Statistica Sinica 4(1994), 617-638

AN EMPIRICAL BAYES CONFIDENCE REPORT

Edward I. George and George Casella

University of Texas at Austin and Cornell University

Abstract: For the problem of estimating a multivariate normal mean, it is known that confidence sets recentered at shrinkage estimators offer strictly larger coverage probability than the usual confidence set. Unfortunately, the conventional frequentist report of a constant confidence coefficient (infimum of the coverage probabilities) fails to communicate the gain of these improved confidence sets. Through an empirical Bayes argument we introduce a confidence report for the recentered confidence set which is strictly larger than the conventional infimum report. This confidence report is shown to dominate the infimum report according to an appropriate risk criterion. By using this new confidence report, the improved confidence region provides an informative frequentist measure of precision for the shrinkage estimator about which it has been recentered.

Key words and phrases: Conditional confidence, loss estimation, shrinkage estimation, Stein estimator.

1. Introduction and Summary

An important companion problem for the point estimation of a multivariate normal mean is to provide associated confidence regions. More precisely, based on the observation of a p-dimensional multivariate normal vector

$$X \sim N_p(\theta, I), \tag{1.1}$$

the classical confidence set for θ is of the form

$$C_0(X) = \{\theta : |\theta - X| \leq c\}. \tag{1.2a}$$

For all θ, the coverage probability of this set estimator is constant,

$$P_\theta(\theta \in C_0(X)) \equiv P(\chi_p^2 \leq c^2) = 1 - \alpha, \tag{1.2b}$$

and the practitioner reports that $C_0(X)$ contains θ with confidence $1-\alpha$. Hwang and Casella (1982, 1984) showed that for $p \geq 3$ the set $C_0(X)$ can be improved by recentering at a positive part Stein estimator

$$\delta_a(X) = u_a(|X|)X, \quad \text{where} \quad u_a(r) = \max\{[1 - (a/r^2)], 0\} \tag{1.3}$$

to obtain

$$C_a(X) = \{\theta : |\theta - \delta_a(X)| \leq c\}. \tag{1.4}$$

This set dominates $C_0(X)$ in the sense that both sets are of equal volume but for a in a certain range, $C_a(X)$ has uniformly higher coverage probability for all θ, that is,

$$P_\theta(\theta \in C_a(X)) > P_\theta(\theta \in C_0(X)) \equiv 1 - \alpha. \tag{1.5}$$

For this situation the statistician is better off reporting $C_a(X)$ than $C_0(X)$ as a $1-\alpha$ confidence region for θ. Table 1 and Figure 1 display values of $P_\theta(\theta \in C_{p-2}(X))$ for $1-\alpha = .90$ and a variety of p and $|\theta|$ so the reader can appreciate the potential improvement offered by $C_a(X)$. Note that the improvement increases as θ gets closer to the shrinkage target 0 where $\delta_a(X)$ provides the most shrinkage.

The conventional frequentist confidence report of the infimum of the coverage probabilities is woefully inadequate for C_a. Unfortunately, both C_a and C_0 have the same infimum confidence report since

$$\inf_\theta P_\theta(\theta \in C_a(X)) = 1 - \alpha. \tag{1.6}$$

As is clear from Table 1 and Figure 1, reporting confidence $1-\alpha$ in C_a is not only misleading, but also fails to communicate to the user the potential improvement offered by C_a. It is the purpose of this paper to propose an alternative (post-experimental) report to $1-\alpha$ which better reflects the coverage of $C_a(X)$. Furthermore, coupled with $C_a(X)$, this new confidence report can provide an informative frequentist measure of precision for the Stein estimator $\delta_a(X)$.

The new confidence report which we propose to accompany $C_a(X)$, is of the form

$$\gamma_{b,d}(X) = P[\chi_p^2 \leq c^2/u_{b,d}(|X|)], \tag{1.7a}$$

where for some constants $d > b > 0$,

$$u_{b,d}(r) = [1 - b/(d + r^2)]. \tag{1.7b}$$

Note that $u_{b,d}$ is similar to the shrinkage factor of the Stein estimator $\delta_a(X)$ in (1.3). However, instead of being truncated at 0 when $|X|^2 = a$, this factor decreases continuously as $|X| \to 0$, where it is bounded by $(d-b)/d > 0$. As opposed to the infimum report in (1.6), this confidence report has the properties that $\gamma_{b,d}(X) > 1-\alpha$, and $\gamma_{b,d}(X)$ increases with the amount of shrinkage provided by $\delta_a(X)$. Table 1 and Figure 1 illustrate how, for appropriate choices of b and d, $\gamma_{b,d}(X)$ will on average be much closer to $P_\theta(\theta \in C_a(X))$ than $1-\alpha$.

Table 1. For $1-\alpha = .9$, coverage probabilities of C_{p-2} and expected values of confidence estimators $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$. The constants used are $a = p-2$, $b = (p-2)^2/p$, and $d = d_{mn}$ from (2.16). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at $\gamma_{\max}$ of (2.15).

	$p=5$			$p=8$			$p=15$		
$\lvert\theta\rvert$	P_θ	$\gamma_{b,d}$	γ_{LB}	P_θ	$\gamma_{b,d}$	γ_{LB}	P_θ	$\gamma_{b,d}$	γ_{LB}
0.0	.988	.957	.959	.998	.978	.969	1.00	.993	.981
1.0	.986	.953	.955	.997	.975	.966	1.00	.992	.980
2.0	.981	.944	.945	.996	.969	.959	1.00	.990	.977
3.0	.969	.934	.934	.991	.959	.950	1.00	.985	.972
4.0	.933	.925	.925	.971	.949	.941	.999	.979	.966
5.0	.921	.918	.918	.952	.940	.934	.993	.972	.960
6.0	.915	.914	.914	.938	.932	.927	.985	.964	.953
7.0	.911	.910	.910	.930	.926	.922	.975	.956	.947
8.0	.909	.908	.908	.923	.921	.918	.963	.949	.942
9.0	.907	.906	.906	.918	.917	.914	.954	.942	.936
10.0	.906	.905	.905	.914	.914	.912	.945	.937	.932
15.0	.902	.902	.902	.906	.906	.905	.917	.917	.916
20.0	.901	.901	.901	.902	.903	.902	.908	.908	.907
25.0	.901	.900	.900	.901	.901	.901	.903	.903	.903
30.0	.900	.900	.900	.900	.900	.900	.900	.900	.900

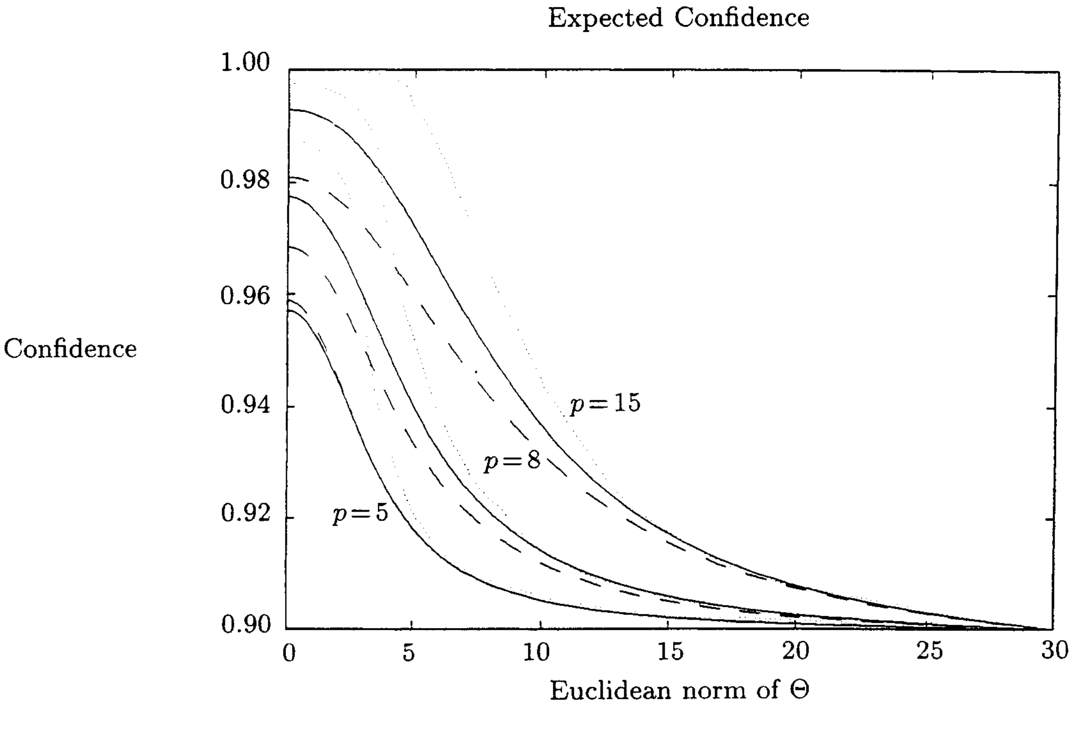

Figure 1. For $1-\alpha = .9$, coverage probabilities of C_{p-2} (dotted lines) and expected values of confidence estimators $\gamma_{b,d}(X)$ (solid lines) and $\gamma_{LB}(X)$ (dashed lines). The constants used are $a = p-2$, $b = (p-2)^2/p$, and $d = d_{mn}$ from (2.16). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at $\gamma_{\max}$ of (2.15).

The formal approach that we take in this paper is to consider a confidence procedure as a pair $\langle C(X), \gamma(X)\rangle$ where $C(X)$ is a set estimator and $\gamma(X)$ is a quoted confidence estimator, the confidence that θ is in the set $C(X)$. Such an approach is natural in both the Bayesian and conditional frequentist setting (see Kiefer (1977) and Berger and Wolpert (1984)). By taking this approach the problem of reporting post-experimental "confidence" can be put in a decision theory framework and treated as an additional estimation problem. In Section 2.1, we argue that the relevant estimation problem for this purpose is the estimation of the indicator of realized coverage under squared-error loss. This is contrasted with the alternative approach of estimating pre-experimental coverage probability.

In Section 2.2, the confidence report $\gamma_{b,d}$ is motivated by an empirical Bayes argument similar to that used by Efron and Morris (1973) and Morris (1983) to motivate $\delta_a(X)$. In this context, $C_a(X)$ can be seen as an empirical Bayes estimate of a highest posterior density credible set. The confidence report $\gamma_{b,d}$ is then obtained as an empirical Bayes estimate of the corresponding posterior probability associated with such a region. In effect, a complete frequentist confidence report is derived as a complement to the complete Bayes credible set report.

In Section 2.3, the report $\gamma_{b,d}$ is supported by a comparison of $P_\theta(\theta \in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$. We show that for an appropriate choice of b, (depending on a), $P_\theta(\theta \in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ will agree for large $|\theta|$. In Section 2.4, choices of d are suggested to guarantee that $\gamma_{b,d}(X) \leq \max_\theta P_\theta(\theta \in C_a(X))$. Otherwise, reporting $\gamma_{b,d}$ could result in negatively biased relevant betting procedures (see Casella (1987, 1988)). The proposed choices of b and d seem to lead to substantial practical improvements as is borne out, for example, in Table 1 and Figure 1. Similar improvements were obtained by Robert and Casella (1994) for the usual confidence set.

In Section 3, we evaluate $\gamma_{b,d}$ in a number of ways. We first consider, in Section 3.1, the evaluation of $\gamma_{b,d}$ according to the risk criteria described in Section 2.1. It is shown that, for certain choices of a, b, d, $\gamma_{b,d}$ uniformly dominates $1-\alpha$ as a confidence report for C_a. In Section 3.2, we show that for certain choices of a, b, d, $\gamma_{b,d}$ has the property of frequentist validity, namely (see Berger (1988)),

$$E_\theta\gamma_{b,d}(X) \leq P_\theta(\theta \in C_a(X)) \quad \text{for all } \theta. \tag{1.8}$$

This property, like the infimum report (1.6), insures that in the long run the confidence report will not overstate the true coverage probability. Although some authors believe this is a crucial property, (for example Lu and Berger (1989), or Hwang and Brown (1991)), we do not view the property of frequency validity as a main desideratum of a confidence estimator, and are really not sure of the overall value of the property. In particular, to deliberately underestimate (on average) the true coverage seems to conflict with our goal of obtaining small risk.

In Section 3.3, the practical potential of $\gamma_{b,d}$ is substantiated by simulating its expectation and risk for some special cases. Furthermore, we also compare $\gamma_{b,d}$ with a confidence report for $C_a(X)$ proposed by Lu and Berger (1989). They showed that a report of the form

$$\gamma_{LB}(X) = 1 - \alpha + \frac{bp}{dp + |X|^2}\alpha, \tag{1.9}$$

for some positive b and d, uniformly dominates the infimum report $1-\alpha$ in terms of communication risk R_I (defined in Section 2). Although it is not obtained by Bayesian considerations, γ_{LB} does have the intuitively desirable properties that $\gamma_{LB}(X) > 1-\alpha$ and $\gamma_{LB}(X)$ increases with the amount of shrinkage provided by $\delta_a(X)$. Lu and Berger show that γ_{LB} also has the property of frequentist validity. However, our numerical results show that $\gamma_{b,d}(X)$ compares favorably with γ_{LB}.

Section 4 contains concluding remarks. Finally, to improve the readability of the paper, the necessary technical lemmas have been placed in an Appendix.

2. Confidence Estimation

As described in the introduction, we consider a confidence procedure as a pair $\langle C(X), \gamma(X)\rangle$ where $C(X)$ is a set estimator and $\gamma(X)$ is a quoted confidence estimator, the confidence that θ is in the set $C(X)$. In this setting, the selection of $\gamma(X)$ can be put in a decision theory framework by identifying the problem of reporting confidence with the estimation of an unknown "parameter". In the next section, we consider which "parameter" best serves this purpose.

2.1. Confidence as an estimate of coverage

In order to treat the choice of a confidence report $\gamma(X)$ as an estimation problem, we note that the key concern of the experimenter is whether, for a given realization $X = x$, the parameter θ is in the set $\{\theta : \theta \in C(x)\}$. This suggests that the indicator function

$$I[\theta \in C(x)] = \begin{cases} 1, & \text{if } \theta \in C(x), \\ 0, & \text{if } \theta \notin C(x), \end{cases} \tag{2.1}$$

is the primary object of interest. Indeed, the only concern of the experimenter (with respect to set estimation) is whether the realized set covers the parameter, which is exactly the information captured by $I[\theta \in C(x)]$. The problem of confidence estimation is thus identified with the problem of estimating the unknown "parameter" $I[\theta \in C(x)]$.

When evaluating an estimator of $I[\theta \in C(x)]$ from a frequentist view, we must evaluate estimation of $I[\theta \in C(X)]$, as X has not yet been realized. A decision theoretic approach to estimation of $I[\theta \in C(X)]$ leads us to consider the risk function

$$R_I(\theta, \langle C(X), \gamma(X)\rangle) = E_\theta \left[I[\theta \in C(X)] - \gamma(X)\right]^2, \tag{2.2}$$

called the communication risk by Lu and Berger (1989). This risk function is also used by Robinson (1979) in considering conditional properties of procedures, and by Godambe (1961).

An alternative to estimating $I[\theta \in C(X)]$ as an approach to confidence estimation, and one we consider less useful, is to estimate instead the coverage probability $P_\theta(\theta \in C(X))$. In doing so, one may consider the risk function

$$R_P(\theta, \langle C(X), \gamma(X)\rangle) = E_\theta \left[P_\theta(\theta \in C(X)) - \gamma(X)\right]^2. \tag{2.3}$$

Compared to $I[\theta \in C(X)]$, the pre-experimental coverage probability $P_\theta(\theta \in C(X))$ is of only secondary importance. Perhaps the most severe criticism of $P_\theta(\theta \in C(X))$ as a confidence measure is that it has no meaning as a post-data measure (after $X = x$ is observed), which is what an experimenter is concerned with. After the data are observed, and the experimenter constructs $C(x)$, the probability $P_\theta(\theta \in C(X))$ becomes irrelevant. Moreover, the post-data analog $P_\theta(\theta \in C(x))$ is meaningless as a probability — it is an indicator function.

On a more technical side it is revealing to consider the identity

$$\begin{aligned} R_I(\theta, \langle C(X), \gamma(X)\rangle) =& R_P(\theta, \langle C(X), \gamma(X)\rangle) - 2\mathrm{Cov}[I[\theta \in C(X)], \gamma(X)] \\ &+ P_\theta(\theta \in C(X))[1 - P_\theta(\theta \in C(X))], \end{aligned} \tag{2.4}$$

where $\mathrm{Cov}[I[\theta \in C(X)], \gamma(X)]$ is the covariance between $I[\theta \in C(X)]$ and $\gamma(X)$. Since the third term in (2.4) is beyond our control, the R_I criterion is composed of the R_P criterion and a covariance criterion. Intuitively, $\mathrm{Cov}[I[\theta \in C(X)], \gamma(X)] \geq 0$ is a desirable property for γ, and R_I imposes a penalty if this is not the case. This quantity is not taken into account by R_P.

Another reason for preferring R_I to R_P is that squared error loss in $I[\theta \in C(X)]$ is a proper scoring rule. In the Bayesian setting, minimizing the posterior risk with this loss would yield the posterior probability of $C(X)$ as the optimal γ. More precisely, for a fixed set estimator $C(X)$, consider the class of Bayes rules against the loss

$$[I[\theta \in C(X)] - \gamma(X)]^2. \tag{2.5}$$

For $X \sim f(x|\theta)$ and prior $\pi(\theta)$, the Bayes rule is the posterior probability

$$\gamma^\pi(x) = \frac{\int_C f(x|\theta)\pi(\theta)d\theta}{\int_\Theta f(x|\theta)\pi(\theta)d\theta}. \tag{2.6}$$

So for this estimation problem, that of estimating $I[\theta \in C(X)]$, the Bayes (and thus some admissible) rules against squared error loss are posterior, or post-data, coverage probabilities. In contrast, estimation of $P_\theta(\theta \in C(X))$ will not yield Bayes rules as posterior probabilities.

In spite of our marked preference for R_I over R_P as an evaluation criterion for a confidence report, we show in Section 3 that the confidence procedure $\langle C_a(X), \gamma_{b,d}(X)\rangle$ for certain choices of a, b, d, performs very well with respect to both criteria. In particular, for some a, b, d, $\langle C_a(X), \gamma_{b,d}(X)\rangle$ dominates $\langle C_a(X), 1-\alpha\rangle$ both in terms of R_I and R_P in the following usual sense. For a set estimator $C(X)$, the confidence report $\gamma_1(X)$ is said to dominate $\gamma_2(X)$ with respect to the risk R if

$$R(\theta, \langle C(X), \gamma_1(X)\rangle) \leq R(\theta, \langle C(X), \gamma_2(X)\rangle) \tag{2.7}$$

for all θ with strict inequality for some θ.

2.2. Empirical Bayes motivation

A popular and often successful approach for discovering estimators with good risk properties is the empirical Bayes approach, (see Berger (1985)). Johnstone (1988) used an empirical Bayes argument to derive an improved loss estimate and establish its minimaxity. In this section, we show that the confidence procedure $\langle C_a(X), \gamma_{b,d}(X)\rangle$ defined by (1.4) and (1.7) can be motivated as an empirical Bayes approximation to a Bayes credible region and its associated posterior coverage probability. The approximated Bayes procedure is obtained by assuming that θ is a realization from a conjugate normal prior

$$\theta \sim N_p(0, \tau^2 I). \tag{2.8a}$$

Following the development in Efron and Morris (1973), the posterior distribution for θ under (2.8a) is

$$\theta | X \sim N_p((1-B)X, (1-B)I), \quad \text{where } B = 1/(1+\tau^2). \tag{2.8b}$$

Based on this distribution the highest posterior density credible interval estimate for θ is

$$C_B(X) = \{\theta : |\theta - (1-B)X| \leq c\} \tag{2.9a}$$

with posterior coverage probability

$$P_{\theta|X}[\theta \in C_B(X)] = P[\chi_p^2 \leq c^2/(1-B)]. \tag{2.9b}$$

Adopting the perspective that the prior (2.8) is correctly specified but that τ^2 is unknown, an empirical Bayes approximation of (2.10) can be obtained by substituting a data based estimate of $(1-B)=\tau^2/(1+\tau^2)$. We consider the class of empirical Bayes estimates of the form

$$(1-\hat{B})=u_{b,d}(|X|)=[1-b/(d+|X|^2)], \tag{2.10}$$

where $d>b>0$. Such estimates may be based on the marginal distribution of X, $X\sim N_p(0,(1+\tau^2)I)$. Note that as opposed to the truncated estimates of $1-B$ typically used in Stein estimators, the estimate (2.10) is continuously decreasing as $|X|\to 0$ where it is bounded by $(d-b)/d>0$. Substitution of $u_{b,d}(|X|)$ into $P_{\theta|X}[\theta\in C_B(X)]$ yields the confidence function $\gamma_{b,d}(X)$ in (1.7).

2.3. Equating $P_\theta(\theta\in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ and the choice of b

In this section, we show that for an appropriate choice of b, (depending only on a), $P_\theta(\theta\in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ will agree for large $|\theta|$. This not only supports the form of $\gamma_{b,d}$ as a promising confidence report for C_a, but provides guidance as to which b will provide the best confidence report. We should point out that the goal of equating $P_\theta(\theta\in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ is not at odds with the goal of estimating $I[\theta\in C_a(x)]$. Indeed, because of the close connection between R_I and R_P in (2.4), agreement of $P_\theta(\theta\in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ is a consequence of a good confidence estimator. We also note that the choice of b in (2.14) below is further supported by risk considerations based on Theorem 3.1 below.

The behavior of $P_\theta(\theta\in C_a(X))$ for large $|\theta|$ was discovered by Hwang and Casella (1984) who, adapting arguments of Berger (1980), showed

$$P_\theta(\theta\in C_a(X))=1-\alpha+(a/p)(2(p-2)-a)c^2f_p(c^2)|\theta|^{-2}+O(|\theta|^{-3}), \tag{2.11}$$

where f_p is the χ^2_p density. Below, we obtain an analogous form for the behavior of $\gamma_{b,d}(X)$ for large $|\theta|$. In order to describe these and another asymptotic results, we use the following modified notation for the order of approximation. Let $h_1(b,\theta)$ and $h_2(b,\theta)$ be two real valued functions of b and θ. We write

$$h_1(b,\theta)=o_b(h_2(b,\theta)) \text{ iff for some } B>0, \quad \lim_{|\theta|\to\infty}\sup_{b\in[0,B]} h_1(b,\theta)/h_2(b,\theta)=0, \tag{2.12a}$$

and

$$h_1(b,\theta)=O_b(h_2(b,\theta)) \text{ iff for some } B>0, \quad \lim_{|\theta|\to\infty}\sup_{b\in[0,B]} h_1(b,\theta)/h_2(b,\theta) \text{ is finite.} \tag{2.12b}$$

Note that o_b and O_b strengthen the usual o and O orders of approximation to be uniform in b over a neighborhood of zero. The following shows that $E_\theta\gamma_{b,d}(X)$ may be uniformly approximated in this sense.

Theorem 2.1. *For $c^2 \geq p-2$,*

$$E_\theta \gamma_{b,d}(X) = 1 - \alpha + bc^2 f_p(c^2)|\theta|^{-2} + O_b(b|\theta|^{-4}). \tag{2.13}$$

Proof. Let $g_{b,d}(|X|^2) = \gamma_{b,d}(X) - (1-\alpha) = P[c^2 \leq \chi_p^2 \leq c^2/u_{b,d}(|X|)]$. The result now follows from the fact that $E_\theta g_{b,d}(|X|^2) = bc^2 f_p(c^2)|\theta|^{-2} + O_b(b|\theta|^{-4})$ which is stated and proved as (A.5a) of Lemma A.3 in the appendix.

Comparison of (2.11) and (2.13) shows that for large $|\theta|$, $\gamma_{b,d}(X)$ with

$$b = (a/p)(2(p-2) - a), \tag{2.14}$$

equates $P_\theta(\theta \in C_a(X))$ and $E_\theta \gamma_{b,d}(X)$ (up to second order). In particular, it shows that for the choice $a = p-2$ recommended by Hwang and Casella, one should choose $b = (p-2)^2/p \approx p-2$.

2.4. The choice of d

For choosing d, we recommend that $\gamma_{b,d}$ should never be larger than the maximum of $P_\theta(\theta \in C_a(X))$. Otherwise such a report would always overestimate coverage with X near 0, and could result in negatively biased relevant betting procedures (see Casella (1987, 1988)). This maximum coverage probability, which we denote $\gamma_{\max}$, occurs when $\theta = 0$, and is given by

$$\begin{aligned} \gamma_{\max} &= \max_\theta P_\theta(\theta \in C_a(X)) = P_0(0 \in C_a(X)) \\ &= P_0(0 \leq |X|^2 \leq (c^2 + 2a + c(c^2+4a)^{1/2})/2), \end{aligned} \tag{2.15}$$

where the last equality in (2.15) is stated and proved as Lemma A.7 in the appendix. It then follows from (1.7) and (2.15) that in order to insure $\gamma_{b,d}(0) \leq P_0(0 \in C_a)$, we should choose

$$d \geq d_{\min} \equiv \frac{2a + c(c^2+4a)^{1/2} + c^2}{2a + c(c^2+4a)^{1/2} - c^2} b. \tag{2.16}$$

The choice $d = d_{\min}$ worked well in the simulations presented in Section 3.3.

3. Evaluation of $\gamma_{b,d}$

3.1. Risk domination

In this section, we focus on the evaluation of $\gamma_{b,d}(X)$ as a confidence report for $C_a(X)$ in terms of the risk R_I in (2.2). Similar results will be seen to follow immediately for R_P of (2.3). Note that although expressions (2.11) and (2.12) show that for $b = (a/p)(2(p-2)-a)$, $P_\theta(\theta \in C_a(X))$ and $E_\theta \gamma_{b,d}(X)$ will agree for large $|\theta|$, this does not guarantee that $\gamma_{b,d}(X)$ will have good risk properties with respect to R_I (or R_P).

Before stating our main risk evaluation results, we point out that we are only interested in risk assessment for values of a where C_a dominates C_0 in terms of coverage probability. Hwang and Casella (1984) show analytically that the coverage dominance of the recentered region $C_a(X)$ over $C_0(X)$ occurs whenever $0 < a \le \min\{a_1, a_2\}$ where a_1 and a_2 are the unique solutions respectively of

$$\left[\frac{c + (c^2+4a)^{1/2}}{2\sqrt{a}}\right]^{p-2} e^{-c\sqrt{a}/2} = 1 \tag{3.1a}$$

and

$$\left[\frac{(c^2+4a)^{1/2} - c}{2\sqrt{a}}\right]\left[\frac{c + (c^2+4a)^{1/2}}{\sqrt{a}}\right]^{p-1} e^{-c\sqrt{a}} = 1. \tag{3.1b}$$

We should add, however, that numerical calculations of Hwang and Casella strongly suggest that domination holds for values of a larger than $p-2$, but not as large as $2(p-2)$, (the bound for point estimation). They recommend the choice $a = p-2$, mainly due to the optimality of the point estimator for this choice.

Because we are most interested in the comparison of the report $\gamma_{b,d}$ with the conservative report $1-\alpha$, we will express our main results in terms of the risk reduction function,

$$H_\theta(a,b,d) = R_I(\theta, \langle C_a(X), 1-\alpha\rangle) - R_I(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle). \tag{3.2}$$

The following result reveals the risk reduction behavior of $\gamma_{b,d}$ for large $|\theta|$.

Theorem 3.1. *Letting f_p be the χ^2_p density,*

$$H_\theta(a,b,d) = [(a/p)(2(p-2)-a) - b/2]\, 2bc^4 f_p^2(c^2)|\theta|^{-4} + O_b(b|\theta|^{-5}). \tag{3.3}$$

Proof. To show (3.3), rewrite (3.2) as

$$H_\theta(a,b,d) = 2E_\theta[I[\theta \in C_a(X)] - (1-\alpha)]g_{b,d}(|X|^2) - E_\theta g_{b,d}^2(|X|^2). \tag{3.4}$$

The result (3.3) now follows directly by combining (3.4) with

$$E_\theta g_{b,d}^2(|X|^2) = b^2c^4 f_p^2(c^2)|\theta|^{-4} + O_b(b|\theta|^{-6}),$$

which is stated and proved as (A.5b) of Lemma A.3 in the appendix, and

$$\begin{aligned} &E_\theta[I[\theta \in C_a(X)] - (1-\alpha)]g_{b,d}(|X|^2) \\ &= \left[(a/p)(2(p-2)-a)c^2 f_p(c^2)|\theta|^{-2}\right] bc^2 f_p(c^2)|\theta|^{-2} + O_b(b|\theta|^{-5}), \end{aligned}$$

which is stated and proved as Lemma A.4 in the appendix.

Theorem 3.1 shows that for $|\theta|$ large, the choice $b = (a/p)(2(p-2)-a)$ in (2.14), which equates (2.11) and (2.13) up to second order terms, also provides the maximum risk reduction over $1-\alpha$ in terms of R_I. It also follows from (3.3) that uniform risk domination of $\gamma_{b,d}(X)$ over $1-\alpha$ with respect to R_I is impossible when $b \geq 2(a/p)(2(p-2)-a)$, as H_θ will then be negative for large $|\theta|$.

Of course, it is of more interest to know if H_θ is strictly positive for all θ, for in that case $\gamma_{b,d}$ will uniformly dominate $1-\alpha$ as a confidence report for C_a. The following results show that for any $C_a(X)$ known to dominate C_0 in coverage probability, there exist choices of b and d such that $\gamma_{b,d}(X)$ will dominate $1-\alpha$ in terms of R_I.

Theorem 3.2. *For $p \geq 5$, $0 < a \leq a^* \equiv \min\{p-4, a_1, a_2\}$ and d large enough, there exists $b^* > 0$ such that for $0 < b \leq b^*$,*

$$R_I(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle) < R_I(\theta, \langle C_a(X), 1-\alpha\rangle) \text{ for all } \theta. \tag{3.5}$$

Proof. Begin by fixing $a' \in (0, a^*]$. We will show that for d large enough, there exists b^* such that for $0 < b \leq b^*$, $H_\theta(a', b, d) > 0$ for all θ. First, a consequence of Theorem 3.1 is that for $0 < b' < 2a'(2(p-2)-a')/p$,

$$\lim_{|\theta|\to\infty} \inf_{b\in(0,b']} b^{-1}|\theta|^4 H_\theta(a', b, d) > 0. \tag{3.6}$$

It then follows that for each d, there exists $M_d > 0$ (which does not depend on b) such that

$$H_\theta(a', b, d) > 0 \text{ for all } b \in (0, b'] \text{ and } |\theta| > M_d. \tag{3.7}$$

Next, Lemma A.5 shows that for d large enough $(\partial/\partial b)H_\theta(a', 0, d) > 0$ for all θ. Furthermore, $(\partial/\partial b)H_\theta(a', b, d)$ is a continuous function of b and θ, and so is uniformly continuous on the closed set $\{(b,\theta) : 0 \leq b \leq b' \text{ and } |\theta| \leq M_d\}$. Thus, there exists $0 < b^* \leq b'$ such that

$$(\partial/\partial b)H_\theta(a', b, d) > 0 \text{ over } \{(b,\theta) : 0 \leq b \leq b^* \text{ and } |\theta| \leq M_d\}. \tag{3.8}$$

But since $H_\theta(a', 0, d) = 0$ for all θ, (3.8) implies

$$H_\theta(a', b, d) > 0 \text{ for all } b \in (0, b^*] \text{ and } |\theta| \leq M_d. \tag{3.9}$$

The desired result now follows from (3.7) and (3.9) since $b^* \leq b'$ by construction.

We should point out that Theorem 3.2 is an existence proof since explicit bounds for b^* are not obtained. This limitation was also a characteristic of the risk dominance result of γ_{LB} in Lu and Berger (1989). Unfortunately, the constructive result remains elusive. Our next result, which is stated and proved as Lemma A.8 in the appendix, shows that for all $b > 0$, $\gamma_{b,d}(X)$ dominates $1-\alpha$ at $\theta = 0$.

Theorem 3.3. *For* $p \geq 5$, $0 < a \leq a^* \equiv \min\{p-4, a_1, a_2\}$, $d > b > 0$ *and* $d \geq d_{\min}$,

$$R_I(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle) < R_I(\theta, \langle C_a(X), 1-\alpha\rangle) \text{ for } \theta = 0.$$

As mentioned before, results established for the risk function R_I in (2.2) can be extended in a straightforward manner to R_P in (2.3), which is not surprising considering the relationship (2.4). The following results for R_P, which are stated without proof, are analogous to those for R_I in Theorems 3.1, 3.2 and 3.3, and can be proved similarly.

Theorem 3.4.

(i) $R_P(\theta, \langle C_a(X), 1-\alpha\rangle) - R_P(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle)$
$= [(a/p)(2(p-2)-a) - b/2]2bc^4 f_p^2(c^2)|\theta|^{-4} + O_b(b|\theta|^{-5})$,
where f_p *is the* χ_p^2 *density.*

(ii) *For* $p \geq 5, 0 < a \leq a^* \equiv \min\{p-4, a_1, a_2\}$ *and d large enough, there exists* $b^* > 0$ *such that for* $b \in (0, b^*]$,

$$R_p(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle) < R_P(\theta, \langle C_a(X), 1-\alpha\rangle) \text{ for all } \theta.$$

(iii) *For* $p \geq 5, 0 < a \leq a^* \equiv \min\{p-4, a_1, a_2\}$ *and* $d > b > 0$,

$$R_P(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle) < R_P(\theta, \langle C_a(X), 1-\alpha\rangle) \text{ for } \theta = 0.$$

3.2. Frequentist validity

In this section, we show that for certain choices of a, b, d that $\gamma_{b,d}$ has the conservative property of frequentist validity for C_a in (1.8). The fact that $\gamma_{b,d}(X)$ is also frequency valid is an interesting added attraction, since the empirical Bayes motivation of the confidence estimator does not take this property into account. Having a confidence estimator that has expectation uniformly smaller than the coverage probability is a conservative tactic, and is a reasonable practical inference property. Unfortunately, it does not lead to optimal behavior against a loss function in unconstrained problems, and is thus not a property of admissible rules. It may also lead to non-coherent procedures since, if the true probability is always underestimated, then relevant sets will exist, and conditional inferences will be conservative.

Theorem 3.5. *For* $0 < a \leq a^* \equiv \min\{p-4, a_1, a_2\}$, *there exists* $b^* > 0$ *such that for all* $b \in (0, b^*]$ *and* $d > b$,

$$E_\theta \gamma_{b,d}(X) \leq P_\theta(\theta \in C_a(X)) \quad \textit{for all } \theta.$$

Proof. We begin as in the proof of Theorem 2.1 in Lu and Berger (1989). It follows from (1.5) and (2.11), that for $a \in (0, a^*]$ and $b < p$, there exists $\epsilon > 0$ such that

$$\inf_\theta \{(|\theta|^2 + p + d - b)[P_\theta(\theta \in C_a(X)) - (1-\alpha)]\} \geq \epsilon.$$

Thus,

$$P_\theta(\theta \in C_a(X)) > 1 - \alpha + \epsilon/(|\theta|^2 + p + d - b) \text{ for all } \theta.$$

Since $\gamma_{b,d}(X) = 1 - \alpha + g_{b,d}(|X|^2)$, it suffices to show that b^* can chosen so that for all $b \in (0, b^*]$,

$$E_\theta g_{b,d}(|X|^2) \leq \epsilon/(|\theta|^2 + p + d - b) \quad \text{for all } \theta.$$

This is stated and proved as Lemma A.6 in the appendix.

3.3. Numerical results

To further investigate the performance of $\gamma_{b,d}$, simulations were carried out. Based on the considerations described in Section 2.3 we investigated the case $a = p - 2$, $b = (p-2)^2/p$, $d = d_{\min}$ (from (2.16)). Although the calculations were performed for many values of $1 - \alpha$ and p, we report here only on the choices $1 - \alpha = .90$ and $p = 5, 8, 15$. For these choices, the actual coverage probability of C_{p-2} was estimated for a large number of $|\theta|$ values.

In Table 1 and Figure 1 these coverage probabilities are compared with $E_\theta(\gamma_{b,d}(X))$ and $E_\theta(\gamma_{LB}(X))$. (We use the choice of constants for γ_{LB} recommended by Lu and Berger (1989), but make one further modification which improves the performance of γ_{LB}. We truncate γ_{LB} at $\gamma_{\max}$ of (2.15).) These show that $\gamma_{b,d}$ and γ_{LB} are close to $P_\theta(\theta \in C_{p-2}(X))$, with $\gamma_{b,d}$ being closer for larger p. Table 1 also show that the particular choices of $\gamma_{b,d}$ and γ_{LB} are frequency valid for $p = 5$, 8 and 15. Table 2 and Figure 2 provide a risk comparison of $\gamma_{b,d}$, γ_{LB} and $1 - \alpha$ based on R_I. We see that both $\gamma_{b,d}$ and γ_{LB} provide substantial risk improvement over $1 - \alpha$, with $\gamma_{b,d}$ better for larger p. These also show that the substantial risk improvement is concentrated where C_{p-2} achieves its greatest improvement in coverage probability, with everything collapsing together as $|\theta|$ becomes large.

Table 2. For $1-\alpha=.9$, risk of the confidence estimators $1-\alpha$, $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ using squared error loss. The constants used are $a=p-2$, $b=(p-2)^2/p$, and $d=d_{mn}$ from (2.16). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at $\gamma_{\max}$ of (2.15).

	$p=5$			$p=8$			$p=15$		
$\|\theta\|$	$1-\alpha$	$\gamma_{b,d}$	γ_{LB}	$1-\alpha$	$\gamma_{b,d}$	γ_{LB}	$1-\alpha$	$\gamma_{b,d}$	γ_{LB}
0.0	.0196	.0123	.0122	.0122	.0030	.0035	.0100	.0001	.0004
1.0	.0210	.0141	.0140	.0123	.0033	.0038	.0100	.0001	.0005
2.0	.0252	.0194	.0193	.0137	.0052	.0058	.0101	.0003	.0007
3.0	.0350	.0310	.0310	.0171	.0096	.0102	.0103	.0006	.0012
4.0	.0633	.0630	.0631	.0336	.0294	.0298	.0109	.0016	.0022
5.0	.0733	.0735	.0735	.0483	.0463	.0465	.0152	.0070	.0076
6.0	.0782	.0783	.0783	.0594	.0585	.0585	.0216	.0149	.0154
7.0	.0813	.0815	.0815	.0660	.0656	.0655	.0300	.0249	.0253
8.0	.0829	.0830	.0830	.0714	.0712	.0712	.0396	.0361	.0362
9.0	.0842	.0842	.0842	.0754	.0753	.0752	.0470	.0444	.0445
10.0	.0854	.0854	.0854	.0785	.0785	.0784	.0542	.0525	.0525
15.0	.0884	.0884	.0884	.0855	.0855	.0855	.0760	.0757	.0757
20.0	.0888	.0888	.0888	.0880	.0880	.0880	.0839	.0838	.0838
25.0	.0895	.0895	.0895	.0890	.0890	.0890	.0874	.0873	.0873
30.0	.0900	.0900	.0900	.0900	.0900	.0900	.0900	.0900	.0900

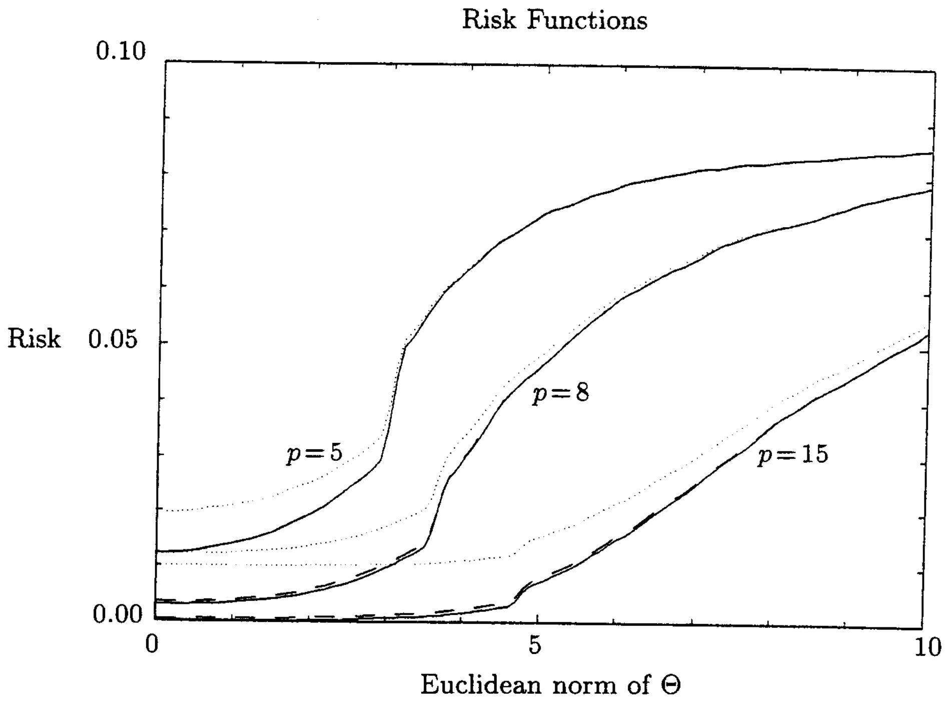

Figure 2. For $1-\alpha=.9$, risk of the confidence estimators $1-\alpha$ (dotted lines), $\gamma_{b,d}(X)$ (solid lines) and $\gamma_{LB}(X)$ (dashed lines) using squared error loss. The constants used are $a=p-2$, $b=(p-2)^2/p$, and $d=d_{mn}$ from (2.16). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at $\gamma_{\max}$ of (2.15).

Table 3. For $1-\alpha = .9$, proportional decrease in risk (3.10) of the confidence estimators $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ over $1-\alpha$. The constants used are $a = p-2$, $b = (p-2)^2/p$, and $d = d_{mn}$ from (2.16). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at $\gamma_{\max}$ of (2.15).

	$p=5$		$p=8$		$p=15$	
$\|\theta\|$	$\gamma_{b,d}$	γ_{LB}	$\gamma_{b,d}$	γ_{LB}	$\gamma_{b,d}$	γ_{LB}
0.0	.374	.380	.753	.715	.989	.956
1.0	.330	.334	.732	.690	.987	.952
2.0	.231	.232	.623	.578	.972	.929
3.0	.114	.115	.440	.401	.937	.884
4.0	.004	.003	.124	.112	.858	.797
5.0	−.002	−.002	.042	.039	.541	.498
6.0	−.002	−.002	.015	.015	.311	.286
7.0	−.002	−.002	.007	.007	.169	.158
8.0	−.001	−.001	.003	.004	.089	.085
9.0	−.000	.000	.002	.002	.054	.053
10.0	−.000	−.000	.001	.001	.032	.032
15.0	−.000	−.000	−.000	.000	.004	.004
20.0	−.000	−.000	.000	.000	.001	.001
25.0	−.000	−.000	.000	.000	.000	.000
30.0	.000	.000	.000	.000	.000	.000

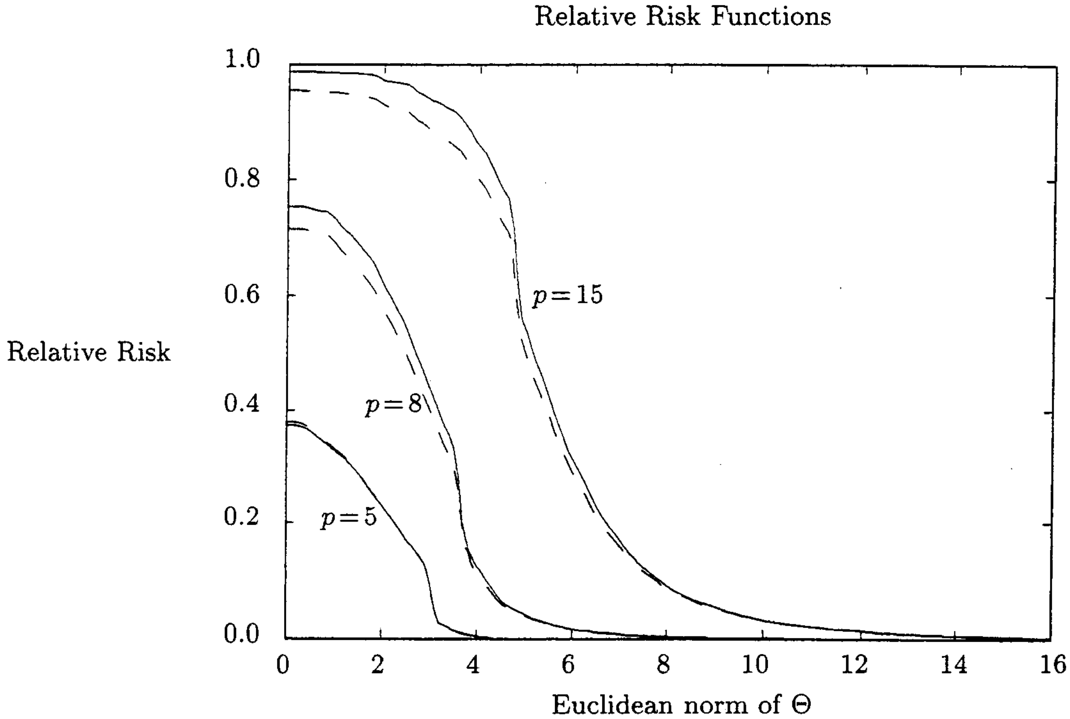

Figure 3. For $1-\alpha = .9$, proportional decrease in risk (3.10) of the confidence estimators $\gamma_{b,d}(X)$ (solid lines) and $\gamma_{LB}(X)$ (dashed lines) over $1-\alpha$. The constants used are $a = p-2$, $b = (p-2)^2/p$, and $d = d_{mn}$ from (2.16). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at $\gamma_{\max}$ of (2.15).

Lastly, Table 3 and Figure 3 show the proportional decrease in communication risk R_I of $\gamma(X)$ over $1-\alpha$, namely

$$\frac{R_I(\theta,\langle C_{p-2}(X),1-\alpha\rangle)-R_I(\theta,\langle C_{p-2}(X),\gamma(X)\rangle)}{R_I(\theta,\langle C_{p-2}(X),1-\alpha\rangle)} \tag{3.10}$$

for $\gamma(X)=\gamma_{b,d}(X)$ and $\gamma(X)=\gamma_{LB}(X)$. The figure shows clearly the improvement of $\gamma_{b,d}$ over both γ_{LB} and $1-\alpha$ for larger p and smaller $|\theta|$.

4. Conclusions

Treating the confidence estimation problem in a decision theory framework has enabled us to do away with the conventional infimum report $1-\alpha=\inf_\theta P_\theta(\theta \in C(X))$, and consider more relevant and useful confidence reports. Furthermore, rather than simply estimate coverage probabilities, we have identified the problem of confidence estimation with that of estimating the indicator function $I[\theta \in C(x)]$, (x is a realization of X). It has been argued that this point of view leads to the appropriate criteria for evaluating a post-data confidence report.

The confidence report $\gamma_{b,d}$ in (1.7) has been introduced as a promising new solution to the confidence estimation problem. The form of $\gamma_{b,d}$ is motivated by an empirical Bayes argument similar to that which motivates the Stein estimator. It is shown that $\gamma_{b,d}$ can be chosen so that $P_\theta(\theta \in C(X))$ and $E_\theta\gamma_{b,d}(X)$ will essentially agree. This is not at odds with estimation of $I[\theta \in C(x)]$, but rather in agreement. The pre-data quantity $E_\theta\gamma_{b,d}(X)$ estimates the pre-data quantity $P_\theta(\theta \in C(X))$, while the post-data quantity $\gamma_{b,d}(x)$ estimates the post-data quantity $I[\theta \in C(x)]$.

In terms of the proposed risk criteria, $\gamma_{b,d}$ fares extremely well. It is seen that $\gamma_{b,d}$ can uniformly dominate $1-\alpha$ as a confidence report for C_a. Numerical evaluation shows that meaningful improvements can be obtained in practice. The fact that $\gamma_{b,d}(X)$ is also frequency valid is an interesting added attraction, since the empirical Bayes motivation of the confidence estimator does not take this property into account. However, as mentioned before, we question the value of this conservative criterion.

In conclusion, this paper has established the merits of using the improved confidence procedure $\langle C_a(X),\gamma_{b,d}(X)\rangle$ over $\langle C_a(X),1-\alpha\rangle$ and certainly $\langle C_0(X), 1-\alpha\rangle$. This procedure not only offers better coverage properties than $C_0(X)$, it also provides a meaningful report of the improvement. Furthermore, just as $\langle C_0(X),1-\alpha\rangle$ provides an associated frequentist precision measure for $\delta_0(X)= X$, the procedure $\langle C_a(X),\gamma_{b,d}(X)\rangle$ provides an associated frequentist precision measure for the Stein estimator $\delta_a(X)$. We should mention that it may be possible to improve further on $\gamma_{b,d}$ as a confidence estimator for C_a; however, the empirical Bayes form of $\gamma_{b,d}$ suggests that we may have a first approximation to an admissible rule. By mimicking a posterior probability, $\gamma_{b,d}(X)$ is a post data confidence estimator that performs quite reasonably against frequentist criteria and, hopefully, is close to admissible.

Appendix: The Lemmas

These technical lemmas form the basis of the results of this paper, and have been placed in this appendix to improve the readability of the main text. Note that many of these results make use of the uniform approximation orders o_b and O_b defined by (2.12).

Lemma A.1. *For X, $\theta \in \mathbb{R}^p$ such that $t\theta + (1-t)X \neq 0$ for any $t \in [0,1]$,*

$$|X|^{-2n} = |\theta|^{-2n} - 2n(X-\theta)'\theta|\theta|^{-2(n+1)} - n(X-\theta)'(X-\theta)|\theta^*|^{-2(n+1)} + 2n(n+1)(X-\theta)'\theta^*\theta^{*\prime}(X-\theta)|\theta^*|^{-2(n+2)},$$

where $\theta^* = t\theta + (1-t)X$ for some $t \in [0,1]$.

Proof. Straightforward application of Taylor's Theorem.

Lemma A.2. *For $p \geq 5$ and $d > b > 0$,*

$$E_\theta(|X|^2 + d - b)^{-1} = |\theta|^{-2} + O_b(|\theta|^{-4}), \tag{A.1a}$$

$$E_\theta(|X|^2 + d - b)^{-2} = |\theta|^{-4} + O_b(|\theta|^{-6}). \tag{A.1b}$$

Proof. First note that for $n = 1, 2$ and $\delta > 0$,

$$E_\theta(|X|^2 + d - b)^{-n} I[|X|^2 \leq \delta] < (d-b)^{-n} P_\theta[|X|^2 \leq \delta],$$

which has an exponential tail and hence can be ignored. Thus it suffices to show that for $n = 1, 2$ and $\delta > 0$,

$$E_\theta(|X|^2 + d - b)^{-n} I[|X|^2 > \delta] = |\theta|^{-2n} + O_b(|\theta|^{-2(n+1)}).$$

Since

$$(|X|^2 + d - b)^{-1} = |X|^{-2} - (d-b)/(|X|^2(|X|^2 + d - b)) = |X|^{-2} + O(|X|^{-4}),$$

it follows that

$$(|X|^2 + d - b)^{-n} = |X|^{-2n} + O(|X|^{-2(n+1)}). \tag{A.2}$$

Thus, it suffices to show that for $n \geq 1$

$$E_\theta|X|^{-2n} I[|X|^2 > \delta] = |\theta|^{-2n} + O(|\theta|^{-2(n+1)}). \tag{A.3}$$

From Lemma A.1, we have

$$E_\theta|X|^{-2n}I[|X|^2>\delta]=E_\theta[|\theta|^{-2n}+h_1(X,\theta)+h_2(X,\theta)]I[|X|^2>\delta], \qquad (A.4)$$

$$\begin{aligned} h_1(X,\theta) &= -2n(X-\theta)'\theta|\theta|^{-2(n+1)},\\ h_2(X,\theta) &= -n(X-\theta)'(X-\theta)|\theta^*|^{-2(n+1)}\\ &\quad +2n(n+1)(X-\theta)'\theta^*\theta^{*\prime}(X-\theta)|\theta^*|^{-2(n+2)}, \end{aligned}$$

where $\theta^*=t\theta+(1-t)X$ for some $t\in[0,1]$. (Note that the exceptional set where Lemma A.1 does not hold has measure zero and so may be ignored in obtaining (A.4)). Clearly $E_\theta|\theta|^{-2n}I[|X|^2>\delta]=|\theta|^{-2n}(1-P_\theta[|X|^2\le\delta])=|\theta|^{-2n}+O(|\theta|^{-2n}e^{-|\theta|^2/2})$. Using Cauchy-Schwarz and the fact that $E_\theta h_1(X,\theta)=0$ and $E_\theta|X-\theta|$ is bounded, it follows that $E_\theta h_1(X,\theta)I[|X|^2>\delta]=-E_\theta h_1(X,\theta)I[|X|^2\le\delta]=O(|\theta|^{-2(n+1/2)}e^{-|\theta|^2/2})$. Finally, using Cauchy-Schwarz, the fact that $E_\theta|X-\theta|^2$ is bounded and $E_\theta|\theta^*|^n=O(|\theta|^n)$, it follows that $E_\theta h_2(X,\theta)I[|X|^2>\delta]=O(|\theta|^{-2(n+1)})$. Coupled with (A.4), this shows (A.3).

Lemma A.3. *Let $g_{b,d}(|X|^2)=P[c^2\le\chi^2_p\le c^2/u_{b,d}(|X|)]$, and let f_p be the χ^2_p density. For $c^2\ge p-2$,*

$$E_\theta g_{b,d}(|X|^2)=bc^2f_p(c^2)|\theta|^{-2}+O_b(b|\theta|^{-4}), \qquad (A.5a)$$

$$E_\theta g^2_{b,d}(|X|^2)=b^2c^4f^2_p(c^2)|\theta|^{-4}+O_b(b|\theta|^{-6}). \qquad (A.5b)$$

Proof. Since $f_p(y)$ is decreasing for $y\ge p-2$, we may expand $g_{b,d}$ for such X as

$$\begin{aligned} g_{b,d}(|X|^2) &= \int_{c^2}^{c^2/u_{b,d}(|X|)} f_p(y)dy\\ &= f_p(c^2)\left[c^2/u_{b,d}(|X|)-c^2\right]+R=bc^2f_p(c^2)(|X|^2+d-b)^{-1}+R, \end{aligned} \qquad (A.6a)$$

where for some fixed K (which does not depend on b),

$$\begin{aligned} |R| &\le [f_p(c^2)-f_p(c^2/u_{b,d}(|X|))]\left[c^2/u_{b,d}(|X|)-c^2\right]\\ &\le Kb^2(|X|^2+d-b)^{-2}. \end{aligned} \qquad (A.6b)$$

(A.5a) then follows from applying (A.1) to (A.6). (A.5b) is obtained similarly by using the square of (A.6a).

Lemma A.4. *Let $g_{b,d}(|X|^2)=P[c^2\le\chi^2_p\le c^2/u_{b,d}(|X|)]$, and let f_p be the χ^2_p density. For $c^2\ge p-2$,*

$$\begin{aligned} &E_\theta[I[\theta\in C_a(X)]-(1-\alpha)]g_{b,d}(|X|^2)\\ &= \left[(a/p)(2(p-2)-a)c^2f_p(c^2)|\theta|^{-2}\right]bc^2f_p(c^2)|\theta|^{-2}+O_b(b|\theta|^{-5}). \end{aligned} \qquad (A.7)$$

Proof. Note that when $I[\theta \in C_a(X)] = 1$, $|X - \theta|$ is bounded by a constant in which case $(|X|^2 + d - b)^{-n} = |\theta|^{-2n} + O_b(|\theta|^{-2n+1})$, which follows from (A.2) and Lemma A.1. Using (A.6), we have, for X satisfying $I[\theta \in C_a(X)] = 1$,

$$g_{b,d}(|X|^2) = bc^2 f_p(c^2)|\theta|^{-2} + O_b(b|\theta|^{-3}). \tag{A.8}$$

Thus,

$$E_\theta I[\theta \in C_a(X)] g_{b,d}(|X|^2) = (E_\theta I[\theta \in C_a(X)])(bc^2 f_p(c^2)|\theta|^{-2} + O_b(b|\theta|^{-3})) \tag{A.9}$$

which, combined with (A.5a), yields

$$\begin{aligned} &E_\theta[I[\theta \in C_a(X)] - (1-\alpha)] g_{b,d}(|X|^2) \\ &= [E_\theta I[\theta \in C_a(X)] - (1-\alpha)](bc^2 f_p(c^2)|\theta|^{-2} + O_b(b|\theta|^{-3})). \end{aligned} \tag{A.10}$$

The result (A.7) is now obtained by combining (A.10) with the fact that

$$[E_\theta I[\theta \in C_a(X)] - (1-\alpha)] = (a/p)(2(p-2) - a)c^2 f_p(c^2)|\theta|^{-2} + O(|\theta|^{-3}), \tag{A.11}$$

(which does not depend on b), a reexpression of (2.11), obtained by Hwang and Casella (1984).

Lemma A.5. *For $H_\theta(a,b,d)$ in (3.2), there exists d^* such that $(\partial/\partial b)H_\theta(a,0,d) > 0$ for $a \in [0,a^*]$, $d \geq d^*$ and all θ.*

Proof. Differentiating (3.4) and making use of $\gamma_{0,d}(X) \equiv 1 - \alpha$ and $(\partial/\partial b)g_{0,d}(|X|^2) \propto (d + |X|^2)^{-1}$, yields

$$(\partial/\partial b)H_\theta(a,0,d) \propto E_\theta[I[\theta \in C_a(X)] - (1-\alpha)]/(d + |X|^2). \tag{A.12}$$

Comparing this expression with (31) of Lu and Berger (1989), the desired result follows directly from their Lemmas 3.1 and 3.2.

Lemma A.6. *For any $\epsilon > 0$, there exists $b^* > 0$ such that for $b \in (0, b^*]$ and $d > b$,*

$$E_\theta g_{b,d}(|X|^2) \leq \epsilon/(|\theta|^2 + p + d - b) \quad \textit{for all } \theta. \tag{A.13}$$

Proof. Based on (A.6), we have, for fixed $K_1, K_2 > 0$,

$$\begin{aligned} E_\theta g_{b,d}(|X|^2) &\leq K_1\,[1/u_{b,d}(|X|) - 1] + K_2\,[1/u_{b,d}(|X|) - 1]^2 \\ &= E_\theta\,[bK_1(|X|^2 + d - b)^{-1} + b^2 K_2(|X|^2 + d - b)^{-2}]. \end{aligned} \tag{A.14}$$

It follows from (26) of Lu and Berger (1989), that

$$E_\theta(|X|^2 + d - b) \leq (1 + \epsilon^*)/(|\theta|^2 + p + d - b), \tag{A.15}$$

where $\epsilon^* = (3+\sqrt{2p})/(d-b)$. A similar argument yields

$$E_\theta(|X|^2+d-b)^{-2} \le 2(1+\epsilon^{*^2})/(|\theta|^2+p+d-b). \tag{A.16}$$

Combined with (A.14), these yield

$$E_\theta g_{b,d}(|X|^2) \le [bK_1(1+\epsilon^*)+2b^2K_2(1+\epsilon^{*^2})]/(|\theta|^2+p+d-b). \tag{A.17}$$

The desired result now follows with $b^* = \min\{1, \epsilon/[K_1(1+\epsilon^*)+2K_2(1+\epsilon^{*^2})]\}$.

Lemma A.7. $P_0(0\in C_a(X)) = P_0(0 \le |X|^2 \le (c^2+2a+c(c^2+4a)^{1/2})/2)$.

Proof.

$$\begin{aligned} P_0(0\in C_a(X)) &= P_0(|\delta_a(X)|^2 \le c^2)\\ &= P_0((1-a/|X|^2)^2|X|^2 \le c^2, |X|^2 > a) + P_0(|X|^2 < a). \end{aligned}$$

The desired result now follows by noting that

$$\begin{aligned} &\{x : (1-a/x^2)^2x^2 \le c^2\}\\ &= \{x : (c^2+2a-c(c^2+4a)^{1/2})/2 \le x^2 \le (c^2+2a+c(c^2+4a)^{1/2})/2\} \end{aligned}$$

and

$$(c^2+2a-c(c^2+4a)^{1/2})/2 \le a \le (c^2+2a+c(c^2+4a)^{1/2})/2.$$

Lemma A.8. *For $p\ge 5$, $0 < a \le a^* \equiv \min\{p-4, a_1, a_2\}$, $d > b > 0$ and $d \ge d_{\min}$,*

$$R_I(\theta, \langle C_a(X), \gamma_{b,d}(X)\rangle) < R_I(\theta, \langle C_a(X), 1-\alpha\rangle) \text{ for } \theta = 0.$$

Proof. For simplicity of notation, let $I_\theta(X) \equiv I[\theta \in C_a(X)]$ and $\gamma(X) \equiv \gamma_{b,d}(X)$. Then

$$\begin{aligned} E_\theta[I_\theta(X)-\gamma(X)]^2 &= E_\theta[[I_\theta(X)-E_\theta I_\theta(X)] + [E_\theta I_\theta(X)-\gamma(X)]]^2\\ &= E_\theta[I_\theta(X)-E_\theta I_\theta(X)]^2 + E_\theta[E_\theta I_\theta(X)-\gamma(X)]^2\\ &\quad + 2E_\theta[I_\theta(X)-E_\theta I_\theta(X)][E_\theta I_\theta(X)-\gamma(X)]. \end{aligned}$$

For the cross-product term above, note that both $I_\theta(X)$ and $\gamma(X)$ are decreasing in $|X|$. Thus, at $\theta = 0$,

$$\begin{aligned} &E_0[I_0(X)-E_0I_0(X)][E_0I_0(X)-\gamma(X)]\\ &\le E_0[I_0(X)-E_0I_0(X)]E_0[E_0I_0(X)-\gamma(X)] = 0. \end{aligned}$$

So, at $\theta = 0$,

$$E_0[I_0(X) - \gamma(X)]^2 = E_0[I_0(X) - E_0 I_0(X)]^2 + E_0[E_0 I_0(X) - \gamma(X)]^2.$$

Similarly (but more easily),

$$E_0[I_0(X) - (1-\alpha)]^2 = E_0[I_0(X) - E_0 I_0(X)]^2 + E_0[E_0 I_0(X) - (1-\alpha)]^2,$$

since the cross-product term is 0 here. Thus a sufficient condition for $\gamma(X)$ to dominate $(1-\alpha)$ at $\theta = 0$ is

$$E_0[E_0 I_0(X) - \gamma(X)]^2 \le E_0[E_0 I_0(X) - (1-\alpha)]^2. \qquad \text{(A.18)}$$

This is satisfied by construction since for $\gamma = \gamma_{b,d}$ above

$$1 - \alpha < \gamma(X) \le \max P_\theta(\theta \in C_a(X)) = E_0 I_0(X).$$

Acknowledgement

The authors would like to thank Dean Foster for many helpful conversations, and the reviewers for crucial suggestions. The research of the first author was performed in part while he was visiting the Mathematical Sciences Institute at Cornell University. The research of the second author was supported by National Science Foundation Grant No. DMS89-0039.

References

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

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer-Verlag, New York.

Berger, J. O. (1988). Discussion of Casella's article. In *Statistical Decision Theory and Related Topics IV, Volume I* (Edited by S. S. Gupta and J. O. Berger), 85-90.

Berger, J. O. and Wolpert, R. W. (1984). *The Likelihood Principle*, 2nd edition. IMS Monograph Series, Institute of Mathematical Statistics, Hayward, CA.

Casella, G. (1987). Conditionally acceptable recentered set estimators. *Ann. Statist.* **15**, 1363-1371.

Casella, G. (1988). Conditionally acceptable frequentist solutions. In *Statistical Decision Theory and Related Topics IV, Volume I* (Edited by S. S. Gupta and J. O. Berger), 73-117.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors — An empirical Bayes approach. *J. Amer. Statist. Assoc.* **68**, 117-130.

Godambe, V. P. (1961). Bayesian shortest confidence intervals and admissibility. *Bull. Inst. Internat. Statist. 33rd Session (Paris)* **125**, 1-5.

Hwang, J. T. and Brown, L. D. (1991). Estimated confidence under and the validity constraint. *Ann. Statist.* **19**, 1964-1977.

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

Hwang, J. T. and Casella, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Dec.* Supplement Issue **1**, 3-16.

Johnstone, I. (1988). On inadmissibility of some unbiased estimates of loss. *Proc. of the 4th Purdue Symp. Statist. Dec. Theory and Rel. Topics* (Edited by S. S. Gupta and J. Berger), 361-379, Springer-Verlag, New York.

Kiefer, J. C. (1977). Conditional confidence statements and confidence estimators (with discussion). *J. Amer. Statist. Assoc.* **72**, 789-808.

Lu, K. L. and Berger, J. O. (1989). Estimated confidence procedures for multivariate normal means. *J. Statist. Plann. Inference* **23**, 1-19.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78**, 47-65.

Robert, C. and Casella, G. (1994). Improved confidence estimators for the usual multivariate normal confidence set. In *Proc. of the 5th Purdue Symposium on Statistical Decison Theory and Related Topics* (Edited by S. S. Gupta and J. Berger), Academic Press.

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

CBA 5.202, MSIS Department, University of Texas at Austin, TX 78712, U.S.A.
Biometrics Unit, Cornell University, Ithaca, NY 14853, U.S.A.

(Received August 1990; accepted January 1994)