Minimum Volume Confidence Regions for a Multivariate Normal Mean Vector

Bradley Efron

Abstract

Since Stein's original proposal in 1962, a series of papers have constructed confidence regions of smaller volume than the standard spheres for the mean vector of a multivariate normal distribution. A general approach to this problem is developed here, and used to calculate a lower bound on the attainable volume. Bayes and fiducial methods are involved in the calculation. Scheffe-type problems are used to show that low volume by itself does not guarantee favorable inferential properties.

Key Words James-Stein estimator, Fisher-von Mises distribution, non-central chi, Scheffe intervals

1. Introduction Stein (1962) conjectured and heuristically demonstrated confidence regions for a multivariate normal mean vector having everywhere smaller volume than the standard ones. A variety of ingenious constructions has verified Stein's conjecture, including those of Faith (1976), Berger (1980), Casella and Hwang (1983), Tseng and Brown (1997), and Samworth (2005). This paper concerns a general approach to multivariate normal mean confidence regions, a bound on the minimally attainable volume, and some related inferential questions.

Let $\mathbf{x}$ be an n-dimensional normal vector with mean $\boldsymbol{\mu}$ and covariance matrix the identity,

$$\mathbf{x} \sim N(\boldsymbol{\mu}, I), \tag{1.1}$$

and define r and ρ as the ordinary Euclidean length of $\mathbf{x}$ and $\boldsymbol{\mu}$,

$$r = \|\mathbf{x}\| \text{ and } \rho = \|\boldsymbol{\mu}\|. \tag{1.2}$$

The standard level α confidence region for $\boldsymbol{\mu}$ given $\mathbf{x}$ is the sphere

$$\mathcal{C}^0_{\mathbf{x}} = \{\boldsymbol{\mu} : \|\boldsymbol{\mu} - \mathbf{x}\| \le \chi^{(\alpha)}_n\}, \tag{1.3}$$

where $\chi^{(\alpha)}_n$ indicates the upper α quantile of r when $\rho = 0$. Figure 1 indicates $\mathcal{C}^0_{\mathbf{x}}$ for $\alpha = 0.95, n = 10$, and $r = 6$, with $\mathbf{x} = (6, 0, 0, \cdots, 0)'$. The vertical axis shows the first coordinate direction, with the nine perpendicular directions represented horizontally; the full sphere is obtained by rotating the "Standard" half-circle around the vertical axis. All regions considered here will enjoy this form of *rotational symmetry*.

Four regions with smaller volumes than $\mathcal{C}^0_{\mathbf{x}}$ are illustrated in the same way, (three with their profiles to the left for graphical clarity). "Tseng-Brown" is an oblate spheroid constructed according to formula (2) of Tseng and Brown (1997), with constants A and B taken from their Table 2. "Casella-Hwang" is a spherical region centered at the positive-part James-Stein estimate "js", $[1 - (n-2)/r^2]_+\mathbf{x}$, with radius from formula (4.7) of Casella and Hwang (1983). Their formula gives conservative coverage greater than .95, so "Adjusted" has a reduced radius mentioned in Section 3. Remark F of Section 5 describes the Tseng-Brown and Casella-Hwang constructions.

"MinVolume" represents a confidence region obtaining, in a sense to be described, a lower bound on the volume of any reasonable rotationally symmetric .95 region. The term *reasonable* is necessary because the Neyman-Pearson construction by itself does not constrain a region's size for any particular observed $\mathbf{x}$. We might, for example, achieve *empty* regions

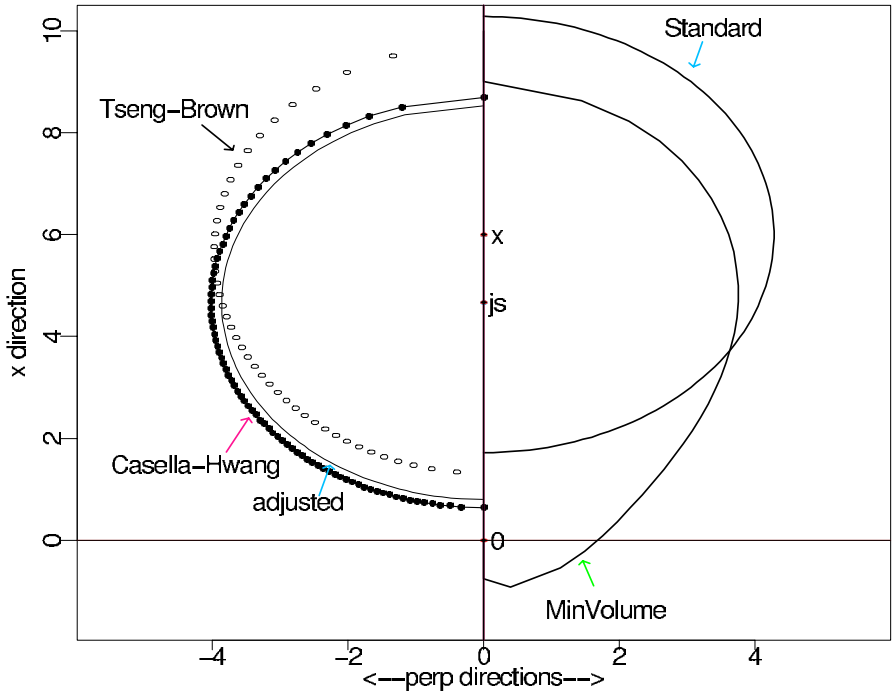

Figure 1: .95 *confidence regions for normal mean vector* $\boldsymbol{\mu}$ *having observed* $\mathbf{x} = (6, 0, 0, \cdots, 0)'$ *in dimension* $n = 10$*; full regions obtained by rotating profile curves around vertical axis. Volumes in units of* 10^6*: Standard 5.24 Tseng-Brown 2.64, Casella-Hwang 2.84, Adjusted CH 1.89, MinVolume 1.65. The Casella-Hwang regions are spheres centered at the James-Stein estimate "*js*".*

for r in $[5.99, 6.01]$ by slightly expanding $\mathcal{C}_x^0$ for other values of r. The Bayesian and fiducial arguments of Section 3 are intended to suppress this kind of special pleading.

Samworth (2005) discusses this last point – ensuring reasonable coverage probabilities for any given observation $\mathbf{x}$ – for regions that generalize and extend the Casella-Hwang construction beyond model (1.1). He provides references to the relevant literature on conditionality, such as Robinson (1979), as well as a more complete bibliography for the reduced volume literature. Another type of minimum volume problem, where conditionality is not a concern, is solved in Brown, Casella, and Hwang (1995).

Reduced volume is an intuitively appealing property but it is not obvious that the alternative regions of Figure 1 improve on $\mathcal{C}_{\mathbf{x}}^0$ for the purposes of statistical inference. Section 4 considers two Scheffe-type inference problems, with mixed conclusions as to the advantages of smaller volume. This paper is *not* an argument in favor of Figure 1's MinVolume region, which in fact does not work well in the Scheffe situations. It does provide a useful lower bound on attainable volume, showing for example that the adjusted Casella-Hwang region comes impressively close to the minimum. A reasonable conclusion would be that volume

reduction has gone about as far as it can, and that increased attention should be paid to specific inferential properties.

The paper concludes in Section 5 with derivations of the previous results as well as some further remarks.

2. A General Construction This paper follows the literature in only considering confidence regions having rotational symmetry around the axis determined by $\mathbf{x}$, as indicated in Figure 1. Rotational symmetry allows us to reduce vectorial situation (1.1) to the scalar problem of finding confidence intervals for ρ based on r, (1.2), as shown next. The rationale for Tseng-Brown and Casella-Hwang constructions becomes more transparent in the (ρ, r) framework.

The distribution of $r = \|\mathbf{x}\|$ given $\rho = \|\boldsymbol{\mu}\|$ is "non-central chi", say $r \sim \chi_n(\rho)$, the square root of a non-central chi-squared distribution $\chi^2_n(\rho^2)$. We denote the density of r given ρ by

$$f_\rho(r), \quad r \geq 0 \tag{2.1}$$

with $F_\rho(r)$ as the corresponding cumulative distribution function (cdf).

Suppose $a_\rho(r)$ is a function satisfying

$$0 \leq a_\rho(r) \leq 1 \tag{2.2}$$

for all non-negative ρ and r, with

$$\int_0^\infty a_\rho(r) f_\rho(r) dr = \alpha \qquad \text{for all } \rho \geq 0. \tag{2.3}$$

Then $a_\rho(r)$ defines level α randomized confidence intervals $\mathcal{I}(r)$ for ρ given r : value ρ exists in $\mathcal{I}(r)$ with probability $a_\rho(r)$. Non-randomized intervals have $a_\rho(r)$ equal 0 or 1. For instance the usual two-sided 0.95 intervals are defined in terms of the quantiles of $\chi_n(\rho)$ by

$$a_\rho(r) = \begin{cases} 1 & r \in [\chi_n(\rho)^{(.025)}, \chi_n(\rho)^{(.975)}] \\ & \text{if} \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

We will call $a_\rho(r)$ satisfying (2.2)-(2.3) an "inclusion function".

Every inclusion function defines a *non-randomized* set of level α confidence regions for $\boldsymbol{\mu}$ given $\mathbf{x} \sim N(\boldsymbol{\mu}, I)$. The construction, which figures implicitly in Casella and Hwang

(1983), begins with the *Fisher-von Mises distribution* for the angle "γ" between $\mathbf{x}$ and $\boldsymbol{\mu}$ (conditional on both $\boldsymbol{\mu}$ and r). In terms of $z = \cos(\gamma)$, the Fisher-von Mises density is

$$c_{r\rho} e^{r\rho z} (1 - z^2)^{\frac{n-3}{2}} \qquad \text{for } -1 \le z \le 1, \tag{2.5}$$

see Remark A of Section 5. The constant

$$c_{r\rho} = \left[\int_{-1}^{1} e^{r\rho z} (1 - z^2)^{\frac{n-3}{2}} dz \right]^{-1} \tag{2.6}$$

can be expressed in terms of modified Bessel functions, as in Downs (1966).

For γ an angle in $[0, \pi]$ define $\mathcal{S}_{\boldsymbol{\mu}}(\gamma, r)$ as the spherical cap of angular radius γ centered at $r \cdot (\boldsymbol{\mu}/\rho)$, (the multiple of $\boldsymbol{\mu}$ having length r),

$$\mathcal{S}_{\boldsymbol{\mu}}(\gamma, r) = \{\mathbf{x} : \|\mathbf{x}\| = r \quad \text{and} \quad \frac{\mathbf{x}'\boldsymbol{\mu}}{r\rho} \ge \cos(\gamma)\}. \tag{2.7}$$

(Notice that $\mathbf{x}'\boldsymbol{\mu}/r\rho$ is the cosine of the angle between $\mathbf{x}$ and $\boldsymbol{\mu}$.) The conditional probability content of $\mathcal{S}_{\boldsymbol{\mu}}(\gamma, r)$ given $\boldsymbol{\mu}$ and r is

$$\text{Prob}_{\boldsymbol{\mu}}\{\mathbf{x} \in \mathcal{S}_{\boldsymbol{\mu}}(\gamma, r) | r\} = \int_{\cos(\gamma)}^{1} c_{r\rho} e^{r\rho z} (1 - z^2)^{\frac{n-3}{2}} dz \tag{2.8}$$

according to (2.5), increasing from 0 to 1 as γ increases from 0 to π. For any value of the inclusion function $a_\rho(r)$ there is a unique angle $\gamma_\rho(r)$, with cosine $z_\rho(r)$, such that

$$a_\rho(r) = \int_{z_\rho(r)}^{1} c_{r\rho} e^{r\rho z} (1 - z^2)^{\frac{n-3}{2}} dz, \tag{2.9}$$

i.e. such that $\mathcal{S}_{\boldsymbol{\mu}}(\gamma_\rho(r), r)$ has conditional probability content $a_\rho(r)$.

We can now define a level α acceptance region $\mathcal{A}_{\boldsymbol{\mu}}$ (for testing the hypothesis that $\mathbf{x}$'s mean vector equals $\boldsymbol{\mu}$) as a union of spherical caps,

$$\mathcal{A}_{\boldsymbol{\mu}} = \bigcup_{r \ge 0} \mathcal{S}_{\boldsymbol{\mu}}(\gamma_\rho(r), r), \tag{2.10}$$

so that

$$\text{Prob}_{\boldsymbol{\mu}}\{\mathbf{x} \in \mathcal{A}_{\boldsymbol{\mu}}\} = \alpha; \tag{2.11}$$

(2.11) follows immediately from (2.3) since the rth spherical cap has conditional probability $a_\rho(r)$ of containing $\mathbf{x}$.

The level α acceptance regions $\mathcal{A}_{\boldsymbol{\mu}}$ can be inverted to give confidence regions $\mathcal{C}_{\mathbf{x}}$ according to the usual Neyman construction:

$$\mathcal{C}_{\mathbf{x}} = \underset{\rho \geq 0}{\cup} \mathcal{S}_{\mathbf{x}}(\gamma_\rho(r), \rho). \tag{2.12}$$

Here $\mathcal{S}_{\mathbf{x}}(\gamma_\rho(r), \rho)$ is a spherical cap of angular radius $\gamma_\rho(r)$ centered at $\rho \cdot (\mathbf{x}/r)$, the multiple of $\mathbf{x}$ having length ρ; the level α confidence region $\mathcal{C}_{\mathbf{x}}$ is the union of such caps over all choices of ρ. The fact that the same function $\gamma_\rho(r)$ figures in both (2.10) and (2.12) reflects the circular geometry of the spherical caps.

To summarize, any inclusion function $a_\rho(r)$, (2.1)-(2.3), generates a set of non-randomized level α acceptance regions $\mathcal{A}_{\boldsymbol{\mu}}$ and corresponding confidence regions $\mathcal{C}_{\mathbf{x}}$. We can "reverse engineer" a given set of confidence regions to find $a_\rho(r)$. Figure 2 shows $a_\rho(r)$ for the Standard, Tseng-Brown, and Casella-Hwang regions, (with the last two plotted in the negative direction for graphical clarity), in the case $\rho = 6, n = 10$, and $\alpha = 0.95$. The noncentral chi density $f_\rho(r)$ is also shown. All three inclusion functions satisfy $\int_0^\infty a_\rho(r) f_\rho(r) dr \geq 0.95$ in accordance with (2.3). The Casella-Hwang integral equals 0.959 while the other two are exactly 0.95.

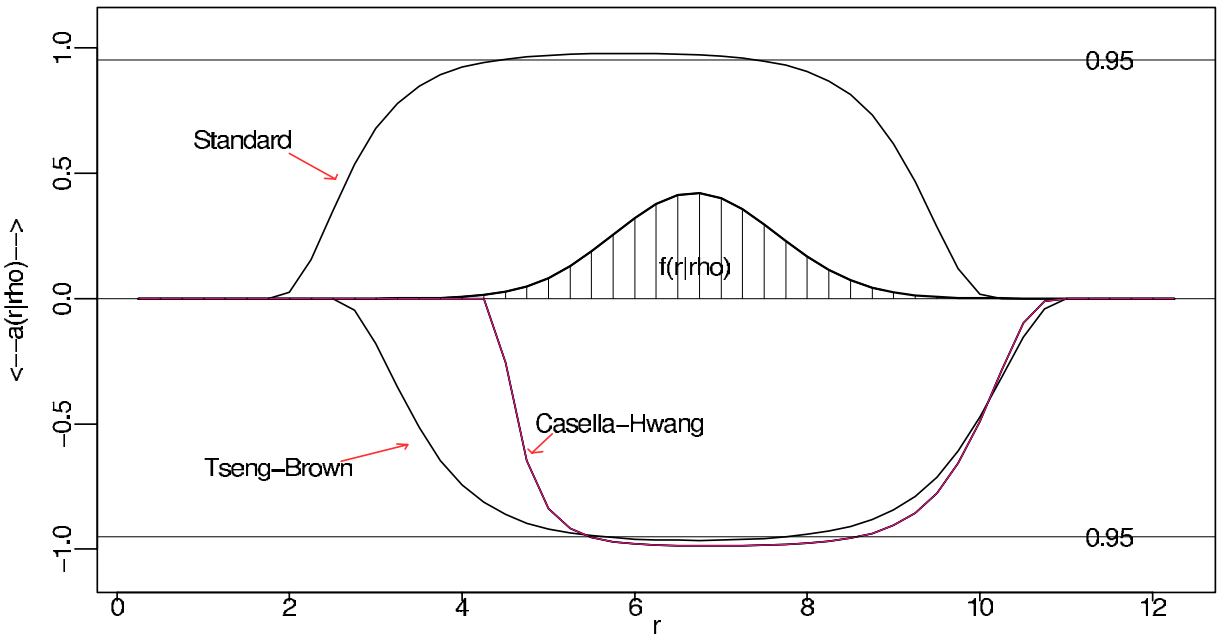

Figure 2: *Inclusion functions $a_\rho(r)$ for the Standard, Tseng-Brown, and Casella-Hwang regions; as a function of $r = \|\mathbf{x}\|$, with $\rho = \|\boldsymbol{\mu}\| = 6$; (T-B and C-H plotted negatively); $f_\rho(r)$ is noncentral chi density; $\int a_\rho(r) f_\rho(r) \geq 0.95$, with equality for Standard and Tseng-Brown.*

We see that the Standard inclusion function does not match up well with $f_\rho(r)$; shifting

it rightwards along the r axis gives a better match, allowing us to reduce $a_\rho(r)$ while still satisfying (2.3). This is what the Tseng-Brown and Casella-Hwang constructions both do. Doing so for all ρ gives smaller confidence regions $\mathcal{C}_{\mathbf{x}}$, in a way made clearer in Section 3.

At this point we might try selecting $a_\rho(r)$ so as to minimize the volume of the acceptance region $\mathcal{A}_{\boldsymbol{\mu}}$ subject to $\int a_\rho(r) f_\rho(r) dr = \alpha$. It turns out that this yields $\mathcal{A}_{\boldsymbol{\mu}} = \{\mathbf{x} : \|\mathbf{x} - \boldsymbol{\mu}\| \leq \chi_n^{(\alpha)}\}$, in other words the standard spherical acceptance region connected with (1.3); see Remark C, Section 5. This, perhaps, sharpens the surprise that Stein's conjecture is true. The nonstandard constructions of Figure 1 increase acceptance regions but decrease confidence regions, as shown in Section 3.

Algorithm (2.2)-(2.12) provides a very flexible recipe for constructing confidence regions for a multivariate normal mean vector, *too* flexible from the point of view of actual application. Section 3 considers, along with Bayesian and fiducial regions, a restricted class of inclusion functions, "percentile rules",

$$a_\rho(r) = A(F_\rho(r)), \tag{2.13}$$

where $F_\rho(r)$ is the noncentral chi cdf, while $A(\cdot)$ is a function satisfying $0 \leq A(F) \leq 1$ and

$$\int_0^1 A(F) dF = \alpha \tag{2.14}$$

The non-randomized interval (2.4) is a percentile rule.

Percentile rules are a function of only one variable, rather than two for general rules $a_\rho(r)$. We see that $a_\rho(r) = A(F_\rho(r))$ satisfies (2.2)-(2.3), the latter from

$$\begin{aligned} \int_0^\infty A(F_\rho(r)) f_\rho(r) dr &= \int_0^\infty A(F_\rho(r)) \frac{\partial F_\rho(r)}{\partial r} dr \\ &= \int_0^1 A(F) dF = \alpha. \end{aligned} \tag{2.15}$$

3. Bayes and Fiducial Regions Figure 2 reduces the construction of acceptance regions for $\mathbf{x} \sim N(\boldsymbol{\mu}, I)$ to that of forming acceptance intervals for $r \sim \chi_n(\rho)$. This section concerns the corresponding theory for confidence regions, from which the MinVolume region of Figure 1 will be derived. Doing so requires Bayesian and fiducial ideas, since frequentist methods by themselves do not constrain the size of a level α region $\mathcal{C}_{\mathbf{x}}$ for any particular $\mathbf{x}$. We continue to denote $r = \|\mathbf{x}\|$ and $\rho = \|\boldsymbol{\mu}\|$.

The *confidence density* for ρ given r, Efron (1993), is defined in terms of the cdf $F_\rho(r)$ of a $\chi_n(\rho)$ distribution,

$$g^\dagger(\rho|r) = -\frac{\partial}{\partial \rho} F_\rho(r), \tag{3.1}$$

in this case $g^\dagger$ being the same as Fisher's fiducial distribution (though Fisherian theory tended to avoid deficiencies of the sort mentioned below.) The name "confidence density" is appropriate since the quantiles of $g^\dagger(\rho|r)$ are the usual confidence limits for ρ given r,

$$\int_{\rho_1}^{\infty} g^\dagger(\rho|r)d\rho = \alpha_1 \;\Rightarrow\; F_{\rho_1}(r) = \alpha_1; \tag{3.2}$$

i.e., the upper α_1 quantile of $g^\dagger(\rho|r)$ is also the upper endpoint of a one-sided level α_1 confidence limit for ρ. One can think of $g^\dagger(\rho|r)$ as an objective Bayes posterior density for ρ given r, as discussed in Efron (1993).

Confidence density (3.1) is deficient in the sense that

$$\int_0^{\infty} g^\dagger(\rho|r)d\rho = F_0(r) < 1. \tag{3.3}$$

Construction (3.1) suggests that we add to $g^\dagger$ an atom

$$\bar{F}_0 \equiv 1 - F_0(r) \tag{3.4}$$

at $\rho = 0$. For $r = 6$ and $n = 10, \bar{F}_0$ is only 0.00008, so we will not worry about the missing atom in our numerical example.

A more direct objective Bayesian approach begins with the improper prior on $\boldsymbol{\mu}$

$$g(\boldsymbol{\mu}) = \rho^{-(n-1)}. \tag{3.5}$$

In work going back to Peers (1965), Stein (1985) and Tibshirani (1989) showed that having observed $\mathbf{x} \sim N(\boldsymbol{\mu}, I)$ with prior (3.5), the posterior density $g(\rho|\mathbf{x})$ of ρ gives Bayesian credible intervals agreeing with the usual confidence intervals to second order; $g(\rho|\mathbf{x})$ depends only on $r = \|\mathbf{x}\|$, and can be expressed as $g(\rho|r)$. It is not surprising that $g(\rho|r)$ and $g^\dagger(\rho|r)$ closely agree, as seen in Figure 3, since they both relate to standard confidence intervals for ρ; $g(\rho|r)$ has a handy formula, (5.7), which we will employ in the MinVolume calculation.

We can define the conditional coverage probability, given r, for the confidence region based on inclusion function $a_\rho(r)$, to be

$$\alpha^\dagger(r) = \int_0^{\infty} a_\rho(r) g^\dagger(\rho|r)d\rho, \tag{3.6}$$

"conditional" here referring to posterior coverage probability based on $g^+(\rho|r)$.

Theorem A percentile rule (2.13) that has $\alpha^\dagger(r) \geq \alpha$ for some value of r generates confidence regions (2.12) having frequentist coverage $\geq \alpha$ for all $\boldsymbol{\mu}$.

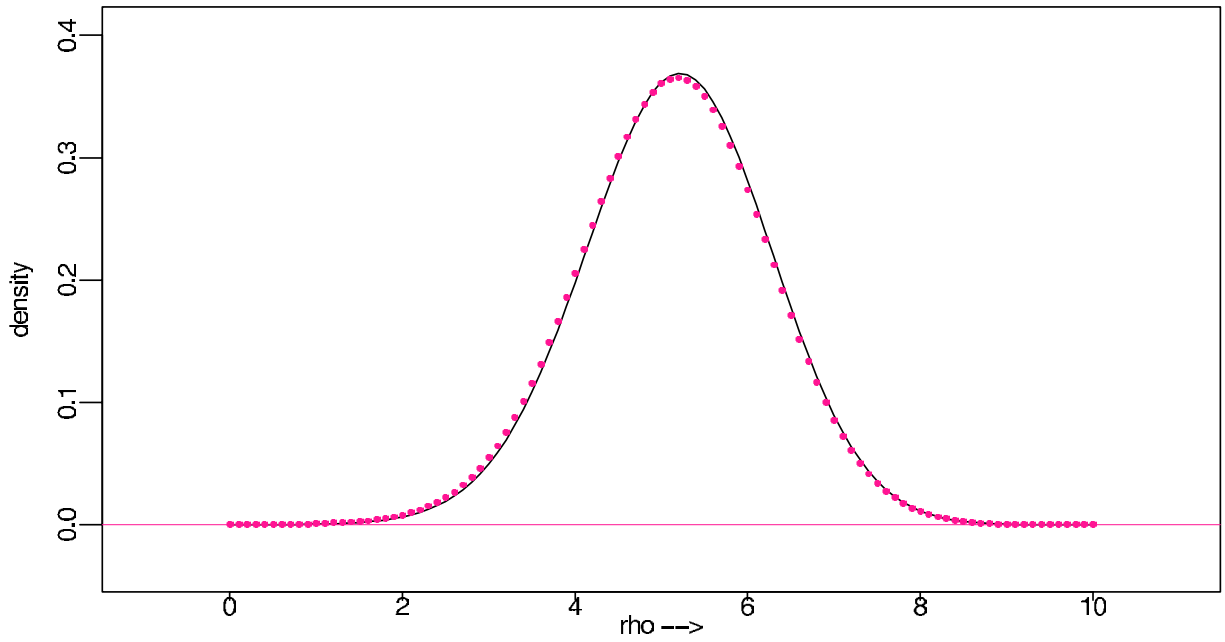

Figure 3: *Confidence density $g^\dagger(\rho|r)$ (3.1) for $r = 6, n = 10$ (solid curve); dots show Bayes posterior density $g(\rho|r)$ from prior (3.5).*

Proof For a percentile rule $a_\rho(r) = A(F_\rho(r))$,

$$\begin{aligned}\alpha \le \alpha^\dagger(r) &= -\int_0^\infty A(F_\rho(r))\frac{\partial F_\rho(r)}{\partial \rho}d\rho = \int_0^{F_0(r)} A(F)dF \\ &\le \int_0^1 A(F)dF = \int_0^\infty a_\rho(r)f_\rho(r)dr\end{aligned} \tag{3.7}$$

for all ρ, as in (2.15); this last quantity is the acceptance probability (2.11). In what follows we will use the Theorem to construct frequentist confidence regions from a rule $a_\rho(r_0)$ that minimizes volume at a given value $r = r_0$.

Prior (3.5) gives a formal Bayes version of the conditional coverage probability (3.6),

$$\alpha(r) = \int_0^\infty a_\rho(r)g(\rho|r)d\rho; \tag{3.8}$$

$\alpha(r)$ nearly equals $\alpha^\dagger(r)$ for situations like that of Figure 3.

We wish to compare various confidence regions for a given value of $\mathbf{x}$, or equivalently (for rotationally symmetric regions) for a given value of $r = \|\mathbf{x}\|$. The idea of what follows is to enforce fair comparison by insisting on equal conditional coverage, say $\alpha^\dagger(r) = 0.95$. Samworth (2005) stresses the importance of good conditional coverage properties.

Figure 4 plots inclusion functions $a_\rho(r)$ for the Standard, Tseng-Brown, Casella-Hwang, and MinVolume regions of Figure 1, now with ρ varying and $r = 6$ fixed ($n = 10$ and

$\alpha = 0.95$.) These are shown in relation to the confidence density $g^\dagger(\rho|r)$ from Figure 3. Again it is clear that the standard region is mismatched to $g^\dagger(\rho|r)$, and that by shifting it leftwards we can reduce $a_\rho(r)$ while maintaining $\alpha^\dagger(r)$ in (3.6). In fact the Tseng-Brown curve is roughly equivalent to shifting the Standard curve 0.6 units leftward, while the Casella-Hwang shift is 1.4 units.

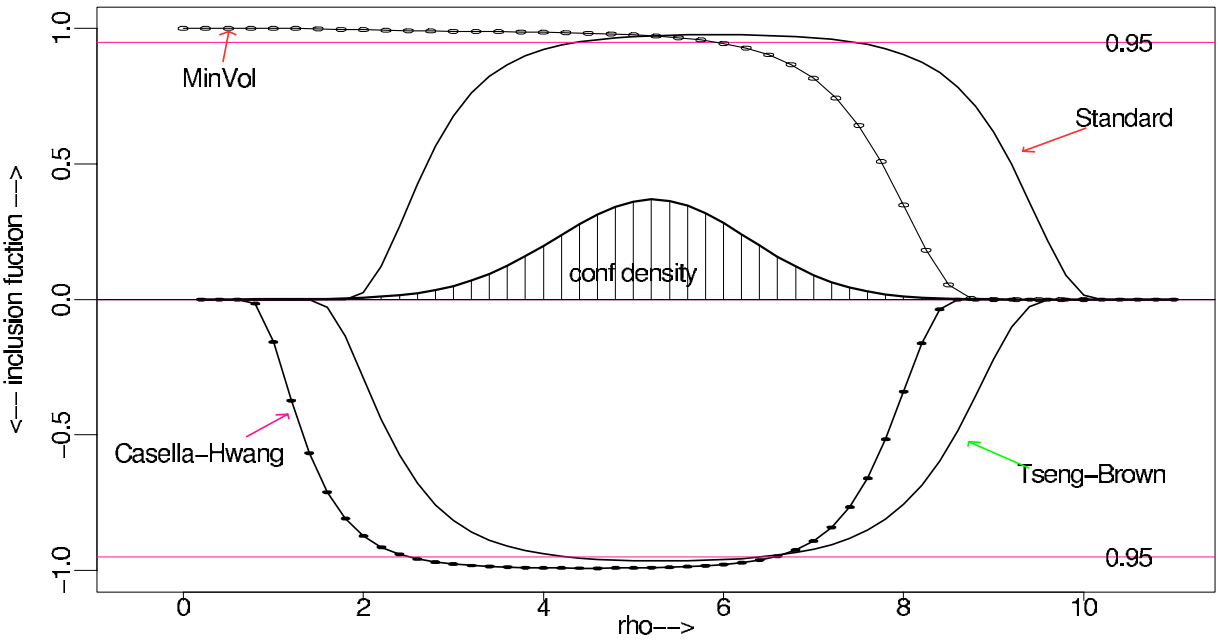

Figure 4: *Inclusion functions $a_\rho(r)$ for Standard, Tseng-Brown, Casella-Hwang and MinVolume regions of Figure 1; as a function of $\rho = \|\boldsymbol{\mu}\|$, with $r = \|\mathbf{x}\| = 6$, $n = 10, \alpha = 0.95$. (T-B and C-H plotted negatively.) Also shown is confidence density $g^\dagger(\rho|r)$ from Figure 3. Conditional coverage probabilities $\alpha^\dagger(r)$: Standard 0.94, T-B 0.94, C-H 0.97, MinVol 0.95. MinVol minimizes volume of confidence region (2.12) subject to $\int a_\rho(r) g^\dagger(\rho|r) d\rho = 0.95$. The adjusted C-H region of Figure 1 reduces the sphere's radius to make $\alpha^\dagger(r)$ exactly 0.95.*

The $(n-1)$-dimensional volume of spherical cap $\mathcal{S}_{\mathbf{x}}(\gamma_\rho(r), \rho)$ in (2.12) is

$$v_\rho(r) = k_n \rho^{n-1} \int_{z_\rho(r)}^1 (1-z^2)^{\frac{n-3}{2}} dz \qquad \left[z_\rho(r) = \cos(\gamma_\rho(r)),\ k_n = 2\pi^{\frac{n-1}{2}} / \Gamma\left(\frac{n-1}{2}\right) \right], \quad (3.9)$$

proportional to $[\rho \cdot \gamma_\rho(r)]^{n-1}$ for small values of $\gamma_\rho(r)$, see Remark A of Section 5. The leftward shifts reduce volume in two ways: by reducing ρ and by reducing how big $\gamma_\rho(r)$, or equivalently $a_\rho(r)$, need be while satisfying $\alpha^\dagger(r) = 0.95$.

For a given value of r_0 let $a_\rho(r_0)$ be a function of ρ taking values in $[0, 1]$; $a_\rho(r_0)$ determines

$z_\rho(r_0)$ as in (2.9) and then a confidence region (2.12) having volume

$$V(r_0) = \int_0^\infty v_\rho(r_0) d\rho. \tag{3.10}$$

The "MinVol" curve in Figure 4 is the function $a_\rho(r_0)$ that minimizes $V(r_0)$ subject to the conditional coverage constraint

$$\alpha^\dagger(r_0) = \int_0^\infty a_\rho(r_0) g^\dagger(\rho|r_0) d\rho = 0.95 \qquad (r_0 = 6). \tag{3.11}$$

The corresponding confidence region $\mathcal{C}_{\mathbf{x}}$ is labeled "MinVolume" in Figure 1.

In order to extend $a_\rho(r_0)$ to other values of r, define the function A by

$$A(F_\rho(r_0)) = a_\rho(r_0). \tag{3.12}$$

Then according to our previous results, (2.15) and the Theorem, $a_\rho(r) = A(F_\rho(r))$ is an inclusion function whose corresponding confidence regions have coverage ≥ 0.95. We have now constructed a rotationally symmetric system of 0.95 confidence regions having minimum volume at $r_0 = 6$ among those with $\alpha^\dagger(r_0) = 0.95$.

The actual calculation of the MinVolume inclusion function $a_\rho(r_0)$ appears in Remark B of Section 5; $a_\rho(r)$ looks peculiar in Figure 4, approaching 1 instead of 0 as ρ goes to 0 (causing the MinVolume region to contain the origin in Figure 1). This is possible because the spherical shell volume $v(\rho)$, (3.9), goes to zero very rapidly near $\rho = 0$.

The MinVolume calculation depends crucially on the constraint $\alpha^\dagger(r) = \alpha, \alpha = 0.95$ in our example. Several arguments can be advanced supporting the use of confidence density $g^\dagger(\rho|r)$ in definition (3.6):

• Section 2 reduced the problem of forming confidence regions for $\boldsymbol{\mu}$ given $\mathbf{x} \sim N(\boldsymbol{\mu}, I)$ to that of confidence intervals for ρ given $r \sim \chi_n(\rho)$, where by definition $g^\dagger(\rho|r)$ is the correct choice.

• The theorem shows $g^\dagger(\rho|r)$ playing a special role for regions based on percentile rules.

• $g^\dagger(\rho|r)$ is closely related to the invariant prior $g(\boldsymbol{\mu}) = \rho^{-(n-1)}$, which is known to have favorable properties for estimating $\boldsymbol{\mu}$. Under squared-error loss, $g(\boldsymbol{\mu})$ leads to minimax, admissible estimates of $\boldsymbol{\mu}$ that resemble the James-Stein estimator, see Remark D of Section 5.

Figure 1's volume bound of $1.65 \cdot 10^6$ is attainable in that there exists a bone-fide set of 0.95 confidence regions having that volume at $\|\mathbf{x}\| = 6$. Moreover the minimizing region is

(almost) Bayes versus a scale-invariant prior that does not obviously favor $r = 6$. There still remains an aspect of "special pleading" to MinVolume in that the construction (3.12) that extends $a_\rho(r_0)$ does not minimize volumes at values of r other than $r_0 = 6$. In this sense, $1.65 \cdot 10^6$ is a lower bound, which we can use to judge general constructions like that of Tseng-Brown and Casella-Hwang. The adjusted Casella-Hwang estimate comes impressively close in Figure 1, exceeding the lower bound by only 15%. (Notice that the adjusted rule of Figure 1 can be extended to give frequentist 0.95 regions by use of the theorem.)

4. Scheffe-type Inferences It is worthwhile asking if reduced volume leads to improved statistical inference. This section considers simple Scheffe-type simultaneous interval problems, when the answer is "yes" or "perhaps not" depending on the specific question.

For $\boldsymbol{\delta}$ a unit n-dimensional vector, define

$$\mu_\delta = \boldsymbol{\delta}'\boldsymbol{\mu}, \tag{4.1}$$

the component of $\boldsymbol{\mu}$ along direction $\boldsymbol{\delta}$. An α-level confidence region $\mathcal{C}_\mathbf{x}$ induces Scheffe-type simultaneous level α confidence intervals $\mathcal{I}_\delta$ for all-linear combinations μ_δ,

$$\mathcal{I}_\delta = [u_\delta, v_\delta] \quad \text{where} \begin{cases} u_\delta = \inf\ \boldsymbol{\delta}'\boldsymbol{\mu} \\ \qquad\qquad \text{for } \boldsymbol{\mu} \in \mathcal{C}_\mathbf{x}. \\ v_\delta = \sup\ \boldsymbol{\delta}'\boldsymbol{\mu} \end{cases} \tag{4.2}$$

For standard region $\mathcal{C}^0_\mathbf{x}$, the length of $\mathcal{I}_\delta$ is $2 \cdot \chi^{(\alpha)}_n$ for all $\boldsymbol{\delta}$. It is less for the Casella-Hwang regions, by a factor of 0.90 for the adjusted version in Figure 1. If average $\mathcal{I}_\delta$ length is the criteria then the Casella-Hwang construction is a definite improvement over $\mathcal{C}^0_\mathbf{x}$.

We might, however, wish to know which linear combinations μ_δ are significantly different than zero, i.e. which intervals $\mathcal{I}_\delta$ do not contain 0. This was the original motivation for the Scheffe construction, as in Section 3.5 of Scheffe (1959).

If $\mathcal{C}_\mathbf{x}$ contains the origin $\mathbf{0}$ then every interval $\mathcal{I}_\delta$ contains 0. If not, define

$$\gamma_0 = \sup_{\mathbf{y} \in \mathcal{C}_\mathbf{x}} \left\{ \cos^{-1}\left(\cdot \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \right) \right\}, \tag{4.3}$$

the maximum angle between $\mathbf{x}$ and a point $\mathbf{y}$ in $\mathcal{C}_\mathbf{x}$. Elementary geometry shows that

$$\mathcal{S}_{\text{app}} = \{\boldsymbol{\delta} : \mathcal{I}_\delta \text{ does not contain } 0\} \tag{4.4}$$

is a spherical cap on the unit sphere of angular radius $\frac{\pi}{2} - \gamma_0$ centered along $\mathbf{x}$,

$$\mathcal{S}_{\text{app}} = \mathcal{S}_{\mathbf{x}}(\pi/2 - \gamma_0, 1)$$

in notation (2.7). Here "app" stands for "apparent", $\mathcal{S}_{\text{app}}$ being the set of apparently interesting linear combinations.

The set of truly interesting linear combinations "$\mathcal{S}_{\text{true}}$" may be defined as those for which u_δ exceeds some threshold value t_0. This again is a spherical cap on the unit sphere, now centered along $\boldsymbol{\mu}$ with angular radius $\cos^{-1}(t_0/\rho)$, $\rho = \|\boldsymbol{\mu}\|$,

$$\mathcal{S}_{\text{true}} = \mathcal{S}_{\boldsymbol{\mu}}(\cos^{-1}(t_0/\rho), 1), \tag{4.5}$$

An ideal confidence procedure would have $\mathcal{S}_{\text{app}} = \mathcal{S}_{\text{true}}$; error rates can be described in terms of departures from this ideal. We define two set differences from the intersection $\mathcal{S}_{\text{int}} = \mathcal{S}_{\text{true}} \cap \mathcal{S}_{\text{app}}$,

$$\mathcal{S}_1 = \mathcal{S}_{\text{app}} - \mathcal{S}_{\text{int}} \quad \text{and} \quad \mathcal{S}_2 = \mathcal{S}_{\text{true}} - \mathcal{S}_{\text{int}}, \tag{4.6}$$

and let $v(\cdot)$ indicate $(n-1)$-dimensional volumes on the unit sphere. Natural definitions for error rates of the first and second kinds, i.e. type I and type II errors, are

$$\text{err}_1 = v(\mathcal{S}_1)/v(\mathcal{S}_{\text{app}}) \quad \text{and} \quad \text{err}_2 = v(\mathcal{S}_2)/v(\mathcal{S}_{\text{true}}); \tag{4.7}$$

err_1 is the proportion not true of $\mathcal{S}_{\text{app}}$, while err_2 is the proportion of $\mathcal{S}_{\text{true}}$ not in $\mathcal{S}_{\text{app}}$.

Given $\boldsymbol{\mu}$, $\mathbf{x}$, and $\mathcal{C}_{\mathbf{x}}$, the volumes and error rates in (4.7) can be computed from (3.9)-(3.10) using a more elaborate geometric calculation pertaining to cap intersections. Figure 5 compares the expectations of $(\text{err}_1, \text{err}_2)$ for the Standard and Casella-Hwang confidence regions, in dimension $n = 10$, $\alpha = 0.95$.The threshold for truly interesting linear combinations μ_δ was set at $t_0 = 4$. Expected error rates were computed by simulation from $\mathbf{x} \sim N(\boldsymbol{\mu}, I)$, with $\rho = \|\boldsymbol{\mu}\|$ varying from 4.5 to 8.

We see that the two methods are incomparable, with the Casella-Hwang regions giving lower errors of the second kind, but higher of the first. There is no clear-cut benefit to lower volume in this context, and if anything we might prefer the Standard regions, especially for smaller values of ρ. One thing is clear: the MinVolume regions are useless for this problem. MinVolume always contains the origin, as in Figure 1, Remark B Section 5, so that $\mathcal{S}_{\text{app}}$ is always empty.

Of course this is just one of many situations that might be used to compare confidence regions. The point is only that reduced volume by itself offers no guarantee of superior performance. The Stein shrinkage phenomenon permits reduced volume, and it is shrinkage itself,

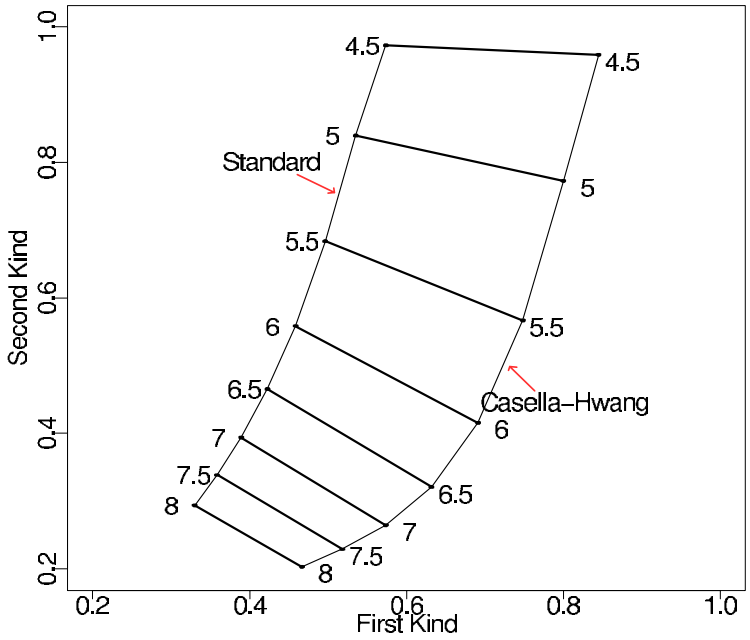

Figure 5: *Expected error rates of the first and second kind, (4.7), Standard and Casella-Hwang confidence regions, $n = 10$, $\alpha = 0.95$. For $\rho = \|\boldsymbol{\mu}\|$ increasing from 4.5 to 8. The C-H procedure gives lower errors of the second kind but higher of the first kind.*

not volume, that determines favorable or unfavorable performance in specific situations, as discussed in the example of Section 8, Efron (1982).

5. Remarks and Derivations This section expands on points raised in the previous discussion, as well as providing detailed derivations.

A. *Polar coordinates and the Fisher-von Mises density* All of our calculations depend on the multivariate polar transformation

$$x_1 = r \cdot \cos(\gamma),\ x_2 = r \cdot \sin(\gamma)\cos(\gamma_2),\ x_3 = r \cdot \sin(\gamma)\sin(\gamma_2)\cos(\gamma_3) \cdots , \tag{5.1}$$

and

$$x_n = r \cdot \sin(\gamma)\sin(\gamma_2)\sin(\gamma_3) \cdots \sin(\gamma_{n-2})\sin(\gamma_{n-1}), \tag{5.2}$$

with $\gamma, \gamma_2, \gamma_3, \cdots, \gamma_{n-2} \in [0, \pi]$ while $\gamma_{n-1} \in [0, 2\pi]$. (The case $n = 3$ is familiar.) Letting $\boldsymbol{\gamma}_2 = (\gamma_2, \gamma_3, \cdots, \gamma_{n-1})$, the Jacobian from $\mathbf{x}$ to $(r, z, \boldsymbol{\gamma}_2)$, $z = \cos(\gamma)$, is

$$J(\mathbf{x} \to r, z, \boldsymbol{\gamma}_2) = r^{n-1}(1 - z^2)^{\frac{n-3}{2}} J_2, \tag{5.3}$$

where $J_2 = \prod_{j=2}^{n-2} \sin(\gamma_j)^{n-1-j}$. The integral of J_2 over the full range of $\boldsymbol{\gamma}_2$ equals k_n, (3.9), the $(n-2)$-dimensional "area" of an $(n-1)$ dimensional unit sphere. (The area interpretation of k_n verifies formula (3.9) using familiar calculus methods.)

Since $\|\mathbf{x}-\boldsymbol{\mu}\|^2 = r^2+\rho^2-2r\rho z$, where $r=\|\mathbf{x}\|$ and $\rho=\|\boldsymbol{\mu}\|$, the n-dimensional normal density of (1.1) becomes

$$f_{\boldsymbol{\mu}}(r,z,\boldsymbol{\gamma}_2) = (2\pi)^{-\frac{n}{2}} r^{n-1} e^{-\frac{1}{2}r^2-\frac{1}{2}\rho^2+r\rho z}(1-z^2)^{\frac{n-3}{2}} J_2 \tag{5.4}$$

in polar coordinates. Integrating out $\boldsymbol{\gamma}_2$ gives

$$f_{\boldsymbol{\mu}}(r,z) = (2\pi)^{-n/2} k_n r^{n-1} e^{-\frac{1}{2}r^2-\frac{1}{2}\rho^2+r\rho z}(1-z^2)^{\frac{n-3}{2}}, \tag{5.5}$$

from which the Fisher-von Mises density (2.5) for $f_{\boldsymbol{\mu}}(z|r)$ follows directly. The primary angle γ is measured relative to $\boldsymbol{\mu}$ in this argument.

Beginning with prior density $g(\boldsymbol{\mu}) = \rho^{-(n-1)}$, the same argument, interchanging the roles of $\mathbf{x}$ and $\boldsymbol{\mu}$, gives

$$g(\rho,z|\mathbf{x}) = g(\rho,z|r) \;\propto\; e^{-\rho^2/2+r\rho z}(1-z^2)^{\frac{n-3}{2}}, \tag{5.6}$$

with $J(\boldsymbol{\mu}\to\rho,z,\boldsymbol{\gamma}_2) = \rho^{n-1}(1-z^2)^{\frac{n-3}{2}}J_2$ canceling out the prior factor $\rho^{-(n-1)}$. (Here the primary angle $\gamma = \cos^{-1}(z)$ is measured relative to the fixed vector $\mathbf{x}$.) Integrating out z in (5.6) produces posterior density

$$g(\rho|r) = \frac{C_0}{c_{r\rho}} e^{-\frac{1}{2}\rho^2} \qquad \left[c_{r\rho}^{-1} = \int_{-1}^{1} e^{r\rho z}(1-z^2)^{\frac{n-3}{2}} dz\right], \tag{5.7}$$

as in (2.6). It can be shown that

$$C_0^{-1} = \int_{-1}^{1} [\Phi(rz)/\varphi(rz)](1-z^2)^{\frac{n-3}{2}} dz, \tag{5.8}$$

where Φ and φ are the standard normal cdf and density.

A Bessel function expression for (2.6) appears in formula 9.6.18 of Abramowitz and Stegun (1964), but the integral form for $C_{r\rho}$ is more convenient for computation than the usual Bessel power series representation.

B. *The MinVolume Region* The MinVol curve in Figure 4 is the inclusion function $a_\rho(r), r = 6$, that minimizes volume of the corresponding confidence region (2.12) subject to $\alpha^\dagger(r) =$

0.95, (3.6). First we consider minimizing volume subject to the Bayesian constraint $\alpha(r) = 0.95$, (3.8).

The volume and $\alpha(r)$ contributions from an infinitesimal interval $[\rho, \rho + d\rho]$ are

$$dV = \left[k_n \rho^{n-1} \int_{z_\rho(r)}^{1} (1-z^2)^{\frac{n-3}{2}} dz\right] d\rho$$
$$\text{and} \tag{5.9}$$
$$d\alpha = \left[g(\rho|r) \int_{z_\rho(r)}^{1} c_{r\rho} e^{r\rho z} (1-z^2)^{\frac{n-3}{2}} dz\right] d\rho.$$

A small change dz in $z_\rho(r)$ produces changes $(dV)'$ and $(d\alpha)'$, with

$$\frac{(d\alpha)'}{(dV)'} = \frac{g(\rho|r) c_{r\rho} e^{r\rho z}}{k_n \rho^{n-1}} = \frac{c_0 e^{r\rho z - \rho^2/2}}{k_n \rho^{n-1}}, \tag{5.10}$$

using (5.7). A standard extremal argument says that (5.10) must be a constant along the volume-minimizing boundary $\widehat{z}_\rho(r)$. Taking logs in (5.10) gives

$$\widehat{z}_\rho(r) = \frac{\rho^2/2 + (n-1)\log(\rho) + \lambda}{r\rho}, \tag{5.11}$$

where λ is a free constant. Here we have not accounted for the constraint $z_\rho(r) \in [-1, 1]$, but Kuhn-Tucker calculations show that truncating (5.11) to $[-1, 1]$ yields the correct MinVolume solution.

Choosing different λ's gives MinVolume regions for different choices of Bayes coverage probability (3.8). Figure 6 shows the optimum boundaries as $\alpha(r)$ varies from 0.005 to 0.99. Notice that the origin $\mathbf{0}$ is always included, as can be seen by letting $\rho \to 0$ in (5.11). For larger values of r the optimum region breaks into two components, one of which contains $\mathbf{0}$. As communicated in Samworth (2005), the James-Stein estimate tends to lie above the regions' midpoints.

Defining $\phi_\rho(r) = \log(g^\dagger(\rho|r)/g(\rho|r))$, the same argument gives

$$\widehat{z}_\rho(r) = \frac{\rho^2/2 + (n-1)\log(\rho) + \partial\phi_\rho(r)/\partial r + \lambda}{r\rho}, \tag{5.12}$$

again truncated to $[-1, 1]$, as the optimal boundary for minimizing volume subject to $\alpha^\dagger(r)$, (3.6), equaling some specific value α; numerical methods are needed to solve (5.12). Figure 3 shows that $\phi_\rho(r)$ is nearly 1 for $r = 6, n = 10$, and in fact the $\alpha = 0.95$ boundary in Figure 6 nearly equals MinVolume in Figure 1, except near the origin. (Different horizontal scales in the two figures distort the visual comparison.)

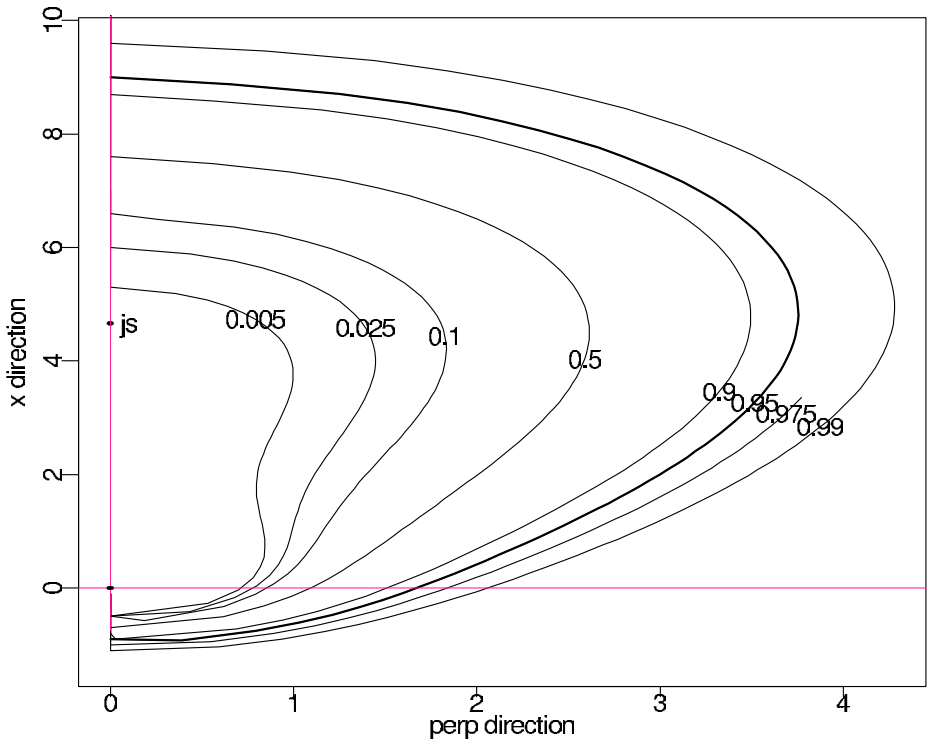

Figure 6: *Bayes optimum boundary (5.11), $r = 6, n = 10$, for $\alpha(r)$, (3.8), ranging from 0.005 to 0.99. The 0.95 curve nearly equals that for MinVolume in Figure 1.*

The Bayes boundary can also be derived from Neyman-Pearson considerations. Let $h_1(\boldsymbol{\mu}) = 1$ and $h_2(\boldsymbol{\mu}) = g(\boldsymbol{\mu}|\mathbf{x}) = \rho^{-(n-1)}$, (3.5). We wish to choose $\mathcal{C}_{\mathbf{x}}$ to minimize $V = \int_{\mathcal{C}_{\mathbf{x}}} h_1(\mu) d\boldsymbol{\mu}$ subject to $\alpha = \int_{\mathcal{C}_{\mathbf{x}}} h_2(\boldsymbol{\mu}) d\boldsymbol{\mu}$. According to the Neyman-Pearson lemma, the optimizing region is

$$\mathcal{C}_{\mathbf{x}} = \{\boldsymbol{\mu} : h_2(\boldsymbol{\mu})/h_1(\boldsymbol{\mu}) \geq \lambda_\alpha\}, \tag{5.13}$$

with constant λ_α chosen to satisfy the α constraint. Our previous results show that

$$h_2(\boldsymbol{\mu}) \propto \rho^{-(n-1)} e^{-\frac{1}{2}(r^2+\rho^2-2r\rho z)}, \tag{5.14}$$

making (5.14) equivalent to (5.11).

C. *Minimum Volume Acceptance Regions* Suppose that we wish to choose $a_\rho(r)$ to be the function of r that minimizes the volume of acceptance region $\mathcal{A}_{\boldsymbol{\mu}}$, (2.10), subject to constraint $\int_0^\infty a_\rho(r) f_\rho(r) dr = \alpha,\ \alpha = 0.95$ in Figure 2. The equivalent of relations (5.9) for

this situation is

$$dV = \left[k_n r^{n-1} \int_{z_\rho(r)}^{1} (1-z^2)^{\frac{n-3}{2}} dz\right] dr$$

and

$$d\alpha = \left[f_\rho(r) \int_{z_\rho(r)}^{1} c_{r\rho} e^{r\rho z} (1-z^2)^{\frac{n-3}{2}} dz\right] dr. \tag{5.15}$$

Integrating out z from (5.5) yields a convenient expression for the non-central chi density $f_\rho(r)$,

$$f_\rho(r) = c e^{-\rho^2/2} r^{n-1} e^{-r^2/2}/c_{r\rho} \qquad \left(c = \left[2^{\frac{n}{2}-1}\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\right]^{-1}\right). \tag{5.16}$$

In this case (5.10) becomes

$$\frac{(d\alpha)'}{(dV)'} = \text{constant} \cdot e^{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2}, \tag{5.17}$$

showing that the standard acceptance spheres $\{\mathbf{x} : \|\mathbf{x}-\boldsymbol{\mu}\| \le \chi_n^{(\alpha)}\}$ are minimum volume. Here, as opposed to Figure 3, it is the standard inclusion function $a_\rho(r)$ that is shifted leftwards, reducing volume by reducing r, in analogy with (3.9).

D. *Hierarchical Models* A familiar hierarchical Bayes model for estimating $\boldsymbol{\mu}$ in (1.2) takes $\boldsymbol{\mu} \sim N(0, CI)$ where C has an improper "hyperprior" density

$$h(C) = C^{-p/2}, \quad C > 0, \tag{5.18}$$

Efron and Morris (1975), Berger and Strawderman (1996). Formal integration, using polar coordinates, gives $\boldsymbol{\mu}$ marginal density

$$g(\boldsymbol{\mu}) = \rho^{-(n-2+p)}, \tag{5.19}$$

so we see that the choice $p = 1$ results in (3.5), which we called the uninformative prior for $\boldsymbol{\mu}$. The resulting squared error Bayes estimate of $\boldsymbol{\mu}$ having observed $\mathbf{x} \sim N(\boldsymbol{\mu}, I)$ is minimax and admissible, Morris (private communication), Berger and Strawderman (1996).

Selecting p different from 1 in (5.18) has only minor effects on our theory, for example changing $n-1$ in (5.10)-(5.11) to $n-2+p$. The choice $p = 0$ makes the Bayes and James-Stein estimates most alike; in this sense $p = 1$ is an "overshrinker", which explains the above-center location of *js* in Figure 6.

E. *The Case* $\mathbf{x} \sim N(\boldsymbol{\mu}, \sigma^2 I), \sigma^2$ *unknown* In most linear model situations (1.1) is generalized to

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \sigma^2 I) \text{ independent of } s^2 \sim \sigma^2 \chi^2_\nu / \nu. \tag{5.20}$$

However the results of this paper do not extend easily to situation (5.20), at least not the frequentist results. (Nor do those in our main references.) The difficulty here lies with the distribution of γ, or $z = \cos(\gamma)$, where there is no obvious pivotal quantity that has its density independent of the nuisance parameter σ. We might, instead, eliminate σ by giving it some sort of objective prior. The scale invariant prior density $\pi(\sigma) = \sigma^{-d}$ produces a version of the Fisher-von Mises density (2.5) for situation (5.20),

$$f_{\boldsymbol{\mu}}(z|r,s) = c(1-z^2)^{\frac{n-3}{2}} \int_0^\infty c_{\hat{r}\hat{\rho}t^2} t^{2(v_2-1/2)} e^{(\hat{r}\hat{\rho}z - v_2)t^2} dt \tag{5.21}$$

where $\widehat{r} = r/s, \widehat{\rho} = \rho/s, v_2 = (v+d-1)/2$, and $c = 2v_2^{v_2}/\Gamma(v_2)$, from which we might construct regions as in Section 2.

F. *The Casella-Hwang and Tseng-Brown Regions* Let c^2 equal the upper α quantile of a χ^2_n variate, $c^2 = \chi^{(\alpha)2}_n$ in (1.3), and define

$$R = 1 - (n-2)/r^2. \tag{5.22}$$

Assuming $r^2 > c^2 > n$ (satisfied in Figure 1, where $r^2 = 36$, $c^2 = 18.3, n = 10$), the Casella-Hwang confidence region is a sphere centered at the James-Stein estimator $R \cdot \mathbf{x}$, with squared radius $R \cdot (c^2 - n\log(R))$. This radius is less than $\chi^{(\alpha)}_n$, the standard region radius in (1.3).

The Tseng-Brown construction is based on spherical *acceptance* regions,

$$\mathcal{A}(\boldsymbol{\mu}) = \left\{ \mathbf{x} : \left\| \mathbf{x} - \boldsymbol{\mu}\left(1 + \frac{1}{A + B\rho^2}\right) \right\|^2 \le \chi^2_n \left(\frac{\rho^2}{(A+B\rho^2)^2} \right)^{(\alpha)} \right\}, \tag{5.23}$$

where $\chi^2_n(\Delta^2)^{(\alpha)}$ indicates the upper α quantile of a noncentral chi-squared distribution with noncentrality parameter Δ^2. Inverting $\mathcal{A}(\boldsymbol{\mu})$ gives oblate confidence regions, as seen in Figure 1. The constants A and B depend on n and α; with $n = 10$ and $\alpha = 0.95$, their Table 2 provides $A = 3.44$, $B = 0.125$.

References

Abramowitz, M. and Stegun, L. (1968) *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series **55**.

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

Berger, J. and Strawderman, W. (1996) "Choice of hierarchical priors: admissability in estimation of normal means", *Ann. Stat.*, **24**, 931-951.

Brown, L., Casella, G., and Hwang, G. (1995) "Optimal confidence sets and the Limacon of Pascal", *JASA*, **90**, 880-889.

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

Downs, T. (1966) "Some relationships among von Mises distributions of different dimensions", *Biometrika*, **53**, 269-272.

Efron, B. and Morris, C. (1975) "Data analysis using Stein's estimator and its competitors", *JASA*, **70**, 311-319.

Efron, B. (1982) "Maximum likelihood and decision theory", *Ann. Stat.*, **10**, 340-356.

Efron, B. (1993) "Bayes and likelihood calculations from confidence intervals", *Biometrika*, **80**, 3-26.

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

Peers, H. (1965) "On confidence points and Bayesian probability points in the case of several parameters", *JRSS-B*, **27**, 9-16.

Robinson, G.K. (1979a) Conditional properties of statistical procedures, *Ann. Stat.*, **7**, 742-755.

Samworth, R. (2005) "Small confidence sets for the mean of a spherically symmetric distribution" *JRSS-B*, **67**, 343-362.

Scheffe, M. (1959) *The Analysis of Variance*, Wiley, N.Y.

Stein, C. (1962) Confidence sets for the mean of a multivariate normal distribution (with discussion). *JRSS-B*, **24**, 265-296.

Stein, C. (1985) "On the coverage probability of confidence sets based on a prior distribution", *Banach Center Publications*, **16**, Warsaw: PWN-Polish Scientific Publishers.

Tibshirani, R. (1989) "Noninformative priors for one parameter of many", *Biometrika*, **76**, 604-608.

Tseng, Y.L. and Brown, L.D. (1997) Good exact confidence sets for a multivariate normal mean. *Ann. Stat.*, **25**, 2228-2258.