

Empirical Bayes Confidence Sets for the Mean
of a Multivariate Normal Distribution

GEORGE CASELLA and JIUNN TZON HWANG*

Through the use of an empirical Bayes argument, a confidence set for the mean of a multivariate normal distribution is derived. The set is a recentered sphere, is easy to compute, and has uniformly smaller volume than the usual confidence set. An exact formula for the coverage probability is derived, and numerical evidence is presented which shows that the empirical Bayes set uniformly dominates the usual set in coverage probability.

KEY WORDS: Confidence sets; Stein estimation; Multivariate normal density.

* George Casella is an Assistant Professor of Biological Statistics and Jiunn Tzon Hwang is an Assistant Professor of Mathematics, both at Cornell University, Ithaca, NY 14853. Dr. Casella's research was supported by National Science Foundation Grant No. MCS81-02541 and Dr. Hwang's research was supported by National Science Foundation Grant No. MCS80-03568. This paper is BU-767-M in the Biometrics Unit Series, Cornell University, Ithaca, NY 14853. The authors wish to thank the Editor, Associate Editor and a referee for their helpful comments which led to a much improved version of this paper.

1. INTRODUCTION

In many situations an experimenter is not only interested in obtaining point estimates of an unknown parameter, but also in associating with these point estimates a confidence set which guarantees a certain probability of covering the true parameter. Although there have been many breakthroughs in the theory of point estimation, little research has been aimed at the problem of set estimation. This is not because the problem is an unimportant one (indeed, many improved point estimators suffer in application from the lack of an associated confidence region), but rather because the set estimation problem involved great technical difficulty.

In this paper we take a step in providing an applicable confidence region for use with an improved point estimator, the positive-part Stein estimator. The region is uniformly smaller than the usual one, and strong evidence is presented to support the claim that the region retains a specified confidence coefficient. The region is developed as an empirical Bayes solution to a decision-theoretic estimation problem. This structure is employed in order to obtain a reasonable form for the confidence region. Our ultimate evaluations of the performance are in terms of familiar frequentist criteria: volume and coverage probability.

The procedure derived is applicable in many instances where simultaneous statements are desired; estimation of many contrasts in the analysis of variance or simultaneous interval estimates for regression coefficients. The applications are limited, however, to cases where the variances are known up to a common scale factor (i.e., a covariance matrix of the form $\sigma^2\Sigma$, where Σ is known). Our techniques, as of yet, do not extend to cases where there are more than one totally unknown variance, which is an important case for future study.

In both the analysis of variance and linear regression, the estimation problem can be reduced to that of estimating the mean vector, θ , of a multivariate normal distribution. For now, we assume that the covariance matrix is in the identity.

The classic (maximum likelihood) point estimator based on one observation, x , is x itself, and the classic $1 - \alpha$ confidence set for θ is

$$C_x^0 = \{\theta : |\theta - x| \leq c\} , \quad (1.1)$$

where c satisfies $P(\chi_p^2 \leq c^2) = 1 - \alpha$, and $|\cdot|$ denotes the Euclidean norm. (The quantity $1 - \alpha$ is called the confidence coefficient. A $1 - \alpha$ confidence procedure C satisfies $\inf_{\theta} P_{\theta}(\theta \in C) = 1 - \alpha$.)

Just as it is possible to improve upon x (in terms of risk) as a point estimator of θ , it is possible to improve upon C_x^0 as a set estimator of θ . We consider a procedure C to be an improvement over C_x^0 if the following are satisfied:

$$\begin{aligned} \text{i) } & P_{\theta}(\theta \in C) \geq P_{\theta}(\theta \in C_x^0) && \text{for all } \theta , \\ \text{ii) } & \text{Volume}(C) \leq \text{Volume}(C_x^0) && \text{for all } x , \end{aligned} \quad (1.2)$$

with strict inequality either in i) for some θ or in ii) for all x in some set with positive Lebesgue measure. (In the terminology of Joshi (1969), C is strongly preferable to C_x^0 .) There is a technical caveat, first noticed by Joshi (1969), which should be mentioned. Since, by adding point sets to any confidence set C , it is possible to increase its coverage probability without increasing its volume, (1.2) is defined only up to an equivalence class, where we define two procedures C_1 and C_2 to be equivalent if their symmetric difference $(C_1 \setminus C_2) \cup (C_2 \setminus C_1)$ has Lebesgue measure zero.

If the dimension of the problem, p , is greater than two, the existence of a dominating procedure was established independently by Brown (1966) and

Joshi (1967). It was shown that by recentering the usual confidence set at a Stein-type estimator (hence keeping the same volume), a uniform improvement in coverage probability can be achieved. The arguments used were existential, however, and did not lead to a usable improved procedure. The problem of exhibiting a confidence set, and proving dominance over C_x^0 is one of enormous difficulty. Although progress has been made, this progress is minuscule when compared with that in the point estimation problem.

Significant progress was made by Berger (1980), although uniform dominance results (according to 1.2) were not obtained. However, strong evidence (both analytic and numerical) was presented which shows that the procedure derived is an improvement over C_x^0 . A major difficulty with this procedure is in implementation. Berger's confidence sets are of the form

$$C^* = \{\theta : [\theta - \delta(x)]' \Sigma^{-1}(x) [\theta - \delta(x)] \leq k^2\} , \quad (1.3)$$

where $\delta(x)$ is an admissible, generalized Bayes estimator of θ , and $\Sigma(x)$ is its posterior covariance matrix. C^* is an ellipse, so its interpretation is straightforward. Although C^* can yield remarkable improvement both in coverage probability and volume, it is fairly difficult to calculate, which limits its practical advantage.

More recently, Morris (1983) has investigated the question of improving upon componentwise interval estimates of each θ_i . Using intervals centered at empirical Bayes estimators, with length determined by the posterior variance, Morris has demonstrated that it is possible to achieve substantial reduction in length while maintaining a confidence coefficient of approximately $1 - \alpha$. (Since the usual one-dimensional confidence interval is admissible (Joshi, 1969), it is impossible to dominate it uniformly.)

A simpler approach was taken by Hwang and Casella (1982). They considered sets of the form

$$C_{\delta}^+ = \{\theta : |\theta - \delta^+(x)| \leq c\} \quad , \quad (1.4)$$

where $\delta^+ = [1 - (a/|x|^2)]^+x$, a positive part James-Stein estimator. C_{δ}^+ is obtained by recentering C_x^0 at $\delta^+(x)$, hence has the same volume. The first analytic dominance results were obtained in Hwang and Casella (1982), where it was proved that, for a specified range of values of a , $P_{\theta}(\theta \in C_{\delta}^+) > P_{\theta}(\theta \in C_x^0)$ for all θ , when $p \geq 4$. The improvement in coverage probability is quite good, yielding values over 99% for some θ and p , when c corresponds to a 90% confidence coefficient.

Although sets of the form (1.4) provide uniform improvement in coverage probability, they have the same volume and confidence coefficient as C_x^0 . From a practical point of view, it would be more desirable to retain the same confidence coefficient as C_x^0 , but decrease the volume of the confidence set. In order to make such confidence sets easy to implement, they should have a simple form such as (1.4). The major goal of this paper is to examine sets of the form

$$C_{\delta}^v = \{\theta : |\theta - \delta^+(x)| \leq v(|x|)\} \quad , \quad (1.5)$$

where $v(|x|)$ is a nondecreasing function, $0 \leq v(|x|) \leq c$. Clearly C_{δ}^v has smaller volume than C_x^0 and, if $v(|x|)$ is a reasonably simple function, C_{δ}^v would be quite easy to compute. There are many problems in dealing with sets of the form (1.5); some mathematical and some statistical. Both problems are centered around the function $v(|x|)$. From a mathematical point of view, there are certain minimal requirements on the function $v(|x|)$ that are needed to obtain a workable formula for the coverage probability of C_{δ}^v . From a statistical point of view, we want the function $v(|x|)$ to be meaningful, since it will be interpreted somewhat like a standard deviation.

In Section 2 we present some preliminaries that are necessary for the complete development of C_{δ}^v . The preliminaries are aimed more at the mathematical, rather than statistical, problem but are of importance when considering the associated statistical problem of hypothesis testing. Section 3 contains the derivation of the formula for the coverage probability of a class of confidence sets which contain C_{δ}^v , along with some other related results. In Section 4 a specific form of the function $v(|x|)$ is derived through the use of a modified empirical Bayes argument. By deriving $v(|x|)$ in this way, we arrive at a functional form which has a meaningful statistical interpretation. The procedure is then evaluated using the criteria of volume and coverage probability. It is shown that this procedure can achieve significant volume reduction, and strong numerical evidence is also presented that shows that this procedure has uniformly higher coverage probability than the usual set. The fact that such a result is not demonstrated analytically, but numerically, is a limitation of our results; however, the formula for the coverage probability of C_{δ}^v (or almost any other variable-radius confidence set) is so complicated that analytic verification of dominance is enormously difficult. Thus, the coverage probabilities have been evaluated numerically (not simulated) for a wide range of $|\theta|$ and p . Section 5 contains some comments and generalizations, including a discussion of the component confidence question and the case of unknown variance, and Section 6 contains remarks about the results presented here and the set estimation problem in general.

2. PRELIMINARIES

Let x be one observation from a p -variate normal distribution with mean θ and covariance matrix I , where $x \in R^p$ and $\theta \in R^p$. A confidence procedure C is defined as a Lebesgue measurable subset of the product space $R^p \times R^p$,

$$\{(x, \theta) : (x, \theta) \in C\} \quad (2.1)$$

Associated with C are two cross sections obtained by fixing either x or θ . The x section of C , which is the confidence set for θ , is defined for each x as

$$C_x = \{\theta : (x, \theta) \in C\} . \quad (2.2)$$

The other cross section, the θ section, is defined for each θ as

$$C_\theta = \{x : (x, \theta) \in C\} . \quad (2.3)$$

It would seem that, if the interest is in set estimation of θ , then C_θ can be ignored. This is not quite true, however, for the following two reasons. Firstly, evaluation of the coverage probability of C_x usually proceeds by employing the tautology $x \in C_\theta$ if and only if $\theta \in C_x$, and hence

$$P_\theta[\theta \in C_x] = P_\theta[x \in C_\theta] . \quad (2.4)$$

Thus, evaluation of the coverage probability depends quite strongly on C_θ .

The confidence set C_δ^v given in (1.5), i.e.,

$$C_\delta^v = \{\theta : |\theta - \delta^+(x)| \leq v(|x|)\} , \quad (2.5)$$

is a sphere of radius $v(|x|)$. Thus, its interpretation as a confidence set is quite straightforward. When evaluating its coverage probability, however, we work with the θ section, which is given by

$$C_\theta^v = \{x : |\theta - \delta^+(x)| \leq v(|x|)\} . \quad (2.6)$$

In C_θ^v , θ is fixed and x is allowed to vary. Since x appears on both sides of the inequality, there is no guarantee that C_θ^v is a sphere (in general it will not be). In fact, there is no guarantee that C_θ^v is even a connected set; it

may be composed of many disjoint regions. The form of $v(|x|)$ will determine the structure of C_θ^v , and this is of great importance in deriving the correct expression for the coverage probability. In particular, whether or not C_θ is a connected set determines whether the integration must be carried out over one or more regions.

Secondly, for fixed $\theta = \theta_0$, the θ section of a confidence procedure,

$$C_{\theta_0} = \{x: (x, \theta_0) \in C\} \quad (2.7)$$

is the acceptance region for a test of the null hypothesis $H_0: \theta = \theta_0$. When seen in this setting, there are certain minimal properties which should be required of any θ section C_θ . The only property that we require of C_θ is that it be connected, i.e., between any two points in C_θ there is a continuous path in C_θ that connects them. This is a necessary condition for the avoidance of logical contradictions in the hypothesis test; for example, if C_{θ_0} is not connected it could be the case that H_0 is accepted for some value x , but would have been rejected for a value closer to θ_0 .

If the primary goal of the experiment is an hypothesis test rather than a confidence set, it should probably be required that C_θ is not only connected, but also convex. The results of Birnbaum (1955) show that a convex acceptance region is a necessary and sufficient condition for a test to be admissible. (More precisely, for testing $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$, tests with convex acceptance regions form a minimal complete class.)

3. COVERAGE PROBABILITIES OF A CLASS OF CONFIDENCE PROCEDURES

In this section the formula for the coverage probability of a general class of confidence sets is derived. We also establish sufficient conditions that insure that the θ sections are connected sets. Our primary goal in establishing these conditions is to facilitate the evaluation of coverage probabilities, but we also consider the associated hypothesis testing problem. General convexity results for the θ -sections are not obtained; however, for an important special case the acceptance regions considered are convex (and, hence, the tests are admissible).

Consider a confidence set of the form

$$C_{\theta}^v = \{\theta : |\theta - \delta(x)| \leq v(|x|)\} , \quad (3.1)$$

where x is an observation from a p -variate normal distribution with mean θ and identity covariance matrix, $\delta(x) = \gamma(|x|)x$, and $\gamma(|x|)$ and $v(|x|)$ are both nonnegative functions. The coverage probability of C_{θ}^v will be evaluated using the identity $P_{\theta}[\theta \in C_{\theta}^v] = P_{\theta}[x \in C_{\theta}^v]$, where C_{θ}^v is the θ section of the confidence procedure,

$$C_{\theta}^v = \{x : |\theta - \delta(x)| \leq v(|x|)\} . \quad (3.2)$$

If we let β be the angle between x and θ , $0 \leq \beta \leq \pi$, then we can write

$$C_{\theta}^v = \{x : |x|^2 \gamma^2(|x|) - 2|x||\theta|\gamma(|x|)\cos\beta + |\theta|^2 \leq v^2(|x|)\} . \quad (3.3)$$

In the following theorem, we derive necessary and sufficient conditions on the functions $\gamma(|x|)$ and $v(|x|)$ which insure that the θ section C_{θ}^v is connected.

Theorem 3.1: The set $C_{\theta}^v = \{x : |\theta - \gamma(|x|)x| \leq v(|x|)\}$ is connected if and only if the set $S_{\theta} = \{t : |\theta - t\gamma(t)| \leq v(t)\}$ is an interval.

Proof: Suppose S_θ is the interval $\{t: r_-(|\theta|) \leq t \leq r_+(|\theta|)\}$. Note that a point $k\theta$, where k is a scalar, is in C_θ^v if and only if $r_-(|\theta|) \leq |k\theta| \leq r_+(|\theta|)$. Thus any two points in C_θ^v that are on the ray through θ can be connected by a line in C_θ^v . We will show that for an arbitrary point $x \in C_\theta^v$, there is a path in C_θ^v to the ray through θ , which implies that there is a path in C_θ^v between any two points in the set, hence C_θ^v is connected.

Fix θ , and let $x \in C_\theta^v$. From (3.3) we have that

$$|x|^2 \gamma^2(|x|) - 2\gamma(|x|)\theta'x + |\theta|^2 \leq v^2(|x|) \quad (3.4)$$

Define the set $T = \{y: |y| = |x|, \theta'y \geq \theta'x\}$. From (3.4) it follows that $T \subset C_\theta^v$. (The set T contains a continuous path from x to a point on the ray through θ , along the surface of a sphere of radius $|x|$ centered at 0.) In particular, the point $y^* = (|x|/|\theta|)\theta \in T$ since

$$v^2(|x|) \geq |x|^2 \gamma^2(|x|) - 2\gamma(|x|)|x||\theta| + |\theta|^2 = [|x|\gamma(|x|) - |\theta|]^2,$$

by the Cauchy-Schwartz inequality. Hence C_θ^v is connected.

To prove that C_θ^v is connected only if S_θ is an interval, we consider the contrapositive. Let $t_1 < t_2 < t_3$ be such that $t_1, t_3 \in S_\theta$ but $t_2 \notin S_\theta$. Let $\theta^* = k\theta$ satisfy $|\theta^*| = t_2$. Then $\theta^* \notin C_\theta^v$. Moreover, if x is any point satisfying $|x| = t_2$, then

$$\begin{aligned} |x|^2 \gamma^2(|x|) - 2\gamma(|x|)\theta'x + |\theta|^2 &\geq |x|^2 \gamma^2(|x|) - 2\gamma(|x|)|\theta||x| + |\theta|^2 \\ &= (|x|\gamma(|x|) - |\theta|)^2 > v^2(|x|), \end{aligned}$$

since $|x| = t_2$ and $t_2 \notin S_\theta$. Thus, the shell $\{x: |x| = t_2\}$ separates C_θ^v into two non-overlapping sets, and C_θ^v cannot be connected. ||

We now establish the convexity of C_θ^v , and hence the admissibility of the associated test, for a particular null hypothesis.

Theorem 3.2: For testing $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$, the acceptance region

$$C_{\theta}^v = \{x: |\gamma(|x|)x| \leq v(|x|)\} \quad (3.5)$$

is convex if $\{t: |t\gamma(t)| \leq v(t)\}$ is an interval.

Proof: The acceptance region is of the form $\{x: \gamma(|x|)|x| \leq v(|x|)\}$, which is equal to the set $\{x: 0 \leq |x| \leq B\}$, for some $B > 0$. This last set is a sphere, which is convex. ||

Since the condition of this theorem is a special case of that of Theorem 3.1, it immediately follows that any set which satisfies Theorem 3.1 also satisfies Theorem 3.2. More importantly, since $\gamma(|x|)$ will usually be chosen to satisfy $0 \leq \gamma(|x|) \leq 1$, the estimator $\delta(x) = \gamma(|x|)x$ shrinks x toward zero. Such an estimator is really only appropriate when there is some prior belief that θ is near zero. (Although $\delta(x)$ is usually chosen to be minimax, and hence will uniformly improve on x , the region of significant risk improvement is centered around $\theta = 0$.) If it is thought that θ is near some value other than 0, say θ_0 , a better estimator is

$$\delta^*(x) = \theta_0 + \gamma(|x - \theta_0|)(x - \theta_0) \quad , \quad (3.6)$$

which retains the good risk properties of $\delta(x)$, but centers the region of significant risk improvement around θ_0 . The following corollary follows immediately from Theorem 3.2:

Corollary 3.1: If C_{θ}^v of (3.5) is convex, then the set

$$\{x: |\theta_0 - \delta^*(x)| \leq v(|x - \theta_0|)\}$$

is convex, and hence provides an admissible test of $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

We now turn to the evaluation of the coverage probability of $C_{\theta}^v = \{\theta: |\theta - \gamma(|x|)x| \leq v(|x|)\}$. The representation given in the following

theorem can be modified to include procedures with disconnected θ sections, but this seems to have little practical value.

Theorem 3.4: Let $p \geq 2$, and $C_{\theta}^v = \{\theta: |\theta - \gamma(|x|)x| \leq v(|x|)\}$, where the functions γ and v satisfy

- i) $\gamma(t) \geq 0$ and $v(t) \geq 0 \quad \forall t > 0$,
- ii) for fixed $|\theta|$, the set $S_{\theta} = \{t: |t\gamma(t) - |\theta|| \leq v(t)\}$ is an interval $[r_{-}(|\theta|), r_{+}(|\theta|)]$.

If $|\theta| > 0$ then

$$P_{\theta}[\theta \in C_{\theta}^v] = k \int_{r_{-}(|\theta|)}^{r_{+}(|\theta|)} r^{p-1} e^{-\frac{1}{2}(r^2 + |\theta|^2)} \int_{h(r)}^1 (1-u^2)^{\frac{p-3}{2}} e^{r|\theta|u} du, \quad (3.7)$$

where $k^{-1} = \sqrt{\pi} \Gamma\left(\frac{p-1}{2}\right) 2^{(p-2)/2}$, and

$$h(r) = \max\left\{\frac{r^2 \gamma^2(r) + |\theta|^2 - v^2(r)}{2r\gamma(r)|\theta|}, -1\right\} \quad \text{if } r\gamma(r) \neq 0$$

$$= -1 \quad \text{if } r\gamma(r) = 0.$$

If $|\theta| = 0$, then $P_{\theta}[\theta \in C_{\theta}^v] = P[\chi_p^2 \leq r_{+}(0)]$.

Proof: If $|\theta| = 0$ then $P_{\theta}[\theta \in C_{\theta}^v] = P_{\theta}[x: |\gamma(|x|)x| \leq v(|x|)]$. Clearly the region $\{|x|: |x|\gamma(|x|) \leq v(|x|)\}$ contains $|x| = 0$, hence $r_{-}(0) = 0$ and the result follows.

If $|\theta| > 0$, the coverage probability of $C_{\theta}^v(\theta)$ is

$$P_{\theta}[\theta \in C_{\theta}^v] = (2\pi)^{-p/2} \int_{\{x: |\delta(x) - \theta| \leq v(|x|)\}} e^{-\frac{1}{2}|x-\theta|^2} dx. \quad (3.8)$$

Transform to the spherical coordinates $r = |x|$, $\cos\beta = x'\theta/|x||\theta|$. In terms of these variables, the region of integration becomes

$$\begin{aligned} & \{(r, \beta): r^2 \gamma^2(r) - 2r|\theta|\gamma(r)\cos\beta + |\theta|^2 \leq v^2(r)\} \\ & = \{(r, \beta): |\theta|^2 \leq v^2(r), r\gamma(r) = 0\} \\ & \cup \{(r, \beta): [r\gamma(r) - |\theta|]^2 \leq v^2(r), r\gamma(r) \neq 0, \cos\beta \geq h(r)\} \\ & = \{(r, \beta): r_-(|\theta|) \leq r \leq r_+(|\theta|), \cos\beta \geq h(r)\} \end{aligned} \quad (3.9)$$

Direct substitution yields

$$\begin{aligned} P_\theta[\theta \in C_\delta^v] &= k \int_{r_-(|\theta|)}^{r_+(|\theta|)} \int_0^{\cos^{-1}[h(r)]} r^{p-1} \sin^{p-2}\beta e^{-\frac{1}{2}(r^2 - 2r|\theta|\cos\beta + |\theta|^2)} d\beta dr \\ &= k \int_{r_-(|\theta|)}^{r_+(|\theta|)} \int_{h(r)}^1 r^{p-1} (1-u^2)^{\frac{p-3}{2}} e^{-\frac{1}{2}(r^2 - 2r|\theta|u + |\theta|^2)} du dr, \end{aligned}$$

where the last equality follows from the substitution $u = \cos\beta$. ||

If p is odd, the inner integral can be evaluated using the binomial formula. If we first apply the transformation $s = r|\theta|(1-u)$, and then use the binomial expansion, we obtain our computational formula

$$P_\theta[\theta \in C_\delta^v] = k \int_{r_-}^{r_+} \left(\frac{r}{|\theta|}\right)^{n+1} e^{-(r-|\theta|)^2/2} \sum_{i=0}^n \left\{ \left[\frac{\binom{n}{i} 2^{n-i} (-1)^i (n+i)!}{(r|\theta|)^i} \right] \left[1 - \sum_{j=0}^{n+i} \frac{e^{-r|\theta|\Delta} (r|\theta|\Delta)^{n+i-j}}{(n+i-j)!} \right] \right\} dr, \quad (3.10)$$

where $n = (p-3)/2$ and $\Delta = \min\{1-h(r), 2\}$.

One difficulty, which causes major problems when dealing with (3.7) analytically, is that, in general, there is no explicit solution for $r_-(|\theta|)$ and $r_+(|\theta|)$, which are the roots of the equation $||\theta| - t\gamma(t)| - v(t) = 0$. If $v(t)$ is constant (as in Hwang and Casella, 1982), then the roots can be explicitly determined; however, such confidence sets do not yield a volume reduction.

For reasonable choices of γ and v , such as those considered in the next

section, very good bounds can be obtained, and $r_{-}(|\theta|)$ and $r_{+}(|\theta|)$ can be calculated quite rapidly. This makes the numerical evaluation of (3.7) a relatively simple task.

4. AN EMPIRICAL BAYES CONFIDENCE SET

In this section we consider specific choices of the functions γ and v which lead to improved confidence sets for θ . For the function γ , we consider $\gamma(|x|) = [1 - (a/|x|^2)]^+$, where a is a constant, which leads us to centering the confidence set at the positive-part James-Stein estimator. This choice is based on both theoretical and practical considerations. From a theoretical point of view, it is known that it is difficult to improve upon the positive-part James-Stein estimator as a point estimator, and, moreover, it has been shown (Hwang and Casella, 1982), that the set

$$\{\theta: |\theta - [1 - (a/|x|^2)]^+x| \leq c\}$$

has, for a range of values of a , higher coverage probability than the set $\{\theta: |\theta - x| \leq c\}$ for all θ . From a practical point of view, this estimator is much easier to calculate than its admissible counterparts, and hence is more likely to be used.

Our main concern here, however, is with confidence sets of variable radii, which leads us to consider specific choices of the function $v(|x|)$, a more difficult task. The major goal is to dominate the confidence set

$$C_x^0 = \{\theta: |\theta - x| \leq c\} ,$$

in both volume and coverage probability, with the set

$$C_\delta^v = \{\theta: |\theta - \delta^+(x)| \leq v(|x|)\} .$$

Thus, we immediately require that $v(|x|) \leq c$ for all $|x|$. Indeed, for large

$|x|$ we must have $v(|x|) \approx c$, so it is the growth rate of $v(|x|)$ which becomes important. We also keep in mind the condition from Section 3 to insure that C_θ^v is connected. With this restriction on $v(|x|)$, we can derive a very simple condition which is necessary and sufficient for C_δ^v to dominate C_x^o in coverage probability at $\theta = 0$ (and hence is necessary for overall dominance).

Theorem 4.1: If the set $\{t: t - a/t \leq v(t), t^2 \geq a\}$ is an interval, say $[a, t_0]$, then a necessary and sufficient condition for C_δ^v to dominate C_x^o in coverage probability at $\theta = 0$ is $v(c) \geq c - a/c$.

Proof: At $\theta = 0$,

$$\begin{aligned} P_\theta[\theta \in C_\delta^v] &= P_\theta(|x|^2 < a) + P_\theta[|x| - (a/|x|) \leq v(|x|), |x|^2 > a] \\ &= P_\theta(|x|^2 \leq t_0^2) . \end{aligned}$$

Since $P_\theta[\theta \in C_x^o] = P_\theta(|x|^2 \leq c^2)$, and $c \in [0, t_0]$ if and only if $v(c) \geq c - (a/c)$, the result follows. ||

The theorem also gives a lower bound on the improvement in volume when $|x| = c$. If, in fact, $v(c) = c - a/c$, then the ratio of the radius of C_δ^v to C_x^o at $\theta = 0$ is $1 - a/c^2$.

Unfortunately, choices of $v(|x|)$ which attain this lower bound were found (numerically) to fail to dominate in coverage probability. One natural choice that was tried was

$$\begin{aligned} v^2(|x|) &= c^2[1 - (a/c^2)] && \text{if } |x| \leq c \\ &= c^2[1 - (a/|x|^2)] && \text{if } |x| > c , \end{aligned} \tag{4.1}$$

but this $v(|x|)$ does not lead to a dominating procedure. It is interesting to note that, of the procedures which failed to dominate in coverage probability, including (4.1), the region of failure was for $|\theta|$ values near c .

Instead of searching for a suitable $v(|x|)$ in a haphazard fashion, a more structured approach, an empirical Bayes approach, was attempted. If we measure the loss of a confidence set C by

$$L(\theta, C) = k \text{Volume}(C) - I_C(\theta) \quad , \quad (4.2)$$

where $I_C(\theta) = 1$ if $\theta \in C$ and 0 otherwise, then the Bayes rules against $L(\theta, C)$ are of the form (Joshi, 1969)

$$C_x^B = \{ \theta : \pi(\theta|x) > k \} \quad , \quad (4.3)$$

where $\pi(\theta|x)$ is the posterior distribution of θ given x . Moreover, if $k = k_0 = e^{-c^2/2}/(2\pi)^{p/2}$, then the usual confidence set $C_x^0 = \{ \theta : |\theta - x| \leq c \}$ is minimax against $L(\theta, C)$.

The Bayes rule against $L(\theta, C)$, assuming that θ is multivariate normal with mean 0 and covariance matrix $\tau^2 I$, is given by

$$C_x^B = \left\{ \theta : |\theta - \delta^B(x)|^2 \leq -\frac{2\tau^2}{\tau^2+1} \log \left[k \left(\frac{2\pi\tau^2}{\tau^2+1} \right)^{p/2} \right] \right\} \quad ,$$

where $\delta^B(x) = [\tau^2/(\tau^2+1)]x$. If we set $k = k_0$, this set can be written

$$C_x^B = \left\{ \theta : |\theta - \delta^B(x)|^2 \leq \frac{\tau^2}{\tau^2+1} \left[c^2 - p \log \left(\frac{\tau^2}{\tau^2+1} \right) \right] \right\} \quad . \quad (4.4)$$

The marginal distribution of x is multivariate normal with mean 0 and covariance matrix $(\tau^2+1)I$. Thus, marginally,

$$E \left(1 - \frac{p-2}{|x|^2} \right) = \frac{\tau^2}{\tau^2+1} \quad ,$$

and the empirical Bayes approach suggests replacing $\tau^2/(\tau^2+1)$ in (4.4) by its unbiased estimate. This would lead to an unreasonable procedure, however,

since the quantity $\left(1 - \frac{p-2}{|x|^2}\right)$ can become negative, and its logarithm would not be defined. We therefore consider a modification of the empirical Bayes estimator given by

$$C_{\delta}^E = \{\theta : |\theta - \delta^+(x)| \leq v_E(|x|)\} , \quad (4.5)$$

where, here,

$$\delta^+(x) = \{1 - [(p-2)/|x|^2]\}^+ x , \quad (4.6)$$

and

$$\begin{aligned} v_E^2(|x|) &= \left(1 - \frac{p-2}{c^2}\right) \left[c^2 - p \log\left(1 - \frac{p-2}{c^2}\right) \right] && \text{if } |x| \leq c \\ &= \left(1 - \frac{p-2}{|x|^2}\right) \left[c^2 - p \log\left(1 - \frac{p-2}{|x|^2}\right) \right] && \text{if } |x| > c . \end{aligned} \quad (4.7)$$

It is straightforward to verify that $v_E^2(|x|)$ is nondecreasing in $|x|$ if $c^2 > p$, and $\lim_{|x| \rightarrow \infty} v_E^2(|x|) = c^2$. Hence C_{δ}^E provides uniform volume improvement over C_x^0 .

The restriction that $c^2 > p$ is really quite minor. The confidence set

$$C_x^0 = \{\theta : |\theta - x|^2 \leq p\}$$

has a confidence coefficient of, roughly, 55%, and hence is of little practical value. Also, as will be seen in the following theorem, the condition that $c^2 > p$ is sufficient to guarantee the connectedness of C_{δ}^E .

Theorem 4.2: If $c^2 > p$, then for all $|\theta|$, the set

$$\left\{ t : \left| t \left(1 - \frac{p-2}{t^2}\right)^+ - |\theta| \right| \leq v_E(t) \right\}$$

is an interval, and hence C_{δ}^E is connected.

Proof: Given in the appendix.

C_{δ}^E also satisfies Corollary 3.1, so the acceptance region

$$\{x: |\delta(x - \theta_0)| \leq v_E(|x - \theta_0|)\},$$

where δ and v are given in (4.6) and (4.7), provides an admissible test of $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

The function $v_E(|x|)$ also satisfies the conditions of Theorem 4.1, so we know that C_{δ}^E dominates C_x^O at $\theta = 0$. Dominance in coverage probability for other values of θ could not be proved analytically, however, but the formula of Theorem 3.4 allows exact computation of these probabilities.

The calculation of $P_{\theta}[\theta \in C_{\delta}^E]$ involves first finding the roots $r_-(|\theta|)$ and $r_+(|\theta|)$ of the equation

$$||\theta| - [t - (p - 2)/t]^+| = v_E(t),$$

and then performing the numerical integration. Although there is no explicit solution for $r_-(|\theta|)$ and $r_+(|\theta|)$, the following expressions yield very tight bounds. Fix $|\theta|$, and define, for $u > 0$,

$$b_{\pm}(u) = \frac{1}{2}\{(|\theta| \pm u) + [(|\theta| \pm u)^2 + 4a]^{\frac{1}{2}}\},$$

then

- i) If $|\theta| \leq v_E(0)$, $r_-(|\theta|) = 0$ and $b_+(0) \leq r_+(|\theta|) \leq b_+(c)$.
- ii) If $v_E(0) < |\theta| \leq v_E(\infty) = c$, then $(p - 2)^{\frac{1}{2}} \leq r_-(|\theta|) \leq b_-[v_E(0)]$
 $b_+[v_E(0)] \leq r_+(|\theta|) \leq b_+(c)$.
- iii) If $|\theta| > v_E(\infty)$, then $b_-(c) \leq r_-(|\theta|) \leq b_-[v_E(0)]$ and $b_+[v_E(0)]$
 $\leq r_+(|\theta|) \leq b_+(c)$.

The coverage probabilities of C_{δ}^E , for selected odd values of p , are shown in Figures 1 and 2. The probabilities were computed using formula (3.10). The numerical evidence is quite strong; with the exception of the

cases $p=3$, $\alpha=.1$, and $p \leq 5$, $\alpha=.05$, C_8^E dominates C_x^O in coverage probability. C_8^E provides a substantial gain in coverage probability for $|\theta|$ near zero, with the gain decreasing as $|\theta|$ increases (since C_8^E is converging to C_x^O as $|\theta|$ increases).

Although C_8^E performs well for all values of p considered, it is not a $1-\alpha$ confidence set for $p=3$, $\alpha=.1$ or $p \leq 5$, $\alpha=.05$. The failure of C_8^E is so slight in these cases, however, that for all practical purposes it can be ignored. The problem is that, for these cases, $v_E(|x|)$ is too small. This problem can be overcome by modifying $v_E(|x|)$. One modification that works is to replace $(p-2)$ by $(p-2)/2$ in (4.7). The resulting procedure is now a $1-\alpha$ confidence procedure for these cases, at the expense of increased volume (but still smaller volume than C_x^O).

The way in which C_8^E fails in the cases cited is also of great interest. For these cases, the coverage probability is not a decreasing function of $|\theta|$, as it is in the case of constant volume, and $P_\theta(\theta \in C_8^E)$ falls below the nominal value for middle values of $|\theta|$ (approximately $|\theta|=p$), where it seems to have a global minimum. This observation adds to the already difficult task of showing dominance analytically; the minimum must be identified and the procedure must be shown to dominate at this value. (This phenomenon is not unique to $v_E(|x|)$. Most of the other radius functions examined exhibited similar behavior.) For larger values of p , this behavior persists, as can be seen in Figures 1 and 2. Thus, it appears that variable radius confidence sets do not, in general, have decreasing coverage probabilities, which only adds to the difficulty of any analytic argument.

As can be seen from Figures 1 and 2, the coverage probability of C_8^E is not a continuous function of $|\theta|$. The one discontinuity occurs at the point

$|\theta| = v_E(0)$. The discontinuity is partially due to the use of $\delta^+(x)$ as the center of the set, but also to the fact that $v_E(|x|)$ is constant for $0 \leq |x| \leq c$. (The sets of Hwang and Casella (1982) exhibit a similar discontinuity at the point $|\theta| = c$.) If $v_E(|x|)$ were replaced by a monotone increasing function, the coverage probability would become continuous. However, we do not view the discontinuity as creating any practical problem, and prefer to work with these simple sets.

The practical gain in using a procedure such as C_δ^E , however, comes in the volume reduction achieved. An experimenter, whether using C_δ^E or C_x^O will report the same confidence coefficient, but use of C_δ^E allows a smaller radius for the same confidence coefficient. Two measures of volume reduction are considered. The first is merely the ratio of the volume of C_δ^E to C_x^O . Since C_δ^E is a sphere of radius $v_E(|x|)$, the ratio of the volumes is given by

$$\frac{\text{Vol}[C_\delta^E]}{\text{Vol}[C_x^O]} = \left(\frac{v_E(|x|)}{c} \right)^p \quad (4.8)$$

The second measure, known as the ratio of the effective radii, is given by the pth root of (4.8). Since C_δ^E is a sphere, this is merely the ratio of the radii.

By construction, $v_E(|x|) \leq c$, so C_δ^E has uniformly smaller volume than C_x^O . For small values of $|x|$, this reduction can be quite significant. From (3.6), if $|x| \leq c$, then

$$\frac{\text{radius of } C_\delta^E}{\text{radius of } C_x^O} = \left[\left(1 - \frac{p-2}{c^2}\right) \left[1 - \frac{p}{c^2} \log\left(1 - \frac{p-2}{c^2}\right)\right] \right]^{\frac{1}{2}},$$

which can result in a substantial improvement, especially for larger values

of p . Moreover, from Theorem 4.1, it follows that the greatest improvement attainable, while still maintaining dominance in coverage probability, is $\{1 - [(p-2)/c^2]\}^{\frac{1}{2}}$ at $|x| = c$. Again C_{δ}^E provides improvement that is close to optimal for large values of p . Figures 3 and 4 show the ratio of the radius of C_{δ}^E to that of C_x^O . C_{δ}^E also compares favorably to the more complicated procedure of Berger (1980). Table 1 gives some comparisons with this procedure. It can be seen that Berger's procedure has smaller volume for $|x|$ near zero, but is similar to C_{δ}^E in volume for moderate and large $|x|$.

As mentioned before, if a prior guess, other than 0, is known, C_{δ}^E should be modified to be centered at this prior guess. If the prior guess is θ_0 , then the preferred confidence set is

$$\{\theta : |\theta - \delta(x^*)| \leq v(|x^*|)\} ,$$

where $x^* = x - \theta_0$ and $\delta(x^*) = \theta_0 + \{1 - [(p-2)/|x^*|^2]\}x^*$. This procedure centers the region of greatest improvement around θ_0 .

5. COMMENTS AND GENERALIZATIONS

It is, no doubt, possible to improve upon C_{δ}^E in the sense of (1.2); however, at this point we view this as an extremely difficult task. The choice of both the point estimator, $\delta(x)$, and the radius function, $v(|x|)$, is mostly based on intuition. Indeed, aside from the few, relatively minor, restrictions on $v(|x|)$ derived here, very little is known about the forms of $v(|x|)$ which would lead to reasonable confidence sets. One obvious way to proceed, which would be somewhat analogous to the development for the case of point estimation, is through a Bayesian argument. By this we mean finding a prior $g(\theta)$ (possibly variations of those considered in Berger, 1980) which yields a posterior distribution $\pi(\theta|x)$ such that the set $\{\theta : \pi(\theta|x) \geq k\}$ is an admissible, minimax confidence set.

The empirical Bayes derivation used in Section 4 is somewhat nonstandard in that it employs a loss function. For those who are uncomfortable with a loss function approach to set estimation, we remark that our use of a loss function is a means to an end; the ultimate evaluations of C_{δ}^E are done for volume and coverage probability separately and are independent of the choice of loss function. An alternate derivation, without the use of a loss function, might be based on the Bayes Highest Posterior Density (HPD) region which, in the notation of Section 4, is given by

$$C_x^H = \{\theta : |\theta - \delta^B(x)|^2 \leq c^2 \tau^2 / (\tau^2 + 1)\} \quad . \quad (5.1)$$

An empirical Bayes version of this region is given by

$$C_x^{EH} = \{\theta : |\theta - \delta^+(x)|^2 \leq v_{EH}^2(|x|)\} \quad , \quad (5.2)$$

where $\delta^+(x)$ is the positive part James-Stein estimator, and

$$\begin{aligned} v_{EH}^2(|x|) &= c^2 \left(1 - \frac{p-2}{c^2}\right) && \text{if } |x| \leq c \\ &= c^2 \left(1 - \frac{p-2}{|x|^2}\right) && \text{if } |x| > c \quad . \end{aligned} \quad (5.3)$$

However, as mentioned in Section 4, numerical evidence has shown that C_{δ}^{EH} fails to dominate C_x^O in coverage probability for a range of middle values of $|\theta|$. Hence, C_{δ}^{EH} is not a (frequentist) $1 - \alpha$ confidence set.

This also points out the necessity of the logarithm term in $v_E(|x|)$. Without it, the radius is too small to guarantee dominance in coverage probability. Thus, although our evaluations of the set C_{δ}^E are independent of the choice of loss function, our interpretation of the procedure is somewhat dependent on it: we are using an empirical Bayes rule based on a decision-

theoretic Bayes set. Whether or not the inclusion of the logarithm term hinders the statistical interpretation is an individual choice. We are comfortable with it, but others may not be. However, the implications are clear in that some extra term (over that in $v_{EH}(|x|)$) must be present in order to achieve dominance in coverage probability.

Perhaps one of the most important practical uses of a procedure such as C_{δ}^E is in the construction of simultaneous confidence intervals for the individual components of θ (or linear combinations of the components). If we are interested in the estimation of linear combinations $b'\theta$, where b is any $p \times 1$ vector, we can construct simultaneous intervals for this linear combinations using C_{δ}^E . In fact, if C_{δ}^E is a $1 - \alpha$ confidence set, then with probability $1 - \alpha$,

$$b'\delta^+(x) - (b'b)^{\frac{1}{2}}v_E(|x|) \leq b'\theta \leq b'\delta^+(x) + (b'b)^{\frac{1}{2}}v_E(|x|) \quad (5.4)$$

for all $p \times 1$ vectors b . The intervals in (5.4) provide a uniform improvement over the usual (Scheffé-type) simultaneous intervals, and are useful, for example, in post-ANOVA data analysis. (The fact that (5.4) provides a set of simultaneous $1 - \alpha$ intervals follows directly from the fact that $\max_b \{ (b'[\theta - \delta^+(x)])^2 / b'b \} = |\theta - \delta^+(x)|^2$.) One can similarly construct simultaneous intervals on the individual components of θ , and assert with probability $1 - \alpha$,

$$\delta_i^+(x) - v_E(|x|) \leq \theta_i \leq \delta_i^+(x) + v_E(x), \quad i = 1, \dots, p \quad (5.5)$$

Although the intervals in (5.4) and (5.5) are smaller than the usual Scheffé intervals, they are wider than the usual one-dimensional $1 - \alpha$ intervals. (Of course, componentwise $1 - \alpha$ intervals will not yield a set of simultaneous $1 - \alpha$ intervals.) Morris (1983) treats the problem of one-dimensional interval estimation, and obtains empirical Bayes intervals which can yield substantial length reductions over the usual intervals. His ultimate probability evaluations,

however, are different from those used here. Morris calculates coverage probabilities by integrating over the joint distribution of x and θ , and requires this probability to exceed $1 - \alpha$ for a class of prior distributions on θ . However, his treatment of the problem and the one given here are ultimately not comparable. Whether an experimenter is interested in individual confidence statement or simultaneous confidence statements (as given in (5.4)) defines two distinct statistical problems.

For the case $x \sim N(\theta, \sigma^2 \mathbf{I})$, σ^2 unknown, very little is known about the behavior of confidence sets other than the usual one. Some simulations were done by Berger (1980) which suggest that his procedure will retain good performance, in terms of volume and coverage probability, if σ^2 is replaced by an independent estimate. For the set C_{δ}^E considered here, a natural modification is

$$C_{x,s}^E = \{ \theta : |\theta - \delta^+(x,s)| \leq sv_E(|x|/s) \} \quad (5.6)$$

where $s^2 \sim (\sigma^2/\nu)\chi_{\nu}^2$ independent of x ,

$$\delta^+(x,s) = \left(1 - \frac{as^2}{|x|^2}\right)^+ x, \quad a = \frac{\nu}{\nu+2}(p-2), \quad (5.7)$$

and

$$\begin{aligned} v_E^2(|x|/s) &= \left(1 - \frac{a}{pF_{\alpha,p,\nu}}\right) \left[pF_{\alpha,p,\nu} - \text{plog}\left(1 - \frac{a}{pF_{\alpha,p,\nu}}\right) \right] \text{ if } \frac{|x|^2}{s^2} \leq pF_{\alpha,p,\nu}, \\ &= \left(1 - \frac{as^2}{|x|^2}\right) \left[pF_{\alpha,p,\nu} - \text{plog}\left(1 - \frac{as^2}{|x|^2}\right) \right] \text{ if } \frac{|x|^2}{s^2} > pF_{\alpha,p,\nu}, \end{aligned} \quad (5.8)$$

where $F_{\alpha,p,\nu}$ is the upper α -level cutoff point from an F distribution with p and ν degrees of freedom. (Recall that the usual $1 - \alpha$ confidence set for θ is $C_{x,s}^0 = \{ \theta : |\theta - x|^2 \leq s^2 pF_{\alpha,p,\nu} \}$.)

It is easy to check that $v_F^2(|x|/s) \leq pF_{\alpha,p,v}$ for all $|x|/s$, so $C_{x,s}^E$ provides a uniform volume reduction over $C_{x,s}^O$. The reduction in volume is minor for small values of v ($v=2,5$) but becomes more substantial for moderate values ($v=10,20$). Also, using the results of Section 3, the exact formula for the coverage probability of $C_{x,s}^E$ can be derived. (The formula is similar to that given in (3.7), except there is now an integral over s^2 .) These probabilities were calculated for selected values of p and v , and the results are presented in Table 2. With a few minor exceptions (at $p=3$), $C_{x,s}^E$ is a $1-\alpha$ confidence set for θ , with coverage probabilities increasing as v increases. We also note that, unlike the case of known variance, the coverage probability of $C_{x,s}^E$ is a continuous function; the integration over s^2 smooths things out.

If $x \sim N(\theta, \sigma^2 \Sigma)$, where σ^2 is unknown and Σ is a known positive definite matrix, the use of the transformation $y = \Sigma^{-\frac{1}{2}} x$ reduces the covariance matrix to $\sigma^2 I$, showing that our results apply to this more general case. For example, in the usual one-way analysis of variance, x = vector of cell means, $\Sigma = \text{diag}(n_1^{-1}, n_2^{-1}, \dots, n_p^{-1})$, where n_i = number of observations in cell i , and the sets and intervals given here apply. The case of estimating the coefficients in a linear regression can be handled in a similar manner.

6. REMARKS

Only recently has significant progress been made in the problem of improving upon the usual confidence regions for a multivariate normal mean. The problem is one of great statistical importance, for the lack of such regions has made it impossible to provide tests and interval estimates associated with improved point estimators, and has limited the applicability of these estimators. Part of the reason why there has been recent progress is due to the increased role of the computer: many results in this area have yet to be established

analytically, and claims must be supported with extensive numerical evidence. Another reason for the recent progress is perhaps our improved understanding of phenomena like the Stein effect. We have observed a synthesis of the approach of the Bayesian and the frequentist, perhaps culminating in an approach like that of an empirical Bayesian. Through this synthesis a better understanding of Stein-type and other related procedures is gained, and new methods are found for constructing estimators. This is important, for it is not enough to exhibit a better procedure, there must be a reasonably sound statistical justification for it: we do not want to merely take advantage of mathematical anomalies.

While we believe that the results presented here represent a step forward in the confidence set problem, there are many limitations to our results. The most obvious one is that analytical dominance results were not obtained for C_{δ}^E . Although analytic dominance results have been obtained for fixed-radius confidence sets, progress has been slower on the statistically more important variable radius sets. The solution to this problem will not only serve to make the theory more complete, but may also point out new and better radius functions, perhaps even different shape for the confidence sets themselves.

The basic question of the proper shape of a confidence set is one that has still not been settled. For the equal variance case all the sets presented here are spheres. There are arguments (given in Berger, 1980, and originating with Stein, 1962) that show that a sphere may not be the optimal shape for these sets. The statistical interpretation of these non-spherical sets is difficult, and, at this time, might be difficult for practitioners to accept. But such sets should not be dismissed, and with better understanding (perhaps through empirical Bayes considerations) these sets could become meaningful alternatives.

In terms of practical applications, the results presented here extend to cases where the variances are known up to a scale factor (the common assumptions in ANOVA and regression), but fail to extend to the case of unequal, unknown variance. This is, again, not due to the lack of importance of the problem, but rather to its enormous difficulty. It is safe to say that, until recently, it was not even known how to properly construct improved point estimators in this case. The work of Morris (1983) is a major step forward in this problem, providing applicable point and interval estimators for the unequal variance case. But more work is needed, particularly in the case of set estimation, on this important practical problem.

APPENDIX: PROOF OF THEOREM 4.2

Theorem A1: Let C be a confidence set defined by

$$C_x = \{\theta: |\theta - \gamma(|x|)x| \leq v(|x|) \},$$

where $\gamma(|x|) = [1 - (a/|x|^2)]^+$. Let $\tau(t, |\theta|) = [|\theta| - t\gamma(t)]^2$, and let $t^*(|\theta|)$ be the positive root of $\tau(t, |\theta|) = 0$. If $|\theta| = 0$, define $t^*(|\theta|) = a$. Suppose $v^2(|x|)$ is a nondecreasing function of the form:

$$\begin{aligned} v^2(|x|) &= v^2(b) \text{ for } |x| \leq b \\ &= \text{concave function for } |x| > b \end{aligned}$$

for some constant $b \geq 0$. Then C_x is connected if either

i) $b \leq t^*(|\theta|)$

or

ii) $b \geq t^*(|\theta|)$ and $[b\gamma(b)]^2 \leq v^2(b)$.

Proof: Define, for each $|\theta|$,

$$\begin{aligned} S_1 &= \{t: \tau(t, |\theta|) \leq v^2(t) \text{ and } t \leq t^*(|\theta|)\} \\ S_2 &= \{t: \tau(t, |\theta|) \leq v^2(t) \text{ and } t \geq t^*(|\theta|)\} \end{aligned}$$

Note that $t^*(|\theta|) \in S_1 \cap S_2$, so if both S_1 and S_2 are connected, it follows that $S = S_1 \cup S_2$ is connected and, Theorem A1 then follows from Theorem 3.1.

It is straightforward to verify that $\tau(t, |\theta|)$ is convex for $t \geq a$, and, for $t \geq a$, $t^*(|\theta|)$ is the unique root of $\tau(t, |\theta|) = 0$. Therefore, for $t \leq t^*(|\theta|)$, the function $\tau(t, |\theta|) - v^2(t)$ is nonincreasing, so S_1 is an interval (and hence is connected).

To see that S_2 is also an interval, consider first the case $b \leq t^*(|\theta|)$. If this is so, then $\tau(t, |\theta|) - v^2(t)$ is convex for $t \geq t^*(|\theta|)$ and the theorem

follows. If $b \geq t^*(|\theta|)$, then let $t_1 \leq t_2$ be any two points in S_2 . We now show that the interval $[t_1, t_2]$ is in S_2 . If $b \notin [t_1, t_2]$ then either $b < t_1$ or $b > t_2$. If $b < t_1$, then $\tau(t, |\theta|) - v^2(t)$ is convex for $t \geq t_1$, and hence $[t_1, t_2] \subset S_2$. If $b > t_2$, then for any t' satisfying $t_1 \leq t' \leq t_2$ we have

$$\begin{aligned} \tau(t', |\theta|) - v^2(t') &\leq \tau(b, |\theta|) - v^2(b) \\ &\leq [b\gamma(b)]^2 - v^2(b) \leq 0, \end{aligned}$$

since $b\gamma(b) \geq |\theta|$. Hence $[t_1, t_2] \subset S_2$.

Finally, if $b \in [t_1, t_2]$, write

$$[t_1, t_2] = [t_1, b] \cup [b, t_2].$$

An argument similar to those above shows that each of these intervals is in S_2 and hence $[t_1, t_2] \subset S_2$. Thus S_2 is connected and the theorem is proved. ||

Corollary A1: If $c^2 > p$, then for $p \geq 3$ the confidence set C_δ^E defined by (4.5), (4.6), and (4.7) has, for all θ , a connected θ section.

Proof: The corollary follows immediately from Theorem A1 if we verify that $[c\gamma(c)]^2 \leq v_E^2(c)$ and the concavity of $v_E^2(t)$ for $t > c$. The inequality is straightforward to verify, and the concavity follows from noting that, for $t > c$,

$$\frac{d^2}{dt^2} v_E^2(t) = \frac{-(p-2)}{t} \left[4p[t^2 - (p-2)]^{-1} + 6 \left[c^2 - p - p \log \left(1 - \frac{p-2}{t^2} \right) \right] \right],$$

which is negative if $c^2 > p$. ||

Table 1. Comparison of the Volume of the Empirical Bayes Confidence Set C_{δ}^E with those of C^* (Berger's Procedure)

$ x $	Ratio of the effective radii of C_{δ}^E to C^*									
	0	1	2	4	6	8	10	20	50	
p										
6	1.080	1.072	1.049	1.039	1.019	1.011	1.007	1.001	1.001	
12	1.075	1.072	1.063	1.019	1.075	1.043	1.027	1.006	1.001	

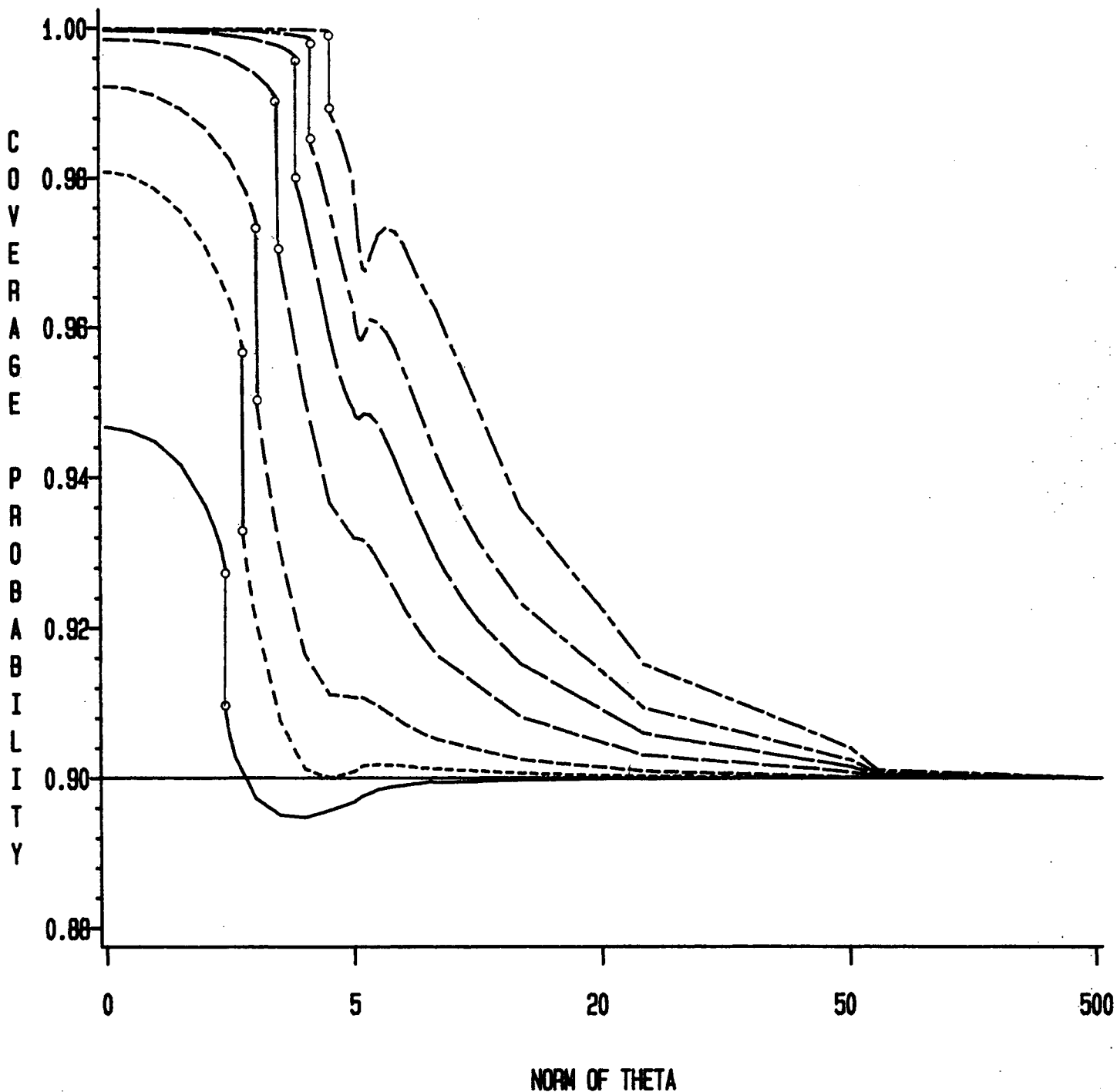
Table 2. Coverage Probabilities for the Set $C_{x,s}^E$ with Squared Radius

$$\sqrt{E}^2(|x|/s) \leq pF_{.1,p,v}, \text{ where } F_{.1,p,v} \text{ Satisfies } P(F_{p,v} \leq F_{.1,p,v}) = .9$$

θ	p = 3					p = 5				
	v					v				
	2	5	10	20	30	2	5	10	20	30
0	.902	.912	.924	.933	.937	.903	.922	.944	.961	.967
1	.902	.911	.923	.932	.936	.903	.920	.941	.958	.964
2	.901	.906	.914	.921	.926	.902	.915	.932	.948	.955
4	.900	.899	.898	.897	.896	.901	.903	.905	.905	.905
6	.900	.899	.899	.899	.898	.900	.901	.901	.902	.902
8	.900	.900	.899	.899	.899	.900	.900	.901	.901	.901
10	.900	.900	.900	.900	.900	.900	.900	.901	.901	.901
15	.900	.900	.900	.900	.900	.900	.900	.900	.900	.900
20	.900	.900	.900	.900	.900	.900	.900	.900	.900	.900

θ	p = 7					p = 9				
	v					v				
	2	5	10	20	30	2	5	10	20	30
0	.904	.927	.953	.973	.979	.904	.930	.959	.979	.986
1	.903	.925	.950	.970	.977	.903	.928	.956	.977	.984
2	.902	.920	.942	.962	.970	.902	.923	.948	.970	.978
4	.901	.908	.916	.920	.920	.901	.911	.925	.935	.938
6	.900	.903	.906	.908	.908	.901	.905	.911	.915	.916
8	.900	.902	.904	.905	.906	.900	.903	.907	.910	.911
10	.900	.901	.902	.904	.904	.900	.902	.904	.907	.908
15	.900	.900	.901	.902	.902	.900	.901	.902	.903	.904
20	.900	.900	.901	.901	.901	.900	.900	.901	.902	.902

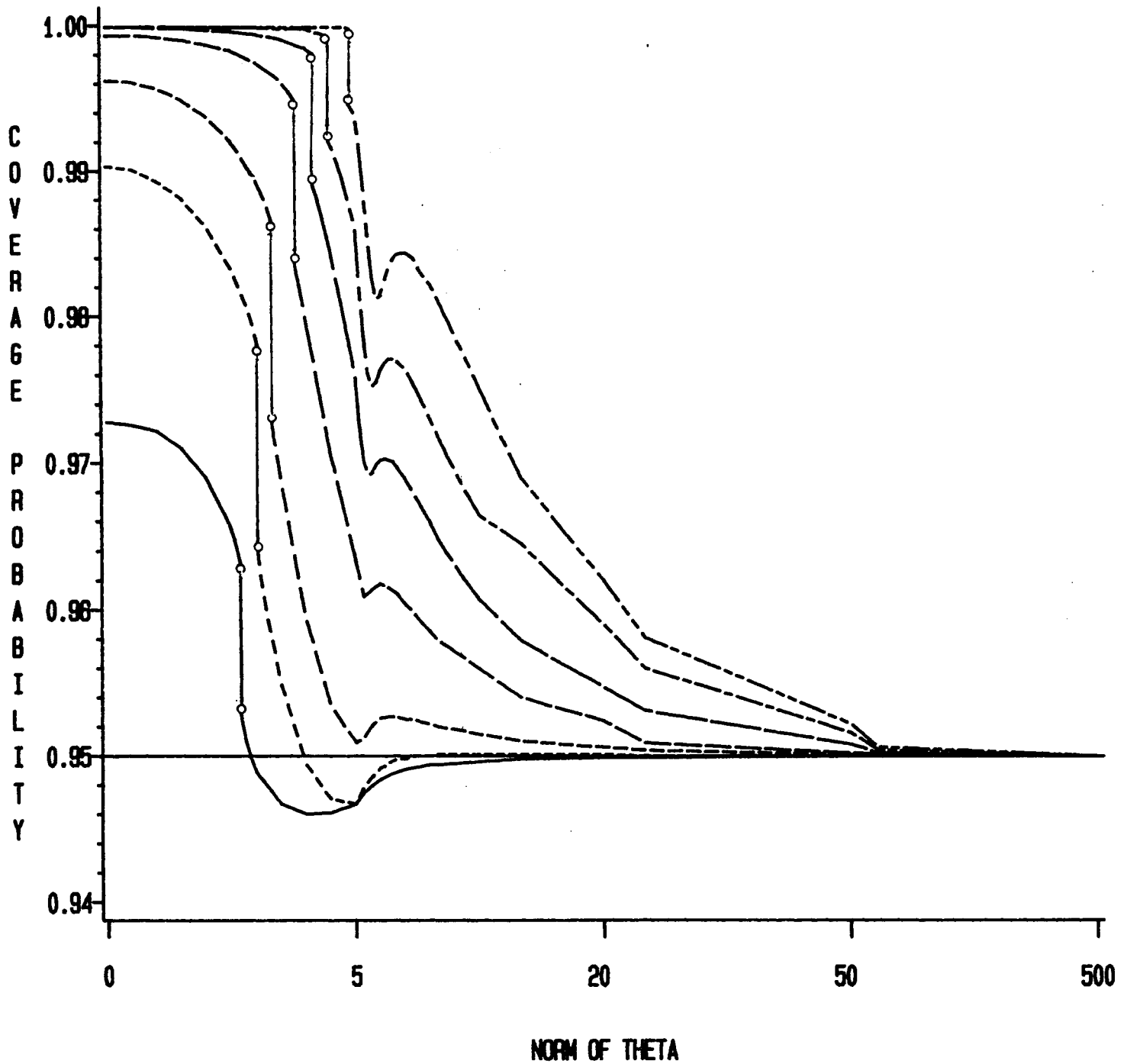
COVERAGE PROBABILITIES FOR EMPIRICAL BAYES CONFIDENCE SETS (ALPHA=.10)



LEGEND: P ——— 3 - - - - 5 - - - - 7 - - - - 11 - - - - 15
 - - - - 19 - - - - 25

Figure 1. Coverage probabilities for the set C_0^E with radius $v_E(|x|) \leq c$, where c satisfies $P(\chi_p^2 \leq c^2) = .9$. The circles mark the jump discontinuities, which occur at the points where $|\theta| = v_E(0)$.

COVERAGE PROBABILITIES FOR EMPIRICAL BAYES
CONFIDENCE SETS (ALPHA=.05)



LEGEND: P ——— 3 - - - - 5 - - - - 7 ——— 11 ——— 15
 - - - - 19 - - - - 25

Figure 2. Coverage probabilities for the set C_{δ}^E with radius $v_E(|x|) \leq c$, where c satisfies $P(\chi_p^2 \leq c^2) = .95$. The circles mark the jump discontinuities, which occur at the points where $|\theta| = v_E(0)$.

RATIO OF THE RADIUS OF THE EMPIRICAL BAYES SET TO THE USUAL SET (ALPHA=.10)

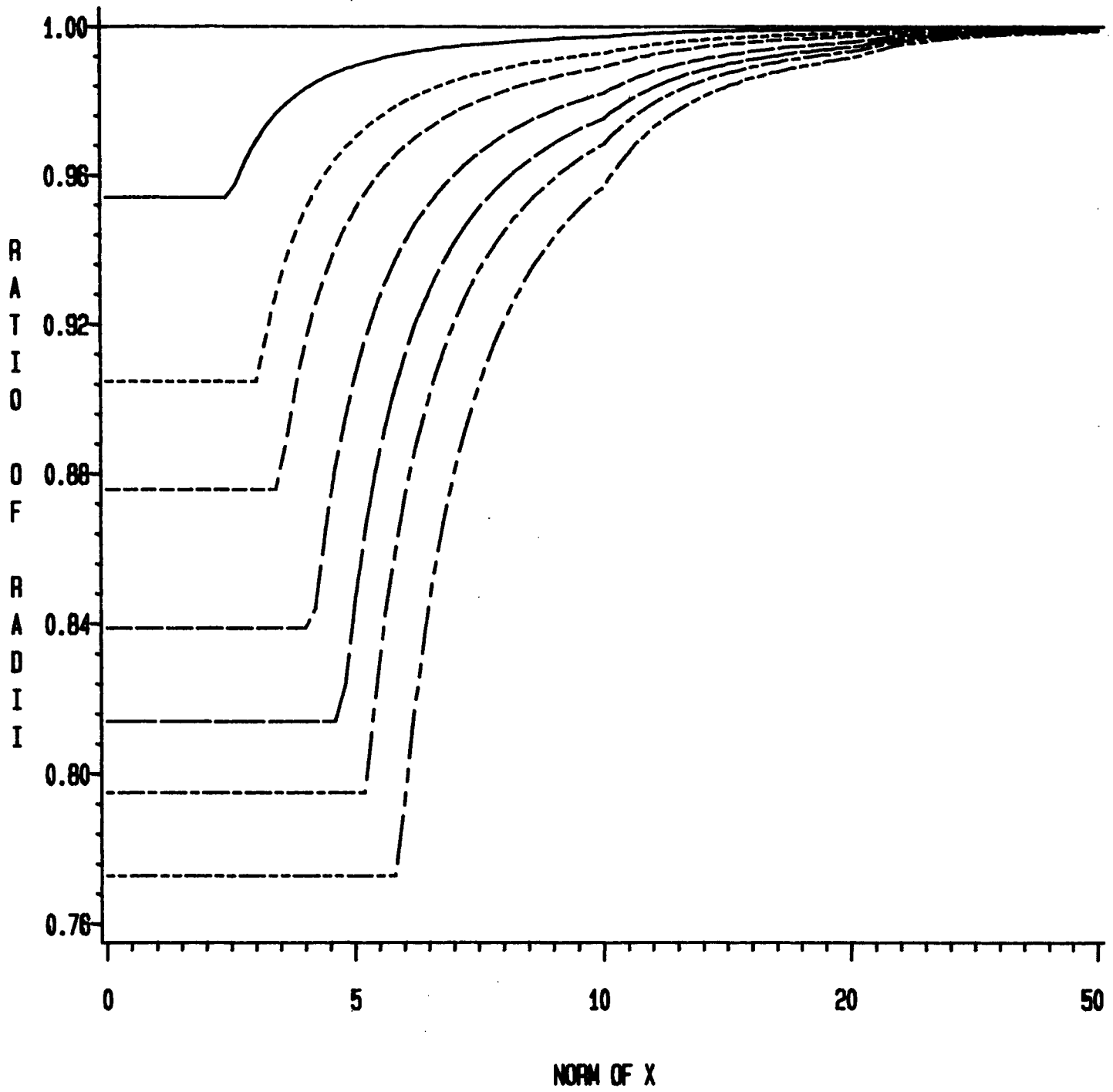
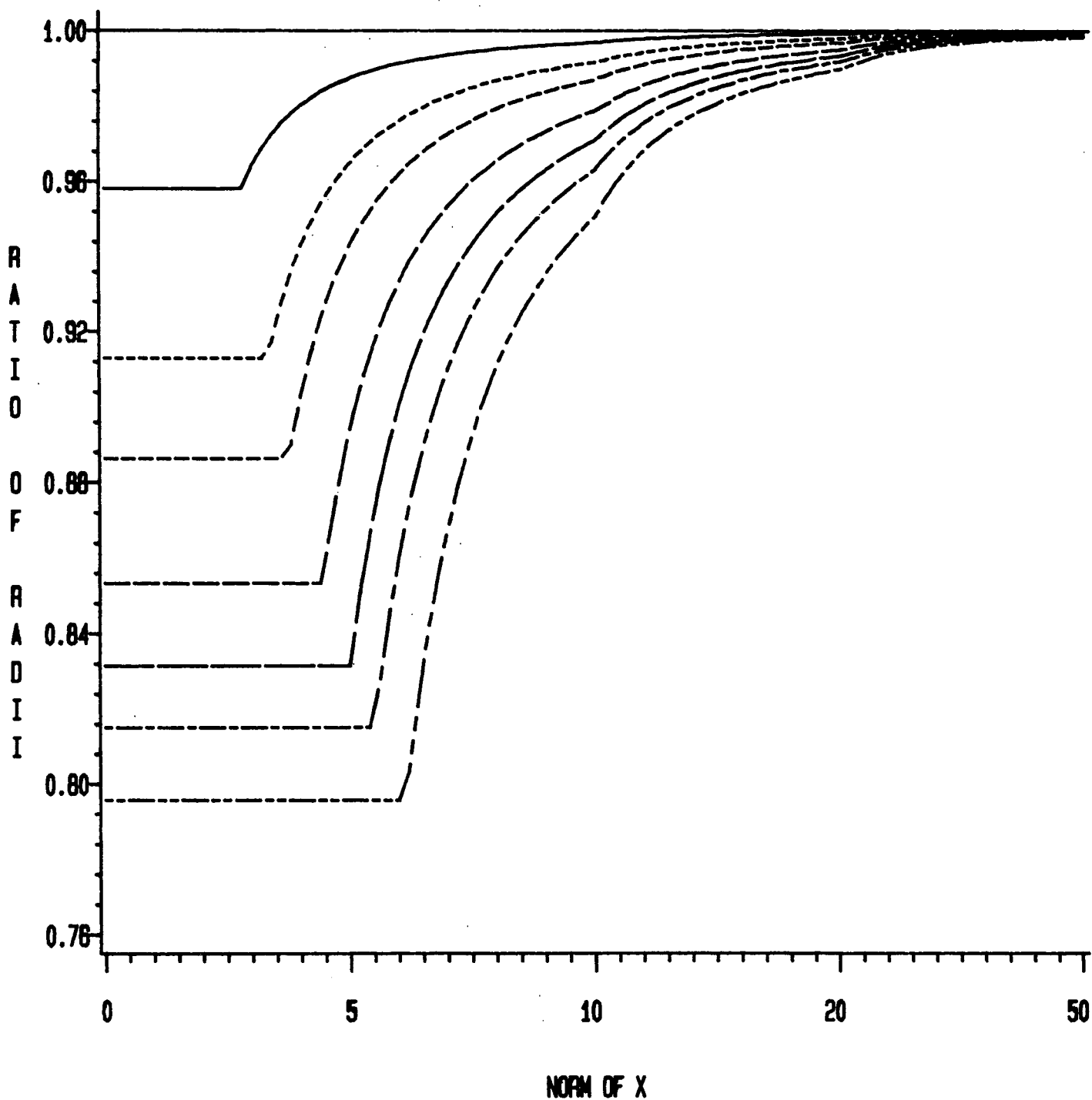


Figure 3. Ratio of the radii $v_E(|x|)/c$ where c satisfies $P(\chi_p^2 \leq c^2) = .9$.

RATIO OF THE RADIUS OF THE EMPIRICAL BAYES SET TO THE USUAL SET (ALPHA=.05)



LEGEND: P ——— 3 - - - - 5 - - - - 7 - - - - 11
 - - - - 15 - - - - 19 - - - - 25

Figure 4. Ratio of the radii $v_E(|x|)/c$ where c satisfies

$$P(\chi_p^2 \leq c^2) = .95.$$

Captions of Tables and Figures:

Tables:

1. Comparison of the Volume of the Empirical Bayes Confidence Set C_{δ}^E with those of C^* (Berger's Procedure)
2. Coverage Probabilities for the Set $C_{x,s}^E$ with squared radius $v_E^2(|x|/s) \leq pF_{.1,p,\nu}$, where $F_{.1,p,\nu}$ satisfies $P(F_{p,\nu} \leq F_{.1,p,\nu}) = .9$

Figures:

1. Coverage probabilities for the set C_{δ}^E with radius $v_E(|x|) \leq c$, where c satisfies $P(\chi_p^2 \leq c^2) = .9$. The circles mark the jump discontinuities, which occur at the points where $|\theta| = v_E(0)$.
2. Coverage probabilities for the set C_{δ}^E with radius $v_E(|x|) \leq c$, where c satisfies $P(\chi_p^2 \leq c^2) = .95$. The circles mark the jump discontinuities, which occur at the points where $|\theta| = v_E(0)$.
3. Ratio of the radii $v_E(|x|)/c$, where c satisfies $P(\chi_p^2 \leq c^2) = .9$.
4. Ratio of the radii $v_E(|x|)/c$, where c satisfies $P(\chi_p^2 \leq c^2) = .95$.

REFERENCES

- Berger, James O. (1980), "A Robust Generalized Bayes Estimate and Confidence Region for a Multivariate Normal Mean," Annals of Statistics, 8, 716-761.
- Birnbaum, Allan (1955), "Characterizations of Complete Classes of Tests of Some Multiparametric Hypotheses, with Applications to Likelihood Ratio Tests," Annals of Mathematical Statistics, 26, 21-36.
- Brown, L. D. (1966), "On the Admissibility of Invariant Estimators of One or More Location Parameters," Annals of Mathematical Statistics, 37, 1087-1136.
- Hwang, Jiunn T., and Casella, George (1982), "Minimax Confidence Sets for the Mean of a Multivariate Normal Distribution," Annals of Statistics, 10, 868-881.
- Joshi, V. M. (1967), "Inadmissibility of the Usual Confidence Sets for the Mean of a Multivariate Normal Population," Annals of Mathematical Statistics, 38, 1868-1875.
- Joshi, V. M. (1969), "Admissibility of the Usual Confidence Sets for the Mean of a Univariate or Bivariate Normal Population," Annals of Mathematical Statistics, 40, 1042-1067.
- Morris, C. N. (1983), "Parametric Empirical Bayes Confidence Inference: Theory and Applications," Journal of the American Statistical Association, 78.
- Stein, C. (1962), "Confidence Sets for the Mean of a Multivariate Normal Distribution," Journal of the Royal Statistical Society, Series B, 24, 265-296.