Improved Confidence Estimators for the Usual
Multivariate Normal Confidence Set

BU-1041-M

July 1989

Christian Robert¹
Université Paris VI

George Casella²
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

\[ C_X = \{ \theta; \| \theta - X \| ^2 \leq c_\alpha^2 \}, \]  

(1.1)

where \( X \sim \mathcal{N}(\theta, I_p) \) and \( P_0(\| X \| ^2 \leq c_\alpha^2) = 1 - \alpha \), 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, \( (C, \gamma(X)) \) where if we observe \( X = x \), \( \gamma(x) \) is the reported confidence statement for the set \( C \). Note that \( (C, \gamma(X) \equiv 1 - \alpha) \) 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

\[ L(\gamma, \theta, x) = (\gamma(x) - I_C(x-\theta))^2, \]  

(1.2)

where \( I_C(t) \) is the usual indicator function. Note that, if a confidence estimator \( \gamma_0 \) is not admissible under (1.2), there exists \( \gamma_1 \neq \gamma_0 \) such that, for every \( \theta \),

\[ E_\theta \left( L(\gamma_1, \theta, X) \right) \leq E_\theta \left( L(\gamma_0, \theta, X) \right), \]

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

\[ E \left[ (I_C(X-\theta) - \gamma_0(X))s(X) \right] \geq \frac{1}{2} E \left[ s^2(X) \right], \]  

(1.3)

which is almost equivalent to the existence of a relevant betting procedure against \( (C, \gamma_0) \), 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,

\[ E_\theta \left[ (I_C(X-\theta) - \gamma_0(X))s(X) \right] \geq \epsilon \ E_\theta \left[ s(X) \right], \]  

(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 \((C,1-\alpha)\) does not behave correctly conditionally. Moreover, the dominating estimator \(\gamma\) 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 $L(\delta(X),\theta)$ of an estimator $\delta$, he should follow a frequentist validity principle, that is, if $\rho(X)$ is an estimator of $L(\delta(X),\theta)$, then $\rho(X)$ should satisfy

$$E_\theta[\rho(X)] \geq E_\theta[L(\delta(X),\theta)].$$ (1.5)

According to this principle, the estimator $\rho$ does not underestimate the loss $L(\delta(X),\theta)$ 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 \( (C_X, 1-\alpha) \) is admissible for \( p \leq 4 \) using the loss (1.2). They also show that \( (C_X, 1-\alpha) \) is inadmissible in any dimension if comparison is restricted to frequency valid estimators. We show here the \( (C_X, 1-\alpha) \) is inadmissible for \( p \geq 5 \) 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 \( \gamma_1 \) 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 \( \gamma_0 \) is inadmissible.

**Proof.** First note that \( R(\gamma_0, \theta) = E_{\theta} \left( 1 - \alpha - I_C(X - \theta) \right)^2 = \alpha(1-\alpha) \). If

\[
\lim_{\|\theta\| \to +\infty} \frac{\|\theta\|^m}{\|\theta\|^m} = a > 0,
\]

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

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

By assumption, \( \alpha(1-\alpha) - R(\gamma_1, \theta) \geq q(\|\theta\|) \) for every \( \theta \), as we may restrict consideration to \( \gamma_1 \) such that \( R(\gamma_1, \theta) \leq 1 \) (otherwise, \( \gamma_1 \) 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

\[
R(\gamma_1^\xi, \theta) = E_{\theta} \left[ \gamma_1(X - \xi) - I_C(X - \theta) \right]^2 = E_{\theta} \left[ \gamma_1(X - \xi) - I_C(X - \theta - (\theta - \xi)) \right]^2 = R(\gamma_1, \theta - \xi).
\]

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

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

where the expectation is taken over the distribution of \( \xi \). Then
\[ R_\gamma(\gamma_2, \theta) = \mathbb{E}_\theta \left[ \mathbb{E}^\xi \left( \gamma_1(X - \xi) - 1_{C(X - \theta)} \right)^2 \right] \]
\[ \leq \mathbb{E}_\theta^X \mathbb{E}^\xi \left[ \gamma_1(X - \xi) - 1_{C(X - \theta)} \right]^2 \]
\[ = \mathbb{E}^\xi \left( R(\gamma, \xi) \right) \]
\[ = \mathbb{E}^\xi \left( R(\gamma_1, \theta - \xi) \right). \]

Furthermore, as \( R(\gamma_0, \theta) = \alpha(1 - \alpha) \),
\[ \alpha(1 - \alpha) - R(\gamma_2, \theta) \geq \mathbb{E}^\xi \left[ R(\gamma_0, \theta - \xi) - R(\gamma_1, \theta - \xi) \right] \]
\[ \geq \mathbb{E}^\xi \left[ q(||\theta - \xi||) \right]. \]

Now, for \( \theta = 0 \),
\[ k^p \mathbb{E}^\xi \left[ q(||\xi||^2) \right] = (2\pi)^{-p/2} \int q(||\xi||^2) e^{-||\xi||^2/2k^2} d\xi \]
\[ \propto \int_0^r (\alpha(1 - \alpha) - 1) e^{-t^2/2u^2} t^{p-1} dt + \frac{3}{2} \int_r^\infty r^{p-m-1} e^{-n^2/2k^2} dr \]
\[ \geq (\alpha(1 - \alpha) - 1) r^p + \frac{3}{2} k^{p-m} \int_{r/u}^\infty z^{p-m-1} e^{-z^2/2} dz, \]
which goes to \( +\infty \) as \( k \) goes to infinity. Therefore, if \( k \) is large enough, \( \mathbb{E}^\xi \left[ q(||\xi||^2) \right] > 0. \)

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

So, to show that \( (C_X, 1 - \alpha) \) is not admissible, we just need to find a \( \gamma_1 \) such that \( \lim_{||\theta|| \to +\infty} \left[ (\alpha(1 - \alpha) - R(\gamma_1, \theta)) ||\theta||^m > 0. \right. \) This is done in the next theorem.

**Theorem 2.2.** For \( p \geq 5 \), \( \gamma_0(X) = 1 - \alpha \) 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 \( \delta \) is a positive constant. Then

\[
\alpha(1-\alpha) - R(\gamma, \theta) = -2\delta E_\theta \left[ (1+\|X\|^2)^{-1} (1-\alpha-I_C(X-\theta)) \right] - \delta^2 E_\theta (1+\|X\|^2)^{-2}
\]

\[
= -2\delta E_0 \left[ (1+\|X+\theta\|^2)^{-1} (1-\alpha-I_C(X)) \right] - \delta^2 E_0 (1+\|X+\theta\|^2)^{-2}.
\]

Using a Taylor expansion, write

\[
(1+\|x+\theta\|^2)^{-1} = (1+\|\theta\|^2)^{-1} + \sum_{i=1}^{\frac{p}{2}} x_i \frac{\partial}{\partial z_i} (1+\|z\|^2)^{-1} \bigg|_{z=\theta} + \frac{1}{2} \sum_{i,j=1}^{p} x_i x_j \frac{\partial^2}{\partial z_i \partial z_j} (1+\|z\|^2)^{-1} \bigg|_{z=\theta} + \epsilon(\theta, x)
\]

\[
= (1+\|\theta\|^2)^{-1} - 2 \sum_{i=1}^{\frac{p}{2}} \frac{x_i^2}{(1+\|\theta\|^2)^{3}} + \frac{4}{\|\theta\|^3} \sum_{i,j=1}^{p} \frac{x_i x_j}{(1+\|\theta\|^2)^{3}} + \epsilon(\theta, x),
\]

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

\[
E_0 \left[ x_i (1-\alpha-I_C(X)) \right] = 0, \quad \text{and} \quad E_0 \left[ x_i x_j (1-\alpha-I_C(X)) \right] = 0 \quad (i \neq j),
\]

we have

\[
E_0 \left[ (1+\|X+\theta\|^2)^{-1} (1-\alpha-I_C(X)) \right] = E_0 \left[ x_i^2 (1-\alpha-I_C(X)) \right] \left( \frac{4\|\theta\|^2}{(1+\|\theta\|^2)^3} - \frac{p}{(1+\|\theta\|^2)^2} \right)
\]

\[
+ \epsilon'(\theta) = E_0 \left[ x_i^2 (1-\alpha-I_C(X)) \right] \left( \frac{4-p}{(1+\|\theta\|^2)^3} \right) + \epsilon'(\theta).
\]

Also, \( E_0 \left[ x_i^2 (1-\alpha-I_C(X)) \right] > 0 \), because both functions are increasing in \( x_i^2 \). Now,

\[
\alpha(1-\alpha) - R(\gamma, \theta) = 2\delta E_0 \left[ x_i^2 (1-\alpha-I_C(X)) \right] \left( \frac{(p-4)\|\theta\|^2+p}{(1+\|\theta\|^2)^3} - \delta^2 \frac{1}{(1+\|\theta\|^2)^2} + \epsilon''(\theta) \right),
\]

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

\[
\|\theta\|^4 \left( \alpha(1-\alpha) - R(\gamma, \theta) \right) \approx 2\delta (p-4) E_0 \left[ x_i^2 (1-\alpha-I_C(X)) \right] - \delta^2 > 0
\]

for \( \delta \) 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. \( \square \)

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 $\gamma_0$ is not admissible, it is possible to uniformly improve upon it. We consider an estimator of the form

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

(3.1)

Note the similarity with the James-Stein correction to the MLE. The estimator $\gamma_a$ 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),

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

(3.2)

with $a \leq ba$. For practical usage, $\gamma_a$ would, of course, be truncated at 1 (see Section 3.3).

As $a \geq 0$, $\gamma_a(x) \geq 1 - \alpha$ for every $x$. Therefore, $\gamma_a$ 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 $\gamma_a$ to dominate $\gamma_0$. We have the following lemmas.

Lemma 3.1. For $\theta = 0$, $\gamma_a$ has a smaller risk than $\gamma_0$ if and only if

$$a \leq 2(p-4)(\alpha - \nu),$$

(3.3)

where $\nu$ satisfies

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

Proof. The difference of the risks of $(C_X, 1-\alpha)$ and $(C_X, \gamma_a(X))$ at $\theta = 0$ is

$$E_0\left(L(\gamma_0, \theta, X)\right) - E_0\left(L(\gamma_a, \theta, X)\right) = a \left(-2 \mathbb{E}_0\left[\frac{1}{\|X\|^2}\left(1 - \gamma_a(X)\right)\right]\right) - \mathbb{E}_0\left[\frac{1}{\|X\|^4}\right].$$

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

$$E_0\left(L(\gamma_0, \theta, X)\right) - E_0\left(L(\gamma_a, \theta, X)\right) = a \left\{ 2 \mathbb{E}_0\left(\frac{l_C(X)}{\|X\|^2}\right) - 2 \frac{1-\alpha}{p-2} - \frac{1}{(p-2)(p-4)} \right\}.$$
\[ = \frac{a}{p-2} \left\{ 2 \left( P\left( \chi_{p-2}^2 \leq c_\alpha^2 \right) - (1-\alpha) \right) - a \frac{1}{p-4} \right\} , \]

which is positive if \( a \leq 2(p-4)(\alpha-\nu) \), establishing the lemma. \qed

**Lemma 3.2.** As \( ||\theta|| \to \infty \), \( \gamma_a \) has a risk smaller than \( \gamma_0 \) asymptotically if

\[ a \leq 2(p-4)(\beta-\alpha), \quad (3.4) \]

where \( \beta \) satisfies

\[ P\left( \chi_{p+2}^2 \leq c_\alpha^2 \right) = 1-\beta . \]

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

\[ \alpha(1-\alpha) - R\left( \gamma_a, \theta \right) \approx a \left\{ 2 E_0 \left[ X_i^2 (1-\alpha-I_C(X)) \right] \right\} \frac{p-4}{||\theta||^4} - \frac{a}{||\theta||^4} \]

\[ = \frac{a}{||\theta||^4} \left\{ 2 \left( (p-4)(1-\alpha) - \frac{p-4}{p} E_0 \left[ ||X||^2 I_C(X) \right] \right) - a \right\} \]

\[ = \frac{a}{||\theta||^4} \left\{ 2 \left( (p-4)(1-\alpha) - \frac{p-4}{p} E_0 \left[ ||X||^2 I_C(X) \right] \right) - a \right\} \]

\[ = \frac{a}{||\theta||^4} \left\{ 2(p-4) \left( 1-\alpha - P\left( \chi_{p+2}^2 \leq c_\alpha^2 \right) \right) - a \right\} , \]

as \( E_0\left( f(||X||^2)||X||^2 \right) = p \ E_0\left[ f\left( \chi_{p+2}^2 \right) \right] \). The last expression is positive as long as \( a \leq 2(p-4)(\beta-\alpha) \). \qed

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 \( \alpha \).

**Lemma 3.3.** The inequality

\[ P\left( \chi_p^2 \leq c \right) - P\left( \chi_{p+2}^2 \leq c \right) \geq P\left( \chi_{p-2}^2 \leq c \right) - P\left( \chi_p^2 \leq c \right) \]

holds if and only if \( c \geq p \).

**Proof.** Integrating by parts shows
\[ P\left( \chi_{p+2}^2 \leq c \right) = -\frac{2^{-p/2}}{\Gamma\left(\frac{p}{2}+1\right)} \frac{c^{p/2}}{e} e^{-c/2} + P\left( \chi_p^2 \leq c \right) . \]

Therefore

\[ P\left( \chi_p^2 \leq c \right) - P\left( \chi_{p+2}^2 \leq c \right) = \frac{2^{-p/2}}{\Gamma\left(\frac{p}{2}+1\right)} c^{-p/2} e^{-c/2} . \]

Deriving a similar expression for \( p\left( \chi_{p-2}^2 \leq c \right) - p\left( \chi_p^2 \leq c \right) \) and taking differences will establish the lemma.

Thus, the bound of Lemma 3.2 is the larger when \( c_{\alpha}^2 < p \), 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 \( \gamma_0 \). However, simulations have shown that, for \( c_{\alpha}^2 \geq p \), 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 \( p \geq 7 \) and \( a \geq (p-4)(\alpha-\eta) \), where \( \eta \) satisfies

\[ P\left( \chi_{p-4}^2 \leq c_{\alpha}^2 \right) = 1-\eta . \]  \hspace{1cm} (3.5)

**Proof.** We will prove that the difference

\[ \Delta(\theta) = \alpha(1-\alpha) - R\left(\gamma_a, \theta\right) \]

is increasing in \( \lambda = ||\theta|| \). We have

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

\[ = -a \left\{ 2 E_0\left[ \frac{1}{||X+\theta||^2} (1-\alpha - \mathbb{I}_C(X)) \right] + a E_0\left[ \frac{1}{||X+\theta||^4} \right] \right\} \]
\[-13-\]

\[
\begin{align*}
= -a \left\{ 2 \mathbb{E}_0 \left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-1} \left( 1 - \alpha - 1_C(\|X\|^2) \right) \\
+ a \mathbb{E}_0 \left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-2} \right\}
\end{align*}
\]

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

\[
\frac{\partial}{\partial \lambda} \Delta(\theta) = 4a \mathbb{E}_0 \left\{ \left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-2} \left( 1 - \alpha - 1_C(\|X\|^2) \right) \\
+ a \left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-3} \left( \|X\|\cos\phi + \lambda \right) \right\}
\]

and

\[
\frac{\partial^2}{\partial \lambda^2} \Delta(\theta) = -8a \mathbb{E}_0 \left\{ 2\left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-3} \left( 1 - \alpha - 1_C(\|X\|^2) \right) \\
+ 3a \left( \|X\|^2 + 2\lambda\cos\phi\|X\| + \lambda^2 \right)^{-4} \left( \|X\|\cos\phi + \lambda \right)^2 \\
+ 4a \mathbb{E}_0 \left( \left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-2} \left( 1 - \alpha - 1_C(\|X\|^2) \right) \\
+ a \left( \|X\|^2 + 2\lambda\|X\|\cos\phi + \lambda^2 \right)^{-3} \right\}.
\]

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

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

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

\[
\int_0^\pi \cos^2\phi \sin^{-2}\phi \, d\phi = \frac{1}{p-1} \int_0^\pi \sin^p\phi \, d\phi = \frac{1}{p} \int_0^\pi \sin^{p-2}\phi \, d\phi.
\]

Then
\[
\frac{\partial^2}{\partial \lambda^2} \Delta(\theta) \bigg|_{\theta=0} = 4a \mathbb{E}_0 \left[ ||X||^{-4} (1-\alpha-1)(||X||^2) (1 - \frac{4}{p}) + a||X||^{-6} (1 - \frac{6}{p}) \right]
\]

\[
= \frac{4a}{p} \left\{ (p-4) \frac{1-\alpha-p(\chi_{p+4}^2 \leq c_\alpha^2)}{(p-2)(p-4)} + a \frac{p-6}{(p-2)(p-4)(p-6)} \right\}
\]

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

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

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 \((a \leq (p-4)(\alpha-\eta))\), or decreasing then increasing \((a \geq (p-4)(\alpha-\eta))\) for \( p \geq 7 \). In both cases, as \( c_\alpha^2 \geq p \), they indicate that the domination of \( \gamma_0 \) by \( \gamma_a \) should be uniform.

### 3.3 An Improved Version of the Confidence Estimator

As noted at the beginning of Section 3.1, the choice of \( \gamma_a \) does not seem quite natural since \( \gamma_a(X) > 1 \) for \( X \) close to \( a \). Therefore, we consider

\[
\gamma_a^+(X) = \left(1 - \alpha + \frac{a}{||X||^2}\right)^+
\]

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

This estimator obviously dominates \( \gamma_a \) for quadratic loss and should also dominate \( \gamma_0 \) for \( a \leq 2(p-4)(\alpha-\nu) \). As, asymptotically, the truncation is negligible, the bound (3.4) is applicable to \( \gamma_a^+ \). But it follows from the simulations that the constant \( a \) can exceed the bound (3.3) and still dominate \( \gamma_0 \). Note that here, given the nature of the loss, we cannot reject \( \gamma_a^+ \) as inadmissible: there is no complete class theorem.
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.

<table>
<thead>
<tr>
<th>p</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( \alpha - \nu )</td>
<td>0.0387</td>
<td>0.0348</td>
<td>0.0321</td>
<td>0.0299</td>
<td>0.0283</td>
</tr>
<tr>
<td>( \alpha - \eta )</td>
<td>0.0491</td>
<td>0.0471</td>
<td>0.0453</td>
<td>0.0436</td>
<td>0.0422</td>
</tr>
</tbody>
</table>

*Table 3.1. Some values of the bounds (3.3) and (3.5)*
4. Conclusion.

This work is obviously only a first step as we have not established formally the domination of \( \gamma_a \) over \( \gamma_0 \). 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 \((C(X), \gamma(X))\), 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 \( C_\delta \) 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 \((C_\delta(X), \gamma(X))\) that dominates the pair \((C_\delta(X), 1 - \alpha)\) in the sense that, for a fixed set \( C_\delta \), \( \gamma(X) \) is a better confidence estimator than \( 1 - \alpha \). Here, we show that for the usual confidence set \( C_X \), there is a pair \((C_X(X), \gamma(X))\) that dominates \((C_X(X), 1 - \alpha)\). 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


Figure 3.1 Risk Ratios $R(y, \theta)/R(1-\alpha, \theta)$ for $\gamma_a$ and $\gamma_{a^*}$ where $a^* = (p-4)(\alpha-\nu)$, and $p = 7$. Calculations based on 1000 simulations.
Figure 3.2: Expected Values of $\gamma_a$ and $\gamma_a^+$ for $a=a^*$, $2a^*$, and $p=7$.
Calculations based on 1000 simulations.