

BU-779-M  
8/82

ZERO-RADIUS CONFIDENCE PROCEDURES

GEORGE CASELLA<sup>1</sup> AND JIUNN TZON HWANG<sup>2</sup>

Cornell University

SUMMARY

A zero-radius confidence procedure is a procedure which, for certain realizations, may produce a confidence set with zero radius. Such a procedure is constructed, and is shown to uniformly dominate the usual confidence procedure in both volume and coverage probability. The question of whether a zero-radius confidence set can have a meaningful interpretation is explored. It is argued that, both to a frequentist and an empirical Bayesian, such a set can have reasonable interpretations. Such interpretations depend quite heavily on a judicious choice of prior input.

---

<sup>1</sup> Research supported by National Science Foundation Grant No. MCS81-02541.

<sup>2</sup> Research supported by National Science Foundation Grant No. MCS80-03568.

AMS 1980 subject classifications. Primary 62F25; secondary 62C20.

Key words and phrases. Multivariate normal mean, Stein estimation, empirical Bayes, minimax set estimation.

1. Introduction. In recent years, the theory of set estimation of a multivariate normal mean has seen many advances. In particular, it seems possible that by taking advantage of the Stein effect, confidence sets can be constructed which uniformly improve upon the usual set in both volume and coverage probability. These improved sets have nonconstant coverage probability which can be quite close to 1 in a small region of the parameter space. By trading volume for coverage probability, such improved sets can have greatly reduced volume (in some region) while maintaining a coverage probability above some nominal, specified level. Such sets have been constructed by Faith (1976), Berger (1980), Morris (1982), and Casella and Hwang (1982a). All these sets, however, share the property that the radii are strictly positive. The concern of this paper is the construction and interpretation of a  $1 - \alpha$  confidence procedure which attains the ultimate volume reduction: a zero-radius set.

We define a zero volume (or radius) confidence procedure in the following way.

DEFINITION 1.1. A confidence procedure,  $C^Z$ , is a  $1 - \alpha$  zero-radius confidence procedure if the coverage probability of  $C^Z$  is at least  $1 - \alpha$  for all parameter values and, for at least one observation,  $C^Z$  has radius equal to zero.

The fact that  $C^Z$  exists is not very surprising; indeed, the fact of the existence of  $C^Z$  is not our major concern. Rather, we are concerned with the interpretation of the procedure  $C^Z$ . In particular, if an observation produces a zero (or near zero) radius confidence set, how can an experimenter make sense out of a frequentist coverage probability?

In practice, if one uses the zero-radius confidence procedure of Section 2, the realized confidence set will almost surely have a positive radius. Thus, one need not really be concerned with the possibility of obtaining a zero-radius set. However, with procedures such as the one described in Section 2, it is quite possible to obtain a  $1 - \alpha$  confidence set with radius "too small to be believed". That is, with respect to a known error variance, the radius of the confidence set seems unreasonably small, and one might tend to not trust any inferences that are based on it. This is the situation to which we refer when we speak of "interpreting a zero-radius confidence set".

The zero-radius confidence set is, perhaps, one of the ultimate demonstrations of the power of the Stein effect. The goal of this paper is to demonstrate that a zero-radius confidence set is not just a mathematical oddity, but rather a useful statistical tool in its own right.

In Section 2 we derive, through the use of a modified empirical Bayes argument, a zero-radius confidence procedure,  $C^Z$ . It is shown that both the volume and expected volume of  $C^Z$  are uniformly smaller than that of the usual confidence procedure. The exact formula for the coverage probability is also derived. Although dominance over the usual confidence procedure is not demonstrated analytically, strong numerical evidence is presented which shows that  $C^Z$  is, in fact, a  $1 - \alpha$  confidence procedure and, moreover, has uniformly higher coverage probability than the usual confidence set.

In Section 3 we consider various interpretations of a zero-radius confidence set. We find that the set  $C^Z$  can have meaningful interpretations to both a frequentist and an empirical Bayesian. Section 4 contains some conclusions and comments.

2. A Zero-Radius Confidence Procedure. In this section we present a confidence procedure which, for a particular observation, may have zero radius. This procedure has uniformly smaller volume than the usual procedure. Also, strong numerical evidence is presented which shows that this procedure uniformly dominates the usual one in coverage probability. The procedure is derived by using a modification of an empirical Bayes argument, similar to that used in Casella and Hwang (1982a).

For the moment, assume the Bayesian model

$$(2.1) \quad \begin{aligned} X|\theta, \sigma^2 &\sim N(\theta, \sigma^2 I) \\ \theta|\mu, \tau^2 &\sim N(\mu, \tau^2 I) \quad , \end{aligned}$$

where  $\mu$ ,  $\sigma^2$  and  $\tau^2$  are all known. An easy calculation shows that the posterior distribution of  $\theta$ ,  $\pi(\theta|X)$ , is given by

$$(2.2) \quad \pi(\theta|X) = \left(\frac{\sigma^2 + \tau^2}{2\pi\sigma^2\tau^2}\right)^{p/2} \exp\left\{-\frac{1}{2}\left(\frac{\sigma^2 + \tau^2}{\sigma^2\tau^2}\right)|\theta - \delta^B(X)|^2\right\} \quad ,$$

where  $\delta^B(X) = E(\theta|X)$  is given by

$$(2.3) \quad \delta^B(X) = (\sigma^2 + \tau^2)^{-1}(\sigma^2\mu + \tau^2X) \quad .$$

A Bayes credible set for  $\theta$ , a region of high posterior density, is given by

$$(2.4) \quad C^B = \{\theta : \pi(\theta|X) \geq k\}$$

for some constant  $k$ .  $C^B$  is also the Bayes rule against the loss function

$$(2.5) \quad L(\theta, C) = k \text{ Volume}(C) - I_\theta(C) \quad ,$$

where  $I_\theta(C) = 1$  if  $\theta \in C$  and zero otherwise.

To avoid trivial cases,  $k$  is restricted to lie in the interval  $0 < k \leq (2\pi\sigma^2)^{-p/2}$ . This is a necessary restriction, because if  $k=0$  the volume

component does not enter into the loss, and the Bayes credible set would be all of  $\mathbb{R}^p$ . Also, if  $k > (2\pi\sigma^2)^{-p/2}$ , it is shown in Casella and Hwang (1982b) that the empty set (or any set of Lebesgue measure zero) is a proper Bayes minimax rule. Thus, since  $0 < k \leq (2\pi\sigma^2)^{-p/2}$ , without loss of generality we can take  $k$  to be of the form

$$(2.6) \quad k = (2\pi\sigma^2)^{-p/2} e^{-c^2/2},$$

for  $0 \leq c^2 < \infty$ . This is a particularly convenient form, for then the usual confidence set

$$(2.7) \quad C^0 = \{\theta : |\theta - X|^2 \leq c^2\sigma^2\}$$

is minimax against the loss (2.5).

For  $k$  of the form (2.6), the Bayes rule against  $L(\theta, C)$  can be written as

$$(2.8) \quad C^B = \{\theta : |\theta - \delta^B(X)|^2 \leq v^B\},$$

where

$$(2.9) \quad v^B = \sigma^2 \left( \frac{\tau^2}{\tau^2 + \sigma^2} \right) \left[ c^2 - p \log \left( \frac{\tau^2}{\tau^2 + \sigma^2} \right) \right].$$

If one is not a Bayesian, or if for some reason,  $\tau^2$  is unknown, the set  $C^B$  is of no use. One way to proceed in such a case is through an empirical Bayesian methodology, trying to replace  $\tau^2$  by an estimate. One technique that has been employed is to notice that, under the model (2.1), the marginal distribution of  $X$  is  $N[\mu, (\sigma^2 + \tau^2)I]$ . It then follows that  $|X - \mu|^2 \sim (\sigma^2 + \tau^2) \chi_p^2$ , and

$$(2.10) \quad E \left[ 1 - \frac{(p-2)\sigma^2}{|X - \mu|^2} \right] = \frac{\tau^2}{\tau^2 + \sigma^2}.$$

The empirical Bayesian would consider replacing  $\tau^2/(\tau^2 + \sigma^2)$  in  $v^B$  by its unbiased estimate. Such a procedure was used by Casella and Hwang (1982a),

where  $v^B$  was replaced by

$$(2.11) \quad v^E = \sigma^2 T^E (c^2 - p \log T^E) \quad \text{and} \quad T^E = 1 - \frac{p-2}{\max\{|X-\mu|^2/\sigma^2, c^2\}}$$

The estimate  $T^E$  was truncated to avoid having the argument of the logarithm negative. The resulting confidence procedure performs quite well; however, it is not a zero-radius procedure.

To obtain a zero-radius procedure, we consider a different modification, and replace  $\tau^2/(\tau^2 + \sigma^2)$  in  $v^B$  by

$$(2.12) \quad T = 1 - \frac{(p-2)\sigma^2}{(p-2)\sigma^2 + |X-\mu|^2}$$

While (2.12) does not provide an unbiased estimate of  $\tau^2/(\tau^2 + \sigma^2)$ , we expect it to provide a reasonable estimate. Also, (2.12) is nonnegative, and solves the logarithm problem in a continuous fashion. Moreover, it produces a zero-radius procedure. Using (2.12), we construct the function

$$(2.13) \quad v^Z = v^Z(|X - \mu|) = \sigma^2 T (c^2 - p \log T)$$

Referring back to the set  $C^B$  of (2.8), we replace the Bayes estimate  $\delta^B(X)$  by

$$(2.14) \quad \delta_{\mu}^+(X) = \mu + \left(1 - \frac{(p-2)\sigma^2}{|X-\mu|^2}\right)^+ (X - \mu),$$

and obtain the zero-radius confidence procedure

$$(2.15) \quad C^Z = \{\theta : |\theta - \delta_{\mu}^+(X)|^2 \leq v^Z\}$$

The function  $v^Z$  is a continuous function of  $|X - \mu|$ , and equals 0 at  $|X - \mu| = 0$ . We now demonstrate that  $C^Z$  is a uniform improvement over  $C^0$ .

We first consider the volume of  $C^Z$ . We have the following theorem:

THEOREM 2.1.  $\text{Volume}(C^Z) < \text{Volume}(C^0)$  for all  $|X - \mu| < \infty$  if and only if  $c^2 \geq p$ .

REMARK. The condition  $c^2 \geq p$  is an extremely minor restriction, and will almost certainly be satisfied in practice. The confidence set  $\{\theta : |\theta - X|^2 \leq p\}$  has coverage probability approximately equal to .55.

PROOF. Note that  $T$  is an increasing function of  $|X - \mu|$ , and  $0 \leq T \leq 1$ . Since  $v^Z = \sigma^2 T(c^2 - p \log T)$ , we have

$$(2.16) \quad \frac{\partial}{\partial T} v^Z = \sigma^2 [c^2 - p(1 + \log T)]$$

If  $c^2 \geq p$  then (2.16) is positive for all  $|X - \mu|$ , and  $v^Z$  is strictly increasing in  $|X - \mu|$  with maximum value equal to  $\sigma^2 c^2$ , attained at  $|X - \mu| = \infty$ . Hence, since both  $C^Z$  and  $C^0$  are spheres, it follows that  $\text{Volume}(C^Z) < \text{Volume}(C^0)$  for all  $|X - \mu| < \infty$ . If  $c^2 < p$ , it is straightforward to check that  $v^Z$  attains a unique maximum at  $T = \exp\{(c^2 - p)/p\}$ . The maximum value of  $v^Z$  is

$$(2.17) \quad v_{\max}^Z = \sigma^2 p \exp\{(c^2 - p)/p\}$$

Now  $v_{\max}^Z > \sigma^2 c^2$  if and only if

$$(2.18) \quad e^{(c^2/p) - 1} > c^2/p,$$

which is always the case if  $c^2 < p$ . ||

The following corollary is an immediate consequence of Theorem 2.1.

COROLLARY 2.1. If  $c^2 > p$ , then  $E_{\theta}[\text{Vol}(C^Z)] < E_{\theta}[\text{Vol}(C^0)]$  for all  $\theta$ .

Therefore, in terms of both volume and expected volume,  $C^Z$  is a uniform improvement over  $C^0$ .

We now turn to the evaluation of the coverage probability of  $C^Z$ ,

$$(2.19) \quad P_{\theta}(\theta \in C^Z) = \int_{\{X: |\theta - \delta_{\mu}^+(X)|^2 \leq v^Z\}} f(X|\theta) dX,$$

where  $f(X|\theta)$  is the density of an  $N(\theta, \sigma^2 I)$ . The technique used is to reduce this  $p$ -dimensional integral to a two-dimensional integral through the transformation  $r = |X - \mu|$ ,  $\cos\beta = \theta'(X - \mu)/|\theta||X - \mu|$ . For odd values of  $p$ , the integral over  $\beta$  can then be evaluated using the binomial formula, and we are left with a one-dimensional integral. Before performing the transformation, however, it is important to check whether the set

$$(2.20) \quad \{X : |\theta - \delta_{\mu}^+(X)|^2 \leq v^Z\}$$

is connected, for this dictates whether the integration with respect to  $r$  will be over an interval. From Theorem 3.1 of Casella and Hwang (1982a), it follows that (2.20) is connected if and only if the one-dimensional set

$$(2.21) \quad \{t : [|\theta| - \gamma(t)t]^2 \leq v^Z\}$$

is an interval, where  $\gamma(t) = [1 - (a\sigma^2/t^2)]^+$ . We have the following lemma:

LEMMA 2.1. The set (2.21) is an interval and, hence, the set (2.20) is connected.

PROOF. The result will be established if we can show that the function

$$h(t^2) = [|\theta| - \gamma(t)t]^2 - v^Z$$

is a convex function of  $t^2$  for  $t^2 > (p-2)\sigma^2$ , since otherwise  $h(t^2)$  is decreasing in  $t^2$ . It is straightforward to verify that the function

$$v^Z = \sigma^2 \left(1 - \frac{(p-2)\sigma^2}{(p-2)\sigma^2 + t^2}\right) \left[ c^2 - p \log \left( \frac{(p-2)\sigma^2}{(p-2)\sigma^2 + t^2} \right) \right]$$

is a concave function of  $t^2$  and, for  $t^2 > (p-2)\sigma^2$ , the function  $[|\theta| - \gamma(t)t]^2$  is a convex function of  $t^2$ . Hence, the convexity of  $h(t^2)$  for  $t^2 > (p-2)\sigma^2$  immediately follows, and the lemma is proved. ||

We note that the lemma is true in greater generality than stated, and holds if  $v^Z$  is replaced by any concave function of  $t^2$  and if  $p-2$  is replaced by any positive constant.

Thus, if we carry out the spherical transformation, we get the following representation for the coverage probability of  $C^Z$ .

THEOREM 2.2. If  $|\theta| > 0$ ,

$$(2.22) \quad P_{\theta}(\theta \in C^Z) = K \int_{r_-}^{r_+} r^{p-1} e^{-\frac{1}{2}(r^2 + |\theta|^2)} \int_{h(r)}^1 (1-u^2)^{\frac{p-3}{2}} e^{r|\theta|u} du dr, \quad ,$$

where  $K^{-1} = \sqrt{\pi} 2^{\frac{p-2}{2}} \Gamma[(p-1)/2]$ ,  $r_+$  and  $r_-$  are the endpoints of the interval (2.21), and

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

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

If  $|\theta| = 0$ , then  $P_{\theta}(\theta \in C^Z) = P(\chi_p^2 \leq r_+)$ .

Note that, if  $p$  is odd, the inner integral in (2.22) can be evaluated using the binomial formula. The coverage probabilities of  $C^Z$  were calculated exactly for a wide range of values of  $p$  and  $|\theta|$ , and are presented in Table 1. The evidence is quite conclusive and shows that, with the exception of the case  $p=3$ ,  $C^Z$  is a  $1-\alpha$  confidence set. The failure of  $C^Z$  for  $p=3$  is so slight that it can be ignored. However, it can also be remedied by replacing  $p-2$  by  $(p-2)/2$  in  $v^Z$ . For this new value, the numerical evidence shows that  $C^Z$  is a  $1-\alpha$  confidence set even for  $p=3$ .

Thus, it has been demonstrated that  $C^Z$  is a uniform improvement, both in volume and coverage probability over the usual confidence set,  $C^0$ . According to frequentist criteria, then,  $C^Z$  should always be preferred over  $C^0$ .

### 3. Interpreting a Zero-Radius Confidence Set

3.1. Frequentist considerations. In this section we deal with the problem of finding a reasonable frequentist interpretation of a zero-radius confidence set. The problem arises when one is forced to believe (or have confidence in) a set that seems to be too small, based on the value of the known variance. Before discussing the interpretations of a near-zero radius confidence set, we thought it reasonable to inquire as to just how rare an event we are dealing with. For  $C^Z$  of (2.15), the most favorable case (when we can expect the greatest volume reduction) is when  $\mu = \theta$ . For various values of  $\gamma$ , the probability  $P[\text{radius}(C^Z) \leq \gamma \text{ radius}(C^O) | \mu = \theta]$  was calculated, and the results presented in Table 2. The results show that although it is reasonable to expect some improvement in volume, reduction in radius beyond 50% is extremely rare. Thus, if a near-zero radius confidence set is realized, one is dealing with an extremely rare event.

Nevertheless, our goal is to provide a reasonable interpretation of a zero-radius set; for no matter how rare the event is, the important point is that such a set can occur. Consider the following example, where  $p = 5$  and  $X \sim N(\theta, I)$  is observed. The experimenter uses  $C^Z$  of (2.15), and decides to set  $\mu = 0$ . The observed data are  $X = (-.059, -.539, .229, -.078, .194)$ . For these data,  $|X| = .625$ , and a 90% confidence set based on  $C^Z$  would have radius 1.519. The usual confidence set,  $C^O$ , has radius 3.039 and, moreover, a univariate 90% interval has radius 1.645. With the variance known and equal to 1, it seems difficult for the frequentist to have faith in  $C^Z$ , even though it has been demonstrated that  $C^Z$  is superior to  $C^O$  based on frequentist criteria.

Of course, this problem can be avoided by merely quoting the definition of a  $1 - \alpha$  confidence procedure, i.e., the property that  $C^O$  and  $C^Z$  share is that,

in many repeated trials,  $100(1-\alpha)\%$  of the realized sets will cover the true value of the parameter. Further, if the radius is unrelievable small, it might be concluded that this particular realization is one of the  $100\alpha\%$  that fail to cover. This argument, however, avoids the central issue of providing a meaningful single sample interpretation of a zero-radius confidence point.

Even though frequentist properties are obtained by averaging over the sample space, in many cases there is a need for a reasonable single sample interpretation. In practice, an experimenter is often forced to make sensible conclusions from only a single realization of a procedure (or, at best, a few repetitions). The strict frequentist model does not have a mechanism to do this. However, there is room for flexibility and, if we allow a slight amount of flexibility, we find very reasonable interpretations of zero-radius confidence sets.

Before we present an interpretation of a zero-radius confidence set, consider the following simple situation. An experimenter observes  $X_i \sim n(\theta_i, 1)$ ,  $i = 1, \dots, n$ , and finds that all the observations are quite close to zero, say in the interval  $(-1, 1)$ . In such a case, it seems reasonable to infer that all of the  $\theta_i$ 's are quite close to zero, indeed they may be equal to zero. If such an inference is accepted, then two conclusions immediately follow. First, one would estimate the common value with  $\bar{X} = \sum X_i / n$ . Second, the variation that is observed in the  $X_i$ 's is not due to any difference in means, but is due solely to error variation. Hence, the appropriate variance to use is a pooled variance,  $1/n$ . Thus, even though the variance in the original problem is known to be 1, the increased flexibility in this approach allows a reduction in error variance, and the construction of a confidence interval of greatly reduced length. One can easily see that, as  $n$  becomes infinite, such an argument leads

to a very reasonable zero or near-zero radius confidence interval. For the example given at the beginning of this section, the radius of a 90% confidence interval on the common value  $\theta$  is .736, a great reduction in radius over the individual intervals, and smaller than the radius of  $C^Z$  with  $|X| = .625$ .

Of course, there are problems with this ad hoc procedure. In particular, it is only reasonable under the model which assumes that the  $\theta_i$ 's are all equal. The point in introducing such a procedure is to show how a procedure such as  $C^Z$  can have a valid interpretation. If the data support the hypothesis that most of the variation is due to error, rather than differences in the means,  $C^Z$  works on this evidence and, in effect, uses a pooled estimate of variance. But  $C^Z$  does this in a formal manner, and retains dominance over the usual confidence set, both in volume and coverage probability, under the full model.

The confidence procedure  $C^Z$  is centered at an estimator which shrinks the observations toward a prior guess,  $\mu$ . This prior guess plays an important, even essential, part in the interpretation of a zero-radius confidence set. With procedures such as  $C^Z$ , which are not translation invariant, it is important to center the procedure at a reasonable prior guess. In fact, some reflection may convince the reader that unless this center is judiciously chosen, the set  $C^Z$  will become identical with  $C^0$ , and therefore no significant improvement is expected.

There is nothing to lose, and a great deal to gain, by taking some care in choosing  $\mu$ .  $C^Z$  obtains its greatest improvements (in terms of expected volume and coverage probability) for those values of  $\theta$  that are near  $\mu$ . Moreover,  $C^Z$  uniformly dominates the usual confidence procedure,  $C^0$ , no matter what is the chosen value of  $\mu$ . Thus, one can only gain by choosing a value for  $\mu$  which represents one's best guess at  $\theta$ . Seen in these terms, even the staunch-

est frequentist would have to agree that the incorporation of prior information can be useful, and, indeed, is a necessity if one wants a chance at realizing the gains that are made possible by using  $C^Z$ .

One might begin to see now that  $C^Z$  acts somewhat like our ad hoc procedure, if we had started with the prior guess  $\mu = 0$ . For this prior guess the radius of  $C^Z$  will be zero only when  $|X| = 0$  (when the data strongly supports the prior guess), and the radius increases as  $|X|$  increases (as the data move away from the prior guess). Thus, a zero or near-zero radius confidence set is realized only when all the  $X_i$ 's are close to zero. In such a situation,  $C^Z$  acts as if the variation in the  $X_i$ 's is due mostly to error, and not to a difference in means. Again, as in our ad hoc procedure, it is reasonable to use a pooled estimate of variance which, in effect, is what is done by  $C^Z$ .

The data given at the beginning of this section,  $X = (-.059, -.539, .229, -.078, .194)$ ' were taken from a table of random standard normal deviates. If one had, after some thought, centered  $C^Z$  at  $\mu = 0$ , the data provide very strong evidence that such a prior guess is very accurate. In this situation it is then reasonable to attribute the variation in the data to random error, and it is easy to accept as meaningful a confidence set with a very small radius. Thus, by incorporating prior information in the frequentist model, the flexibility of the model is increased so that there is now a reasonable interpretation of a confidence set with radius seemingly smaller than the original error variance will allow.

Of course, it is possible that  $C^Z$  may produce a confidence set with radius smaller than the ad hoc  $\bar{X}$  procedure. At first this might seem disconcerting, but it seems that, by way of incorporation of prior information,  $C^Z$  can make better use of the data (and take advantage of lucky data). By this we mean that, even though the variance is known, the radius of  $C^Z$  is sensitive to the

variation in the data, since it is dependent on the quantity  $|X - \mu|$ . If the prior guess is accurate, then  $|X - \mu|$  is measuring variance and  $C^Z$  uses this information. The usual confidence interval based on  $\bar{X}$  has no provision for "updating" the variance in the way  $C^Z$  does, so it is not surprising that there are samples for which  $C^Z$  can produce a smaller interval. It is interesting to note that, even though we are dealing with the known variance case,  $C^Z$  acts somewhat like an interval estimator based on Student's  $t$  distribution. The classic  $t$  interval has a radius that is an increasing function of  $\Sigma(X_i - \bar{X})^2$ , and hence is sensitive to the variation in the data. Note also that, according to Definition 1.1, the  $t$  interval is a zero-radius confidence procedure.

Therefore, we have seen that, even if the variance is a known constant, a zero-radius confidence set can be a reasonable set for a frequentist to consider. Such a set takes advantage of the evidence in the data, and working under a slightly more flexible model can produce better estimates than the usual procedure. Moreover, all this is done in a formal manner, so that  $C^Z$  retains dominance over  $C^0$  in the original model.

3.2. Bayesian considerations. Another means of finding a reasonable interpretation of a zero-radius confidence set is through a Bayesian argument. If there is a prior distribution on  $\theta$  for which the Bayes credible set is a zero-radius set, then we have a coherent structure in which the zero-radius confidence set is plausible.

Such an argument, which makes inferences backwards, from the sample to the prior is not really a strict Bayesian argument. (We might say that a strict Bayesian is one who knows his prior distribution exactly.) Rather, our argument uses the more flexible empirical Bayesian methodology, where it is not assumed that the prior distribution is known exactly. In the face of

uncertainty about the prior distribution, an empirical Bayes model seems to be the more reasonable course.

Since the confidence procedure  $C^Z$  was derived from empirical Bayes considerations, it is natural to expect it to lend itself to such an interpretation. However, we must also realize that, given there exists a prior distribution for which a zero-radius set is reasonable, we are still confronted with the problem of believing that such a prior is correct. But, at that point, we are now on more familiar ground, and are no longer questioning the validity of a zero-radius confidence set, but rather are questioning the validity of the model.

Recall the Bayesian model (2.1), and the credible set  $C^B$  of (2.8),

$$(3.1) \quad C^B = \{\theta : |\theta - \delta^B(X)|^2 \leq v^B\} \quad ,$$

where  $v^B$  is given in (2.9). It is easy to check that

$$\lim_{\tau^2 \rightarrow 0} v^B = 0 \quad ,$$

and, hence,  $C^B$  is a zero-radius set for  $\tau^2 = 0$ . We now see if the zero-radius procedure  $C^Z$  behaves in a fashion similar to  $C^B$  as  $\tau^2 \rightarrow 0$ .

We compare  $C^B$  and  $C^Z$  according to their probability content under the posterior distribution  $\pi(\theta|X)$ , defined in (2.2). Let  $P_X(A)$  denote the probability of  $A$  under  $\pi(\theta|X)$ . It then follows that

$$(3.2) \quad P_X(\theta \in C^B) = P \left[ \chi^2_p \leq c^2 - p \log \left( \frac{\tau^2}{\tau^2 + \sigma^2} \right) \right] \quad ,$$

and hence

$$(3.3) \quad \lim_{\tau^2 \rightarrow 0} P_X(\theta \in C^B) = 1 \quad \text{for all } X \quad .$$

Thus, the Bayes set has, at  $\tau^2 = 0$ , zero radius and posterior probability one.

Or, for arbitrarily small  $\tau^2$ ,  $C^B$  has arbitrarily small radius and high posterior probability. We now show that, in terms of posterior probability,  $C^Z$  behaves similarly to  $C^B$ .

THEOREM 3.1

$$(3.4) \quad \lim_{\tau^2 \rightarrow 0} P_X(\theta \in C^Z) = \begin{cases} 1 & \text{if } 0 < |X - \mu| < c^* \sigma \\ 0 & \text{if } |X - \mu| > c^* \sigma \end{cases},$$

where  $c^* = t/\sigma$  is the unique positive solution to  $[1 - (p-2)\sigma^2/t^2]^2 t^2 = v^Z(t)$ .

PROOF. Recall the definition of the terms in  $C^Z$ , given in (2.12) - (2.15).

We then have

$$\begin{aligned} P_X(\theta \in C^Z) &= P_X(|\theta - \delta_\mu^+(X)| \leq v^Z) \\ &= P_X \left[ \left| \left( \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2} \right)^{\frac{1}{2}} Z + \delta^B(X) - \delta_\mu^+(X) \right|^2 \leq v^Z \right], \end{aligned}$$

where  $Z \sim N(0, I_p)$ . This last step follows from the fact that  $\theta | X \sim N\{\delta^B(X), [\sigma^2 \tau^2 / (\sigma^2 + \tau^2)] I\}$ , where  $\delta^B(X)$  is given in (2.3). Now, as  $\tau^2 \rightarrow 0$ , it is clear that

$$\left( \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2} \right)^{\frac{1}{2}} Z \rightarrow 0 \text{ in probability},$$

hence, since  $\delta^B(X) \rightarrow \mu$  as  $\tau^2 \rightarrow 0$ ,

$$(3.5) \quad \lim_{\tau^2 \rightarrow 0} P_X(\theta \in C^Z) = \begin{cases} 1 & \text{if } |\mu - \delta_\mu^+(X)|^2 < v^Z \\ 0 & \text{if } |\mu - \delta_\mu^+(X)|^2 > v^Z \end{cases}$$

A little algebra will verify that (3.5) implies the theorem. ||

The value  $c^*$  is quite large, and it is easy to check that  $c^* > c$ , where  $c\sigma$  is the radius of the usual confidence set. From Theorem 3.1 we can conclude that  $C^Z$  behaves very much like  $C^B$ . In fact,  $C^Z$  behaves exactly as one would want an empirical Bayes set to behave.  $C^Z$  gives posterior probability 1 to all samples which satisfy  $0 < |X - \mu|/\sigma < c^*$ . If we calculate the conditional distribution of  $|X - \mu|/\sigma$ , we find

$$|X - \mu|^2/\sigma^2 \sim \chi_p^2\left(\frac{1}{2}|\theta - \mu|^2/\sigma^2\right),$$

where  $\chi_p^2(\cdot)$  denotes a noncentral chi-squared random variable. Small values of  $|X - \mu|/\sigma$  are evidence that  $|\theta - \mu|/\sigma$  is small, which in turn is evidence that  $\tau^2$  is small. Thus,  $C^Z$  behaves as a Bayes set if the sample information is consistent with the prior, and does not mimic the Bayes set if the sample information is inconsistent with the prior.

It is unfortunate, but Theorem 3.1 fails to cover the endpoints of the interval  $[0, c^*\sigma]$ . If  $|X - \mu| = c^*\sigma$ , the limiting probability (as  $\tau^2 \rightarrow 0$ ) is quite difficult to calculate. It appears that, if  $|X - \mu| = c^*\sigma$ , the limiting value of  $P_X(\theta \in C^Z)$  will be neither zero nor one, but some value in between. It is perhaps more unfortunate, and somewhat annoying, that if  $|X - \mu| = 0$ , then  $P_X(\theta \in C^Z) = 0$ . Thus, if the realized confidence set has radius exactly equal to zero, the posterior probability of coverage is also zero. Although this fact seems somewhat dissatisfying from a theoretical point of view, it will have virtually no effect in practice. For example, if  $v^Z$  were truncated so that  $v^Z = \epsilon$  if  $|X - \mu| < \delta$ , for some  $\epsilon$  and  $\delta$ , then we would have  $\lim_{\tau^2 \rightarrow 0} P_X(\theta \in C^Z) = 1$  at  $|X - \mu| = 0$ . The constant  $\epsilon$  could be chosen arbitrarily small (say  $10^{-100}$ ), so that, for all practical purposes we would have a zero-radius procedure. Alternatively, we could calculate the case  $|X - \mu| = 0$  as a limit, and it is easy to check that

$$\lim_{|X-\mu| \rightarrow 0} \lim_{\tau^2 \rightarrow 0} P_X(\theta \in C^Z) = 1 \quad ,$$

even when  $v^Z = 0$  at  $|X - \mu| = 0$  .

4. Comments and Conclusions. Through the power of both the Stein effect and the empirical Bayes methodology, it has been demonstrated that a  $1 - \alpha$  confidence procedure with a zero-radius realization can be constructed. Thus, the ultimate reduction in volume (with no loss in coverage probability) can be achieved.

Although the existence of such a procedure is surprising, it is probably not that surprising to researchers who have worked in this area. The more important issue raised by such procedures is that of reasonable interpretation, especially within the frequentist methodology. Indeed, at first it seems that the existence of zero-radius confidence procedures cast doubt on the validity of the frequentist methodology.

In Section 3.1 it was argued that this is not the case: zero-radius confidence procedures do fit in, and can have reasonable interpretations, within a frequentist theory. One important point made in Section 3.1, which we reiterate, is that it is important to consider prior input when using such procedures. Indeed, when using any Stein-type procedure, improvement over the usual procedure will be minimal unless the estimator is centered according to some reasonable prior guess. The same holds for any procedure which is not translation invariant.

Since prior input costs nothing in terms of risk, and can be of great help in placing the region of improvement where it will do the most good, we see no basis for any argument against the use of prior input. Once prior

input is incorporated into the frequentist model, the zero-radius confidence procedure can be interpreted reasonably.

From a Bayesian, or more appropriately, empirical Bayesian view, a zero-radius confidence set has a straightforward interpretation as a Bayes set arising from a prior concentrated at one point. As mentioned before, the empirical Bayesian is confronted with the problem of whether he believes that the concentrated prior is accurate, but he is at least provided with a coherent structure in which the zero-radius set makes sense.

Thus, with a judicious (lucky?) choice of prior input, it is possible to achieve great gains over the classic confidence procedure. And, again, by incorporating the prior input into the model, and into the single-sample interpretation, an experimenter can have as much faith in a zero-radius confidence set as he has in the usual one.

Acknowledgement. We would like to thank James Berger, whose Bayesian harangues prompted us to write this paper. We also thank Leon Gleser, with whom we have had discussions which helped us better understand this problem.

REFERENCES

- BERGER, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. **8** 716-761.
- CASELLA, G. and HWANG, J. T. (1982a). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. Technical Report BU-767-M, Biometrics Unit, Cornell University, Ithaca, NY.
- CASELLA, G. and HWANG, J. T. (1982b). Decision-theoretic set estimation. Technical Report BU-773-M, Biometrics Unit, Cornell University, Ithaca, NY.
- FAITH, R. E. (1976). Minimax Bayes set and point estimators of a multivariate normal mean. Technical Report #66, University of Michigan, Ann Arbor, MI.
- MORRIS, C. N. (1982). Parametric empirical Bayes confidence intervals. Technical Report, Department of Mathematics, University of Texas, Austin, TX. To appear in Scientific Inference, Data Analysis, and Robustness.



TABLE 2

Distribution of radius reduction possible using the procedure  $C^Z$  of (2.15).

Entries are the probability  $P[\text{radius}(C^Z) \leq \gamma \text{radius}(C^0) | \mu = \theta]$ ,

where  $C^Z$  and  $C^Z$  both have confidence coefficient .9

| $\gamma$ | 3    | 5    | 7    | 9                 | 11                | 13                | 15                |
|----------|------|------|------|-------------------|-------------------|-------------------|-------------------|
| .5       | .014 | .004 | .001 | .0 <sup>3</sup> 3 | .0 <sup>4</sup> 6 | .0 <sup>4</sup> 1 | .0 <sup>5</sup> 3 |
| .75      | .110 | .113 | .087 | .063              | .044              | .031              | .021              |
| .85      | .275 | .386 | .406 | .400              | .386              | .367              | .246              |
| .90      | .465 | .682 | .754 | .789              | .808              | .820              | .827              |
| .95      | .812 | .975 | .994 | .998              | .999              | .999              | .999              |