Statistical Science
2010, Vol. 0, No. 00, 1–13
DOI: 10.1214/10-STS319

Shrinkage Confidence Procedures

George Casella and J. T. Gene Hwang

Abstract. The possibility of improving on the usual multivariate normal confidence was first discussed in Stein (1962). Using the ideas of shrinkage, through Bayesian and empirical Bayesian arguments, domination results, both analytic and numerical, have been obtained. Here we trace some of the developments in confidence set estimation.

Key words and phrases: Stein effect, coverage probability, empirical Bayes.

1. INTRODUCTION

In estimating a multivariate normal mean, the usual p-dimensional $1-\alpha$ confidence set is

$$C^0_{x,\sigma} = \{\theta : |\theta - x| \le c\sigma\}, \tag{1}$$

where we observe $X = x$, where X is a random variable with a p-variate normal distribution with mean θ and covariance matrix $\sigma^2 I$, $X \sim N_p(\theta, \sigma^2 I)$, I is the $p \times p$ identity matrix, and c^2 is the upper α cutoff of a chi-squared distribution, satisfying $P(\chi^2_p \le c^2) = 1-\alpha$.

Although the above formulation looks somewhat naive, it is very relevant in applications of the linear model, still one of the most widely-used statistical models. For such models, typical assumptions lead to $\hat\beta \sim N(\beta, \sigma^2\Sigma)$, where $\hat\beta$ is the least squares estimator (and MLE under normality), β is the vector of regression slopes and Σ is a known covariance matrix (typically depending on the design matrix). The usual confidence set for β is

$$\{\beta : (\hat\beta - \beta)'\Sigma^{-1}(\hat\beta - \beta) \le c^2\sigma^2\}. \tag{2}$$

Letting $x = \Sigma^{-1/2}\hat\beta$ and $\theta = \Sigma^{-1/2}\beta$ reduces (2) to (1).

In theoretical investigations of confidence sets and procedures, we often first take σ^2 known. When σ^2 is unknown, the usual strategy is to replace it by some usual estimator, such as the sample variance s^2. Under normality, if s^2 has ν degrees of freedom, then $s^2 \sim \sigma^2\chi^2_\nu$, independent of $\hat\beta$. For example, the usual F confidence set for the regression parameters based on a linear model can be reduced to $C^0_{x,\sigma}$ with the usual unbiased estimator s^2 substituted for σ^2. This is the usual Scheffé confidence set. Unfortunately, contrary to the point estimation case, there are few theoretical results for unknown σ^2. However, there is continued numerical evidence that the usual confidence set can be dominated in the unknown variance case (see, for example, Casella and Hwang, 1987). Moreover, Hwang and Ullah (1994) argue that the domination of the alternative fixed radius confidence spheres for the unknown σ^2 case, over Scheffé's set, holds with a larger shrinkage factor.

Since we are assuming that σ^2 is known, we take it equal to 1 and (1) becomes

$$C^0_x = \{\theta := : |\theta - x| \le c\}. \tag{3}$$

We now ask the question of whether it is possible to improve on C^0_x in the sense of finding a confidence set C' such that, for all θ and x,

(i) $P_\theta(\theta \in C') \ge P_\theta(\theta \in C^0_x)$;
(ii) volume of $C' \le$ volume of C^0_x;

with strict inequality holding in either (i) or (ii) for a set θ or x with positive Lebesgue measure. The answer to this question may be yes for higher dimensional cases, as suggested by the work of Stein.

The celebrated work of James and Stein (1961) shows that the estimator

$$\delta^{\mathrm{JS}}(x) = \left(1 - \frac{a}{|x|^2}\right)x \tag{4}$$

George Casella is Distinguished Professor, Department of Statistics, University of Florida, Gainesville, FL 32611, USA (e-mail: casella@stat.ufl.edu). J. T. Gene Hwang is Professor, Department of Mathematics, Cornell University, Ithaca, NY 14853, USA, and Adjunct Professor, Department of Statistics, Cheng Kung University, Tainan, Taiwan (e-mail: hwang@math.cornell.edu).

dominates X with respect to squared error loss if $0 < a < 2(p-2)$, that is,

$$\mathrm{E}_\theta|\delta^{\mathrm{JS}}(X)-\theta|^2 \begin{cases} \le \mathrm{E}_\theta|X-\theta|^2 & \text{for all } \theta, \\ < \mathrm{E}_\theta|X-\theta|^2 & \text{for some } \theta. \end{cases} \tag{5}$$

In practice, this estimator has the deficiency of a singularity at 0 in that $\lim_{|x|\to 0}\delta^{\mathrm{JS}}(x) = -\infty$. This deficiency can be corrected with the positive part estimator (appearing in Baranchik, 1964, and mentioned as Example 1 in Baranchik, 1970)

$$\delta^+(x) = \left(1 - \frac{a}{|x|^2}\right)^+ x, \tag{6}$$

where $(b)^+ = \max\{0, b\}$. This estimator actually improves on $\delta^{\mathrm{JS}}(x)$ and is so good that, even though it was known to be inadmissible, it took 30 years to find a dominating estimator (Shao and Strawderman, 1994). The removal of the singularity makes $\delta^+(x)$ a more attractive candidate for centering a confidence set.

A simple proof of (5) can be found in Stein (1981); see also Lehmann and Casella (1998), Chapter 5. Therefore, it seems reasonable to conjecture that we can use a Stein estimator to dominate the confidence set C_x^0. Although this turns out to be the case, it is a very difficult problem.

2. RECENTERING

Stein (1962) gave heuristic arguments[1] that showed why recentered sets of the form

$$C_\delta^+ = \{\theta : |\theta - \delta^+(\mathbf{x})| \le c\} \tag{7}$$

would dominate the usual confidence set (3) in the sense that $P_\theta(\theta \in C_\delta^+(\mathbf{X})) > P_\theta(\theta \in C_x^0(\mathbf{X}))$ for all θ, where $\mathbf{X} \sim N_p(\theta, I)$, $p \ge 3$. (Note that this set has the same volume as C_x^0, but is recentered δ^+. Dominance would thus be established if we can show that C_δ^+ has higher coverage probability than C_x^0.) Stein's argument was heuristic, but Brown (1966) and Joshi (1967) proved the inadmissibility of C_x^0 if $p \ge 3$ (without giving an explicit dominating procedure). Joshi (1969) also showed that C_x^0 was admissible if $p \le 2$.

The existence results of Brown and Joshi are based on spheres centered at

$$\left(1 - \frac{a}{b + |x|^2}\right)x \tag{8}$$

[compare to (6)] where a is made arbitrarily small and b is made arbitrarily large. But these existence results fall short of actually exhibiting a confidence set that dominates C_x^0.

The first analytical and constructive results were established by (surprise!) Hwang and Casella (1982), who studied the coverage probability of C_δ^+ in (7). Since C_δ^+ and C_x^0 have the same volume, domination will be established if it can be shown that C_δ^+ has higher coverage probability for every value of θ. It is easy to establish that:

- $P_\theta(\theta \in C_\delta^+(X))$ is only a function of $|\theta|$, the Euclidean norm of θ, and
- $\lim_{|\theta|\to\infty} P_\theta(\theta \in C_\delta^+(X)) = 1 - \alpha$, the coverage probability of C_x^0.

Therefore, to prove the dominance of C_δ^+, it is sufficient to show that the coverage probability is a nonincreasing function of $|\theta|$. Hwang and Casella (1982) derived a formula for $(d/d|\theta|)P_\theta(\theta \in C_\delta^+(X))$ and found a constant a_0 (independent of θ) such that if $0 < a < a_0$, C_δ^+ dominates C_x^0 in coverage probability for $p \ge 4$. Using a slightly different method of proof, Hwang and Casella (1984) extended the dominance to cover the case $p = 3$. This proof is outlined in Appendix A. The analytic proof was generalized to spherical symmetric distributions by Hwang and Chen (1986).

There is an interesting geometrical oddity associated with the Stein recentered confidence set. To see this, we first formalize our definitions of confidence sets. Note that for any confidence set we can speak of the x-section and the θ-section. That is, if we define a *confidence procedure* to be a set $C(\theta, x)$ in the product space $\Theta \times \mathcal{X}$, then:

(1) The x-section, $C_x = \{\theta : \theta \in C(\theta, x)\}$, is the confidence set.
(2) The θ-section, $C_\theta = \{x : x \in C(\theta, x)\}$, the acceptance region for the test $H_0 : \{\theta\}$.

We then have the tautology that $\theta \in C_x$ if and only if $x \in C_\theta$ and, thus, we can evaluate the coverage probability $P_\theta(\theta \in C_X)$ by computing $P_\theta(X \in C_\theta)$, which is often a more straightforward calculation.

For the usual confidence set, both C_x^0 and C_θ^0 are spheres, one centered at x and one centered at θ. Although the confidence set C_x^+ is a sphere, the associated θ-section C_θ^+ is not, and has the shape portrayed in Figure 1. Notice the flattening of the set in the side closer to 0 in the direction perpendicular to θ, and the slight expansion away from 0. Stein (1962) knew of this flattening phenomenon, which he noted can be

[1] Stein's paper must be read carefully to appreciate these arguments. He uses a large p argument and the fact that X and $X - \theta$ are orthogonal as $p \to \infty$.

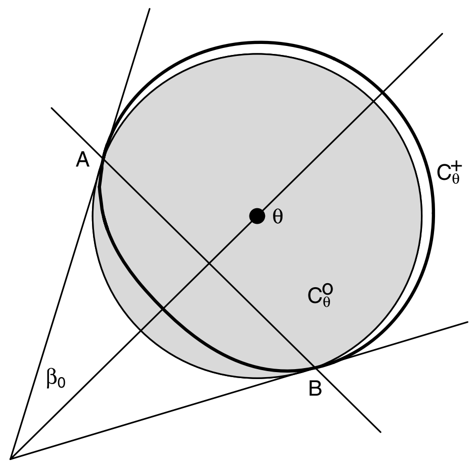

FIG. 1. *Two-dimensional representation for C_θ^+ and C_θ^0 for $|\theta| > c$, where C_θ^0 is the sphere of radius c centered at θ (shaded). The set C_θ^+ intersects C_θ^0 at point A and B (details on the points of intersection are in Hwang and Casella, 1982). Note the flattening of C_θ^+ on the side toward the origin and the decrease in volume over C_θ^0.*

achieved in any fixed direction. What is interesting is that this reshaping of the θ-section of the recentered set leads to a set with higher coverage probability than C_x^0 when $p \geq 3$.

3. RECENTERING AND SHRINKING THE VOLUME

The improved confidence sets that we have discussed thus far have the property that their coverage probability is uniformly greater than that of C_x^0, but the infimum of the coverage probability (the *confidence coefficient*) is equal to that of C_x^0. For example, recentered sets such as C_δ^+ will present the same volume and confidence coefficient to an experimenter so, in practice, the experimenter has not gained anything. (This is, of course, a fallacy and a shortcoming of the frequentist inference, which requires the reporting of the infimum of the coverage probability.)

However, since the coverage probability of C_δ^+ is uniformly higher than the infimum $\inf_\theta P_\theta(\theta \in C_X^0) = 1-\alpha$, it should be possible to reduce the radius of the recentered set and maintain dominance in coverage probability.

In this section we describe some approaches to constructing improved confidence sets, approaches that not only result in a recentering of the usual set, but also try to reduce the radius (or, more generally, the volume). Some of these constructions are based on variations of Bayesian highest posterior density regions, and thus share the problem of trying to describe exactly what the x-section, the confidence set, looks like. Others are more of an empirical Bayes approach, and tend to have more transparent geometry.

3.1 Reducing the Volume–Bayesian Approaches

The first attempt a constructing confidence sets with reduced volume considered sets with the same coverage probability as C_X^0, but with uniformly smaller volume. One of the first attempts was that of Faith (1976), who considered a Bayesian construction based on a two-stage prior where

$$\theta \sim N(0, t^2 I), \quad t^2 \sim \text{Inverted Gamma}(a, b),$$

which is similar (but not equal) to the prior used by Strawderman (1971) in the point estimation problem (Appendix B). The two-stage prior amounts to a proper prior with density

$$\pi(\theta) \propto (2b + |\theta|^2)^{-(a+p/2)},$$

the multivariate t-distribution with $2a$ degrees of freedom. Faith then derived the Bayes decision against a linear loss, but modified it to the more explicitly defined region

$$C_F = \left\{\theta : \left(\frac{\exp(c^2)}{\exp(|x-\theta|^2)}\right)^{1/(p+2a)} \geq \frac{2b+\theta^2}{2b+|x|^2}\right\},$$

where c is the radius of C_x^0. It may happen that C_F is not convex. However, if $a > -p/2$ and $b > (a + p/2)/8$, the convexity of C_F was established. Unfortunately, little else was established except when $p = 3$ or $p = 5$, where for some ranges of a and b it was shown that C_F has smaller volume and higher coverage probability than C_x^0.

Berger (1980) took a different approach. Using a generalization of Strawderman's prior, he calculated the posterior mean $\delta_B(x)$ and posterior covariance matrix $\Sigma_B(x)$ and recommended

$$C_B = \{\theta : (\theta - \delta_B(x))' \Sigma_B(x)^{-1} (\theta - \delta_B(x)) \leq \chi^2_{p,\alpha}\},$$

where $\chi^2_{p,\alpha}$ is the upper α cutoff point from a chi-square distribution with p degrees of freedom. The posterior coverage probability would be exactly $1-\alpha$ if the posterior distribution were normal, but this is not the case (and the posterior coverage is not the frequentist coverage). However, Berger was able to show that his set has very attractive coverage probability and small expected volume based on partly analytical and partly numerical evidence.

3.2 Reducing the Volume–Empirical Bayes Approaches

A popular construction procedure for finding good point estimators is the empirical Bayes approach (see Lehmann and Casella, 1998, Section 4.6, for an introduction), and proves to also be a useful tool in confidence set construction. However, unlike the point estimation problem, where a direct application of empirical Bayes arguments led to improved Stein-type estimators (see, for example, Efron and Morris, 1973), in the confidence set problem we find that a straightforward implementation of an empirical Bayes argument would not result in a $1-\alpha$ confidence set. Modifications are necessary to achieve dominance of the usual confidence set.

Suppose that we begin with a traditional normal prior at the first stage, and have the model

$$X \sim N(\theta, I), \quad \theta \sim N(0, \tau^2 I),$$

which results in the Bayesian Highest Posterior Density (HPD) region

$$C^\pi = \{\theta : |\theta - \delta^\pi(x)|^2 \le c^2 M\}, \tag{9}$$

where $M = \tau^2/(\tau^2+1)$ and $\delta^\pi(x) = Mx$ is the Bayes point estimator of θ. This follows from the classical Bayesian result that $\theta|x \sim N(Mx, MI)$.

However, for a fixed value of τ, the set C^π cannot have frequentist coverage probability above $1-\alpha$ for all values of θ. This is easily seen, as the posterior coverage is identically $1-\alpha$ for all x, and, hence, the double integral over x and θ is equal to $1-\alpha$. This means that the frequentist coverage is either equal to $1-\alpha$ for all θ, or goes above and below $1-\alpha$. Since the former case does not hold (check $\theta = 0$ and a nonzero value), the coverage probability of C^π is not always above $1-\alpha$.

Consequently, if we take a naive approach and replace τ^2 by a reasonable estimate, an empirical Bayes approach, we cannot expect that such a set would maintain frequentist coverage above $1-\alpha$. This is because such a set would have coverage probabilities converging to those of C^π (as the sample size increases) and, hence, such an empirical Bayes set would inherit the poor coverage probability of C^π. This phenomenon has been documented in Casella and Hwang (1983).

As an alternative to the naive empirical Bayes approach, consider a decision-theoretic approach with a loss function to measure the loss of estimating the parameter θ with the set C:

$$L(\theta, C) = k\,\mathrm{vol}(C) - I(\theta \in C), \tag{10}$$

where k is a constant, $\mathrm{vol}(C)$ is the volume of the set C, and $I(\cdot)$ is the indicator function. Starting with a prior distribution $\pi(\theta)$, the Bayes rule against $L(\theta, C)$ is the set

$$\{\theta : \pi(\theta|x) > k\}, \tag{11}$$

where $\pi(\theta|x)$ is the posterior distribution. This is a highest posterior density (HPD) region.

The choice of k is somewhat critical, and we chose it to coincide with properties of C^0. Specifically, if we chose $k = \exp(-c^2/2)/(2\pi)^{p/2}$, then C^0 is minimax for the loss (10). An alternative explanation of this choice of k is based on the reasoning that as $\tau \to \infty$, (11) would converge to C^0, which insures that the alternative intervals would not become inferior to C^0 for large τ^2. (See He, 1992; Qiu and Hwang, 2007; and Hwang, Qiu and Zhao, 2010.) Applying this choice of k with the normal prior $\theta \sim N(0, \tau^2 I)$ yields the Bayes set

$$C^\pi_{x,k} = \{\theta : |\theta - \delta^\pi(x)| \le M[c^2 - p\log M]\},$$

where $\delta^\pi(x)$ and M are as in (9). By estimating the hyperparameters, this is then converted to an empirical Bayes set

$$C^E_x = \{\theta : |\theta - \delta^+(x)| \le v_E(x)\},$$

where $\delta^+(x)$ is the positive part estimator of (6), and $v_E(x)$ is given by

$$\begin{aligned} v_E(x) &= \left(1 - \frac{p-2}{\max(|x|^2, c^2)}\right) \\ &\quad \cdot \left[c^2 - p\log\left(1 - \frac{p-2}{\max(|x|^2, c^2)}\right)\right]. \end{aligned} \tag{12}$$

Note that $M[c^2 - p\log M] \to 0$ as $b \to 0$, but $v_E(x)$ is bounded away from zero. This is important in maintaining coverage probability. Extensive numerical evidence was given to support the claim that C^E_x is a uniform improvement over C^0_x.

Confidence sets with exact $1-\alpha$ coverage probability, with uniformly smaller volume, have also been constructed by Tseng and Brown (1997), adapting results from Brown et al. (1995). These confidence sets are shown, numerically, to typically have smaller volume that those of Berger (1980).

Brown et al. (1995), working on the problem of bioequivalence, start with the inversion of an α-level test and derive a $1-\alpha$ confidence interval that minimizes a Bayes expected volume, that is, the volume averaged with respect to both x and θ. Tseng and Brown (1997), using a normal prior $\theta \sim N(0, \tau^2 I)$, show that the corresponding set of Brown et al. (1995) becomes

$$C^{\mathrm{B}} = \left\{\theta : \left|x - \theta\left(\frac{1+\tau^2}{\tau^2}\right)\right|^2 \le k(|\theta|^2/\tau^4)\right\},$$

where $k(\cdot)$ is chosen so that C^{B} has exactly $1-\alpha$ coverage probability for every θ. A simple calculation shows that the squared term in C^{B} has a noncentral chi squared distribution, so $k(\cdot)$ is the appropriate α cutoff point. In doing this, Tseng and Brown avoided the problem of Casella and Hwang (1983), and the radius does not need to be truncated.

Of course, to be usable, we must estimate τ^2. The typical empirical Bayes approach would be to replace τ^2 with an estimate, a function of x. However, Tseng and Brown take a different approach and replace τ^2 with a function of θ, thereby maintaining the $1-\alpha$ coverage probability. They argue that θ is more directly related to τ than is x, and should provide a better "estimator." Examples where this "pseudo-empirical Bayes" approach was used are discussed in Hwang (1995) and Huwang (1996).

The set proposed by Tseng and Brown is

$$C^{\mathrm{TB}} = \left\{\theta : \left|x - \theta\left(1 + \frac{1}{A + B|\theta|^2}\right)\right|^2 \le k\left(\left(\frac{|\theta|}{A + B|\theta|^2}\right)^2\right)\right\}$$

for constants $A \ge 0$ and $B > 0$, and has coverage exactly equal to $1-\alpha$ for every θ. Combining analytical results and numerical calculations, these sets are shown to have uniformly smaller volume that C_x^0. Moreover, Tseng and Brown also demonstrate volume reductions over the sets of Berger (1980) and Casella and Hwang (1983). The only quibble with their approach is that the exact form of the set is not explicit, and can only be solved numerically.

3.3 Reducing Volume and Increasing Coverage

The first confidence set analytically proven to have smaller volume and higher coverage than C_x^0 is that of Shinozaki (1989). Shinozaki worked with the x-section of the confidence set, starting with the set C_x^0. Consider Figure 1, but drawn as the x-section centered at θ. By moving the two intersecting lines toward the center, he is able to construct a new set with the same coverage probability as C_x^0 but smaller volume. These sets can have a substantial improvement over C_x^0, but smaller improvements compared to Berger (1980) and Casella and Hwang (1983) (especially when p is large and $|\theta|$ is small). Moreover, there is no point estimator that is explicitly associated with this set.

3.4 Other Constructions

Samworth (2005) looked at confidence sets of the form

$$\{\theta : |\theta - \delta^+|^2 \le w_\alpha(\theta)\},$$

where δ^+ is the positive part estimator (6), $w_\alpha(\theta)$ is the appropriate α-level cutoff to give the confidence set coverage probability $1-\alpha$ for all θ, and X has a spherically symmetric distribution. He then replaced $w_\alpha(\theta)$ by its Taylor expansion

$$w_\alpha(\theta) \approx w_\alpha(0) + \tfrac{1}{2} w''_\alpha(0)|\theta|^2,$$

and, replacing θ with x, arrived at the confidence set

$$\{\theta : |\theta - \delta^+|^2 \le \min(w_\alpha(0) + \tfrac{1}{2} w''_\alpha(0)|x|^2, c^2)\}.$$

Samworth noted the importance of the quantity $f'(c^2)/f(c^2)$, where f is the density of x (the relative increasing rate of f at c^2). The radius of the analytic confidence set only depends on the density through c^2 and $f'(c^2)/f(c^2)$. This point was previously noted by Hwang and Chen (1986) and Robert and Casella (1990).

This confidence set compares favorably with that of Casella and Hwang (1983), having smaller volume especially when $|x|$ is small. Numerical results were given not only for the normal distribution, but also for other spherically symmetric distributions such as the multivariate t and the double exponential. Furthermore, a parametric bootstrap confidence set is also proposed, which also performs well.

Efron (2006) studies the problem of confidence set construction with the goal of minimizing volume. He ultimately shows that seeking to minimize volume may not be the best way to improve inferences, and that relocating the set is more important than shrinking it. Using a unique construction based on a polar decomposition of the normal density, Efron derived a "confidence density" which he used to construct sets with $1-\alpha$ coverage probability, and ultimately a minimum volume confidence set with $1-\alpha$ posterior probability.

The confidence density, which plays a large part in Efron's paper, is used to show the importance of locating the confidence set properly. The sets of Tseng and Brown (1997) and Casella and Hwang (1983) perform well on this evaluation. A minimum volume construction is also derived, and it is shown that the resulting set is not optimal in any inferential sense. Inferential properties, similar to type I and type II errors, are explored. It is also seen that as the relocated sets decrease volume of the confidence set, they *increase* the acceptance regions.

4. SHRINKING THE VARIANCE

Thus far, we have only addressed the problem of improving confidence regions for the mean. However, there is also a Stein effect for the estimation of the variance, and this can be exploited to produce improved confidence intervals for the variance.

Stein (1964) was the first to notice this (of course!). Specifically, let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, univariate, where both μ and σ are unknown, and calculate $\bar{X} = (1/n)\sum_i X_i$ and $S^2 = \sum_i (X_i - \bar{X})^2$. Against squared error loss, the best estimator of σ^2, of the form cS^2, has $c = (n+1)^{-1}$. This is also the best equivariant estimator [with the location-scale group and the equivariant loss $(\delta - \sigma^2)^2/\sigma^4$], and is minimax. Stein showed that the estimator

$$\delta^S(\bar{X}, S^2) = h(\bar{X}^2/S^2)S^2,$$

$$h(\bar{X}^2/S^2) = \min\left\{\frac{1}{n+1}, \frac{1 + n\bar{X}^2/S^2}{n+2}\right\},$$

uniformly dominates $S^2/(n+1)$. Notice that $\delta^S(\bar{X}, S^2)$ converges to $S^2/(n+1)$ if $\bar{X}^2/S^2$ is big, but shrinks the estimator toward zero if it is small. Stein's proof was quite innovative (and is reproduced in the review paper by Maatta and Casella, 1990). The proof is based on looking at the conditional expectation of the risk function, conditioning on $\bar{X}/S$, and showing that moving the usual estimator toward zero moves to a lower point on the quadratic risk surface. This approach was extended by Brown (1968) to establish inadmissibility results, and by Brewster and Zidek (1974), who found the best scale equivariant estimator. Minimax estimators were also found by Strawderman (1974), using a different technique.

Turning to intervals, building on the techniques developed by Stein and Brown, Cohen (1972) exhibited a confidence interval for the variance that improved on the usual confidence interval. If $(S^2/b, S^2/a)$ is the shortest $1-\alpha$ confidence interval based on S^2 (Tate and Klett, 1959), Cohen (1972) considered the confidence interval

$$(S^2/b, S^2/a)I(\bar{X}^2/S^2 > k) + (S^2/b', S^2/a')I(\bar{X}^2/S^2 \le k),$$

where $I(\cdot)$ is the indicator function, $1/a - 1/b = 1/a' - 1/b'$, so each piece has the same length, but $1/a' < 1/a$ and $1/b' < 1/b$. So if $\bar{X}^2/S^2$ is small, the interval is pulled toward zero, analogous to the behavior of the Stein point estimator. Shorrack (1990) built on this argument, and those of Brewster and Zidek (1974), to construct a generalized Bayes confidence interval that smoothly shifts toward zero, keeping the same length as the usual interval but uniformly increasing coverage probability. Building further on these arguments, Goutis and Casella (1992) constructed generalized Bayes intervals that smoothly shifted the usual interval toward zero, reducing its length but maintaining the same coverage probability. For more recent developments on variance estimation see Kubokawa and Srivastava (2003) and Maruyama and Strawderman (2006).

5. CONFIDENCE INTERVALS

In some applications there may be interest in making inference individually for each θ_i. One example is the analysis of microarray data in which the interest is to determine which genes are differentially expressed (that is, having θ_i, the difference of the true expression between the treatment group and the control group, different from zero). Although the confidence sets of the previous section can be projected to obtain confidence intervals, that will typically lead to wider intervals than a direct construction.

If X_i are i.i.d. $N(\theta_i, \sigma_i^2)$, $i = 1, \ldots, p$, the usual one-dimensional interval is

$$I^0_{X_i} = X_i \pm c\sigma_i,$$

where c is chosen so that the coverage probability is $1-\alpha$. Hence, c is the $\alpha/2$ upper quantile of a standard normal.

5.1 Empirical Bayes Intervals

If a frequentist criterion is used, it is not possible to simultaneously improve on the length and coverage probability of $I^0_{X_i}$ in one dimension. However, it is possible to do so if an empirical Bayes criterion is used. Morris (1983) defined an empirical Bayes confidence region with respect to a class of priors Π, having confidence coefficient $1-\alpha$ to be a set $C(X)$ satisfying

$$\begin{aligned} P_\pi\big(\theta \in C(X)\big) &= \int P_\theta\big(\theta \in C(X)\big)\pi(\theta)\,d\theta \\ &\ge 1-\alpha \quad \text{for all } \pi(\theta) \in \Pi. \end{aligned}$$

Note that $P_\pi(\theta \in C(X))$ is the Bayes coverage probability in that both X and θ are integrated out. Using normal priors with both equal and unequal variance, Morris went on to construct $1-\alpha$ empirical Bayes confidence intervals that have average (across i) squared lengths smaller than I^0_X. Bootstrap intervals based on Morris' construction are also proposed in Laird and Louis (1983).

STS stspdf v.2010/06/14 Prn:2010/07/01; 12:08 F:sts319.tex; (Ausra) p. 7

In the canonical model

$$X_i \sim \text{i.i.d. } N(\theta_i, 1) \quad \text{and} \tag{13}$$
$$\theta_i \sim \text{i.i.d. } N(0, \tau^2),$$

He (1992) proved that there exists an interval that dominates I_X^0. Precisely, for $\delta^+(X)$ of (6), it was shown that there exists $a > 0$ such that the interval $\delta_i^+(X) \pm c$ has higher Bayes coverage probability for any $\tau^2 > 0$.

The approach He took is similar to the approach of Casella and Hwang (1983), using a one-dimensional loss function similar to the linear loss (10) except that θ is replaced by only the component θ_i of interest. As in the discussion following (10), k and c need to be properly linked. With such a choice of k, the decision Bayes interval is then approximated by its empirical Bayes counterpart:

$$C_X^{\text{He}} = \{\theta_i : |\theta_i - \delta_i^+(X)|^2 \le \nu(|X|)\}.$$

Here $\delta_i^+(X)$ is the ith component of the James–Stein positive part estimator (6) with $a = p - 2$,

$$\nu(|X|) = \hat{M}(c^2 - \log \hat{M}), \tag{14}$$
$$\hat{M} = \max\left\{\left(1 - \frac{p-2}{|X|^2}\right)^+, \frac{1}{p-1}\right\}.$$

Note the resemblance to (12). There is also a truncation carried out in the definition of $\hat{M}$ so that $\nu(|X|)$ is bounded away from zero.

It can be shown that the length of C_X^{He} is always smaller than that of I_X^0 for each individual coordinate, i as long as $c > 1$, or, equivalently, $1 - \alpha > 68\%$. In contrast, in Morris (1983) only the average length across i was made smaller.

Numerical studies in He (1992) demonstrated that his interval is an empirical Bayes confidence interval with $1 - \alpha$ confidence coefficient. Also, on average, it has shorter length than the intervals of Morris (1983) or Laird and Louis (1983) when $\alpha = 0.05$ or 0.1. He concluded that his interval is recommended only if $\alpha \le 0.1$. Interestingly, in modern application with the concerns of multiple testings, a small value of α is more important.

5.2 Intervals for the Selected Mean

An important problem in statistics is to address the confidence estimation problem after selecting a subset of populations from a larger set. This is especially so if the number p of populations is huge and the number of selected populations, k, is relatively small, a scenario typical in microarray experiments. For example, ignoring the selection and just estimating the parameters of the selected populations by the sample means would have serious bias, especially if the populations selected are the ones with largest sample means. In such a situation, intuition would suggest that some kind of shrinkage approach is very much needed.

Specifically, we consider the canonical model

$$X_i \sim \text{i.i.d. } N(\theta_i, \sigma_i^2) \quad \text{and} \tag{15}$$
$$\theta_i \sim \text{i.i.d. } N(\mu, \tau^2).$$

Let $\theta_{(i)}$ be the parameter of the selected population, that is, it is the θ_j such that $X_j = X_{(i)}$ where

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(p)} \tag{16}$$

are the order statistics of $(X_1, \ldots, X_p)$. In particular, $\theta_{(p)}$ is the θ that corresponds to the largest observation $X_{(p)} = \max_j X_j$. Note that it is *not true* that $\theta_{(1)} \le \theta_{(2)} \le \cdots \le \theta_{(p)}$. In particular, $\theta_{(p)}$ is not necessarily the largest of the θ_j's. It is just that θ_j happens to have produced the largest observations among the X_i's.

In the point estimation problem, the naive estimator of $\theta_{(p)}$ is $X_{(p)}$, which can be intuitively seen to be an overestimate, especially if all θ_i are equal. A shrinkage estimator adapted to this situation would seem more reasonable. Hwang (1993) was able to show that for estimating $\theta_{(p)}$, a variation of the positive-part estimator (6), with X_i replaced by $X_{(i)}$, has smaller Bayes risk than $X_{(i)}$ with respect to one-dimensional squared error loss.

For the construction of confidence intervals, Qiu and Hwang (2007) adapted the approach of Casella and Hwang (1983) and He (1992) to this problem. For any selection, they constructed $1 - \alpha$ empirical Bayes confidence intervals for $\theta_{(i)}$ which are shown numerically to have confidence coefficient $1 - \alpha$ when $\sigma_i = \sigma$ is either known or estimable. Moreover, the interval is everywhere shorter than even the traditional interval, $X_{(i)} \pm c\sigma$, which does not maintain $1 - \alpha$ coverage in this case.

Interestingly, in one microarray data set, Qiu and Hwang (2007) found that the normal prior did not fit the data as well as a mixture of a normal prior and a point mass at zero. For the mixture prior, an empirical Bayes confidence interval for $\theta_{(i)}$ was constructed and shown (numerically and asymptotically as $p \to \infty$) to have empirical Bayes confidence coefficient at least $1 - \alpha$.

Further, combining k empirical Bayes $1 - \alpha/k$ confidence intervals for $\theta_{(i)}$, $i \in S$, where S consists of k indices of the selected $\theta_{(i)}$'s, yields a simultaneous confidence set (rectangle) that has empirical Bayes coverage probability above the nominal $1 - \alpha$ level. Furthermore, their sizes could be much smaller than even the naive rectangles (which ignore selection and hence have poor coverage). This can also lead to a more powerful test.

5.3 Shrinking Means and Variances

Thus far, we have only discussed procedures that shrink the sample means, however, confidence sets can also be improved by shrinking variances. In Section 4 we saw how to construct improved intervals for the variance. In Berry (1994) it was shown that using an improved variance estimator can slightly improve the risk of the Stein point estimator (but not the positive-part). Now we will see that we can substantially improve intervals for the mean by using improved variance estimates, when there are a large number of variances involved.

Hwang, Qiu and Zhao (2010) constructed empirical Bayes confidence intervals for θ_i where the center and the length of the interval are found by shrinking both the sample means and sample variances. They took an approach similar to He (1992), except that the task is complicated by putting yet another prior on σ_i^2. The prior assumption is that $\log \sigma_i^2$ is distributed according to a normal distribution (or σ_i^2 has an inverted gamma distribution). In both cases, their proposed double shrinkage confidence interval maintains empirical Bayes coverage probabilities above the nominal level, while the expected length are always smaller than the t-interval or the interval that only shrinks means. Simulations show that the improvements could be up to 50%.

The confidence intervals constructed are shown to have empirical Bayes confidence coefficient close to $1-\alpha$. In all the numerical studies, including extensive simulation and the application to the data sets, the double shrinkage procedure performed better than the single shrinkage intervals (intervals that shrink only one of the sample means or sample variances but not both) and the standard t interval (where there is no shrinkage).

6. DISCUSSION

The confidence sets that we have discussed broadly fall into two categories: those that are explicitly defined by a center and a radius (such as Berger, 1980, or Casella and Hwang, 1983), and those that are implicit (such as Tseng and Brown, 1997). For experimenters, the explicitly defined intervals may be slightly preferred.

The improved confidence sets typically work because they are able to reduce the volume of the x-section (the confidence set) without reducing the volume of the θ-section (the acceptance region). As the coverage probability results from the θ-section, the result is an improved set in terms of volume and coverage.

Another point to note is that most of the sets presented are based on shrinking toward zero. Moreover, the improved sets will typically have greatest coverage improvement near zero, that is, near the point to which they are shrinking. The point zero is, of course, only a convenience, as we can shrink toward any point μ_0 by translating the problem to $x-\mu_0$ and $\theta-\mu_0$, and then obtain the greatest confidence improvement when $x-\mu_0$ is small. Moreover, we can shrink toward any linear subset of the parameter space, for example, the space where the coordinates are all equal, by translating to $x-\bar{x}\mathbf{1}$ and $\theta-\bar{\theta}\mathbf{1}$, where $\mathbf{1}$ is a vector of 1s. This is developed in Casella and Hwang (1987).

The Stein effect, which was discovered in point estimation, has had far-reaching influence in confidence set estimation. It has shown us that by taking into account the structure of a problem, possibly through an empirical Bayes model, improved point and set estimators can be constructed.

APPENDIX A: PROOF OF DOMINANCE OF C^+

Hwang and Casella (1982) show that $(\partial/\partial|\theta|)P_\theta(\theta \in C^+)$ is decreasing in $|\theta|$, and hence has minimum $1-\alpha$ at $|\theta|=\infty$. The proof is somewhat complex, and only holds for $p \ge 4$. Hwang and Casella (1984) found a simpler approach, which extended the result to $p=3$. We outline that approach here.

For the set $C^+ = \{\theta : |\theta - \delta^+(\mathbf{x})| \le c\}$, the following lemma shows that we do not have to worry about $|\theta| < c$.

Lemma A.1. *For $X \sim N(\theta, I)$ and every $a > 0$ and $|\theta| < c$,*

$$P_\theta\big(\theta : |\theta - \delta^+(X)| \le c\big) \ge P_\theta(\theta : |\theta - X| \le c).$$

Proof. The assumption $|\theta| < c$ implies that $0 \in C_\theta^0$, the θ-section (acceptance region). Therefore, by the convexity of C_θ^0,

$$x \in C_\theta^0 \quad \Longrightarrow \quad \delta^+(x) \in C_\theta^0$$

since $\delta^+(x)$ is a convex combination of 0 and x. Finally, since $\delta^+(x) \in C_\theta^0$, we then have $|\delta^+(x) - \theta| \le c$ so $C_\theta^0 \in C_\theta^+$ and the theorem is proved. □

It is interesting that, even though the confidence sets, the x-sections have exactly the same volume; for small

θ the θ-section of the δ^+ procedure contains the θ-section of the usual procedure.

In addition to not needing to worry about $|\theta| < c$, there is a further simplification if $|\theta| \geq c$. If $|\theta| \geq c$, the inequality $|\theta - \delta^+(x)| \leq c$ is equivalent to

$$|\theta - \delta^+(x)| \leq c \quad \text{and} \quad |x|^2 \geq a,$$

which allows us to drop the "+." Note that if $|\theta| > c$ and $|x|^2 < a$, then $|\theta - \delta^+(x)| > c$.

Last, we note that if $a = 0$, then the two procedures are exactly the same and, thus, a sufficient condition for domination of C_x^0 by C_δ^0 is to show that

$$\text{(A.1)} \qquad \frac{d}{da} P_\theta(\theta \in C_\delta^+) > 0$$

for every $|\theta| > c$ and a in an interval including 0. The inequality (A.1) was established in Hwang and Casella (1984) through the use of the polar transformation $(x, \theta) \to (r, \beta)$, where $r = |x|$ and $x'\theta = |x||\theta| \cos(\beta)$, so β is the angle between x and θ. The polar representation of the coverage probability is differentiable in a, and the following theorem was established.

THEOREM A.2. *For $p \geq 3$, the coverage probability of C_δ^+ is higher than that of C_x^0 for every θ provided $0 < a \leq a^*$, where a^* is the unique solution to*

$$\left(\frac{c^2 + (c^2 + a^*)^{1/2}}{a^*}\right)^{p-2} e^{-c\sqrt{a^*}} = 1.$$

Solutions to this equation are easily computed, and it turns out that $a^* \approx 0.8(p-2)$, which does not quite get to the value $p - 2$, the optimal value for δ^{JS} and the popular choice for δ^+. However, the coverage probabilities are very close. Moreover, the theorem provides a sufficient condition, and it is no doubt the case that $a = p - 2$ achieves dominance.

APPENDIX B: THE STRAWDERMAN PRIOR

The first proper Bayes minimax point estimators were found by Strawderman (1971) using a hierarchical prior of the form

$$X|\theta \sim N_p(\theta, I),$$
$$\theta|\lambda \sim N_p\left(0, \frac{1-\lambda}{\lambda} I\right),$$
$$\lambda \sim (1-a)\lambda^{-a}, \quad 0 < \lambda \leq 1,\ 0 \leq a < 1.$$

The Bayes estimator for this model is $\mathrm{E}(\theta|x) = [1 - \mathrm{E}(\lambda|x)]x$. The function $\mathrm{E}(\lambda|x)$ is a bounded increasing function of $|x|$, and Strawderman was able to show, using an extension of Baranchik's (1970) result, that for $p \geq 5$ the Bayes estimator is minimax. An interesting point about this hierarchy is that the unconditional prior on θ is approximately $1/|\theta|^{p+2-2a}$, giving it t-like tails. (The prior is proper if $p + 2 - 2a \geq 6$.) These are the types of priors that lead to Bayesian posterior credible sets with good coverage probabilities.

Faith (1978) used a similar hierarchical model with $\theta \sim N(0, t^2 I)$ and $t^2 \sim$ Inverted Gamma(a, b), leading to an unconditional prior on θ of the form $\pi(\theta) \approx (2b + |\theta|^2)^{-(p/2+a)}$, the multivariate t distribution. In his unpublished Ph.D. thesis, Faith gave strong evidence that the Bayesian posterior credible sets had good coverage properties.

Berger (1980) used a generalization of Strawderman's prior, which is more tractable than the t prior of Faith, to allow for input on the covariance structure.

ACKNOWLEDGMENTS

Thanks to the Executive Editor, Editor and Referee for their careful reading and thoughtful suggestions, which improved the presentation of the material. Supported by National Science Foundation Grants DMS-0631632 and SES-0631588.

REFERENCES

BARANCHIK, A. J. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Technical Report No. 51, Department of Statistics, Stanford University.

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645. MR0253461

BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761. MR0572619

BERRY, J. C. (1994). Improving the James–Stein estimator using the Stein variance estimator. *Statist. Probab. Lett.* **20** 241–245. MR1294111

BREWSTER, J. and ZIDEK, J. (1974). Improving on equivariance estimators. *Ann. Statist.* **2** 21–38. MR0381098

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136. MR0216647

BROWN, L. D. (1968). Inadmissibility of usual estimators of scale parameters in problems with unknown location and scale. *Ann. Math. Statist.* **39** 29–42. MR0222992

BROWN, L. D., CASELLA, G. and HWANG, J. T. G. (1995). Optimal confidence sets, bioequivalence, and the limaçon of Pascal. *J. Amer. Statist. Assoc.* **90** 880–889. MR1354005

CASELLA, G. and HWANG, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78** 688–697. MR0721220

CASELLA, G. and HWANG, J. T. (1987). Employing vague prior information in the construction of confidence sets. *J. Multivariate Anal.* **21** 79–104. MR0877844

COHEN, A. (1972). Improved confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **67** 382–387. MR0312636

EFRON, B. (2006). Minimum volume confidence regions for a multivariate normal mean vector. *J. Roy. Statist. Soc. Ser. B* **68** 655–670. MR2301013

EFRON, B. and MORRIS, C. N. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

FAITH, R. E. (1976). Minimax Bayes point and set estimators of a multivariate normal mean. Unpublished Ph.D. thesis, Department of Statistics, University of Michigan.

GOUTIS, C. and CASELLA, G. (1991). Improved invariant confidence intervals for a normal variance. *Ann. Statist.* **19** 2015–2031. MR1135162

FAITH, R. E. (1978). Minimax Bayes point estimators of a multivariate normal mean. *J. Multivariate Anal.* **8** 372–379. MR0512607

HE, K. (1992). Parametric empirical Bayes confidence intervals based on James–Stein estimator. *Statist. Decisions* **10** 121–132. MR1165708

HUWANG, L. (1996). Asymptotically honest confidence sets for structured errors-in variables models. *Ann. Statist.* **24** 1536–1546. MR1416647

HWANG, J. T. (1993). Empirical Bayes estimation for the mean of the selected populations. *Sankhyā A* **55** 285–311. MR1319130

HWANG, J. T. (1995). Fieller's problem and resampling techniques. *Statist. Sinica* **5** 161–172. MR1329293

HWANG, J. T. and CASELLA, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10** 868–881. MR0663438

HWANG, J. T. and CASELLA, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Decisions* (Suppl. 1) 3–16. MR0785198

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. MR0840508

HWANG, J. T. and ULLAH, A. (1994). Confidence sets centered at James–Stein estimators. A surprise concerning the unknown variance case. *J. Econometrics* **60** 145–156. MR1247818

HWANG, J. T., QIU, J. and ZHAO, Z. (2010). Empirical Bayes confidence intervals shrinking both means and variances. *J. Roy. Statist. Soc. Ser. B*. To appear.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 311–319. Univ. California Press, Berkeley, CA. MR0133191

JOSHI, V. M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38** 1868–1875. MR0220391

JOSHI, V. M. (1969). Admissibility of the usual confidence set for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40** 1042–1067. MR0264811

KUBOKAWA, T. and SRIVASTAVA, M. S. (2003). Estimating the covariance matrix: A new approach. *J. Multivariate Anal.* **86** 28–47. MR1994720

LAIRD, N. M. and LOUIS, T. A. (1983). Empirical Bayes confidence intervals based on bootstrap.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York. MR1639875

MAATTA, J. M. and CASELLA, G. (1990). Developments in decision-theoretic variance estimation (with discussion). *Statist. Sci.* **5** 90–101. MR1054858

MORRIS, C. N. (1983). Parametric empirical Bayes inference: Theory and applications (with discussion). *J. Amer. Statist. Assoc.* **78** 47–65. MR0696849

MARUYAMA, Y. and STRAWDERMAN, W. E. (2006). A new class of minimax generalized Bayes estimators of a normal variance. *J. Statist. Plann. Inference* **136** 3822–3836. MR2299167

ROBERT, C. and CASELLA, G. (1990). Improved confidence sets in spherically symmetric distributions. *J. Multivariate Anal.* **32** 84–94. MR1035609

QIU, J. and HWANG, J. T. (2007). Sharp simultaneous confidence intervals for the means of selected populations with application to microarray data analysis. *Biometrics* **63** 767–776. MR2395714

SAMWORTH, R. (2005). Small confidence sets for the mean of a spherically symmetric distribution. *J. Roy. Statist. Soc. Ser. B* **67** 343–361. MR2155342

SHAO, P. Y.-S. and STRAWDERMAN, W. E. (1994). Improving on the James–Stein positive-part estimator. *Ann. Statist.* **22** 1517–1538. MR1311987

SHINOZAKI, N. (1989). Improved confidence sets for the mean of a multivariate distribution. *Ann. Inst. Statist. Math.* **41** 331–346. MR1006494

SHORROCK, G. (1990). Improved confidence intervals for a normal variance. *Ann. Statist.* **18** 972–980. MR1056347

STEIN, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–296. MR0148184

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160. MR0171344

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388. MR0397939

STRAWDERMAN, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.* **2** 190–198. MR0343442

TATE, R. F. and KLETT, G. W. (1959). Optimal confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **54** 674–682. MR0107926

TSENG, Y. and BROWN, L. D. (1997). Good exact confidence sets and minimax estimators for the mean vector of a multivariate normal distribution. *Ann. Statist.* **25** 2228–2258. MR1474092

THE LIST OF SOURCE ENTRIES RETRIEVED FROM MATHSCINET

The list of entries below corresponds to the Reference section of your article and was retrieved from MathSciNet applying an automated procedure. Please check the list and cross out those entries which lead to mistaken sources. Please update your references entries with the data from the corresponding sources, when applicable. More information can be found in the support page:
http://www.e-publications.org/ims/support/mrhelp.html.

Not Found!

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645. MR0253461 (40 #6676)

BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761. MR572619 (82f:62064)

BERRY, J. C. (1994). Improving the James-Stein estimator using the Stein variance estimator. *Statist. Probab. Lett.* **20** 241–245. MR1294111

BREWSTER, J. F. AND ZIDEK, J. V. (1974). Improving on equivariant estimators. *Ann. Statist.* **2** 21–38. MR0381098 (52 #1995)

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist* **37** 1087–1136. MR0216647 (35 #7476)

BROWN, L. (1968). Inadmissiblity of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist* **39** 29–48. MR0222992 (36 #6041)

BROWN, L. D., CASELLA, G., AND HWANG, J. T. G. (1995). Optimal confidence sets, bioequivalence, and the limaçon of Pascal. *J. Amer. Statist. Assoc.* **90** 880–889. MR1354005 (96d:62042)

CASELLA, G. AND HWANG, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78** 688–698. MR721220 (85g:62054)

CASELLA, G. AND HWANG, J. T. (1987). Employing vague prior information in the construction of confidence sets. *J. Multivariate Anal.* **21** 79–104. MR877844 (88a:62084)

COHEN, A. (1972). Improved confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **67** 382–387. MR0312636 (47 #1192)

EFRON, B. (2006). Minimum volume confidence regions for a multivariate normal mean vector. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 655–670. MR2301013

EFRON, B. AND MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597 (52 #9433)

Not Found!

GOUTIS, C. AND CASELLA, G. (1991). Improved invariant confidence intervals for a normal variance. *Ann. Statist.* **19** 2015–2031. MR1135162 (92m:62035)

FAITH, R. E. (1978). Minimax Bayes estimators of a multivariate normal mean. *J. Multivariate Anal.* **8** 372–379. MR512607 (81j:62023)

HE, K. (1992). Parametric empirical Bayes confidence intervals based on James-Stein estimator. *Statist. Decisions* **10** 121–132. MR1165708 (93d:62014)

HUWANG, L. (1996). Asymptotically honest confidence sets for structural errors-in-variables models. *Ann. Statist.* **24** 1536–1546. MR1416647 (98d:62050)

HWANG, J. T. (1993). Empirical Bayes estimation for the means of the selected populations. *Sankhyā Ser. A* **55** 285–304. MR1319130 (96a:62030)

HWANG, J. T. G. (1995). Fieller's problems and resampling techniques. *Statist. Sinica* **5** 161–171. MR1329293 (96c:62084)

HWANG, J. T. AND CASELLA, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10** 868–881. MR663438 (83m:62019)

HWANG, J. T. AND CASELLA, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Decisions* suppl. 1 3–16. Recent results in estimation theory and related topics. MR785198 (86j:62123)

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. MR840508 (87i:62067)

HWANG, J. T. G. AND ULLAH, A. (1994). Confidence sets centered at James-Stein estimators: a surprise concerning the unknown-variance case. *J. Econometrics* **60** 145–156. MR1247818 (94k:62113)

Not Found!

JAMES, W. AND STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*. Univ. California Press, Berkeley, Calif., 361–379. MR0133191 (24 #A3025)

JOSHI, V. M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38** 1868–1875. MR0220391 (36 #3451)

JOSHI, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40** 1042–1067. MR0264811 (41 #9402)

KUBOKAWA, T. AND SRIVASTAVA, M. S. (2003). Estimating the covariance matrix: a new approach. *J. Multivariate Anal.* **86** 28–47. MR1994720 (2004e:62015)

Not Found!

LEHMANN, E. L. AND CASELLA, G. (1998). *Theory of point estimation*, Second ed. Springer Texts in Statistics. Springer-Verlag, New York. MR1639875 (99g:62025)

MAATTA, J. M. AND CASELLA, G. (1990). Developments in decision-theoretic variance estimation. *Statist. Sci.* **5** 90–120. With comments and a rejoinder by the authors. MR1054858 (91h:62008)

MORRIS, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.* **78** 47–65. With discussion. MR696849 (84e:62015)

MARUYAMA, Y. AND STRAWDERMAN, W. E. (2006). A new class of minimax generalized Bayes estimators of a normal variance. *J. Statist. Plann. Inference* **136** 3822–3836. MR2299167 (2008e:62029)

ROBERT, C. AND CASELLA, G. (1990). Improved confidence sets for spherically symmetric distributions. *J. Multivariate Anal.* **32** 84–94. MR1035609 (91h:62053)

QIU, J. AND HWANG, J. T. G. (2007). Sharp simultaneous confidence intervals for the means of selected populations with application to microarray data analysis. *Biometrics* **63** 767–776. MR2395714

SAMWORTH, R. (2005). Small confidence sets for the mean of a spherically symmetric distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 343–361. MR2155342

SHAO, P. Y.-S. AND STRAWDERMAN, W. E. (1994). Improving on the James-Stein positive-part estimator. *Ann. Statist.* **22** 1517–1538. MR1311987 (95k:62064)

SHINOZAKI, N. (1989). Improved confidence sets for the mean of a multivariate normal distribution. *Ann. Inst. Statist. Math.* **41** 331–346. MR1006494 (90j:62081)

SHORROCK, G. (1990). Improved confidence intervals for a normal variance. *Ann. Statist.* **18** 972–980. MR1056347 (91j:62039)

STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–296. MR0148184 (26 #5692)

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160. MR0171344 (30 #1575)

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR630098 (83a:62080)

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388. MR0397939 (53 #1794)

STRAWDERMAN, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.* **2** 190–198. MR0343442 (49 #8183)

TATE, R. F. AND KLETT, G. W. (1959). Optimal confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **54** 674–682. MR0107926 (21 #6648)

TSENG, Y.-L. AND BROWN, L. D. (1997). Good exact confidence sets for a multivariate normal mean. *Ann. Statist.* **25** 2228–2258. MR1474092 (98m:62081)

META DATA IN THE PDF FILE

Following information will be included as pdf file Document Properties:

Title : Shrinkage Confidence Procedures
Author : George Casella, J. T. Gene Hwang
Subject : Statistical Science, 2010, Vol.0, No.00, 1-13
Keywords: Stein effect, coverage probability, empirical Bayes
Affiliation:

THE LIST OF URI ADRESSES

Listed below are all uri addresses found in your paper. The non-active uri addresses, if any, are indicated as ERROR. Please check and update the list where necessary. The e-mail addresses are not checked – they are listed just for your information. More information can be found in the support page:
http://www.e-publications.org/ims/support/urihelp.html.

```
200 http://www.imstat.org/sts/ [2:pp.1,1] OK
200 http://dx.doi.org/10.1214/10-STS319 [2:pp.1,1] OK
200 http://www.imstat.org [2:pp.1,1] OK
--- mailto:casella@stat.ufl.edu [2:pp.1,1] Check skip
--- mailto:hwang@math.cornell.edu [2:pp.1,1] Check skip
```