Improved Confidence Estimators for the Usual Multivariate Normal Confidence Set

BU-1041-M July 1989

Christian Robert[1]
Université Paris VI

George Casella[2]
Cornell University

Key Words and Phrases: Conditional confidence, loss estimation

AMS 1980 Subject Classification: 62C99, 62F25

[1] Research performed while Professor Robert was visiting the Mathematical Sciences Institute at Cornell University.

[2] Research supported by National Science Foundation Grant No. DMS89-0039

Summary

The usual multivariate normal confidence set has reported confidence $1-\alpha$, which is equal to its coverage probability. If we take a decision-theoretic view, and attempt to estimate the coverage, we find that $1-\alpha$ is an inadmissible estimator in more than four dimensions. We establish this fact and, moreover, exhibit a confidence estimator that appears to dominate $1-\alpha$. This new confidence statement allows us to attach confidence that is uniformly greater than $1-\alpha$. We provide necessary conditions, and strong numerical evidence to support our domination claim.

1. Introduction.

1.1 *The Model.* For the confidence set

$$C_X = \left\{\theta;\ \|\theta-X\|^2 \le c_\alpha^2\right\}, \tag{1.1}$$

where $X \sim \mathcal{N}(\theta, I_p)$ and $P_0\left(\|X\|^2 \le c_\alpha^2\right) = 1-\alpha$, the reported confidence statement is usually $1-\alpha$. Kiefer (1977) pointed out the conditional defects of such an approach and advocated the use of estimated confidence statements (somewhat analogous to using a *p-value* instead of a fixed confidence level in hypothesis testing problems).

Therefore, rather than considering a confidence region, C, to have a *confidence procedure*, $\langle C, \gamma(X)\rangle$ where if we observe $X = x$, $\gamma(x)$ is the reported confidence statement for the set C. Note that $\langle C, \gamma(X) \equiv 1-\alpha\rangle$ is a special case of this family. This approach is related to the theory of *conditional inference*, formalized by Robinson (1979a,b).

For a given set C, confidence statements will be compared according to the squared error loss

$$L(\gamma,\theta,x) = \left(\gamma(x) - \mathbb{I}_C(x-\theta)\right)^2, \tag{1.2}$$

where $\mathbb{I}_C(t)$ is the usual indicator function. Note that, if a confidence estimator γ_0 is not admissible under (1.2), there exists $\gamma_1 \neq \gamma_0$ such that, for every θ,

$$\mathbb{E}_\theta\left(L(\gamma_1,\theta,X)\right) \le \mathbb{E}_\theta\left(L(\gamma_0,\theta,X)\right),$$

and if we write $\gamma_1 = \gamma_0 + s(x)$, we have

$$\mathbb{E}\left[\left(\mathbb{I}_C(X-\theta) - \gamma_0(X)\right)s(X)\right] \ge \tfrac{1}{2}\, \mathbb{E}_\theta\left[s^2(X)\right], \tag{1.3}$$

which is almost equivalent to the existence of a relevant betting procedure against $\langle C, \gamma_0\rangle$, as introduced in Robinson (1979a). In fact, it can even be argued that (1.3) is a reasonable alternative to the definition of relevant betting, namely,

$$\mathbb{E}_\theta\left[\left(\mathbb{I}_C(X-\theta) - \gamma_0(X)\right)s(X)\right] \ge \epsilon\, \mathbb{E}_\theta\left[|s(X)|\right], \tag{1.4}$$

as the two definitions are equivalent for indicator functions.

In any case, if the confidence statement $1-\alpha$ is dominated under (1.2) by a variable confidence statement, it shows that $\langle C, 1-\alpha \rangle$ does not behave correctly conditionally. Moreover, the dominating estimator γ has an interest in itself, since it may be used with more flexibility than the fixed statement, $1-\alpha$.

1.2 *Related results.* This problem is a special case of the estimation loss problem, where associated with an estimator is an estimated degree of accuracy of this estimator. Bayesian theory generally provides straightforward estimates, while frequentists have to work harder to obtain such accuracy estimates (see, e.g., Berger and Robert, 1988).

The natural candidate for an estimator of loss, the "unbiased estimator of the risk," was excluded by Johnstone (1988) as generally inadmissible for squared error loss if $p \geq 5$. He obtained some dominating estimators for the squared error loss of the MLE and the original James-Stein estimator. Lu and Berger (1987) extended his result, while Rukhin (1988) introduces a mixed loss to compare estimators and their estimated losses simultaneously. A drawback of an approach as general as Rukhin's is that very few results can be derived in a sufficiently large framework. For instance, the complete class theorem is only satisfied under some continuity conditions on the losses, and in particular, does not apply for (1.2).

In the domain of confidence estimation, apart from the innovative paper of Kiefer (1977), a few major papers can be distinguished: Berger (1985b, 1988), Lu and Berger (1989) and Brown and Hwang (1989). The first papers consider some general ideas, while the work of Lu and Berger concentrates on improved confidence statements for sets recentered at positive-part James-Stein estimates. It has been known since Hwang and Casella (1982) that these recentered regions dominate the usual confidence region in the sense that they have the same volume and higher coverage probability. Therefore, the need for an estimated confidence is particularly striking in this case, as the gain possible by using the recentered set cannot be realized if $1-\alpha$ is the quoted (post-data) confidence level. Lu and Berger (1989) have established that, for $p \geq 5$, the constant confidence, $1-\alpha$, is inadmissible. Furthermore, they have exhibited a class of dominating confidence estimators.

Brown and Hwang (1989) use the same model as in Section 1.1 and establish that $1-\alpha$ is an admissible confidence estimator for C_X of (1.1) if $p \leq 4$. This result extends that of Robinson (1979b), as he proved the admissibility for p=1 and the inadmissibility for p=5.

1.3 *Frequentist validity.* Berger (1985a, 1985b, 1988) argues that, when a frequentist is estimating the loss $L\Big(\delta(X),\theta\Big)$ of an estimator δ, he should follow a *frequentist validity principle*, that is, if $\rho(X)$ is an estimator of $L\Big(\delta(X),\theta\Big)$, then $\rho(X)$ should satisfy

$$\mathbb{E}_\theta[\rho(X)] \geq \mathbb{E}_\theta\Big(L(\delta(X),\theta)\Big) . \tag{1.5}$$

According to this principle, the estimator ρ does not underestimate the loss $L\Big(\delta(X),\theta\Big)$ in the long run or, in other words, acts conservatively. Although statisticians tend to react favorably to conservative procedures, being conservative is not necessarily a good property. In particular, always knowingly underestimating that true confidence is not particularly desirable. This can be formalized using the theory of relevant betting in that it shows the existence of a *positively-biased* betting procedure. Furthermore, it has been often argued (with some reason) from a Bayesian point of view that frequentist concepts such as minimaxity were too conservative. Frequentist validity carries such conservativism to confidence estimation. Moreover, without requiring frequentist validity we should be able to provide better confidence estimators under the loss (1.2).

Lu and Berger (1989) succeeded in showing that their improving estimator was frequency valid. However, Brown and Hwang (1989), following the same principle, ended up with the result that $1-\alpha$ is admissible in the class of frequency valid estimators for any dimension. If the restriction to frequency valid estimators is dropped, this result is not true for $p \geq 5$ (see Section 2). The dominating estimators are necessarily not frequency valid but this does not seem to us of any importance, as we only agree on the use of the loss (1.2). An external limitation such as frequentist validity seems irrelevant for the decision-theoretic approach. A more relevant criticism could be about the choice of the loss, which is somewhat arbitrary. But the boundedness of parameters and action space lessens the drawbacks of squared error loss.

2. Inadmissibility Results.

Brown and Hwang (1989) have proved that $\langle C_X, 1-\alpha\rangle$ is admissible for $p \le 4$ using the loss (1.2). They also show that $\langle C_X, 1-\alpha\rangle$ is inadmissible in any dimension if comparison is restricted to frequency valid estimators. We show here the $\langle C_X, 1-\alpha\rangle$ is inadmissible for $p \ge 5$ by using the following theorem, which relies on the technique of *randomization of the origin* (see, e.g., Brown, 1975).

Theorem 2.1. *If there exists γ_1 and $p > m > 0$ such that*

$$\liminf_{\|\theta\| \to +\infty} \left[R(\gamma_0,\theta) - R(\gamma_1,\theta)\right]\|\theta\|^m > 0 ,$$

with $\gamma_0(X) = 1-\alpha$, then γ_0 is inadmissible.

Proof. First note that $R(\gamma_0,\theta) = \mathbb{E}_\theta\left(1 - \alpha - \mathbb{I}_C(X-\theta)\right)^2 = \alpha(1-\alpha)$. If

$$\lim_{\|\theta\| \to +\infty} \left[R(\gamma_0,\theta) - R(\gamma_1,\theta)\right]\|\theta\|^m = a > 0 ,$$

then there exists r such that $\alpha(1-\alpha) - R(\gamma_1,\theta) \ge \frac{a}{2\|\theta\|^m}$ for $\|\theta\| > r$. Define

$$q(t) = \begin{cases} \alpha(1-\alpha) - 1 & \text{if } t \le r \\ \frac{a}{2t^m} & \text{otherwise} . \end{cases}$$

By assumption, $\alpha(1-\alpha) - R(\gamma_1,\theta) \ge q(\|\theta\|)$ for every θ, as we may restrict consideration to γ_1 such that $R(\gamma_1,\theta) \le 1$ (otherwise, γ_1 is dominated by its truncation). Now, for $\xi \in \mathbb{R}^p$, consider the confidence estimator $\gamma_1^\xi(x) = \gamma_1(x - \xi)$. We have

$$\begin{aligned} R\left(\gamma_1^\xi,\theta\right) &= \mathbb{E}_\theta\left[\gamma_1(X-\xi) - \mathbb{I}_C(X-\theta)\right]^2 \\ &= \mathbb{E}_\theta\left[\gamma_1(X-\xi) - \mathbb{I}_C\left(X-\xi - (\theta-\xi)\right)\right]^2 \\ &= R(\gamma_1,\theta-\xi) . \end{aligned}$$

Now let $\xi \sim \mathcal{N}(0,k^2 I_p)$ and define

$$\gamma_2(X) = \mathbb{E}^\xi\left[\gamma_1^\xi(X)\right] = \mathbb{E}^\xi\left[\gamma_1(X - \xi)\right] ,$$

where the expectation is taken over the distribution of ξ. Then

$$R_\theta(\gamma_2,\theta) = \mathbb{E}_\theta\Big[\mathbb{E}^\xi\Big(\gamma_1(X-\xi)\Big) - \mathbb{I}_C(X-\theta)\Big]^2$$
$$\le \mathbb{E}_\theta^X\, \mathbb{E}^\xi\big[\gamma_1(X-\xi) - \mathbb{I}_C(X-\theta)\big]^2$$
$$= \mathbb{E}^\xi\Big(R(\gamma_1^\xi,\theta)\Big)$$
$$= \mathbb{E}^\xi\Big(R(\gamma_1,\theta-\xi)\Big)\,.$$

Furthermore, as $R(\gamma_0,\theta) = \alpha(1\text{-}\alpha)$,

$$\alpha(1\text{-}\alpha) - R(\gamma_2,\theta) \ge \mathbb{E}^\xi\big[R(\gamma_0,\theta-\xi) - R(\gamma_1,\theta-\xi)\big]$$
$$\ge \mathbb{E}^\xi\Big[q\Big(\|\theta-\xi\|\Big)\Big]\,.$$

Now, for $\theta=0$,

$$k^p\mathbb{E}^\xi\Big[q\Big(\|\xi\|^2\Big)\Big] = (2\pi)^{-p/2}\int q\Big(\|\xi\|^2\Big)\, e^{-\|\xi\|^2/2k^2}\, d\xi$$

$$\propto \int_0^r (\alpha(1\text{-}\alpha)-1)e^{-t^2/2u^2}t^{p-1}dt + \tfrac{a}{2}\int_r^{+\infty} r^{p-m-1}e^{-n^2/2k^2}dr$$

$$\ge (\alpha(1\text{-}\alpha)-1)r^p + \tfrac{a}{2}\, k^{p-m}\int_{r/u}^{+\infty} z^{p-m-1}e^{-z^2/2}dz,$$

which goes to $+\infty$ as k goes to infinity. Therefore, if k is large enough, $\mathbb{E}^\xi\big[q(\|\xi\|^2)\big] > 0$.

Finally, given θ, $\|\xi-\theta\|^2 \sim k^2\, \chi_p^2(\|\theta\|^2)$, which has the monotone likelihood property in $\|\theta\|^2$. Therefore, $\mathbb{E}^\xi\big[q(\|\theta-\xi\|^2)\big]$ can cross 0 at most once, from negative to positive. However, this is impossible because $\mathbb{E}^\xi\Big(q(\|\xi\|^2)\Big) > 0$. □

So, to show that $\langle C_X,1-\alpha\rangle$ is not admissible, we just need to find a γ_1 such that $\liminf_{\|\theta\|\to+\infty} \big[\alpha(1\text{-}\alpha) - R(\gamma_1,\theta)\big]\|\theta\|^m > 0$. This is done in the next theorem.

Theorem 2.2. For $p \ge 5$, $\gamma_0(X) = 1-\alpha$ is not an admissible estimator of confidence for the set C_X.

Proof. Consider the estimator

$$\gamma(X) = 1-\alpha + \frac{\delta}{1+\|X\|^2}\,,$$

where δ is a positive constant. Then

$$\alpha(1\text{-}\alpha) - R(\gamma,\theta) = -2\delta\, \mathbb{E}_\theta\Big[(1+\|X\|^2)^{-1}\big(1-\alpha-\mathbb{I}_C(X-\theta)\big)\Big] - \delta^2\, \mathbb{E}_\theta(1+\|X\|^2)^{-2}$$
$$= -2\delta\, \mathbb{E}_0\Big[\big(1+\|X+\theta\|^2\big)^{-1}\big(1-\alpha-\mathbb{I}_C(X)\big)\Big] - \delta^2\, \mathbb{E}_0\big(1+\|X+\theta\|^2\big)^{-2} .$$

Using a Taylor expansion, write

$$\big(1+\|x+\theta\|^2\big)^{-1} = (1+\|\theta\|^2)^{-1} + \sum_{i=1}^{p} x_i \frac{\partial}{\partial z_i}(1+\|z\|^2)^{-1}\Big|_{z=\theta}$$
$$+ \tfrac{1}{2} \sum_{i,j=1}^{p} x_i x_j \frac{\partial^2}{\partial z_i \partial z_j} (1+\|z\|^2)^{-1}\Big|_{z=\theta} + e(\theta,x)$$
$$= (1+\|\theta\|^2)^{-1} - 2\sum_{i=1}^{p} \frac{x_i\theta_i}{(1+\|\theta\|^2)^2}$$
$$+ 4\sum_{i,j=1}^{p} x_i x_j \frac{\theta_i\theta_j}{(1+\|\theta\|^2)^3} - \sum_{i=1}^{p} \frac{x^2}{(1+\|\theta\|^2)^2} + e(\theta,x),$$

where $e(\cdot,\cdot)$ is the error term. Since

$$\mathbb{E}_0\Big[X_i\big(1-\alpha-\mathbb{I}_C(X)\big)\Big] = 0 \text{ , and } \mathbb{E}_0\Big[X_iX_j\big(1-\alpha-\mathbb{I}_C(X)\big)\Big] = 0 \qquad (i\neq j) ,$$

we have

$$\mathbb{E}_0\Big[(1+\|X+\theta\|^2)^{-1}\big(1-\alpha-\mathbb{I}_C(X)\big)\Big] = \mathbb{E}_0\Big[X_1^2\big(1-\alpha-\mathbb{I}_C(X)\big)\Big]\left(\frac{4\|\theta\|^2}{(1+\|\theta\|^2)^3} - \frac{p}{(1+\|\theta\|^2)^2}\right)$$
$$+ e'(\theta)$$
$$= \mathbb{E}_0\Big\{X_1^2\big(1-\alpha-\mathbb{I}_C(X)\big)\Big\} \frac{(4-p)\|\theta\|^2-p}{(1+\|\theta\|^2)^3} + e'(\theta) .$$

Also, $\mathbb{E}_0\big(X_1^2[1-\alpha-\mathbb{I}_C(X)]\big) > 0$, because both functions are increasing in X_1^2. Now,

$$\alpha(1\text{-}\alpha) - R(\gamma,\theta) = 2\delta\, \mathbb{E}_0\Big[X_1^2\big(1-\alpha-\mathbb{I}_C(X)\big)\Big] \frac{(p-4)\|\theta\|^2+p}{(1+\|\theta\|^2)^3} - \delta^2 \frac{1}{(1+\|\theta\|^2)^2} + e''(\theta) ,$$

where $e''(\theta)$ is an error term resulting from the two approximations, and

$$\|\theta\|^4\big(\alpha(1\text{-}\alpha) - R(\gamma,\theta)\big) \approx 2\delta\, (p-4)\, \mathbb{E}_0\Big[X_1^2\big(1-\alpha-\mathbb{I}_C(X)\big)\Big] - \delta^2 > 0$$

for δ small enough . The fact that the error term can be ignored follows from arguments similar to Brown (1975). As $p \geq 5$, we can apply Theorem 2.1 to get the conclusion. □

For the case of $\alpha < \frac{1}{2}$, the result of Theorem 2.2 can be deduced from Lu and Berger (1989). They have obtained their result by another method, which is also not constructive, and relies heavily on some algebraic manipulations. It seems that the above proof has a greater potential for leading more easily to some generalizations.

Here, as in Johnstone (1988), the separating dimension is p=5 (while it is p=3 for the normal mean estimation problem), as conjectured by Robinson (1979b) and established by Brown and Hwang (1989). Johnstone's argument for the higher dimension was mainly technical, namely that two integrations-by-parts, rather than one, are necessary. Another justification could be that, as we are dealing with loss estimation, we are, in a way, using an L_4 loss rather than an L_2 loss. (We might conjecture that q+1 is the separating dimension for an L_q loss.) Errors are then more strongly penalized and then it requires a higher dimension to overcome them by using the other components.

3. A Potential Dominating Estimator

As γ_0 is not admissible, it is possible to uniformly improve upon it. We consider an estimator of the form

$$\gamma_{\mathrm{a}}(X) = 1-\alpha + \frac{a}{\|X\|^2} \quad (a \geq 0)\ . \tag{3.1}$$

Note the similarity with the James-Stein correction to the MLE. The estimator γ_{a} also shares a similar defect to the James-Stein estimator, namely that it behaves poorly if X is near 0, as it grows larger than 1 when X goes to 0. However, the simple form of (3.1) will allow easier computations than the bounded type introduced in Lu and Berger (1989),

$$\gamma_{\mathrm{a,b}}(X) = 1 - \alpha + \frac{a}{b+\|X\|^2} \tag{3.2}$$

with $a \leq b\alpha$. For practical usage, γ_{a} would, of course, be truncated at 1 (see Section 3.3).

As $a \geq 0$, $\gamma_{\mathrm{a}}(x) \geq 1-\alpha$ for every x. Therefore, γ_{a} is not a *frequency valid* estimator. In fact, its behavior is exactly opposite to that of a frequency valid estimator, since in this case a frequency valid estimator would have to be uniformly smaller than $1-\alpha$.

3.1 *Necessary Conditions for Dominance* We first investigate conditions necessary for γ_{a} to dominate γ_0. We have the following lemmas.

Lemma 3.1. *For $\theta=0$, γ_a has a smaller risk than γ_0 if and only if*

$$a \leq 2(p-4)\ (\alpha-\nu)\ , \tag{3.3}$$

where ν satisfies

$$P\left(\chi^2_{p\text{-}2} \leq c^2_\alpha\right) = 1-\nu\ .$$

Proof. The difference of the risks of $\langle C_X, 1\text{-}\alpha\rangle$ and $\langle C_X, \gamma_{\mathrm{a}}(X)\rangle$ at $\theta=0$ is

$$\mathbb{E}_0\Big(L(\gamma_0,\theta,X)\Big) - \mathbb{E}_0\Big(L(\gamma_{\mathrm{a}},\theta,X)\Big) = a\left(-2\ \mathbb{E}_0\left[\frac{1}{\|X\|^2}\Big(1-\alpha-\mathbb{I}_C(X)\Big)\right] - a\ \mathbb{E}_0\left[\frac{1}{\|X\|^4}\right]\right).$$

Now using the fact that $\mathbb{E}_0\Big(f(\|X\|^2)/\|X\|^2\Big) = \frac{1}{p-2}\ \mathbb{E}f\Big(\chi^2_{p-2}\Big)$, we have

$$\mathbb{E}_0\Big(L(\gamma_0,\theta,X)\Big) - \mathbb{E}_0\Big(L(\gamma_{\mathrm{a}},\theta,X)\Big) = a\left\{2\ \mathbb{E}_0\left(\frac{\mathbb{I}_C(X)}{\|X\|^2}\right) - 2\ \frac{1-\alpha}{p-2} - \frac{1}{(p-2)(p-4)}\right\}$$

$$= \frac{a}{p-2}\left\{2\left(P\left(\chi^2_{p-2} \le c^2_\alpha\right) - (1-\alpha)\right) - a\,\frac{1}{p-4}\right\},$$

which is positive if $a \le 2(p\text{-}4)(\alpha\text{-}\nu)$, establishing the lemma. □

Lemma 3.2. As $\|\theta\| \to \infty$, γ_a has a risk smaller than γ_0 asymptotically if

$$a \le 2(p-4)(\beta-\alpha), \tag{3.4}$$

where β satisfies

$$P\left(\chi^2_{p+2} \le c^2_\alpha\right) = 1-\beta\ .$$

Proof. Consider again a Taylor development as in the proof of Theorem 2.1. We have

$$\begin{aligned}
\alpha(1\text{-}\alpha) - R\left(\gamma_a, \theta\right) &\approx a\left\{2\ \mathbb{E}_0\left[X_1^2\left(1-\alpha-\mathbb{I}_C(X)\right)\right]\frac{p-4}{\|\theta\|^4} - \frac{a}{\|\theta\|^4}\right\} \\
&= \frac{a}{\|\theta\|^4}\left\{2\left((p-4)(1-\alpha) - \frac{p-4}{p}\ \mathbb{E}_0\left[\|X\|^2\mathbb{I}_C(X)\right]\right) - a\right\} \\
&= \frac{a}{\|\theta\|^4}\left\{2\left((p-4)(1-\alpha) - \frac{p-4}{p}\ \mathbb{E}_0\left[\|X\|^2\mathbb{I}_C(X)\right]\right) - a\right\} \\
&= \frac{a}{\|\theta\|^4}\left\{2(p-4)\left(1-\alpha - P\left(\chi^2_{p+2} \le c^2_\alpha\right)\right) - a\right\},
\end{aligned}$$

as $\mathbb{E}_0\left(f(\|X\|^2)\|X\|^2\right) = p\ \mathbb{E}_0\left[f\left(\chi^2_{p+2}\right)\right]$. The last expression is positive as long as $a \le 2(p\text{-}4)(\beta\text{-}\alpha)$. □

We now have two bounds on the constant a. The following lemma indicates which one is the largest, a condition depending on the value of α.

Lemma 3.3. The inequality

$$P\left(\chi^2_p \le c\right) - P\left(\chi^2_{p+2} \le c\right) \ge P\left(\chi^2_{p-2} \le c\right) - P\left(\chi^2_p \le c\right)$$

holds if and only if $c \ge p$.

Proof. Integrating by parts shows

$$P\left(\chi^2_{p+2} \le c\right) = -\frac{2^{-p/2}}{\Gamma\left(\frac{p}{2}+1\right)} c^{p/2} e^{-c/2} + P\left(\chi^2_p \le c\right).$$

Therefore

$$P\left(\chi^2_p \le c\right) - P\left(\chi^2_{p+2} \le c\right) = \frac{2^{-p/2}}{\Gamma\left(\frac{p}{2}+1\right)} c^{-p/2} e^{-c/2}.$$

Deriving a similar expression for $p\left(\chi^2_{p-2} \le c\right) - p\left(\chi^2_p \le c\right)$ and taking differences will establish the lemma. □

Thus, the bound of Lemma 3.2 is the larger when $c^2_\alpha < p$, which corresponds approximately to $\alpha < \frac{1}{2}$.

3.2 *Local Dominance* Of course, Lemmas 3.1 and 3.2 are only necessary conditions for domination of γ_0. However, simulations have shown that, for $c^2_\alpha \ge p$, the bound of Lemma 3.1 was also sufficient. The following result gives one more argument in favor of this statement.

Lemma 3.4. *The risk $R(\gamma_a, \theta)$ is decreasing for $\theta \approx 0$ if $p \ge 7$ and $a \ge (p-4)(\alpha-\eta)$, where η satisfies*

$$P\left(\chi^2_{p-4} \le c^2_\alpha\right) = 1-\eta \,. \tag{3.5}$$

Proof. We will prove that the difference

$$\Delta(\theta) = \alpha(1\text{-}\alpha) - R\left(\gamma_a, \theta\right)$$

is increasing in $\lambda = ||\theta||$. We have

$$\Delta(\theta) = a\left\{-2\, \mathbb{E}_\theta\left[\frac{1}{||X||^2}\left(1-\alpha-\mathbb{I}_C(X\text{-}\theta)\right)\right] - a\, \mathbb{E}_\theta\left[\frac{1}{||X||^4}\right]\right\}$$

$$= -a\left\{2\, \mathbb{E}_0\left[\frac{1}{||X+\theta||^2}\left(1-\alpha-\mathbb{I}_C(X)\right)\right] + a\, \mathbb{E}_0\left[\frac{1}{||X+\theta||^4}\right]\right\}$$

$$= -a\left\{2\,\mathbb{E}_0\Big(\|X\|^2+2\lambda\|X\|\cos\phi+\lambda^2\Big)^{-1}\Big(1-\alpha-\mathbb{I}_C(\|X\|^2)\Big)\right.$$

$$\left.+\,a\,\mathbb{E}_0\bigg(\Big(\|X\|^2+2\lambda\|X\|\cos\varphi+\lambda^2\Big)^{-2}\bigg)\right\}$$

where φ is the angle between a and X. Now, by symmetry, $(\partial/\partial\lambda)\Delta(\theta)\big|_{\theta=0}=0$, so the risk of γ_a will be decreasing if the second derivative of $\Delta(\theta)$ is positive. The first and second derivatives of $\Delta(\theta)$ are given by

$$\frac{\partial}{\partial\lambda}\Delta(\theta) = 4a\,\mathbb{E}_0\left\{\Big(\|X\|^2+2\lambda\|X\|\cos\varphi+\lambda^2\Big)^{-2}\Big(1-\alpha-\mathbb{I}_C(\|X\|^2)\Big)\right.$$

$$\left.+\,a\Big(\|X\|^2+2\lambda\|X\|\cos\varphi+\lambda^2\Big)^{-3}\right\}\Big(\|X\|\cos\varphi+\lambda\Big)$$

and

$$\frac{\partial^2}{\partial\lambda^2}\Delta(\theta) = -8a\,\mathbb{E}_0\left\{2(\|X\|^2+2\lambda\|X\|\cos\varphi+\lambda^2)^{-3}\Big(1-\alpha-\mathbb{I}_C(\|X\|^2)\Big)\right.$$

$$\left.+\,3a\Big(\|X\|^2+2\lambda\cos\varphi\|X\|+\lambda^2\Big)^{-4}\right\}\Big(\|X\|\cos\varphi+\lambda\Big)^2$$

$$+\,4a\,\mathbb{E}_0\,\Big(\|X\|^2+2\lambda\|X\|\cos\varphi+\lambda^2\Big)^{-2}\Big(1-\alpha-\mathbb{I}_C(\|X\|^2)\Big)$$

$$+\,a\Big(\|X\|^2+2\lambda\|X\|\cos\varphi+\lambda^2\Big)^{-3}.$$

Evaluating the second derivative at $\theta=0$, using the fact that $(\partial/\partial\lambda)\Delta(\theta)\big|_{\theta=0}=0$, we obtain

$$\frac{\partial^2}{\partial\lambda^2}\Delta(\theta)\bigg|_{\theta=0} = 4a\bigg(-\mathbb{E}_0\Big[\Big(4\|X\|^{-4}\Big(1-\alpha-\mathbb{I}_C(\|X\|^2)\Big)+6a\|X\|^{-6}\Big)\cos^2\varphi\Big]$$

$$+\,\mathbb{E}_0\Big[\|X\|^{-4}\Big(1-\alpha-\mathbb{I}_C(\|X\|^2)\Big)+a\,\|X\|^{-6}\Big]\bigg).$$

Note that, when $\theta=0$, $\|X\|^2$ and φ are independent. Therefore, evaluating the integral over φ yields

$$\int_0^\pi \cos^2\varphi\,\sin^{p-2}\varphi\,d\varphi = \frac{1}{p-1}\int_0^\pi \sin^p\varphi\,d\varphi = \frac{1}{p}\int_0^\pi \sin^{p-2}\varphi\,d\varphi\,.$$

Then

$$\left.\frac{\partial^2}{\partial\lambda^2}\Delta(\theta)\right|_{\theta=0} = 4a\, \mathbb{E}_0\Big[\|X\|^{-4}\big(1-\alpha-\mathbb{I}_C(\|X\|^2)\big)\big(1 - \tfrac{4}{p}\big) + a\|X\|^{-6}\big(1 - \tfrac{6}{p}\big)\Big]$$

$$= \frac{4a}{p}\left\{(p-4)\,\frac{1-\alpha-P\big(\chi^2_{p+4} \le c^2_\alpha\big)}{(p-2)(p-4)} + a\,\frac{p-6}{(p-2)(p-4)(p-6)}\right\}$$

$$= \frac{4a}{p(p-2)(p-4)}\left\{a - (p-4)(\alpha-\eta)\right\}.$$

The second derivative is then positive for $a \ge (p-4)(\alpha-\eta)$, proving the theorem. □

From this result, we deduce that the condition (3.3) is sufficient in a neighborhood of $t=0$, as long as $\alpha-\eta < 2(\alpha-\nu)$. (Table 1 gives a comparison of $\alpha-\eta$ and $\alpha-\nu$ for several values of p and $\alpha = 0.05$.) Simulations (see Figure 3.1 and 3.2) have suggested that the risk $R(\gamma_a,\theta)$ should be either increasing $\big(a \le (p-4)(\alpha-\eta)\big)$, or decreasing then increasing $\big(a \ge (p-4)(\alpha-\eta)\big)$ for $p \ge 7$. In both cases, as $c^2_\alpha \ge p$, they indicate that the domination of γ_0 by γ_a should be uniform.

3.3 *An Improved Version of the Confidence Estimator.* As noted at the beginning of Section 3.1, the choice of γ_a does not seem quite natural since $\gamma_a(X) > 1$ for X close to a. Therefore, we consider

$$\gamma^+_a(X) = \left(1-\alpha + \frac{a}{\|X\|^2}\right)^+$$

$$= \begin{cases} 1 & \text{if} \quad \|X\|^2 < \frac{a}{\alpha} \\ \gamma_a(X) & \text{otherwise}. \end{cases}$$

This estimator obviously dominates γ_a for quadratic loss and should also dominate γ_0 for $a \le 2(p-4)(\alpha-\nu)$. As, asymptotically, the truncation is negligible, the bound (3.4) is applicable to γ^+_a. But it follows from the simulations that the constant a can exceed the bound (3.3) and still dominate γ_0. Note that here, given the nature of the loss, we cannot reject γ^+_a as inadmissible: there is no complete class theorem.

p	5	7	9	11	13
α	0.05	0.05	0.05	0.05	0.05
$\alpha-\nu$	0.0387	0.0348	0.0321	0.0299	0.0283
$\alpha-\eta$	0.0491	0.0471	0.0453	0.0436	0.0422

Table 3.1. – Some values of the bounds (3.3) and (3.5)

As can be seen in Figures 3.1 and 3.2, the improvement in risk obtained by truncation is substantial. However, the improvement in risk is smaller than in Lu and Berger (1989) because there is no Stein effect. As such, there is less to gain.

4. Conclusion.

This work is obviously only a first step as we have not established formally the domination of γ_a over γ_0. However, it demonstrates the viability of a data-dependent reported confidence statement for the usual confidence set. This statement is more satisfactory than the pre-data statement, from a conditional point of view, and should be acceptable to a frequentist.

Note that the results of Lu and Berger (1989) are different from our results in that they attached improved confidence statements to recentered sets. In fact, even though the recentered confidence sets of Hwang and Casella (1982) dominate the usual confidence set with respect to volume and coverage probability, consideration of an appropriate confidence estimator makes the two confidence procedure incomparable. (Unless some sort of mixed loss, as in Rukhin 1988, is used.)

The preceding paragraph shows the importance of considering confidence procedures as a pair $\langle C(X), \gamma(X)\rangle$, where $\gamma(X)$ is the *stated confidence* in the set C(X). Whether $\gamma(X)$ is a good estimator of $P_\theta(\theta \in C(X))$ is a separate concern from whether C(X) is a good set estimator when evaluated according to $P_\theta(\theta \in C(X))$. Hwang and Casella(1982) show that a recentered set C_δ is better than the usual set C_X when evaluated according to the coverage probability $P_\theta(\theta \in C(X))$. Lu and Berger (1989) exhibit a pair $\langle C_\delta(X), \gamma(X)\rangle$ that dominates the pair $\langle C_\delta(X), 1 - \alpha\rangle$ in the sense that, for a fixed set C_δ, $\gamma(X)$ is a better confidence estimator than $1-\alpha$. Here, we show that for the usual confidence set C_X, there is a pair $\langle C_X(X), \gamma(X)\rangle$that dominates $\langle C_X(X), 1-\alpha\rangle$. Thus, we have the subtle difference between *set estimation* (where we try to find a good set) and *confidence estimation* (where, for a fixed set, we try to find good estimates of its confidence).

A possible and interesting extension of this study is to the t confidence set. We know from Brown (1967) that, even in the unidimensional case, the constant statement, $1-\alpha$, is inadmissible (as there exists a relevant set). Furthermore, in this case, no alternative has yet been proposed, even though it seems likely that the recentered confidence set has a higher coverage probability (Robert and Casella, 1987; Hwang and Ullah, 1989).

References

Berger, J.O. (1985a). *Statistical Decision Theory and Bayesian Analysis*; Springer (2nd edition).

Berger, J.O. (1985b). The frequentist viewpoint and conditioning. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. LeCam and R. Olshen, eds.). Wadsworth, Monterey, CA.

Berger, J.O. (1988). Discussion of "Conditionally acceptable frequentist solutions." In *Statistical Decision Theory and Related Topics, Vol. IV* (S. Gupta and J. Berger, eds.). Springer-Verlag, NY.

Berger, J. and Robert, C. (1988). Subjective hierarchical Bayes estimation of a normal mean: on the frequentist interface. Technical Report #88-24C, Statist. Dept., Purdue University. To appear in *The Annals of Statistics.*

Brown, L.D. (1967). The Conditional level of Student's test. *Annals Statist. 38*, 1068-1071.

Brown, L.D. (1975). Estimation with incompletely specified loss functions, *J. Amer. Statist. Assoc.* **70**, 417–26.

Brown, L.D. and Hwang, J.T. (1989). Admissibility of the usual confidence statement. Technical Report, Statistics Center, Cornell University.

Hwang, J.T. and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Stat.* **10**, 868–881.

Hwang, J.T. and Ullah, A. (1989). Confidence sets centered at James-Stein estimators - a surprise concerning the unknown variance case. Research Report No. 8090, Department of Economics, University of Western Ontario.

Johnstone, I. (1988). On the inadmissibility of Stein's unbiased estimate of loss. In *Statistical Decision Theory and Related Topics, Vol. IV* (S. Gupta and J. Berger, eds.). Springer-Verlag, NY.

Lu, K. and Berger, J. (1989a). Estimated confidence procedures for multivariate normal means. *J. Stat. Plann. Inf.* (to appear).

Lu, K. and Berger, J. (1989b). Estimation of normal means: Frequentists estimators of loss approach. *Ann. Stat.* 17, 890–907.

Robert, C. and Casella, G. (1987). Improved confidence sets for spherically symmetric distributions. Technical Report #87-51, Dept. of Statist., Purdue University.

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

Robinson, G.K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Stat.* 7, 756–771.

Rukhin, A. (1988). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics, Vol. IV* (S. Gupta and J. Berger, eds.). Springer-Verlag, NY.

Figure 3.1 Risk Ratios $R(\gamma_a,\theta)/R(1-\alpha,\theta)$ for γ_a and γ_a^+, $a=a^*$ and $2a^*$ where $a^* = (p-4)(\alpha-\nu)$, and $p = 7$. Calculations based on 1000 simulations.

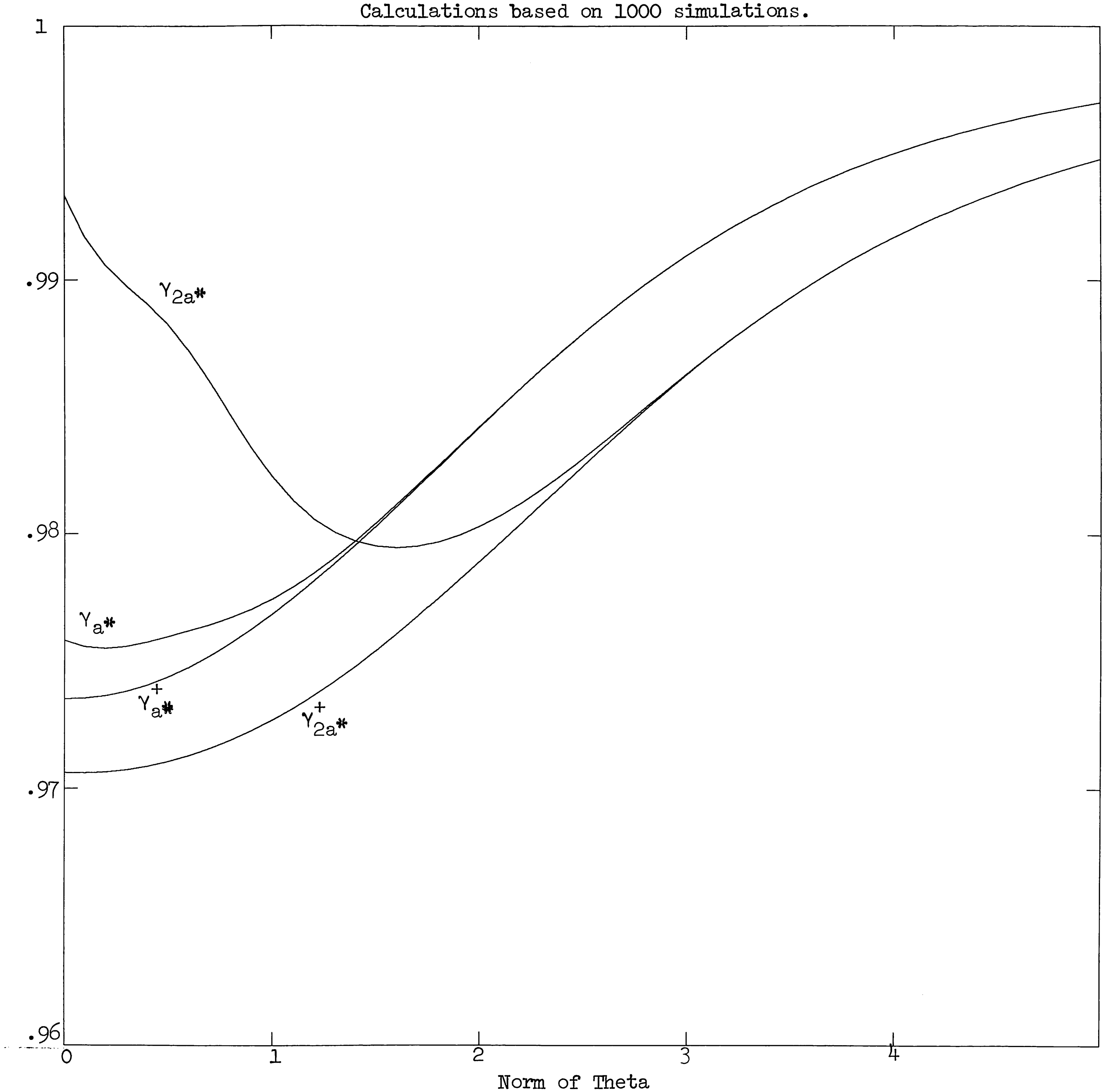

Figure 3.2: Expected Values of γ_a and γ_a^+ for $a=a^*$, $2a^*$, and $p=7$.
Calculations based on 1000 simulations.