Improved Confidence Sets
in Spherically Symmetric
Distributions

by

Christian Robert[1]
Purdue University and Université de Rouen

George Casella[2]
Cornell University and Purdue University

Technical Report # 87-51

Department of Statistics
Purdue University

November, 1987

△

IMPROVED CONFIDENCE SETS
IN SPHERICALLY SYMMETRIC
DISTRIBUTIONS.

by

Christian Robert and George Casella
Purdue University and Cornell University

ABSTRACT

The usual confidence set for a multivariate mean vector can be improved upon by recentering the set at a Stein-type estimator: this fact is known to be true under many different distributional assumptions. Thus far, however, the case of unknown variance has not been dealt with analytically. In this paper we prove that recentered set estimators dominate the usual set estimator when sampling is from any of a class of spherically symmetric distributions with unknown variance.

I. Introduction.

The problem of estimating the mean of a spherically symmetric distribution (s.s.d.) has begun to receive much attention recently. In particular, the work of Hwang and Chen (1986), in set estimation, and Cellier, Fourdrinier and Robert (CFR) (1987), in point estimation, has greatly added to our knowledge of the problem. This paper is, in a sense, a synthesis of the previously mentioned ones. We adapt the techniques of Hwang and Chen (1986) (which, themselves, are adaptions of the technique of Hwang and Casella, 1984), to the more general case of s.s.d. distributions with unknown variance, as considered by CFR in the point estimation case.

Obtaining results that are valid for a class of s.s.d. distributions has important practical implications in that an experimenter is freed from the restrictive normality assumption. Instead, the experimenter need only assume that sampling is from any of a class of s.s.d. distributions.

The general case we consider is that of $Z = (X', Y')'$, an observation from a $p + \nu$ dimensional arbitrary s.s.d. with location parameter $(\theta', 0)'$ and dispersion matrix $\sigma^2 I_{p+\nu}$. The dimensionality of X and θ are both p (assumed ≥ 3). For the point estimation problem with σ^2 unknown, CFR have shown that a rather large class of shrinkage estimators, generalizing those of Judge and Bock (1978), was minimax for every s.s.d. In particular, the positive-part James-Stein estimator

$$\delta^+(Z) = \left(1 - \frac{a||Y||^2}{||X||^2}\right)^+ X \qquad (1.1)$$

is minimax for every s.s.d. if $a \leq 2 \, (p - 2)$.

It is then tempting to try to show that this robustness of shrinkage estimators carries over to the case of confidence intervals. Although Casella and Hwang (1983, 1987), and Hwang and Casella (1982, 1984) have very general results in the normal case with σ^2 known, they only have numerical evidence that the usual confidence interval can be dominated in the unknown variance case (see Casella and Hwang (1987)). For a large class of s.s.d., including the usual ones (normal, multivariate t, double exponential), Hwang and Chen (1986) have established some sufficient conditions on a (in 1.1) for the domination

of the usual set estimator in the known variance case. (These conditions depend on the s.s.d.)

In this paper, our aim is therefore more modest: we establish that, for a given class of s.s.d. (which includes the multivariate t), there exists a_o such that, for $a\epsilon(0, a_o]$, the usual confidence set can be dominated, no matter which s.s.d. in the class is sampled. Unfortunately, this class does not contain the normal distribution with unknown variance, a point that we discuss in more detail in the last section of the paper. Recent work by Kim (1987), however, has made some progress on this problem.

II. Domination in the Class of Spherically Symmetric Distributions.

When sampling from a s.s.d., the usual confidence set, based on the F-distribution, maintains its coverage probability no matter what s.s.d. is sampled. This follows from the results of Kelker (1970) on the characterizations of s.s.d.'s. A direct consequence of his work is the following result:

Proposition 1. *Let $F_{\alpha,p,\nu}$ be the $(1-\alpha)$-quantile of an F-distribution with p and ν degrees of freedom. Let $Z = (X', Y')'$ be a $(p+\nu)$ vector following an arbitrary s.s.d. with location parameter $(\theta', 0)'$ and dispersion matrix $\sigma^2 I_{p+\nu}$. Then*

$$C^\circ = \left\{ \theta : ||\theta - X||^2 \leq \frac{p}{\nu} F_{\alpha,p,\nu} ||Y||^2 \right\} \tag{2.1}$$

is a $(1-\alpha)$-confidence interval.

Proof: From Theorem 11 of Kelker (1970), it follows that $\frac{||\theta-X||^2/p}{||Y||^2/\nu}$ has an F-distribution with parameters p and ν, independently of the s.s.d. $\square$

Note that this result is true whether or not σ^2 is known.

Using (1.1), we construct the recentered set estimator

$$C^+ = \left\{ \theta : ||\theta - \delta^+(Z)||^2 \leq \frac{p}{\nu} F_{\alpha,p,\nu} ||Y||^2 \right\}. \tag{2.2}$$

It is clear, by construction, that Volume (C^+) = Volume (C°). Thus, in order to establish dominance, we need to show dominance in coverage probability; that $P_\theta(\theta \epsilon C^+) \geq P_\theta(\theta \epsilon C^\circ)$ for all θ.

We have

$$P_\theta(\theta \epsilon C^+) = \int_{C^+} f(||x - \theta||^2 + ||y||^2) dz,$$

where f is the density of Z. Thus, defining $k^2 = \frac{p}{\nu} F_{\alpha,p,\nu}$, we can write

$$P_\theta(\theta \epsilon C^+) = \int_{\mathbf{R}^\nu} \int_{\{x \epsilon \mathbf{R}^P : ||\theta - \delta^+(z)||^2 \leq k^2 ||y||^2\}} f(||x - \theta||^2 + ||y||^2) dx\, dy.$$

Following the argument in Theorem 2.1 of Hwang and Casella (1982), we have

$$\int_{\frac{||\theta||}{k}}^{+\infty} (2\pi)^{\nu-1} s^{\nu-1} \int_{\{x : ||\theta - \delta^+(z)||^2 \leq k^2 s^2\}} f(||x - \theta||^2 + s^2) dx\, ds$$

4

$$= \int_{\frac{||\theta||}{k}}^{+\infty} (2\pi)^{\nu-1} P_\theta(||\theta - \delta^+(Z)||^2 \le k^2 s^2 | s) s^{\nu-1} \phi(s) ds$$

$$\ge \int_{\frac{||\theta||}{k}}^{+\infty} (2\pi)^{\nu-1} P_\theta(||\theta - X||^2 \le k^2 s^2 | s) s^{\nu-1} \phi(s) ds, \qquad (2.3)$$

where $s = ||y||$ and ϕ is its density function. In fact, due to the convexity of C°, $\{x : ||\theta - \delta^+(z)||^2 \le k^2 s^2\}$ contains $\{x : ||\theta - x||^2 \le k^2 s^2\}$ if $||\theta||/s < k$. Integration over s^2 then gives inequality (2.3).

Therefore, we only need to consider the integral

$$I(a) = \int_o^{\frac{||\theta||}{k}} (2\pi)^{\nu-1} s^{\nu-1} \int_{\{x : ||\theta - \delta^+(z)||^2 \le k^2 s^2\}} f(||x - \theta||^2 + s^2) dx \, ds, \qquad (2.4)$$

and establish conditions under which it is an increasing function of a. At $a = 0, I(a)$ is equal to the integral obtained for the usual set estimator (2.1). Thus, if we establish a range of a for which $I(a) > I(0)$ this, together with (2.3), will establish dominance of C^+ over C°.

Using the notation

$$\alpha(r) = r^2 - 2||\theta|| r \cos \beta + ||\theta||^2,$$

$$r_\pm^\circ = ||\theta|| \cos \beta \pm \sqrt{k^2 s^2 - ||\theta||^2 \sin^2 \beta},$$

$$r_\pm = (r_\pm^\circ + \sqrt{(r_\pm^\circ)^2 + 4as^2})/2 \qquad (2.5)$$

$$\sin \beta_\circ = \frac{ks}{||\theta||} \quad \text{and} \quad 0 \le \beta_\circ < \frac{\pi}{2},$$

we can write

$$I(a) = \Omega \int_o^{\frac{||\theta||}{k}} (2\pi)^{\nu-1} s^{\nu-1} \int_o^{\beta_\circ} \int_{r_-}^{r_+} r^{p-1} \sin^{p-2}(\beta) f(\alpha(r) + s^2) dr d\beta ds,$$

where Ω is a positive constant, and can be ignored. Differentiating $I(a)$, we obtain

$$\frac{\partial}{\partial a}(I(a)) \propto \int_o^{\frac{||\theta||}{k}} s^{\nu-1} \int_o^{\beta_\circ} \sin^{p-2} \beta \left[r_+^{p-1} f(\alpha(r_+) + s^2) \frac{\partial}{\partial a}(r_+) \right.$$

$$\left. - r_-^{p-1} f(\alpha(r_-) + s^2) \frac{\partial}{\partial a}(r_-) \right] d\beta ds \qquad (2.6)$$

$$= \int_o^{\frac{||\theta||}{k}} s^{\nu+1} \int_o^{\beta_\circ} \sin^{p-2} \beta \left[\frac{r_+^p f(\alpha(r_+) + s^2)}{r_+^2 + as^2} - \frac{r_-^p f(\alpha(r_-) + s^2)}{r_-^2 + as^2} \right] d\beta ds$$

5

by simple algebraic manipulations.

Define $f_1(\cdot)$ by $f(t) = f_1(\frac{t}{\sigma^2})$ $\forall$ t. Using the new notation

$$\xi = \frac{\theta}{\sigma} \quad , \quad \omega = \frac{s}{\sigma}$$

and modifying the old ones accordingly as

$$r_{\pm}^{\circ} = \|\xi\| \cos \beta \pm \sqrt{k^2\omega^2 - \|\xi\|^2 \sin^2\beta},$$

$$r_{\pm} = (r_{\pm}^{\circ} + \sqrt{r_{\pm}^{\circ 2} + 4a\omega^2})/2,$$

$$\alpha(r) = r^2 - 2\|\xi\|r \cos \beta + \|\xi\|^2, \quad \text{and} \quad \sin \beta_{\circ} = \frac{k\omega}{\|\xi\|},$$

we see that $\frac{\partial}{\partial a}I(a)$ is proportional to

$$\int_0^{\|\xi\|/k} \omega^{\nu+1} \int_0^{\beta_{\circ}} \sin^{p-2}\beta \left[\frac{r_+^p f_1(\alpha(r_+) + \omega^2)}{r_+^2 + a\omega^2} - \frac{r_-^p f_1(\alpha(r_-) + \omega^2)}{r_-^2 + a\omega^2} \right] d\beta \, d\omega.$$

As we want a condition uniform in $\|\theta\|$, we have only to consider the case $\sigma^2 = 1$.

To show that

$$\frac{r_+^p f_1(\alpha(r_+) + \omega^2)}{r_+^2 + a\omega^2} > \frac{r_-^p f_1(\alpha(r_-) + \omega^2)}{r_-^2 + a\omega^2}, \tag{2.7}$$

for each β and ω^2, and hence that dominance can be attained, we can apply the result of Hwang and Chen (1986). A sufficient condition for (2.7) to hold is

$$\inf_{\alpha_0 \leq v \leq \alpha_1} \frac{f_1'(v + \omega^2)}{f_1(v + \omega^2)} \geq -\frac{p - 2}{2\omega^2 k\sqrt{a}} \ell n \frac{k + \sqrt{k^2 + a}}{\sqrt{a}} \tag{2.8}$$

where

$$\begin{cases} \alpha_0 = (k - \sqrt{a})^2 \omega^2 \\ \alpha_1 = (k^2 + a)\omega^2 \end{cases}$$

Therefore, we have established

Theorem 2.1: *If a is such that*

$$\inf_{\omega} \left\{ \omega^2 \inf_{\alpha_0 \leq v \leq \alpha_1} \frac{f_1'(v + \omega^2)}{f_1(v + \omega^2)} \right\} \geq -\frac{p - 2}{2k\sqrt{a}} \ell n \left(\frac{k + \sqrt{k^2 + a}}{\sqrt{a}} \right),$$

then C^+ dominates C° for every θ.

6

III. Discussion

It is clear from expression (2.8) that there is a restriction on the *s.s.d.*s to which Theorem 2.1 applies. In particular, the theorem can be applied only if the first term of the inequality is finite, i.e. if

$$\omega^2 \inf_{\alpha_o \leq v \leq \alpha_1} \frac{f_1'(v + \omega^2)}{f_1(v + \omega^2)} \tag{3.1}$$

is bounded from below for every ω^2. In the normal case, $\frac{f_1'(t)}{f_1(t)} = -\frac{1}{2}$; therefore we cannot apply Theorem 2.1 to show dominance in this case.

For the double-exponential distribution, we have

$$f(z|\theta, \sigma^2) \alpha \exp\left\{-\frac{\|z - \theta\|}{\sigma}\right\}$$

and thus $f_1(t) \propto \exp\left(-\sqrt{t}\right)$. Therefore, we get

$$\frac{f_1'(t)}{f_1(t)} = -\frac{1}{2\sqrt{t}} \quad \text{and} \quad \frac{f_1'(v + \omega^2)}{f_1(v + \omega^2)} = -\frac{1}{2} \frac{1}{\sqrt{(v + \omega^2)}}$$

Thus (3.1) cannot be bounded from below. Similar to the normal distribution, the double exponential has tails that are not flat enough to satisfy to this condition. We will see below that the multivariate $-t$ is a limiting case with respect to this criterion.

Consider now the multivariate-t distribution with parameters θ, σ^2 and N (degrees of freedom):

$$f(z|\theta, \sigma^2, N) \propto \left\{1 + \frac{1}{N\sigma^2}(\|x - \theta\|^2 + \|y\|^2)\right\}^{-(N+p+\nu)/2} \tag{3.2}$$

For this distribution we have

$$\frac{f_1'(t)}{f_1(t)} = -\frac{N + p + \nu}{2(N + t)} \quad ,$$

an increasing function of t. Thus

$$\inf_{\alpha_o \leq v \leq \alpha_1} \frac{f'(v + \omega^2)}{f(v + \omega^2)} = -\frac{N + p + \nu}{2(N + \alpha_o + \omega^2)} = -\frac{N + p + \nu}{2(N + (1 + (k - \sqrt{a})^2)\omega^2)},$$

and

$$\inf_\omega \left\{-\frac{\omega^2(N + p + \nu)/2}{N + (1 + (k - \sqrt{a})^2)\omega^2}\right\} = -\frac{N + p + \nu}{2(1 + (k - \sqrt{a})^2)}.$$

7

We now just have to solve the equation

$$\frac{N+p+\nu}{1+(k-\sqrt{a})^2} = \frac{p-2}{k\sqrt{a}}\ell n\left(\frac{k+\sqrt{k^2+a}}{\sqrt{a}}\right) \qquad (3.3)$$

to get an upper bound a_o. And, applying Theorem 2.1 for every $a\epsilon[0,a_o], C^+$ dominates C°.

Note that, as $N \to \infty$, the solution to (3.3) goes to zero, again showing that we cannot deduce anything for the normal distribution with unknown variance. The reason why the normal case cannot be covered by Theorem 2.1 is that it cannot be handled by any technique that operates conditionally on S^2. That is, in order to prove dominance in the normal case one cannot work with the square-bracketed expression in (2.6), but rather with the entire integrand. This present an analytical problem of great difficulty, as the entire integrand is extremely unwieldy.

The fact that the proof conditional on S^2 would not work for the normal is contained in the results of Hwang (1985), which is concerned with universal domination of point estimators. Informally speaking, one estimator will universally dominate another only if its confidence sets dominate for all confidence levels. This can be translated in domination for all S^2, and applying Theorem 3.7 of Hwang (1985), we see that it is possible to dominate the usual confidence set when sampling from a multivariate t. Hwang's Theorem 3.10, however, shows that a proof conditional on S^2 cannot succeed in the normal case.

Table 1 gives some values of a_o for $\alpha = 0.05$. As one can see, the obtained values are far from the 'optimal choice'

$$a^* = \frac{\nu}{\nu+2}(p-2) \qquad (3.4)$$

The bounds given in Table 1 are logically decreasing functions of ν and N since, as ν or N goes to infinity, the multivariate$-t$ converges toward a normal distribution. Although the decrease is not very rapid in N, if we compare our values with those of Hwang and Chen (Table 1, 1986), we can see that our values are much more smaller: working conditionally on S^2 gives restrictive conditions.

Hwang and Chen (1986) also note that the general solution for a_o yields a rather small

upper bound, one that can be improved upon in special cases. It is clear that their result (Theorem 3.1) can be generalized in the following way.

Proposition 3.1 *If*

(a) $g = \ell n f_1$ *is convex,*

(b) $\{t \geq 1; \frac{p-2}{t} + as^2 g'(s^2 a(t - t^{-1}) + (1 + (k - \sqrt{a})^2)s^2)(1 + t^2) \geq 0\}$ *is an interval (possibly degenerate),*

(c) $g(as^2(t^* - \frac{1}{t^*}) + (1 + (k - \sqrt{a})^2)s^2) - g((1 + (k - \sqrt{a})^2)s^2) \geq -(p - 2)\ell n\ t^*$, *with* $t^* = \frac{k+\sqrt{k^2+a}}{\sqrt{a}}$, *for every* s^2,

then C^+ *dominates* C° *for all* θ.

In the *normal case*, we have $g(t) = -\frac{t}{2}$. Even if conditions *(a)* and *(b)* of Proposition 3.1 are satisfied, we have

$$g(as^2(t^* - t^{*-1}) + (1 + (k - \sqrt{a})^2)s^2) - g((1 + (k - \sqrt{a})^2)s^2) = -as^2\frac{t^* - t^{*-1}}{2}$$

which cannot be bounded from below.

For the *double-exponential distribution*, we get the same conclusion as before ($g(t) = \ell n M - \sqrt{t}$ where M is a normalizing constant).

Consider now the *multivariate - t distribution*; we have (see (3.2))

$$g(t) = -\frac{N + p + \nu}{2}\ell n\left(1 + \frac{t}{N}\right)$$

It has been proved in Hwang and Chen (1986) that conditions *(a)* and *(b)* are satisfied for this function. We have

$$g((a(t^* - \frac{1}{t^*}) + 1 + (k - \sqrt{a})^2)s^2) - g((1 + (k - \sqrt{a})^2)s^2)$$

$$= -\frac{N + p + \nu}{2}\ell n\left\{1 + \sqrt{a}s^2\frac{k + \sqrt{k^2 + a})^2 - a}{(k + \sqrt{k^2 + a})(N + (1 + (k - \sqrt{a})^2)s^2)}\right\}$$

As this expression is a decreasing function of s^2, we get the sufficient condition

$$\frac{N + p + \nu}{2(p - 2)}\ell n\left(1 + \sqrt{a}\frac{(k + \sqrt{k^2 + a})^2 - a}{(k + \sqrt{k^2 + a})(1 + (k - \sqrt{a})^2)}\right) = \ell n\left(\frac{k + \sqrt{k^2 + a}}{\sqrt{a}}\right) \tag{3.5}$$

9

Therefore we have established

Theorem 3.2. *If $a \epsilon (0, a_o)$ where a_o is solution of (3.5), C^+ dominates C° for multivariate $-t$ distribution with N degrees of freedom.*

The bounds obtained in Table 2 are greater than their counterparts in Table 2. Yet they still remain significantly inferior to the "optimal" bound for point estimates, given in (3.4). This fact definitely shows the need for methods which do not work conditionally on S^2 but which, roughly speaking, stay "inside the integrals".

N = 2

ν p	5	10	20	25
5	0.50	0.19	0.09	0.07
10	1.40	0.49	0.23	0.19
20	3.21	1.06	0.52	0.39
30	5.02	1.60	0.70	0.56
60	10.42	3.40	1.31	1.03

N = 5

ν p	5	10	20	25
5	0.40	0.16	0.08	0.07
10	1.23	0.44	0.21	0.18
20	2.97	0.99	0.45	0.36
30	4.76	1.53	0.67	0.54
60	10.14	3.12	1.28	1.00

N = 10

ν p	5	10	20	25
5	0.3	0.12	0.07	0.06
10	1.02	0.37	0.19	0.16
20	2.65	0.89	0.41	0.34
30	4.37	1.41	0.63	0.50
60	9.68	2.98	1.22	0.96

Table 1. Solutions of (3.3) for $\alpha = 0.05$.

N = 2

p \ ν	5	10	20	25
5	0.74	0.25	0.11	0.09
10	2.41	0.79	0.33	0.29
20	6.00	1.91	0.8	0.62
30	9.68	3.02	1.24	0.96
60.	20.80	6.32	2.52	1.94

N = 5

p \ ν	5	10	20	25
5	0.58	0.21	0.10	0.08
10	2.03	0.68	0.30	0.24
20	5.42	1.73	0.73	0.57
30	8.98	2.81	1.17	0.91
60	19.84	6.05	2.42	1.87

N = 10

p \ ν	5	10	20	25
5	0.41	0.16	0.08	0.07
10	1.60	0.56	0.26	0.21
20	4.66	1.51	0.65	0.52
30	8.03	2.56	1.06	0.83
60	18.64	5.72	2.32	1.80

Table 2. Solutions of (3.4) for $\alpha = 0.05$.

References

[1] Casella G., Hwang, J.T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution, *J.A.S.A.* 78, 688-698.

[2] Casella G., Hwang, J.T. (1987). Employing Vague Prior Information in the Construction of Confidence Sets, *J. Mult. Anal.* 21(1), 79-104.

[3] Cellier D., D. Fourdrinier, C. Robert (1987). Estimateurs à rétrécisseur du paramètre de position d'une loi à symétrie sphérique *Note aux C.R.A.S.* 304, série I, 439-442.

[4] Hwang J.T. (1985). Universal Domination and Stochastic Domination: Estimation Simultaneously under a Broad Class of Loss Functions *Ann. Stat.* 13(1), 295-314.

[5] Hwang J.T., Casella, G. (1982). Minimax Confidence Sets for the Mean of a Multivariate Normal Distribution. *Ann. Stat.* 10(3), 868-881.

[6] Hwang J.T., Casella, G. (1984). Improved Set Estimators for a Multivariate Normal Mean. *Stat. Dec.* Supplement Issue 1, 3-16.

[7] Hwang J.T., Chen, J. (1986). Improved Confidence Sets for the Coefficients of a Linear Model with Spherically Symmetric Errors. *Ann. Stat.* 14(2), 444-460.

[8] Judge G., M.E. Bock (1978). The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics, North-Holland.

[9] Kelker D. (1970). Distribution Theory of Spherical Distributions and a Location-scale Parameter Generalization. *Sankhyā A* 32, 419-430.

[10]   Kim, P.T. (1987). Recentered Confidence Sets for the Mean of a Multivariate Distribution when the Scale Parameter is Unknown. Ph.D. Thesis, Department of Mathematics, University of California at San Diego.