confidence procedures tend to be compared according to four criteria: shape, coverage probability, volume and conditional properties. It is the need to ensure good performance in all of these respects that makes the problem so demanding.

It is difficult to make concrete statements regarding the shape of a reasonable confidence set. At first sight, it is difficult to look beyond a sphere when dealing with a spherically symmetric distribution. However, Berger (1980) has given a heuristic argument suggesting that this choice may not be so clear cut. In fact, Faith (1976), Shinozaki (1989) and Tseng and Brown (1997) have all also proposed non-spherical confidence regions. There is a consensus that an acceptable confidence set should be at least connected, though this still seems to be quite a weak requirement. We suspect that most practitioners would be reluctant to use a confidence set unless its geometry were fairly well understood.

Fortunately, coverage probability and volume can be treated in a more satisfactory way, and they are of course intimately linked. According to Joshi (1969), a confidence set $C(X)$ strictly dominates $C^0(X)$ if

(a) $\mathbb{P}_\theta\{C(X) \ni \theta\} \geqslant \mathbb{P}_\theta\{C^0(X) \ni \theta\}$, for all $\theta \in \mathbb{R}^k$, and
(b) $\mathrm{Vol}\{C(x)\} \leqslant \mathrm{Vol}\{C^0(x)\}$, for all $x \in \mathbb{R}^k$,

with strict inequality either in

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

Joshi also pointed out that two confidence sets should be considered equivalent if their symmetric difference has zero volume. Of course, the practitioner is more interested in a reduction in volume, provided that the coverage probability does not drop below the nominal level, than in increased coverage probability at a fixed volume.

Appreciation of the importance of the conditional properties of confidence sets began with Fisher (1956, 1959). Rules for satisfactory conditional performance were formalized by Buehler (1959) and Robinson (1979a, b) in terms of a betting game between two players. Casella and Hwang (1986) were the first to consider the conditional properties of confidence sets for the mean of a multivariate normal distribution. Robinson (1979b), Lu and Berger (1989), Robert and Casella (1994) and Wang (2000) all took a complementary approach and discussed how to improve the reported confidence statements for the usual confidence set $C^0(X)$.

Tseng and Brown (1997) have given an excellent review of the earlier literature on the multivariate normal confidence set problem, which, in addition to those references already given, includes Stein (1962, 1981) and Hwang and Casella (1982, 1984). Tseng and Brown themselves proposed somewhat egg-shaped sets which have exact coverage probability and they also found sufficient conditions under which their sets uniformly dominate $C^0(X)$ in terms of volume. Unfortunately, as they themselves admitted, these sufficient conditions do not appear to be entirely satisfactory, and it seems difficult to choose an optimal set from within the class that they studied. In addition, the shape of the sets can be quite complicated, although certain results concerning the geometry are obtained.

Previous work on the spherically symmetric case, such as Ki and Tsui (1985), Hwang and Chen (1986) and Robert and Casella (1990), has focused on proving that confidence sets of the same radius as $C^0(X)$ have higher coverage probability when recentred at a positive part Stein estimator. In this paper, we recognize that, once a spherical confidence set centred at the positive part Stein estimator has been decided on, the ideal, exact, $(1-\alpha)$-level confidence set would be

$$\{\theta \in \mathbb{R}^k : \|T_S^+(X) - \theta\|^2 \leqslant w_\alpha(\theta)\},$$

where $w_\alpha(\theta)$ is the upper α-point of the sampling distribution of $\|T_S^+(X) - \theta\|^2$. Of course, this is not a feasible confidence set as the radius depends on the unknown θ. The approach that is taken in Section 2 is a direct analytic estimation of $w_\alpha(\theta)$. Specifically, we compute the first two non-zero terms in the Taylor series of $w_\alpha(\theta)$ about the origin, allowing us to write

$$w_\alpha(\theta) = w_\alpha(0) + \tfrac{1}{2} w_\alpha''(0) \|\theta\|^2 + o(\|\theta\|^2)$$

as $\|\theta\| \to 0$. Ignoring the $o(\|\theta\|^2)$ term, we estimate $\|\theta\|^2$ by $\|X\|^2$ and obtain the confidence set

$$C(X) = \{\theta \in \mathbb{R}^k : \|T_S^+(X) - \theta\|^2 \leqslant \min\{w_\alpha(0) + \tfrac{1}{2} w_\alpha''(0)\|X\|^2, c^2\}\}.$$

We are motivated by the knowledge that $T_S^+(X)$ performs best as an estimator of θ when $\|\theta\|$ is small, which suggests that this is the region of the parameter space where we would expect a spherical confidence set centred at the positive part Stein estimator with radius $v(\|X\|) = c$ to show the greatest improvement, in terms of coverage probability, over $C^0(X)$. Simulations in Hwang and Casella (1982) support this intuition. More importantly, this suggests that it is for small values of $\|X\|$ that we can hope to see the greatest reduction in volume while maintaining at least the same coverage probability as $C^0(X)$. The radius function $v(r)$ that we propose attains the value $c - a/c$ at $r = 0$, which is rather smaller than the suggestion in Casella and Hwang (1983).

In Section 3, we make use of the simple analytic form of the radius function to prove some results about the properties of the confidence set. A particularly interesting feature of the work from a theoretical point of view is that the radius of the analytic confidence set depends on the density f only through c^2, and a quantity $f'(c^2)/f(c^2)$, called the relative increasing rate (RIR) of f at c^2. Both Hwang and Chen (1986) and Robert and Casella (1990) have noted the importance of this quantity in establishing dominance of their recentred sets over $C^0(X)$. Simulations of the coverage probabilities are provided for three spherically symmetric densities: the k-variate normal, the multivariate t- and the double-exponential densities. These distributions were studied in Hwang and Chen (1986). The multivariate t-distribution with N degrees of freedom has density

$$f(\|x\|^2) \propto \left(1 + \frac{\|x\|^2}{N}\right)^{-(N+k)/2}.$$

Relative to the normal model, it gives more flexibility to the practitioner, through the choice of the number of degrees of freedom, but is a close approximation to normality when the number of degrees of freedom is large (see Zellner (1976)). The double-exponential distribution with parameter d has density

$$f(\|x\|^2) \propto \exp(-d\|x\|),$$

and the parameter choice $d = (k+1)^{1/2}$ ensures that each component of X has unit variance.

As was suggested by the work in Samworth (2003a), another appealing approach to the problem for the modern statistician involves a parametric bootstrap procedure. In the related problem where we have independent random vectors $X_1, \ldots, X_n$, each having the same spherically symmetric distribution as X, we encounter an inconsistency problem like the one described in Samworth (2003a). Nevertheless, as Samworth (2003a) discussed, inconsistency does not preclude the bootstrap from performing successfully at finite sample sizes. We investigate the parametric bootstrap confidence set in Section 4, and Section 5 consists of some comments and generalizations. Section 6 is devoted to the baseball data example, and the proofs of proposition 1 and theorem 2 are given in Appendix A. For all other proofs, which mainly involve some fairly detailed computations, the interested reader is referred to Samworth (2003b).

2. Constructing the analytic confidence set

We say that X has a k-variate spherically symmetric distribution about θ if $X-\theta$ has the same distribution as $P(X-\theta)$ for all $k \times k$ orthogonal matrices P. We assume that $k \geqslant 3$ and that the distribution $\mathbb{P}_\theta$ of X has a density with respect to Lebesgue measure on $\mathbb{R}^k$, whose value at a point $x \in \mathbb{R}^k$ is denoted by $f(\|x-\theta\|^2)$. We begin with a useful result concerning all estimators of θ of the form $\gamma(\|X\|)X$, where $\gamma : [0,\infty) \to \mathbb{R}$ is a measurable function.

Proposition 1. For $\alpha \in (0,1)$, the upper α-point of the sampling distribution of $\|\gamma(\|X\|)X-\theta\|^2$ depends on θ only through $\|\theta\|$.

The positive part Stein estimator

$$T_{\mathrm{S}}^{+}(X) = \left(1 - \frac{a}{\|X\|^2}\right)_{+} X \tag{2.1}$$

is of the form $\gamma(\|X\|)X$, and we let $w_\alpha(\|\theta\|)$ denote the upper α-point of the sampling distribution of $\|T_{\mathrm{S}}^{+}(X)-\theta\|^2$. The theorem below is the main theorem of this section.

Theorem 1. Suppose that $w_\alpha(0) > 0$, and that the spherically symmetric density f is twice continuously differentiable. Then

$$w_\alpha(\|\theta\|) = w_\alpha(0) + \tfrac{1}{2} w_\alpha''(0)\|\theta\|^2 + o(\|\theta\|^2)$$

as $\|\theta\| \to 0$, where $w_\alpha(0) = (c - a/c)^2$ and

$$\frac{1}{2} w_\alpha''(0) = \frac{1}{k}\left(1 - \frac{a}{c^2}\right)\left\{\frac{a(k-1)}{c^2+a} - \frac{2ac^2}{(c^2+a)^2} - \frac{2a^2}{c^2+a}\frac{f'(c^2)}{f(c^2)}\right\} + \frac{a(k-1)}{c^2 k}.$$

The condition that $w_\alpha(0) > 0$ is equivalent to requiring that $\alpha < \mathbb{P}_0(\|X\|^2 > a)$, which in turn is equivalent to $c^2 > a$; this will rarely be restrictive in practice. For instance, when f is the standard k-variate normal density, James and Stein (1961) showed that the ordinary Stein estimator

$$T_{\mathrm{S}}(X) = \left(1 - \frac{a}{\|X\|^2}\right) X$$

strictly dominates X in the point estimation problem with squared error loss for $a \in (0, 2(k-2))$, and that $a = k-2$ is the optimal choice. In this case, the confidence set

$$\{\theta \in \mathbb{R}^k : \|X-\theta\|^2 \leqslant k-2\}$$

has only about 50% coverage probability.

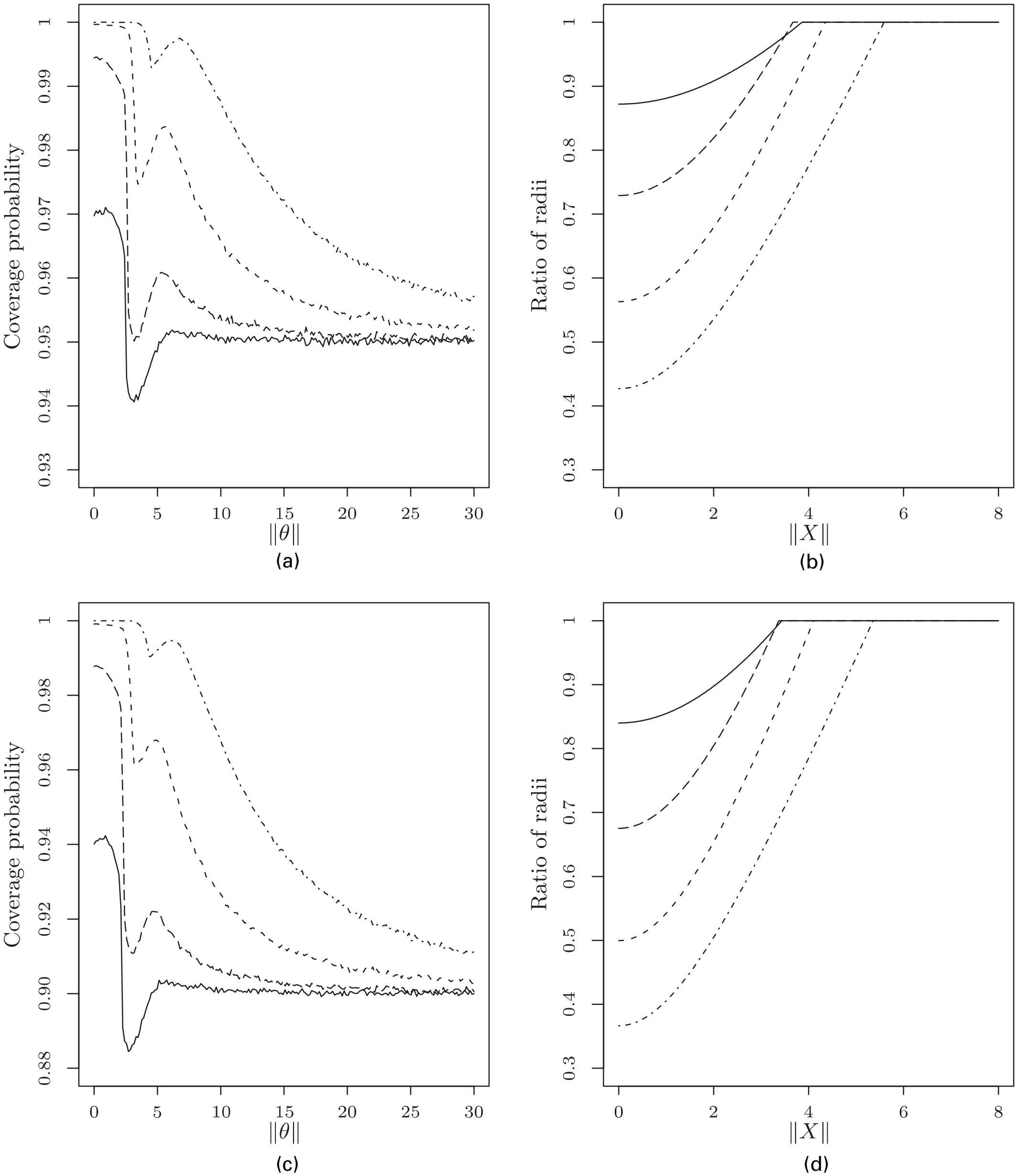

Fig. 1. Coverage probabilities of the confidence set (2.2) for the k-variate normal distribution, and the ratio of each radius to the corresponding radius of $C^0(X)$ (———, $k=3$; – – –, $k=5$; -------, $k=10$; - · - · - ·, $k=20$ ($a=k-2$)): (a), (b) $\alpha=0.05$; (c), (d) $\alpha=0.1$

Having computed $w_\alpha(0)$ and $w''_\alpha(0)$, we estimate $\|\theta\|^2$ by $\|X\|^2$ and therefore construct the confidence set

$$C(X)=\{\theta\in\mathbb{R}^k : \|T_{\mathrm{S}}^+(X)-\theta\|^2 \leqslant \min\{w_\alpha(0)+\tfrac{1}{2}w''_\alpha(0)\|X\|^2, c^2\}\}. \tag{2.2}$$

As noted in Section 1, an interesting feature of this confidence set is that it depends on the

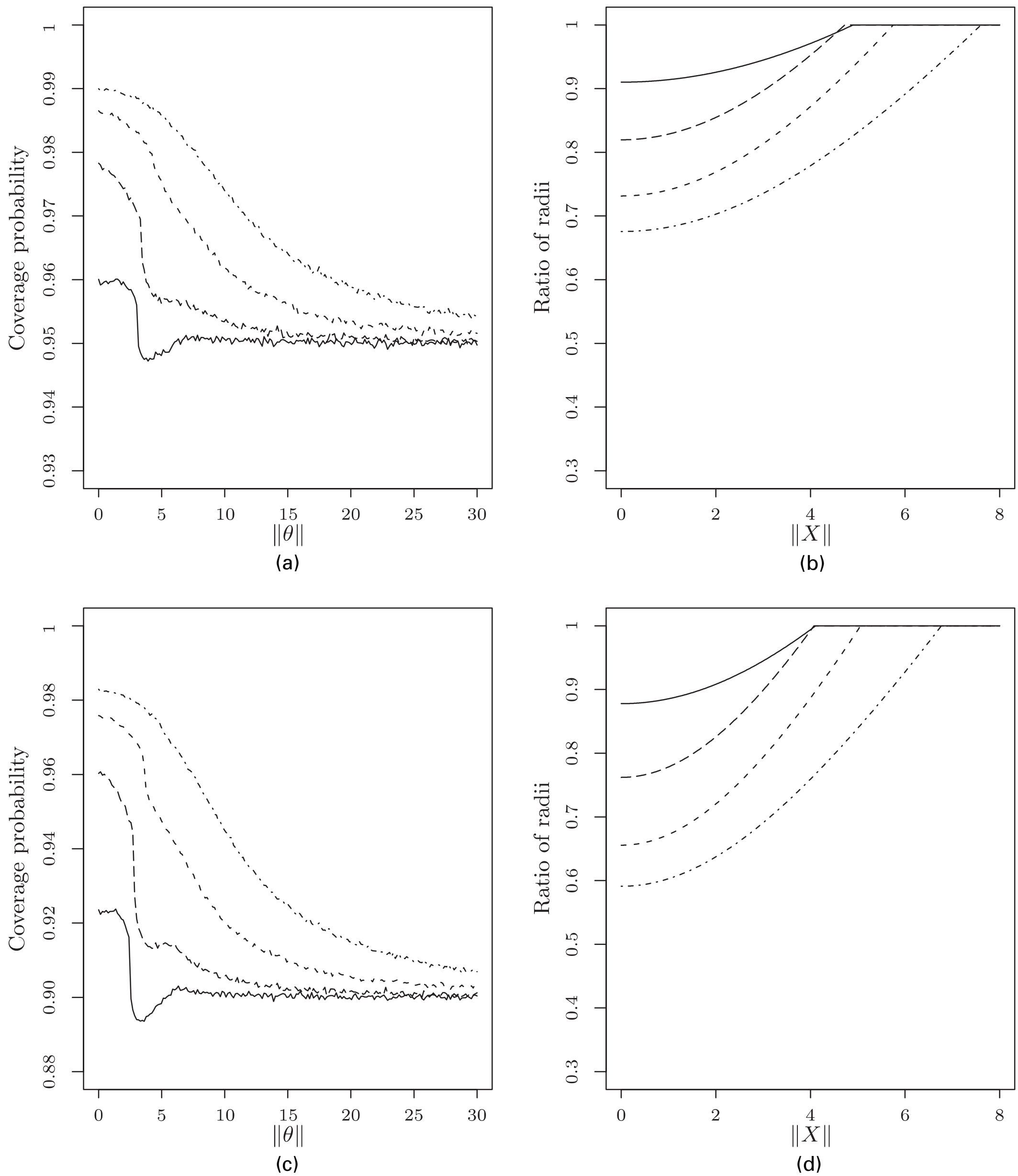

Fig. 2. Coverage probabilities of the confidence set (2.2) for the multivariate *t*-distribution with $N = 10$ degrees of freedom, and the ratio of each radius to the corresponding radius of $C^0(X)$ (———, $k = 3$; – – –, $k = 5$; -------, $k = 10$; - · - · - ·, $k = 20$ ($a = k - 2$)): (a), (b) $\alpha = 0.05$; (c), (d) $\alpha = 0.1$

density f only through c^2 and the RIR of f at c^2. Typically, c^2 will be sufficiently large to ensure that the RIR at c^2 is negative, with very negative values indicating that the distribution has light tails. For the three distributions that were mentioned in Section 1, namely the standard multivariate normal, the multivariate *t*- with N degrees of freedom and the double-exponential distribution with parameter d, the RIRs at c^2 are

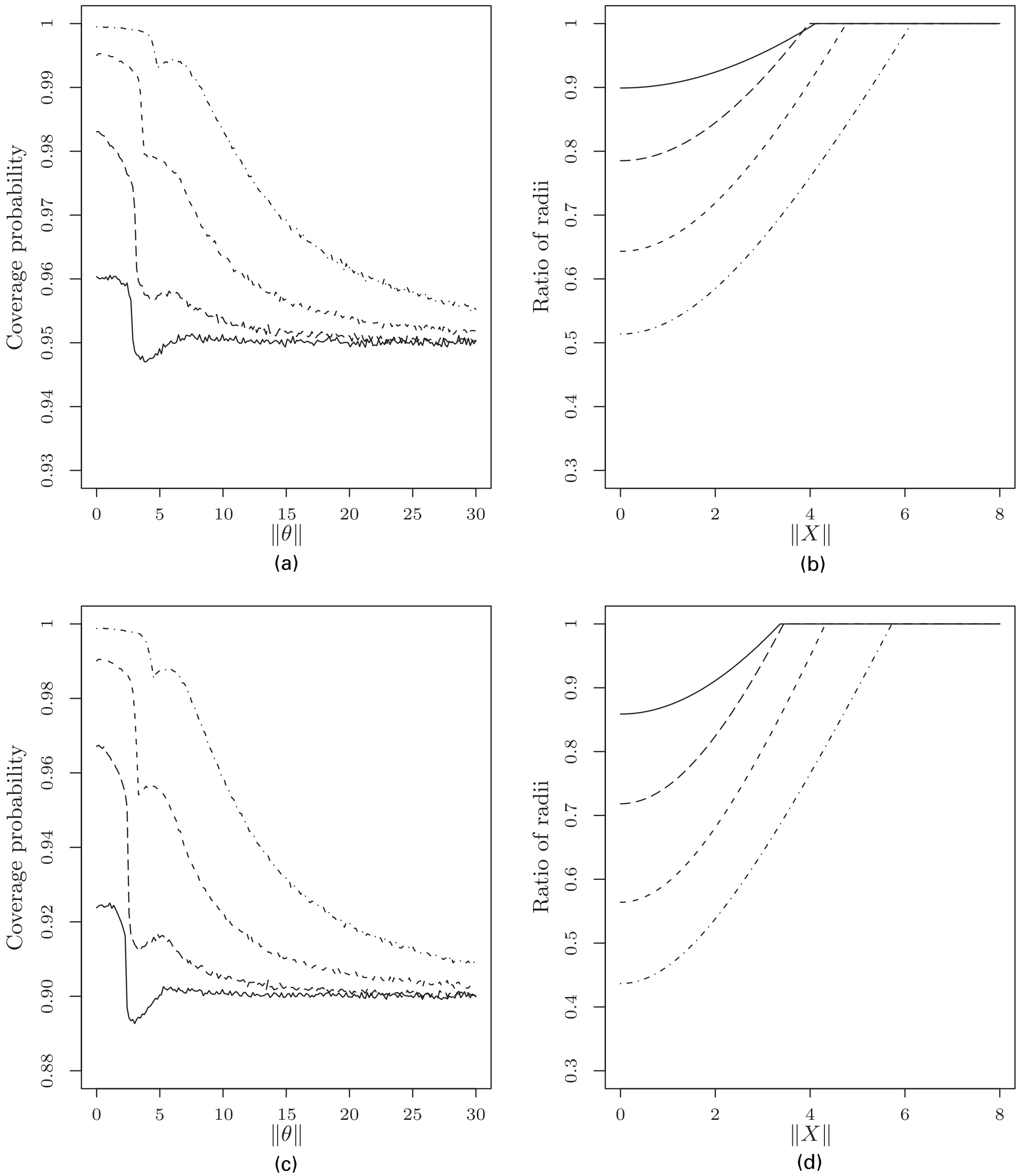

Fig. 3. Coverage probabilities of the double-exponential confidence set (2.2), and the ratio of each radius to the corresponding radius of $C^0(X)$ (———, $k=3$; – – –, $k=5$; -------, $k=10$; - · - · - ·, $k=20$ ($a=k-2$; $d=(k+1)^{1/2}$)): (a), (b) $\alpha=0.05$; (c), (d) $\alpha=0.1$

$$-\frac{1}{2}, \qquad -\frac{N+k}{2(N+c^2)} \quad \text{and} \quad -\frac{d}{2c}$$

respectively.

Since $C^0(X)$ is minimax (Stein, 1962), a necessary condition for the confidence set $C(X)$ in equation (2.2) to dominate $C^0(X)$ in coverage probability is that $w''_\alpha(0)>0$. Perhaps surprisingly in view of the results of Hwang and Chen (1986) and Robert and Casella (1990), this condition corresponds to the RIR at c^2 being *less* than some positive bound depending on a, c^2 and k. One of the themes of these previous works is that a confidence set of the same radius as $C^0(X)$ has uniformly higher coverage probability when recentred at a positive part Stein estimator, provided that the RIR at c^2 is *greater* than some negative bound. As mentioned in the previous paragraph, however, this positive bound will almost certainly be unrestrictive in practice.

The choice of a is more delicate in the spherically symmetric case than the multivariate normal; see Samworth (2003b) for a discussion. However, for simplicity and to ease comparison between different distributions, we take $a=k-2$ throughout in our numerical studies of the confidence set (2.2), presented in Figs 1, 2 and 3. In each figure, 400000 Monte Carlo repetitions were used to approximate the coverage probability at each value of θ, giving a simulation error standard deviation of about 0.0005 at each point.

It appears that the confidence set (2.2) dominates $C^0(X)$ in terms of coverage probability for all of the distributions that are considered, apart from possibly in a narrow middle range of values of $\|\theta\|$ for small values of k. These exceptions are similar to those which are found in Casella and Hwang (1983) and the problems are sufficiently small that they can be ignored in most practical contexts. In view of the point estimation results of Brandwein and Strawderman (1978), the choice of a is almost certainly too large for these small values of k, although we do not pursue this matter further here.

However, for all values of k, the radii show a great improvement over those of $C^0(X)$, especially for small $\|X\|$, as we would expect from their construction. In fact, in the k-variate normal case, the radii tend to be considerably smaller than those of Casella and Hwang (1983), with which they are directly comparable, for small values of $\|X\|$, at the expense of being slightly larger for larger values of $\|X\|$ (Table 1). The ratio of the radii of confidence set (2.2) and $C^0(X)$ in the best case, i.e. when $\|X\|=0$, is $1-a/c^2$. Thus, for fixed a, the maximum improvement in volume is greater for distributions with lighter tails. Table 2 gives estimates of the probability that the radius of the confidence set (2.2) is less than c, which, of course, is a decreasing function of $\|\theta\|$.

Table 1. Ratio of the radii of the confidence set (2.2) to the corresponding radii of Casella and Hwang (1983) in the k-variate normal case and with $\alpha=0.05$

k	*Ratios of the radii for the following values of* $\|X\|$*:*									
	0	*1*	*2*	*3*	*4*	*5*	*6*	*8*	*10*	*20*
5	0.64	0.68	0.79	0.98	1.13	1.07	1.05	1.03	1.02	1.00
10	0.43	0.48	0.62	0.85	1.18	1.21	1.13	1.07	1.04	1.01

Table 2. Estimates of the probability that the radius of the confidence set (2.2) is less than c in the k-variate normal case and with $\alpha=0.05$

k	*Probability estimates for the following values of* $\|\theta\|$*:*						
	0	*1*	*2*	*3*	*4*	*5*	*6*
5	0.99	0.97	0.87	0.58	0.24	0.05	0.01
10	0.97	0.95	0.83	0.58	0.26	0.06	0.01

3. Properties of the analytic confidence set

We have already seen that the RIR of f at c^2 must be negative for the confidence sets $C(X)$ in equation (2.2) to have adequate coverage probability. In this section we shall see that the sets have several desirable properties provided also that the RIR at c^2 is not *too* negative. These results are more in line with the works of Hwang and Chen (1986) and Robert and Casella (1990), where dominance occurs provided that the tails of the distribution are sufficiently heavy. Throughout this section, we assume that $k \geqslant 3$, and that f is twice continuously differentiable. We start with some elementary bounds, which give simple yet general conditions under which proposition 2 and lemma 2 hold.

Lemma 1.

(a) Let $a \in (0, k-1]$ and $\alpha \in (0, \mathbb{P}_0(\|X\|^2 > a))$, and suppose that $f'(c^2)/f(c^2) \geqslant -\frac{1}{2}$. Then

$$\frac{1}{2} w''_\alpha(0) \leqslant \frac{k-1}{k}.$$

(b) Let $a > 0$ and $\alpha \in (0, \mathbb{P}_0(\|X\|^2 > a))$, and suppose that $f'(c^2)/f(c^2) \leqslant 0$. Then

$$\frac{1}{2} w''_\alpha(0) \geqslant \frac{a(k-1)}{c^2 k}.$$

The next three results concern the θ-section that is associated with the confidence set (2.2), given by

$$C(\theta) = \{x \in \mathbb{R}^k : \|T_S^+(x) - \theta\|^2 \leqslant \min\{w_\alpha(0) + \tfrac{1}{2} w''_\alpha(0)\|x\|^2, c^2\}\}.$$

The first is an extension of theorem A1 of Casella and Hwang (1983).

Proposition 2. Let $a > 0$, and suppose that $0 < w''_\alpha(0)/2 < 1$. Then $C(\theta)$ is connected, for all $\theta \in \mathbb{R}^k$.

In fact, $C(\theta)$ can have a stronger property when $\|\theta\|$ lies in a range which is of particular importance to us (Section 5).

Lemma 2. Let $a > 0$, and suppose that $0 < w''_\alpha(0)/2 < 1$. For $\|\theta\| \leqslant c - a/c$, if $x \in C(\theta)$, then so is tx, for all $t \in [0, 1]$.

To present the main theorem of this section, we let

$$C^0(\theta) = \{x \in \mathbb{R}^k : \|x - \theta\|^2 \leqslant c^2\}$$

denote the θ-section corresponding to the usual confidence set $C^0(X)$.

Theorem 2. Let $a \in (0, k-1]$, $\alpha \in (0, \mathbb{P}_0(\|X\|^2 > a))$ and also suppose that $-\frac{1}{2} \leqslant f'(c^2)/f(c^2) \leqslant 0$. If

$$\|\theta\|^2 \leqslant \min\left[w_\alpha(0), \frac{c^2 - a}{2 w''_\alpha(0) c^4}\{2 w''_\alpha(0) c^4 - (c^2 - a)a\}\right],$$

then $C^0(\theta) \subseteq C(\theta)$.

The upper bound on the range of values of $\|\theta\|$ for which the conclusion of theorem 2 holds is the best possible, and theorem 2 is clearly non-vacuous since it holds for $\|\theta\| = 0$. In fact, the upper bound corresponds to a point just before the sharp drop in coverage probability that is seen in Figs 1–3. For instance, when $k = 5$, $\alpha = 0.05$ and f is the k-variate normal density, we have $C^0(\theta) \subseteq C(\theta)$ for $\|\theta\| \leqslant 2.7$.

An obvious corollary of theorem 2 is that $C(X)$ dominates $C^0(X)$ in terms of coverage probability for the particular range of values of $\|\theta\|$ above. Since the confidence set $C(X)$ will have constant radius c for large $\|X\|$ (provided that $c^2 > a$), it follows immediately from theorem 5.1 of Hwang and Chen (1986) that we have strict dominance in coverage probability of $C(X)$ over $C^0(X)$ for sufficiently large $\|\theta\|$, under their mild conditions on f and provided that

$$0 < a < \frac{(k-2)\, f(c^2)}{f'(c^2)}.$$

Further, for large $\|\theta\|$, the greatest coverage probability is attained with

$$a = \frac{(k-2)\, f(c^2)}{2 f'(c^2)}.$$

Theorem 2 also has implications for the conditional properties of $C(X)$, which we now describe.

When making an assertion of the form

$$\mathbb{P}_\theta\{C(X) \ni \theta\} = 1 - \alpha,$$

the statistician is averaging (integrating) over the sample space. However, the confidence set must be specified on the basis of observing $X = x$, say. The statistician should, therefore, question whether the probability assertion is still valid in the light of the data. For instance, such considerations provide a strong criticism of confidence sets that are centred at the ordinary, as opposed to positive part, Stein estimator. For, if $\|x\|$ were very small, the confidence set would presumably be well away from the origin, and the statistician would be unable to justify the probability statement in the light of the data, whatever the true value of θ. Put another way, a hypothetical opponent of the statistician could specify a very small sphere A centred at the origin, staking an amount α to win $1 - \alpha$ that $C(x)$ does not contain θ if $x \in A$, and not making a bet otherwise. Under infinitely many hypothetical repetitions of the experiment with a referee who knows the true value of θ, the opponent would win almost surely.

More formally, Buehler (1959) and Robinson (1979a) introduced various criteria for judging the conditional performance of a confidence set. In our situation, if A is a subset of $\mathbb{R}^k$ of positive Lebesgue measure, Robinson called A a negatively biased relevant subset for $C(X)$ if there is an $\varepsilon > 0$ such that

$$\mathbb{P}_\theta\{C(X) \ni \theta | X \in A\} \leqslant 1 - \alpha - \varepsilon$$

for all $\theta \in \mathbb{R}^k$ and advocated that we should not use a confidence set if there is a negatively biased relevant subset.

Another simple corollary of theorem 2 is that, for any subset A of $\mathbb{R}^k$ of positive Lebesgue measure, we have

$$\mathbb{P}_\theta\{C(X) \ni \theta | X \in A\} \geqslant \mathbb{P}_\theta\{C^0(X) \ni \theta | X \in A\} \tag{3.1}$$

for $\|\theta\|$ in the given range. Casella and Hwang (1986) showed that, for any $u > 0$, there is a $\delta = \delta(u) > 0$ such that

$$\mathbb{P}_\theta\{C^0(X) \ni \theta | \|X\|^2 \leqslant u\} > 1 - \alpha$$

for all $\|\theta\|^2 < \delta$, and a very similar argument shows that, for any $\xi \in \mathbb{R}^k$ and $u > 0$, there is a $\delta = \delta(u) > 0$ such that

$$\mathbb{P}_\theta\{C^0(X) \ni \theta | \|X - \xi\|^2 \leqslant u\} > 1 - \alpha \tag{3.2}$$

for all $\|\theta-\xi\|^2<\delta$. Combining inequalities (3.1) and (3.2), we see that there are no negatively biased relevant spheres centred at ξ for $C(X)$, provided that

$$\|\xi\|^2<\min\left[w_\alpha(0),\ \frac{c^2-a}{2\,w''_\alpha(0)c^4}\{2\,w''_\alpha(0)c^4-(c^2-a)a\}\right].$$

4. The bootstrap confidence set

Here we investigate another way of approximating the ideal confidence set

$$\{\theta\in\mathbb{R}^k:\|T_{\mathrm{S}}^{+}(X)-\theta\|^2\leqslant w_\alpha(\|\theta\|)\}. \tag{4.1}$$

In a parametric bootstrap procedure, we estimate θ by $\hat{\theta}$, say, and approximate expression (4.1) by

$$\{\theta\in\mathbb{R}^k:\|T_{\mathrm{S}}^{+}(X)-\theta\|^2\leqslant w^*_\alpha(\|\hat{\theta}\|)\}, \tag{4.2}$$

where $w^*_\alpha(\|\hat{\theta}\|)=\inf[x\in\mathbb{R}:\mathbb{P}_*\{\|T_{\mathrm{S}}^{+}(X^*)-\hat{\theta}\|^2\leqslant x\}\geqslant 1-\alpha]$. Here, the conditional density of X^* given X is $f(x-\hat{\theta})$, and $\mathbb{P}_*$ denotes the probability under this conditional distribution. In practice, $w^*_\alpha(\|\hat{\theta}\|)$ is still unavailable explicitly, but we can approximate it to any required degree of accuracy (in probability) by Monte Carlo simulation. An algorithm which first approximates the radius of the bootstrap confidence set at a fixed number of equally spaced points, and then uses linear interpolation to find the radius for the observed value of $\|X\|$, greatly improves the computational efficiency.

It is possible to generate random vectors from many spherically symmetric distributions as

$$X=RU+\theta,$$

where R has the same density as $\|X\|$, and U is independent of R and has a uniform distribution on the unit sphere $S=\{x\in\mathbb{R}^k:\|x\|=1\}$. It follows that R has density proportional to $r^{k-1}f(r^2)$ (Fang *et al.* (1989), page 35), whereas U has the same distribution as $Y/\|Y\|$, where $Y\sim N_k(0,I)$. For the double-exponential distribution with parameter d, we have $R\sim\Gamma(k,d)$. We can simulate random vectors from a multivariate t-distribution with N degrees of freedom as follows: generate $Z\sim n/\chi^2_n$, and, conditional on Z, generate $X\sim N_k(0,ZI)$ (Zellner, 1976).

The results of simulating the coverage probabilities of the bootstrap confidence sets for the multivariate normal distribution and using $\|\hat{\theta}\|=\|X\|$ are given in Fig. 4. The corresponding simulations for the multivariate t- and double-exponential distributions are similar but have a slightly less severe undercoverage problem for small k and moderate $\|\theta\|$; see Samworth (2003b).

The coverage probabilities and radii exhibit many of the same features as those of the analytic confidence set (2.2) for small $\|\theta\|$ and $\|X\|$ respectively. However, we find that it is possible to achieve an even smaller radius for larger $\|X\|$ by bootstrapping, while retaining coverage probability at the nominal level. Of course, it is much more difficult to prove any results concerning the properties of the bootstrap confidence set, such as those presented in Section 3 for the analytic confidence set, as the radius is given in a less explicit form. Nevertheless, Beran (1995) has studied the large k asymptotics of similar bootstrap confidence sets centred at the positive part Stein estimator in the multivariate normal case, using a different approach involving a geometrical risk criterion as well as coverage probability. Beran obtained the radii for his confidence sets in a different way, however, and his simulation results suggest greater undercoverage problems, which persist for larger values of k.

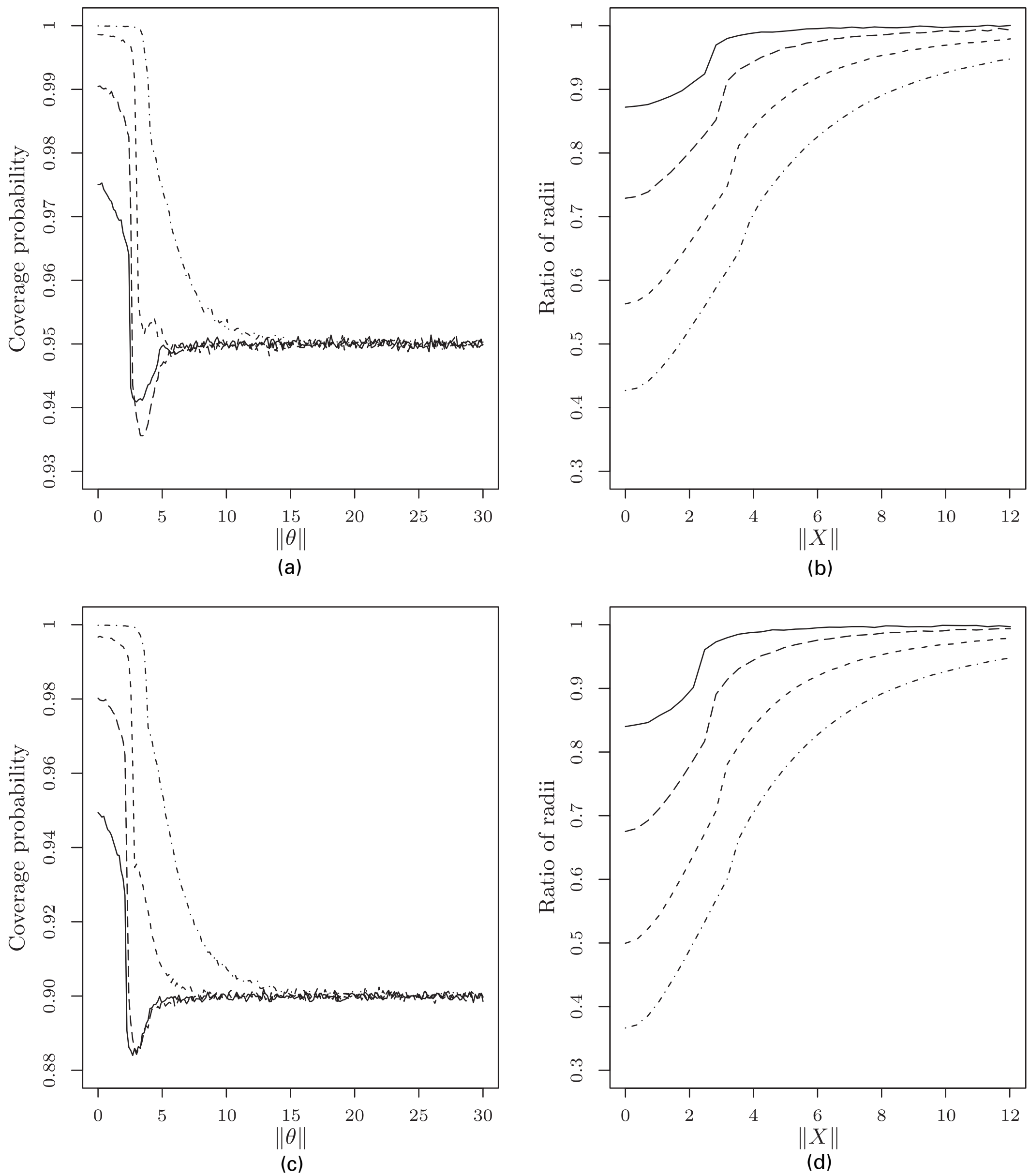

Fig. 4. Coverage probabilities of the bootstrap confidence set (4.2) for the k-variate normal distribution, and the ratio of each radius to the corresponding radius of $C^0(X)$ (———, $k=3$; – – –, $k=5$; -------, $k=10$; - · - · - ·, $k=20$ ($a=k-2$)): (a), (b) $\alpha=0.05$; (c), (d) $\alpha=0.1$

4.1. The unknown scale factor case

Recall the linear model (1.1) that was introduced in Section 1:

$$\underset{n\times 1}{X} = \underset{n\times k}{A}\,\underset{k\times 1}{\theta} + \sigma \underset{n\times 1}{\varepsilon}.$$

In this section, we consider the problem where σ^2 is unknown but can be estimated from the data. The canonical model is where $Z=(X^{\mathrm{T}}, Y^{\mathrm{T}})^{\mathrm{T}}$ has a $(k+\nu)$-dimensional spherically symmetric density with location parameter $\theta' = (\theta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$ and covariance matrix $\sigma^2 I_{k+\nu}$. Here, $I_{k+\nu}$ denotes the $(k+\nu) \times (k+\nu)$ identity matrix, and, despite the increased dimension, we write the density of Z as $f_{\sigma^2}(\|z-\theta'\|^2)$. The appropriate version of the positive part Stein estimator in this set-up is

$$T_{\mathrm{S}}^{+}(X, Y) = \left(1 - \frac{a\|Y\|^2/\nu}{\|X\|^2}\right)_{+} X.$$

In the multivariate normal case,

$$\frac{\|Y\|^2}{\nu} \sim \frac{\sigma^2}{\nu} \chi^2_\nu$$

and is independent of X; James and Stein (1961) showed that $a = \nu(k-2)/(\nu+2)$ is the optimal choice with respect to quadratic loss for a point estimate of θ. As mentioned in Section 1, analytic theory for confidence sets is very difficult when σ^2 is unknown (though we expect a good approximation to the known σ^2 case in the limit as $\nu \to \infty$). Bootstrapping, however, remains a viable possibility, and in Table 3 we present some coverage probabilities of the confidence set

$$\{\theta \in \mathbb{R}^k : \|T_{\mathrm{S}}^{+}(X, Y) - \theta\|^2 \leqslant w_\alpha^*(\|X\|, \|Y\|)\}, \tag{4.3}$$

where

$$w_\alpha^*(\|X\|, \|Y\|) = \inf[x \in \mathbb{R} : \mathbb{P}_*\{\|T_{\mathrm{S}}^{+}(X^*, Y^*) - X\|^2 \leqslant x\} \geqslant 1-\alpha].$$

Here, the conditional density of (X^*, Y^*) given (X, Y) is $f_{\|Y\|^2/\nu}(\|z - X'\|^2)$, where $X' = (X^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$, and $\mathbb{P}_*$ denotes the corresponding probability measure. The usual confidence set in this situation is

$$\left\{\theta \in \mathbb{R}^k : \|X - \theta\|^2 \leqslant \frac{k}{\nu} \|Y\|^2 F_\alpha(k, \nu)\right\}, \tag{4.4}$$

where $F_\alpha(k, \nu)$ is the upper α-point of an F-distribution with k and ν degrees of freedom. This is an exact $(1-\alpha)$-level confidence set, since it follows from theorem 11 of Kelker (1970) that

$$\frac{\|X-\theta\|^2/k}{\|Y\|^2/\nu}$$

has an F-distribution with k and ν degrees of freedom, regardless of the spherically symmetric distribution. Table 4 gives the ratios of the radii of expression (4.3) to the corresponding radii of expression (4.4). We find that it is possible to achieve similar gains in volume to those which

Table 3. Coverage probabilities of the confidence set (4.3) in the k-variate normal case†

ν	*Coverage probabilities for the following values of* $\|\theta\|/\sigma$:								
	0	*1*	*2*	*3*	*4*	*8*	*12*	*16*	*20*
100	0.996	0.996	0.994	0.974	0.944	0.938	0.936	0.937	0.937
1000	0.998	0.998	0.997	0.974	0.953	0.949	0.948	0.950	0.948
10000	0.999	0.998	0.997	0.975	0.953	0.950	0.950	0.950	0.950

†Parameter values: $\alpha = 0.05$; $a = \nu(k-2)/(\nu+2)$; $k = 10$.

Table 4. Ratio of the radii of the confidence set (4.3) to the corresponding radii of confidence set (4.4) in the k-variate normal case†

$\|Y\|^2/\nu\sigma^2$	*Ratios for the following values of* $\|X\|$:							
	0	*1*	*2*	*3*	*4*	*6*	*10*	*20*
0.6	0.57	0.61	0.69	0.82	0.87	0.93	0.96	0.97
0.8	0.57	0.60	0.67	0.77	0.85	0.91	0.95	0.97
1	0.57	0.60	0.66	0.72	0.83	0.90	0.94	0.97
1.2	0.57	0.59	0.65	0.70	0.80	0.88	0.94	0.97
1.4	0.57	0.59	0.63	0.69	0.77	0.87	0.94	0.96

†Parameter values: $\alpha=0.05$; $k=10$; $a=\nu(k-2)/(\nu+2)$; $\nu=100$.

were observed in Section 4 for the known covariance matrix case, but that undercoverage is more a problem, and ν needs to be very large before it disappears entirely.

5. Comments and generalizations

We have seen that the confidence sets (2.2), (4.2) and (4.3) successfully harness the power of the positive part Stein estimator to produce confidence sets which can be much smaller than the usual set $C^0(X)$, while still maintaining adequate coverage probability. Unfortunately, we could not provide a proof that the confidence set (2.2) strictly dominates $C^0(X)$ in terms of coverage probability for sufficiently large k.

The assumption of a spherically symmetric density may be generalized as follows. If $Y=\mu+B^{\mathrm{T}}X$, where $\mu\in\mathbb{R}^m$, B is a $k\times m$ matrix with $B^{\mathrm{T}}B=\Sigma$ having rank k and X has a k-dimensional spherically symmetric density f about the origin, we say that Y has an m-dimensional elliptically symmetric distribution with parameters μ and Σ. If $m=k$, it follows from theorem 2.16 of Fang *et al.* (1989) that $\Sigma^{-1/2}Y$ has spherically symmetric density f about $\theta=\Sigma^{-1/2}\mu$, so here the problem reduces to the spherically symmetric case provided that Σ is known. In particular, if $Y\sim N_k(\mu,\sigma^2\Sigma)$, where σ is an unknown scale factor and Σ is a $k\times k$ known, positive definite matrix, then the transformed vector $X=\Sigma^{-1/2}Y$ satisfies $X\sim N_k(\theta,\sigma^2 I_k)$, where $\theta=\Sigma^{-1/2}\mu$. As an application, consider a one-way analysis of variance with k cells and n_i observations in cell i, for $i=1,\ldots,k$. If Y denotes the vector of cell means, then Σ is a diagonal matrix with diagonal entries $n_1^{-1},\ldots,n_k^{-1}$.

One extension which is especially important for our work is the choice of origin for the Stein estimator. Both confidence set (2.2) and confidence set (4.2) only represent a significant improvement over $C^0(X)$ if θ is reasonably near the origin. If a prior estimate of θ, say θ_0, is available, then we should redefine the positive part Stein estimator as

$$T^+_{\mathrm{S},\theta_0}(X)=\theta_0+\left(1-\frac{a}{\|X-\theta_0\|^2}\right)_+(X-\theta_0)$$

and replace the radius function $v^2(\|X\|)$ by $v^2(\|X-\theta_0\|)$. The region of greatest improvement is then near θ_0, so our confidence sets will perform particularly well if the prior guess is nearly correct.

6. Baseball data example

The data in Table 5 give the baseball batting averages (the number of hits divided by the num-

Table 5. Number of times at bat n_i, batting average Z_i in 1990 and career batting average p_i, of 10 baseball players

Player	n_i	Z_i	p_i
Baines	415	0.284	0.289
Barfield	476	0.246	0.256
Bell	583	0.254	0.265
Biggio	555	0.276	0.287
Bonds	519	0.301	0.297
Bonilla	625	0.280	0.279
Boggs	619	0.302	0.328
Brett	544	0.329	0.305
Brooks, Jr	568	0.266	0.269
Browne	513	0.267	0.271

ber of times 'at bat') of $k=10$ players, all of whom were active players in 1990. The source was `http://www.baseball-reference.com`. For $i=1,\ldots,k$, let n_i and Z_i respectively denote the number of times at bat and the batting average of the ith player during the 1990 season. Further, let p_i denote the player's true batting average, taken to be his career batting average. (Each player had at least 3000 at bats in his career.) We consider the model where $Z_1,\ldots,Z_k$ are independent, with

$$Z_i \sim \frac{1}{n_i}\mathrm{Bin}(n_i, p_i).$$

Making the variance stabilizing transformation

$$X_i = \sqrt{n_i}\sin^{-1}(2Z_i - 1),$$

and letting $\theta_i = n_i^{1/2}\sin^{-1}(2p_i - 1)$, then we have, approximately,

$$X \sim N_k(\theta, I),$$

with $X=(X_1,\ldots,X_k)$ and $\theta=(\theta_1,\ldots,\theta_k)$. In fact, since $\min_i(n_i) \geqslant 400$, an exact calculation gives that the variance of each X_i is between 1 and 1.005 for $p_i \in [0.2, 0.8]$. For our prior guess $\theta_0 = (\theta_{0,1},\ldots,\theta_{0,k})$, we take

$$\theta_{0,i} = n_i^{1/2}\sin^{-1}(2p_0 - 1), \qquad i=1,\ldots,k,$$

with $p_0 = 0.275$. Letting $p=(p_1,\ldots,p_k)$ and recalling that $\theta=\theta(p)$, we may write the analytic confidence set for p as

$$\{p \in [0,1]^k : \|T^+_{\mathrm{S},\theta_0}(X) - \theta\|^2 \leqslant 16.0\}.$$

This compares with the bootstrap confidence set, which has 16.0 replaced with 12.8 above, and the usual confidence set $\{p : \|X-\theta\|^2 \leqslant 18.3\}$. Numerical integration using the algorithm of Arthur Stroud, which is available at `http://www.csit.fsu.edu/burkardt/f_src/stroud/stroud.html`, gives the analytic to usual and bootstrap to usual volume ratios as 0.51 and 0.17 respectively.

In this example, we had $\|X-\theta_0\|^2 = 15.5$, but almost half of the contribution to this squared norm comes from the exceptionally good 1990 season of Brett. If his records are removed and the process is repeated using the nine other players, the volume ratios above are reduced to 0.13 and 0.07 respectively, with $\|X-\theta_0\|^2=8.0$.

Acknowledgements

The author is grateful for helpful conversations with Alastair Young, James Norris, Peter Hall and Iain Johnstone, as well as to the referees for their constructive comments on an earlier draft.

Appendix A

A.1. Proof of proposition 1

The upper α-point w satisfies

$$\int_{\mathbb{R}^k} f(\|x-\theta\|^2)\, \mathbb{1}_{\{\|\gamma(\|x\|)x-\theta\|^2\leqslant w\}}\,\mathrm{d}x=1-\alpha.$$

If P is a $k\times k$ orthogonal matrix, then

$$\int_{\mathbb{R}^k} f(\|x-P\theta\|^2)\, \mathbb{1}_{\{\|\gamma(\|x\|)x-P\theta\|^2\leqslant w\}}\,\mathrm{d}x=\int_{\mathbb{R}^k} f(\|P^{\mathrm{T}}x-\theta\|^2)\, \mathbb{1}_{\{\|\gamma(\|x\|)P^{\mathrm{T}}x-\theta\|^2\leqslant w\}}\,\mathrm{d}x,$$

from which the result follows, on substituting $y=P^{\mathrm{T}}x$.

A.2. Proof of theorem 2

In view of lemma 2, it suffices to show that the boundary $\partial C^0(\theta)$ of $C^0(\theta)$ lies inside $C(\theta)$. Suppose that $x\in\partial C^0(\theta)$, let $r=\|x\|$ and, for $r>0$ and $\|\theta\|>0$, define $\cos(\beta)=x^{\mathrm{T}}\theta/\|x\|\,\|\theta\|$. Then

$$r=\|\theta\|\cos(\beta)+\{c^2-\|\theta\|^2\sin^2(\beta)\}^{1/2}.$$

In their theorem 2.1, Hwang and Casella (1982) proved that $C^0(\theta)$ is contained in the set

$$\{x\in\mathbb{R}^k : \|T_{\mathrm{S}}^{+}(x)-\theta\|^2\leqslant c^2\},$$

for $\|\theta\|\leqslant c$, and the result is trivial if either $\|\theta\|=0$ or $r\leqslant a^{1/2}$. Therefore, for $0<\|\theta\|\leqslant w_\alpha(0)$ and $r>a^{1/2}$, we let

$$\begin{aligned} f(\beta,\|\theta\|)&=\gamma^2(r)r^2-2\|\theta\|\,\gamma(r)r\cos(\beta)+\|\theta\|^2-w_\alpha(0)-\frac{1}{2}w''_\alpha(0)r^2\\ &=-\frac{a^2}{c^2}+\frac{a^2}{r^2}+\frac{2a\|\theta\|\cos(\beta)}{r}-\frac{1}{2}w''_\alpha(0)r^2\\ &=-\frac{a^2}{c^2}+\frac{a^2}{r^2}+\frac{a(r^2+\|\theta\|^2-c^2)}{r^2}-\frac{1}{2}w''_\alpha(0)r^2, \end{aligned}$$

where $r=\|\theta\|\cos(\beta)+\{c^2-\|\theta\|^2\sin^2(\beta)\}^{1/2}$, so it is enough to show that $f(\beta,\|\theta\|)\leqslant 0$ for all $\beta\in[0,\pi]$ and $\|\theta\|$ in the given range.

Since $\partial r/\partial\beta=-\|\theta\|r\sin(\beta)/\{c^2-\|\theta\|^2\sin^2(\beta)\}^{1/2}$, we find that

$$\frac{\partial f}{\partial\beta}=\frac{-2\|\theta\|\sin(\beta)}{\{c^2-\|\theta\|^2\sin^2(\beta)\}^{1/2}}\left\{\frac{a(c^2-\|\theta\|^2-a)}{r^2}-\frac{1}{2}w''_\alpha(0)r^2\right\},$$

from which we deduce that f has turning-points at $\beta=0,\pi$ and possibly at β^*, where

$$[\|\theta\|\cos(\beta^*)+\{c^2-\|\theta\|^2\sin^2(\beta^*)\}^{1/2}]^4=\frac{2a(c^2-\|\theta\|^2-a)}{w''_\alpha(0)}.$$

Since r is a decreasing function of $\beta \in [0, \pi]$, a solution to this last equation exists if and only if

$$(c-\|\theta\|)^4 \leqslant \frac{2a(c^2-\|\theta\|^2-a)}{w_\alpha''(0)} \leqslant (c+\|\theta\|)^4.$$

Observe first that

$$f(\pi, \|\theta\|) = \left(c - \frac{a}{c-\|\theta\|}\right)^2 - \left(c-\frac{a}{c}\right)^2 - \frac{1}{2} w_\alpha''(0)(c-\|\theta\|)^2 \leqslant 0.$$

Next,

$$\begin{aligned} f(0, \|\theta\|) &\leqslant -\frac{a^2}{c^2} + \frac{a^2}{(c+\|\theta\|)^2} + \frac{2a\|\theta\|}{c+\|\theta\|} - \frac{a(k-1)}{c^2 k}(c+\|\theta\|)^2 \\ &\leqslant \frac{1}{c^2(c+\|\theta\|)^2}\left[\|\theta\|\left\{2ac^3 - \frac{4a(k-1)c^3}{k}\right\} + \|\theta\|^2\left\{2ac^2 - \frac{6ac^2(k-1)}{k}\right\}\right] \\ &\leqslant 0. \end{aligned}$$

Finally,

$$\begin{aligned} f(\beta^*, \|\theta\|) &= \frac{1}{r^2}\left[a\left(1-\frac{a}{c^2}\right)\left\{\frac{2a(c^2-\|\theta\|^2-a)}{w_\alpha''(0)}\right\}^{1/2} - 2a(c^2-\|\theta\|^2-a)\right] \\ &= \frac{a(c^2-\|\theta\|^2-a)^{1/2}}{r^2}\left[\left(1-\frac{a}{c^2}\right)\left\{\frac{2a}{w_\alpha''(0)}\right\}^{1/2} - 2(c^2-\|\theta\|^2-a)^{1/2}\right]. \end{aligned}$$

Thus we find $f(\beta^*, \|\theta\|)$ is non-positive for

$$\|\theta\|^2 \leqslant c^2 - a - \left(1-\frac{a}{c^2}\right)^2 \frac{a}{2\, w_\alpha''(0)} = \frac{c^2-a}{2\, w_\alpha''(0) c^4}\{2\, w_\alpha''(0)c^4 - (c^2-a)a\}.$$

References

Beran, R. J. (1995) Stein confidence sets and the bootstrap. *Statist. Sin.*, **5**, 109–127.

Beran, R. J. (1997) Diagnosing bootstrap success. *Ann. Inst. Statist. Math.*, **49**, 1–24.

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

Box, G. E. P. (1953) Spherical distributions. *Ann. Math. Statist.*, **24**, 687–688.

Brandwein, A. R. (1979) Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss. *J. Multiv. Anal.*, **9**, 579–588.

Brandwein, A. R. and Strawderman, W. E. (1978) Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Ann. Statist.*, **6**, 377–416.

Brandwein, A. and Strawderman, W. (1990) Stein estimation: the spherically symmetric case. *Statist. Sci.*, **5**, 356–369.

Brown, L. D. and Low, M. G. (1996) Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.

Brown, L. D., Low, M. G. and Zhao, L. H. (1997) Superefficiency in nonparametric function estimation. *Ann. Statist.*, **25**, 2607–2625.

Buehler, R. J. (1959) Some validity criteria for statistical inferences. *Ann. Math. Statist.*, **8**, 716–761.

Casella, G. and Hwang, J. T. (1983) Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Am. Statist. Ass.*, **78**, 688–698.

Casella, G. and Hwang, J. T. (1986) Confidence sets and the Stein effect. *Communs Statist. Theory Meth.*, **15**, 2043–2063.

Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia (with discussion)? *J. R. Statist. Soc.* B, **57**, 301–369.

Evans, S. and Stark, P. (1996) Shrinkage estimators, Skorokhod's problem and stochastic integration by parts. *Ann. Statist.*, **24**, 809–815.

Faith, R. E. (1976) Minimax Bayes set and point estimators of a multivariate normal mean. *Technical Report 66*. University of Michigan, Ann Arbor.

Fang, K. T., Kotz, S. and Ng, K. W. (1989) *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall.
Fisher, R. A. (1956) *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.
Fisher, R. A. (1959) Mathematical probability in the natural sciences. *Technometrics*, **1**, 21–29.
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. Decsns*, 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. M. (1961) Estimation with quadratic loss. In *Proc. 4th Berkeley Symp. Mathematical Statistics*, vol. 1, pp. 361–379. Berkeley: University of California Press.
Joshi, V. M. (1969) Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal distribution. *Ann. Math. Statist.*, **38**, 1868–1875.
Kelker, D. (1970) Distribution theory of spherical distributions and a location-scale parameter generalisation. *Sankhya* A, **32**, 419–430.
Ki, F. and Tsui, K.-W. (1985) Improved confidence set estimators of a multivariate normal mean and generalizations. *Ann. Inst. Statist. Math.*, **37**, 487–498.
Lu, K. L. and Berger, J. O. (1989) Estimated confidence procedures for multivariate normal means. *J. Statist. Planng Inf.*, **23**, 1–19.
Robert, C. and Casella, G. (1990) Improved confidence sets for spherically symmetric distributions. *J. Multiv. Anal.*, **32**, 84–94.
Robert, C. and Casella, G. (1994) Improved confidence statements for the usual multivariate normal confidence set. In *Statistical Decision Theory and Related Topics*, vol. V, pp. 351–368. New York: Academic Press.
Robinson, G. K. (1979a) Conditional properties of statistical procedures. *Ann. Statist.*, **7**, 742–755.
Robinson, G. K. (1979b) Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist.*, **7**, 756–771.
Samworth, R. J. (2003a) A note on methods of restoring consistency to the bootstrap. *Biometrika*, **90**, 985–990.
Samworth, R. J. (2003b) Some mathematical and theoretical aspects of the bootstrap. *PhD Thesis*. University of Cambridge, Cambridge.
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. In *Proc. 3rd Berkeley Symp. Mathematical Statistics*, vol. 1, pp. 197–206. Berkeley: University of California Press.
Stein, C. M. (1962) Confidence sets for the mean of a multivariate normal distribution (with discussion). *J. R. Statist. Soc.* B, **24**, 265–296.
Stein, C. (1981) Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135–1151.
Thomas, D. H. (1970) Some contributions to radial probability distributions. *PhD Dissertation*. Wayne State University, Detroit.
Tseng, Y.-L. and Brown, L. D. (1997) Good exact confidence sets for a multivariate normal mean. *Ann. Statist.*, **25**, 2228–2258.
Wang, H. (2000) Improved confidence estimators for the multivariate normal confidence set. *Statist. Sin.*, **10**, 659–664.
Zellner, A. (1976) Bayesian and non-Bayesian analysis of the regression model with multivariate student-*t* error terms. *J. Am. Statist. Ass.*, **71**, 400–405.