ESTIMATED CONFIDENCE PROCEDURES FOR MULTIVARIATE NORMAL MEANS

by

K. L. Lu and James Berger
Purdue University

Technical Report #87-21

Department of Statistics
Purdue University

May 1987

Title : Estimated Confidence Procedures For Multivariate Normal Means

Authors : K. L. Lu and James Berger

Purdue University

Short Title : Estimated Confidence for Normal Means

Correspondence to : James Berger

Statistics Department

Purdue University

West Lafayette, IN 47907

1980 AMS Subject Classifications : Primary 62J07, 62C10; secondary 62C15.

Key Words and Phrases : estimated confidence, communication loss, communication risk, Stein estimation.

ESTIMATED CONFIDENCE PROCEDURES FOR MULTIVARIATE NORMAL MEANS [1]

by

K. L. Lu and James O. Berger

Purdue University

Abstract

In estimation of a p-variate normal mean with identity covariance matrix, confidence sets recentered at Stein-type estimators have larger coverage probability than the usual confidence ellipsoids (see Hwang and Casella (1982)). However, the minimum coverage probability (say $1-\alpha$) of these improved sets is identical to that of the usual sets, so that only $1-\alpha$ can be actually reported. Data dependent estimated confidence coefficients, $1-\hat{\alpha}(X)$, can be found which (I) have frequentist validity, (II) are always larger than $1-\alpha$, and (III) dominate the report of $1-\alpha$ via quadratic scoring.

[1] Research supported by National Science Foundation grant DMS-8401996 and by a David Ross fellowship.

1 Introduction

If data $\mathbf{x}$ is to be observed from a distribution with unknown parameter $\boldsymbol{\theta} \in \Theta$, the performance of a confidence procedure $C(\mathbf{x})$ (a subset of Θ for each $\mathbf{x}$) is typically reported to be

$$1 - \alpha = \inf_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(C(\mathbf{X}) \textit{ contains } \boldsymbol{\theta}). \tag{1}$$

A well known problem with this report it is that it is an average measure of performance over all possible data; for the actual observed data, the performance of $C(\mathbf{x})$ can be quite different (cf. Berger and Wolpert (1984)). A relatively unexplored alternative is the *estimated confidence* approach, introduced formally in Kiefer (1977). The idea here is to report a data-dependent measure of accuracy, $1 - \hat{\alpha}(\mathbf{x})$, for the confidence set $C(\mathbf{x})$ when $\mathbf{x}$ is observed. Data dependent measures of accuracy are, of course, common in Bayesian and conditional frequentist approaches (see Kiefer (1977) and Berger and Wolpert (1984) for references). Here, however, we consider the "pure" frequentist motivation (as described in Berger (1985a,b)) of desiring a reported accuracy which is conservatively accurate in long run repeated use. Precisely, we require that $1 - \hat{\alpha}(\mathbf{x})$ satisfy:

(I) Frequentist Validity:

$$E_{\boldsymbol{\theta}}(1 - \hat{\alpha}(\mathbf{X})) \leq P_{\boldsymbol{\theta}}(C(\mathbf{X}) \textit{ contains } \boldsymbol{\theta}), \quad \textit{for all } \boldsymbol{\theta}. \tag{2}$$

(A related notion is the *guaranteed conditional confidence* of Brown (1978).)

Introduction of estimated confidence introduces a new problem: for any given confidence procedure $\{C(\mathbf{x})\}$, there are many possible estimated confidences that have frequentist validity. One approach to selecting among them is based on recognizing that $1 - \hat{\alpha}(\mathbf{x})$ is, in some sense, an estimate of $I_{C(\mathbf{x})}(\boldsymbol{\theta})$, the indicator function on $C(\mathbf{x})$. Clearly the ideal (though generally unattainable) report would be $1 - \hat{\alpha}(\mathbf{x}) = 1$ if $I_{C(\mathbf{x})}(\boldsymbol{\theta}) = 1$, and $1 - \hat{\alpha}(\mathbf{x}) = 0$ if $I_{C(\mathbf{x})}(\boldsymbol{\theta}) = 0$. To capture this notion it is natural to introduce the decision problem of estimating $I_{C(\mathbf{x})}(\boldsymbol{\theta})$ by $1 - \hat{\alpha}(\mathbf{x})$, using (say) quadratic loss

$$L^*(I_C, 1 - \hat{\alpha}) = (1 - \hat{\alpha} - I_C)^2. \tag{3}$$

This has the important property of being a proper scoring rule. Using scoring rules in this setting is appealing; for instance, in a Bayesian setting, scoring rules insure that the optimal $1 - \hat{\alpha}$ is the posterior probability of $C(\mathbf{x})$. We will, however, be adopting the frequentist perspective, and so will evaluate $1-\hat{\alpha}$ by its *communication risk*

$$R_L^*(\boldsymbol{\theta}, 1-\hat{\alpha}) = E_{\boldsymbol{\theta}} L^*(I_{C(\mathbf{X})}(\boldsymbol{\theta}), 1-\hat{\alpha}(\mathbf{X})). \tag{4}$$

One could apply many classical decision-theoretic concepts, based on R_L^*, to the choice of $1-\hat{\alpha}$. Here we will focus on a rather simple condition:

(II) Improved Reported Confidence:

$$R_L^*(\boldsymbol{\theta}, 1-\hat{\alpha}) \leq R_L^*(\boldsymbol{\theta}, 1-\alpha) \text{ for all } \boldsymbol{\theta}, \text{ with strict inequality for some } \boldsymbol{\theta}. \tag{5}$$

The motivation for (5) is that, in considering alternatives to the classical report of $1-\alpha$, we seek to guarantee improved accuracy in the estimate of the confidence in $C(\mathbf{x})$. Without this guarantee, use of any alternative may not be compelling. Note that this criterion was employed by Robinson (1979b).

Note that we will not be considering choice of $\{C(\mathbf{x})\}$ here. The confidence procedure will be considered to be given, and our sole concern will be that of reporting our confidence in it. See Rukhin (1987) for an example where choice of the procedure is combined with choosing the reported confidence.

The specific problem to be addressed in this paper concerns confidence sets for a p-variate normal mean $\boldsymbol{\theta}$, based on $\mathbf{X} = (X_1, \cdots, X_p)^t \sim N_p(\boldsymbol{\theta}, I)$, where I is the $p \times p$ identity matrix. Confidence sets have been proposed which are based on the positive part James-Stein estimator

$$\boldsymbol{\delta}_a(\mathbf{x}) = \left(1 - \frac{a}{||\mathbf{x}||^2}\right)^+ \mathbf{x}; \tag{6}$$

here "+" denotes the positive part, $||\mathbf{x}||^2 = \sum_{i=1}^p x_i^2$, and $0 \leq a \leq 2(p-2)$, with $a = p-2$ being the usual constant considered. Indeed, Hwang and Casella (1982) consider the sets

$$C_a(\mathbf{x}) = \{\boldsymbol{\theta} \ : \ ||\boldsymbol{\theta} - \boldsymbol{\delta}_a(\mathbf{x})||^2 \leq c^2\}, \tag{7}$$

where $P(\chi_p^2 \leq c^2) = 1-\alpha$, χ_p^2 being a central chi-squared random variable with p degrees of freedom, and show that these sets have larger coverage probability than the usual confidence ellipsoid

$$C_o(\mathbf{x}) = \{\boldsymbol{\theta} \ : \ ||\boldsymbol{\theta} - \mathbf{x}||^2 \leq c^2\}, \tag{8}$$

for certain choices of a and $p \geq 3$. We henceforth assume that $p \geq 3$.

The interest in estimated confidence for this problem is that

$$\inf_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(C_a(\mathbf{X}) \ contains \ \boldsymbol{\theta}) = 1 - \alpha, \tag{9}$$

even though $P_{\boldsymbol{\theta}}(C_a(\mathbf{X}) \ contains \ \boldsymbol{\theta}) > 1 - \alpha$ for all $\boldsymbol{\theta}$. Thus the classical frequentist can only report $1 - \alpha$ as the confidence, though he knows that it is really larger than $1 - \alpha$. It is thus natural to seek a reported confidence, $1 - \hat{\alpha}(\mathbf{x})$, which satisfies (2) and for which

$$1 - \hat{\alpha}(\mathbf{x}) > 1 - \alpha \quad for \ all \ \mathbf{x}; \tag{10}$$

then we will have a report which has frequentist validity and reflects the improvement obtained by use of $C_a(\mathbf{x})$. Satisfaction of (5) is also desirable. The construction of such $1 - \hat{\alpha}(\mathbf{x})$ will be discussed in this paper.

The idea of estimated confidence was clearly presented by Kiefer (1977). Kiefer also considered conditional confidence as a frequentist approach to conditioning. Trying to develop a decision theory for correct choice of the collection of conditioning sets gets very involved, however (cf. Kiefer (1977) or Brown (1978)), in contrast with the more standard form of the decision theory for choice of $[1 - \hat{\alpha}(\mathbf{x})]$.

Robinson (1979a,b) gave a proof of inadmissibility of $1-\alpha$ as a point estimator of $I_{C_o(\mathbf{x})}(\boldsymbol{\theta})$ in estimating a five dimensional normal mean with respect to squared-error loss. Theorem 2.2 below can be considered to be a generalization of this result to any $p \geq 5$ and to C_a (i.e. to confidence spheres centered about James-Stein estimators, not just confidence spheres centered about $\mathbf{x}$.) Robinson also illustrated the relationship between existence of relevant betting procedures and admissibility properties of estimated confidence. For location and scale parameters, a sufficient condition was given under which estimated confidence for the Pitman estimator is admissible with respect to squared-error loss.

Estimated confidence can be considered to be a special case of estimated loss, discussed in Lu and Berger (1987). Other work includes Sandved (1968), Kiefer (1975, 1976, 1977), Berger and Wolpert (1984), Berger (1985a,b,d), Rukhin (1987), and Johnstone (1987).

It should be mentioned that we approach this work from the perspective of scientific curiosity. In estimation of a multivariate normal mean, excellent confidence sets have been found using Bayesian methodology (see Berger (1985c) for references), with $1 - \hat{\alpha}(\mathbf{x})$ being the resulting posterior probability of the confidence set. Such procedures do not, however, have guaranteed frequentist properties, and such properties are nice to have, if possible. (Among the large literature showing the impossibility of always having good frequentist properties, Gleser and Hwang (1987) stands out). There is reason to expect that, in this problem, it is possible to simultaneously have good conditional and good frequentist properties (cf. Lu and Berger (1987)). The results here provide strong support for this conjecture, and will hopefully encourage the use of data dependent confidence measures for this problem.

2 Estimated Confidence for Recentered Sets

2.1 Existence of Improved Estimated Confidence

Consider the estimated confidence $1-\hat{\alpha}(\mathbf{x})$ for $C_a(\mathbf{x})$, where

$$\hat{\alpha}(\mathbf{x}) = \left(1 - \frac{bp}{dp + ||\mathbf{x}||^2}\right)\alpha \tag{11}$$

for certain positive numbers b and d. It is obvious that, for all $\mathbf{x}$,

$$1-\hat{\alpha}(\mathbf{x}) > 1-\alpha, \tag{12}$$

so that (10) is satisfied. The following result establishes the frequentist validity of $1-\hat{\alpha}(\mathbf{x})$ for appropriate constants b and d.

Theorem 2.1 *For $0 < a \le a_h$, there exists $b_o > 0$ such that, for $0 < b \le b_o$ and any $d > 0$,*

$$E_{\boldsymbol{\theta}}(1-\hat{\alpha}(\mathbf{X})) < P_{\boldsymbol{\theta}}(C_a(\mathbf{X}) \text{ contains } \boldsymbol{\theta}) \quad \text{for all } \boldsymbol{\theta}; \tag{13}$$

here a_h is the maximum value of a such that

$$P_{\boldsymbol{\theta}}(C_a(\mathbf{X}) \text{ contains } \boldsymbol{\theta}) > 1-\alpha, \quad \text{for all } \boldsymbol{\theta}. \tag{14}$$

(A lower bound for a_h is given below.)

Proof. Given in Section 3. □

A lower bound for a_h, given in Hwang and Casella (1984), is $\min\{a_1, a_2\}$ where a_1, a_2 are unique solutions, respectively, of

$$\left[\frac{c+\sqrt{c^2+4a}}{2\sqrt{a}}\right]^{p-2} e^{-c\sqrt{a}/2} = 1. \tag{15}$$

and

$$\left[\frac{\sqrt{c^2+4a}-c}{2\sqrt{a}}\right]\left[\frac{c+\sqrt{c^2+a}}{\sqrt{a}}\right]^{p-1} e^{-c\sqrt{a}} = 1. \tag{16}$$

The next theorem shows that $1-\hat{\alpha}(\mathbf{x})$ will also be an improved reported confidence with respect to quadratic communication risk.

Table 1: Values of a_o for Values p and $1-\alpha$

p	5	6	7	8	9	10	15	20	25	30
$1-\alpha = 0.90$	1.000	2.000	3.000	4.000	5.000	6.000	10.835	15.060	19.301	23.552
$1-\alpha = 0.95$	1.000	2.000	3.000	4.000	5.000	6.000	10.613	14.803	19.014	23.238

Theorem 2.2 *For $0 < a < a_o = \min\{p-4, a_1, a_2\}$, there exist $b_1 > 0$ and $d_o > 0$ such that, for $0 < b \le b_1$ and $d \ge d_o$,*

$$E_{\boldsymbol{\theta}}(1-\hat{\alpha}(\mathbf{X}) - I_{C_a(\mathbf{X})}(\boldsymbol{\theta}))^2 < E_{\boldsymbol{\theta}}(1-\alpha - I_{C_a(\mathbf{X})}(\boldsymbol{\theta}))^2, \quad \textit{for all } \boldsymbol{\theta}. \tag{17}$$

Proof. Given in Section 3. □

Table 1 gives values of a_o for $1-\alpha = 0.90$ and for $1-\alpha = 0.95$.

2.2 Choice of constants.

In the above section we only proved the existence of suitable constants b and d. Here we suggest usable values of b and d. In Section 2.3, evidence is presented which suggests that these choices of b and d do result in satisfaction of (2) and (5).

Suppose that a is specified and that c is the radius of the confidence set $C_a(\mathbf{x})$. A suggested choice for b and d is

$$b = d = A(p, a, \alpha)/p\alpha, \tag{18}$$

where

$$A(p, a, \alpha) = \frac{a(2p-4-a)c^p e^{-c^2/2}}{p 2^{p/2}\Gamma(p/2)}. \tag{19}$$

The number $A(p, a, \alpha)$ appears in the asymptotic expression for coverage probability given by Hwang and Casella (1984), i.e.,

$$P_{\boldsymbol{\theta}}(C_a(\mathbf{X}) \ \textit{contains} \ \boldsymbol{\theta}) = 1-\alpha + A(p, a, \alpha)/||\boldsymbol{\theta}||^2 + O(||\boldsymbol{\theta}||^{-3}). \tag{20}$$

The choice in (18) is motivated by the following considerations. Note that the estimated confidence can be written as

$$1-\hat{\alpha}(\mathbf{x}) = 1-\alpha + \frac{bp\alpha}{dp + ||\mathbf{x}||^2}, \tag{21}$$

and that $E||X||^2 = ||\boldsymbol{\theta}||^2 + p$. Thus, as $||\boldsymbol{\theta}||^2 \to \infty$,

$$1 - \hat{\alpha}(\mathbf{x}) \to 1 - \alpha + \frac{bp\alpha}{||\boldsymbol{\theta}||^2}. \tag{22}$$

Equating (22) with (20) yields the choice of b in (18). The choice of d in (18) is simply to ensure that $1 - \hat{\alpha}(\mathbf{x})$ in (21) never exceeds 1.

Values of b and d corresponding to the choice $a = p - 2$ are given in Table 2. Note that this value of a exceeds a_o, and hence we do not have theoretical assurance of frequentist validity or improved communication risk for this choice. It is common in this area, however, that theoretical results establish improvement only for values of a which are substantially smaller than those which actually work. Such is the case here; the numerical results suggest that $a = p - 2$ does yield frequentist validity and improved communication risk. We choose $a = p - 2$, of course, because it corresponds to the choice typically made in James-Stein estimation.

2.3 Numerical Results

The numerical results below present the performance of $1 - \hat{\alpha}(\mathbf{x})$, where b and d are chosen as in (18), and $a = p - 2$. Figure 1, for the case $p = 8$, $a = 6$, and for $1 - \alpha = 0.90$ and $1 - \alpha = 0.95$, graphs the estimated confidences $1 - \hat{\alpha}(\mathbf{x})$ with respect to $||\mathbf{x}||$. Clearly, substantial reported improvement is available for small $||\mathbf{x}||$.

Figure 2, for the same case, graphs the coverage probabilities of $C(\mathbf{x})$ and the expectations of the estimated confidence, termed the expected confidences, with respect to $||\boldsymbol{\theta}||$. These figures indicate that the frequentist validity condition (2) is indeed satisfied by $1 - \hat{\alpha}$, for this choice of b and d. Figure 3, for the same case, graphs $R_L^*(\boldsymbol{\theta}, 1 - \alpha)$ and $R_L^*(\boldsymbol{\theta}, 1 - \hat{\alpha})$. The communication risk of $1 - \hat{\alpha}$ definitely appears to be superior to that of $1 - \alpha$. Further evidence of this is provided by Table 3, which presents the proportional decrease in communication risk in use of $1 - \hat{\alpha}$ instead of $1 - \alpha$ (i.e. $[R_L^*(\boldsymbol{\theta}, 1 - \alpha) - R_L^*(\boldsymbol{\theta}, 1 - \hat{\alpha})]/R_L^*(\boldsymbol{\theta}, 1 - \alpha)$) for $p = 5, 6, 7, 8, 10, 15, 20, 25, 30$, $a = p - 2$, b and d as in (18), and various values of $||\boldsymbol{\theta}||$.

Table 2: Suggested Values of b and d When $a = p - 2$

	$1-\alpha=0.90$			$1-\alpha=0.95$		
p	c^2	$a=p-2$	b,d	c^2	$a=p-2$	b,d
5	9.236	3.000	1.225	11.070	3.000	1.541
6	10.645	4.000	1.635	12.592	4.000	2.045
7	12.017	5.000	2.006	14.067	5.000	2.499
8	13.362	6.000	2.343	15.507	6.000	2.909
9	14.684	7.000	2.652	16.919	7.000	3.283
10	15.987	8.000	2.938	18.307	8.000	3.628
11	17.275	9.000	3.204	19.675	9.000	3.948
12	18.549	10.000	3.454	21.026	10.000	4.247
13	19.812	11.000	3.689	22.362	11.000	4.529
14	21.064	12.000	3.912	23.685	12.000	4.795
15	22.307	13.000	4.123	24.996	13.000	5.048
16	23.542	14.000	4.325	26.296	14.000	5.289
17	24.769	15.000	4.519	27.587	15.000	5.520
18	25.989	16.000	4.705	28.869	16.000	5.741
19	27.204	17.000	4.883	30.144	17.000	5.953
20	28.412	18.000	5.057	31.410	18.000	6.159
21	29.615	19.000	5.224	32.671	19.000	6.357
22	30.813	20.000	5.386	33.924	20.000	6.550
23	32.007	21.000	5.542	35.172	21.000	6.736
24	33.196	22.000	5.695	36.415	22.000	6.916
25	34.382	23.000	5.843	37.652	23.000	7.093
26	35.563	24.000	5.989	38.885	24.000	7.264
27	36.741	25.000	6.130	40.113	25.000	7.431
28	37.916	26.000	6.267	41.337	26.000	7.594
29	39.087	27.000	6.403	42.557	27.000	7.754
30	40.256	28.000	6.534	43.773	28.000	7.910

Table 3: The Proportional Decrease In Communication Risk

$\|\|\boldsymbol{\theta}\|\|$	$p=5$	$p=6$	$p=7$	$p=8$	$p=10$	$p=15$	$p=20$	$p=25$	$p=30$
0.000	0.379	0.524	0.649	0.744	0.862	0.952	0.971	0.978	0.982
0.500	0.367	0.512	0.637	0.735	0.857	0.951	0.970	0.977	0.982
1.000	0.335	0.477	0.605	0.708	0.840	0.946	0.968	0.976	0.981
1.500	0.287	0.424	0.553	0.663	0.812	0.938	0.965	0.974	0.979
2.000	0.232	0.358	0.485	0.600	0.769	0.927	0.960	0.971	0.978
3.000	0.119	0.214	0.323	0.437	0.640	0.885	0.944	0.963	0.972
4.000	0.004	0.029	0.070	0.131	0.300	0.807	0.917	0.949	0.963
5.000	0.001	0.007	0.024	0.050	0.136	0.525	0.866	0.928	0.950
6.000	0.000	0.002	0.009	0.021	0.063	0.306	0.660	0.850	0.932
7.000	0.000	0.000	0.004	0.010	0.032	0.167	0.443	0.730	0.873
8.000	0.000	0.000	0.002	0.005	0.017	0.095	0.271	0.534	0.765
9.000	0.000	0.000	0.001	0.003	0.010	0.058	0.167	0.356	0.591
10.000	0.000	0.000	0.000	0.002	0.006	0.037	0.107	0.234	0.420

Figure 1: Estimated Confidence

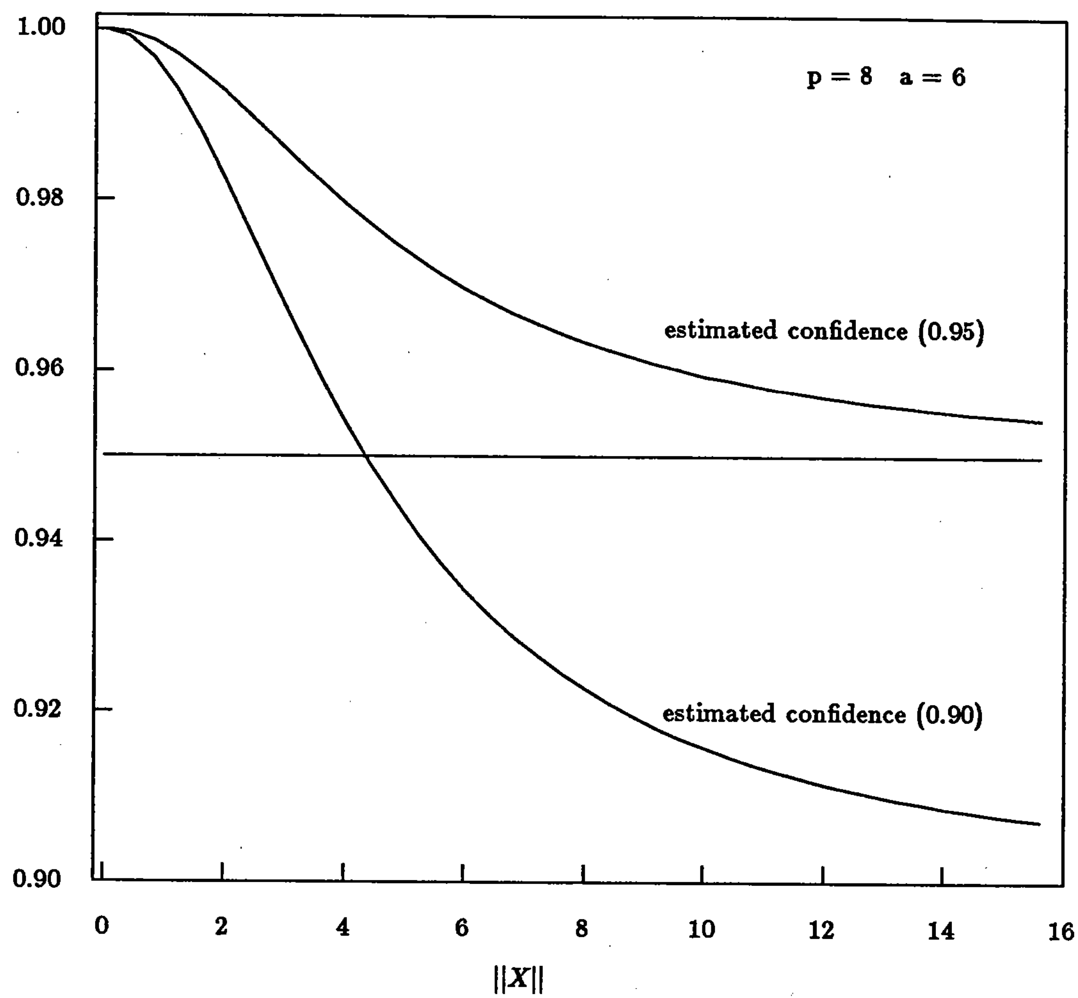


Figure 2: Coverage Probability

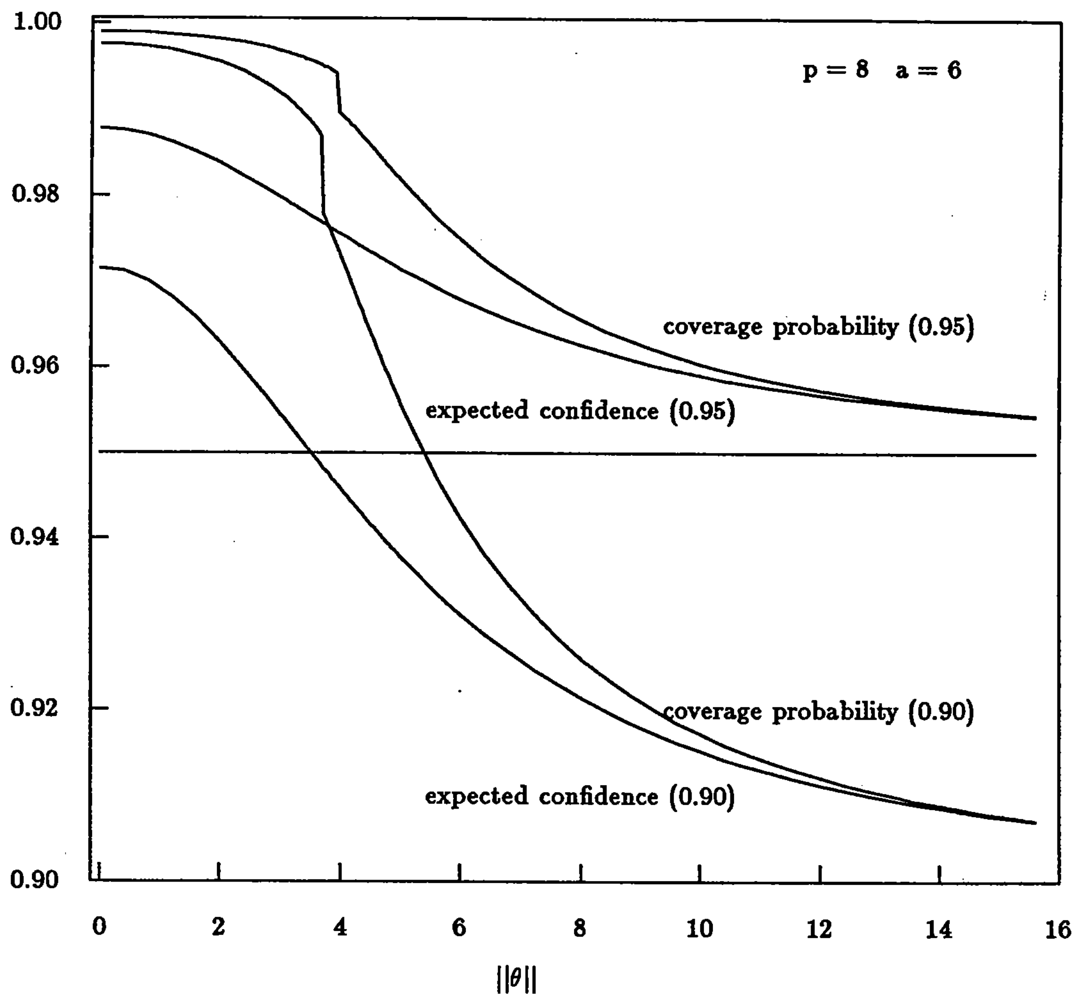
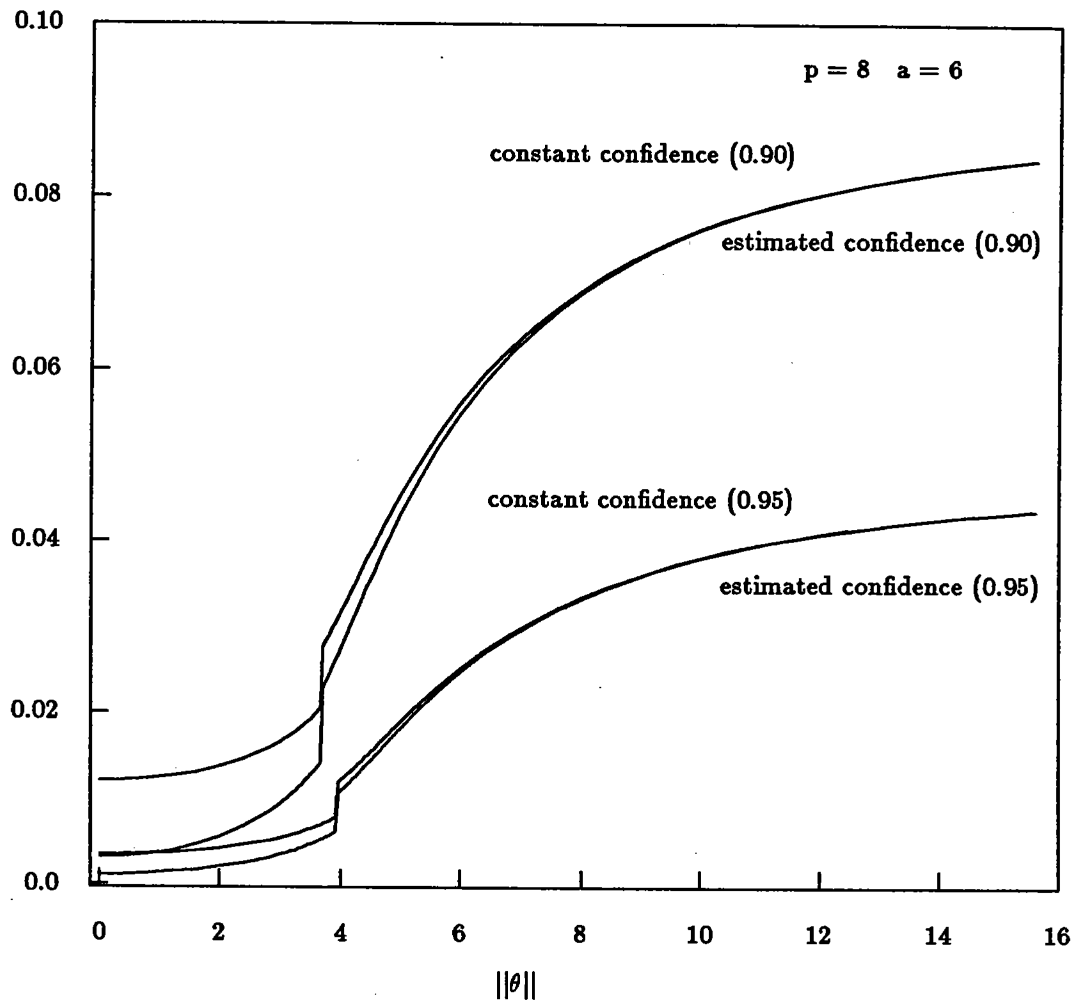


Figure 3: Communication Risks



3 Proof of Theorems

Proof of Theorem 2.1 Theorem 3.1 in Hwang and Casella (1984) establishes the asymptotic expression (20). Thus it is clear that

$$\liminf_{||\boldsymbol{\theta}||\to\infty} ||\boldsymbol{\theta}||^2[P_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in C_a(\mathbf{X}))-(1-\alpha)]>0. \tag{23}$$

From Hwang and Casella (1982), we know that for $||\boldsymbol{\theta}||>c$, $P_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in C_a(\mathbf{X}))$ is greater than $1-\alpha$ and is a continuous and decreasing function of $\boldsymbol{\theta}$. Also, it is continuous when $||\boldsymbol{\theta}||\le c$. Because of these facts and (20) it follows that there is a positive number ϵ such that

$$\inf_{\boldsymbol{\theta}}\{(||\boldsymbol{\theta}||^2+p)[P_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in C_a(\mathbf{X}))-(1-\alpha)]\}=\epsilon. \tag{24}$$

Next, we evaluate the $E_{\boldsymbol{\theta}}(1-\hat{\alpha}(\mathbf{X}))$. From (11),

$$1-\hat{\alpha}(\mathbf{X})=1-\alpha+\frac{bp\alpha}{dp+||\mathbf{X}||^2}. \tag{25}$$

To calculate the expectation of the last item in (25), without loss of generality we suppose that $\boldsymbol{\theta}=(\theta_1,0,...,0)'$. Then

$$\begin{aligned}
E_{\boldsymbol{\theta}}\frac{1}{dp+||\mathbf{X}||^2} &= E_0\frac{1}{dp+||\mathbf{X}+\boldsymbol{\theta}||^2}\\
&= \frac{1}{dp+p+||\boldsymbol{\theta}||^2}(1-E_0\frac{||\mathbf{X}||^2-p+2\mathbf{X}'\boldsymbol{\theta}}{dp+||\mathbf{X}+\boldsymbol{\theta}||^2})\\
&= \frac{1}{dp+p+||\boldsymbol{\theta}||^2}(1-E_0\frac{||\mathbf{X}||^2-p+2X_1(\theta_1+X_1)-2X_1^2}{dp+||\mathbf{X}+\boldsymbol{\theta}||^2})\\
&\le \frac{1+\epsilon_1}{dp+p+||\boldsymbol{\theta}||^2},
\end{aligned} \tag{26}$$

where $\epsilon_1=(3+\sqrt{2p})/dp$; the last inequality follows from

$$\begin{aligned}
|E_0\frac{||\mathbf{X}||^2-p+2X_1(\theta_1+X_1)-2X_1^2}{dp+||\mathbf{X}+\boldsymbol{\theta}||^2}| &\le \frac{E_0|||\mathbf{X}||^2-p|}{dp}+\frac{E_0 2|X_1|}{2dp}+\frac{E_0 2|X_1|^2}{dp}\\
&\le \frac{\sqrt{2p}}{dp}+\frac{1}{dp}+\frac{2}{dp}\\
&= \frac{3+\sqrt{2p}}{dp}=\epsilon_1.
\end{aligned} \tag{27}$$

Let $b_o=\frac{\epsilon}{(1+\epsilon_1)p\alpha}$. By (24) and (26), for $0<b<b_o$ and for all $\boldsymbol{\theta}$,

$$\begin{aligned}
P_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in C_a(\mathbf{X})) &> 1-\alpha+\frac{\epsilon}{p+||\boldsymbol{\theta}||^2}\\
&\ge 1-\alpha+E_{\boldsymbol{\theta}}\frac{bp\alpha}{dp+||\mathbf{X}||^2}\\
&= 1-\hat{\alpha}(\mathbf{X}).
\end{aligned} \tag{28}$$

Thus Theorem 2.1 is proved. □

Proof of Theorem 2.2. For convenience, we define $\beta = 1 - \alpha$ and $\hat{\beta}(\mathbf{x}) = 1 - \hat{\alpha}(\mathbf{x})$. Thus

$$\hat{\beta}(\mathbf{x}) = \beta + \frac{bp\alpha}{dp + ||\mathbf{x}||^2}. \tag{29}$$

Then

$$\begin{aligned}
&E_{\boldsymbol{\theta}}(1 - \alpha - I_{C_a(\mathbf{X})}(\boldsymbol{\theta}))^2 - E_{\boldsymbol{\theta}}(1 - \hat{\alpha}(\mathbf{X}) - I_{C_a(\mathbf{X})}(\boldsymbol{\theta}))^2 \\
&\quad = E_{\boldsymbol{\theta}}(\beta - I_{C_a(\mathbf{X})}(\boldsymbol{\theta}))^2 - E_{\boldsymbol{\theta}}(\hat{\beta}(\mathbf{X}) - I_{C_a(\mathbf{X})}(\boldsymbol{\theta}))^2 \\
&\quad = E_{\boldsymbol{\theta}}\{\beta^2 - \hat{\beta}(\mathbf{X})^2 - 2(\beta - \hat{\beta}(\mathbf{X}))I_{C_a(\mathbf{X})}(\boldsymbol{\theta})\} \\
&\quad = 2bp\alpha E_{\boldsymbol{\theta}}\{\frac{I_{C_a(\mathbf{X})}(\boldsymbol{\theta})}{dp + ||\mathbf{X}||^2} - \frac{\beta}{dp + ||\mathbf{X}||^2} - \frac{bp\alpha}{2(dp + ||\mathbf{X}||^2)^2}\} \\
&\quad = 2b\alpha E_{\boldsymbol{\theta}} g_a(\mathbf{X}),
\end{aligned} \tag{30}$$

where

$$g_a(\mathbf{x}) = \frac{I_{C_a(\mathbf{x})}(\boldsymbol{\theta}) - \beta}{d + ||\mathbf{x}||^2/p} - \frac{b\alpha}{2(d + ||\mathbf{x}||^2/p)^2}. \tag{31}$$

In the special case when $a = 0$, define

$$g_o(\mathbf{x}) = \frac{I_{C_o(\mathbf{x})}(\boldsymbol{\theta}) - \beta}{d + ||\mathbf{x}||^2/p} - \frac{b\alpha}{2(d + ||\mathbf{x}||^2/p)^2}. \tag{32}$$

We want to show that, for all $\boldsymbol{\theta}$, $E_{\boldsymbol{\theta}} g_a(\mathbf{X}) > 0$. This proof is provided by the following two lemmas.

Lemma 3.1 *For $p \geq 5$ and $\alpha < \frac{1}{2}$, there exist $d_2 > 0$ and $b_2 > 0$ such that, for $d \geq d_2$ and $0 < b \leq b_2$,*

$$E_{\boldsymbol{\theta}} g_0(\mathbf{X}) > 0 \quad \textit{for all } \boldsymbol{\theta}. \tag{33}$$

(Note that Lemma 3.1 is a generalization of Robinson (1979b), which established the result for $p = 5$.)

Proof. For convenience let us suppose $\boldsymbol{\theta} = (\theta_1, 0, ..., 0)'$ and define

$$A = pd + p + ||\boldsymbol{\theta}||^2, \quad S = pd + ||\mathbf{x} + \boldsymbol{\theta}||^2, \quad B = ||\mathbf{x}||^2 - p + 2x_1\theta_1, \tag{34}$$

and

$$\psi(\mathbf{x}) = I_{C_o(\mathbf{x})}(\mathbf{0}), \quad \phi(\mathbf{x}) = \frac{I_{C_o(\mathbf{x})}(\boldsymbol{\theta}) - \beta}{d + ||\mathbf{x}||^2/p}. \tag{35}$$

In the following proof, we use the identities

$$\frac{1}{S} = \frac{1}{A}(1 - \frac{B}{S}) \quad and \quad \frac{1}{S} = \frac{1}{A^2}(A - B + \frac{B^2}{S}), \tag{36}$$

where

$$B^2 = (||\mathbf{x}||^2 - p)^2 + 4x_1\theta_1(||\mathbf{x}||^2 - p) + 4x_1^2\theta_1^2. \tag{37}$$

Now we have, by (36) and by the symmetry of the distribution of $\mathbf{X}$,

$$\begin{aligned} E_{\boldsymbol{\theta}}\phi(\mathbf{X}) &= pE_0\frac{\alpha - \psi(\mathbf{X})}{S} \\ &= \frac{p}{A^2}E_0(\alpha - \psi(\mathbf{X}))(A - B + \frac{B^2}{S}) \\ &= \frac{p}{A^2}E_0(B - \frac{B^2}{S})(\psi(\mathbf{X}) - \alpha) \\ &\geq \frac{p}{A^2}E_0\{||\mathbf{X}||^2(\psi(\mathbf{X}) - \alpha) - \frac{B^2}{S}(\psi(\mathbf{X}) - \alpha)\}. \end{aligned} \tag{38}$$

To evaluate this last expression, we use the identity

$$(||\mathbf{X}||^2 - p)^2 = (||\mathbf{X}||^2 - p)(\rho - \frac{||\mathbf{X}||^2 - p}{2\rho S})^2 - \rho^2(||\mathbf{X}||^2 - p) - \frac{(||\mathbf{X}||^2 - p)^3}{4\rho^2 S^2}; \tag{39}$$

here ρ is a constant which a later computation will show to satisfy $\rho^2 \sim 0.112$. Then (37), (38), and (39) yield

$$\begin{aligned} \frac{A^2}{p}E_{\boldsymbol{\theta}}\phi(\mathbf{X}) \geq\; & E_0\{||\mathbf{X}||^2(\psi(\mathbf{X}) - \alpha) + (||\mathbf{X}||^2 - p)(\rho - \frac{||\mathbf{X}||^2 - p}{2\rho S})^2\psi(\mathbf{X}) \\ & -\rho^2(||\mathbf{X}||^2 - p)\psi(\mathbf{X}) - \frac{(||\mathbf{X}||^2 - p)^3}{4\rho^2 S^2}\psi(\mathbf{X}) \\ & -\frac{4X_1\theta_1(||\mathbf{X}||^2 - p)}{S}\psi(\mathbf{X}) - \frac{4X_1^2\theta_1^2}{S}(\psi(\mathbf{X}) - \alpha)\}. \end{aligned} \tag{40}$$

Since the last term in (40) does not go to zero as $||\boldsymbol{\theta}|| \to \infty$, we evaluate it as

$$\begin{aligned} E_0\frac{4X_1^2\theta_1^2}{S}(\psi(\mathbf{X}) - \alpha) &= E_0 4X_1^2\theta_1^2(\psi(\mathbf{X}) - \alpha)(\frac{1}{A} - \frac{B}{AS}) \\ &= E_0\frac{4X_1^2\theta_1^2}{A}(\psi(\mathbf{X}) - \alpha) - E_0\frac{4X_1^2\theta_1^2(||\mathbf{X}||^2 - p + 2X_1\theta_1)}{AS}(\psi(\mathbf{X}) - \alpha) \\ &= E_0\frac{4||\mathbf{X}||^2\theta_1^2}{pA}(\psi(\mathbf{X}) - \alpha) + J_3, \end{aligned} \tag{41}$$

where

$$\begin{aligned} J_3 &= -E_0\frac{4X_1^2\theta_1^2(||\mathbf{X}||^2 - p)}{AS}(\psi(\mathbf{X}) - \alpha) - E_0\frac{8X_1^3\theta_1^3}{AS}(\psi(\mathbf{X}) - \alpha) \\ &\leq \frac{4\alpha}{pd}E_0X_1^2|||\mathbf{X}||^2 - p| - E_0\frac{8X_1^3\theta_1^2(\theta_1 + X_1)}{AS}(\psi(\mathbf{X}) - \alpha) + E_0\frac{8X_1^4\theta_1^2}{AS}(\psi(\mathbf{X}) - \alpha) \\ &\leq \frac{4}{dp}m_4^{\frac{1}{2}}\mu_2^{\frac{1}{2}}(p)\alpha + \frac{4}{\sqrt{dp}}m_6^{\frac{1}{2}}\alpha^{\frac{1}{2}} + \frac{4}{\sqrt{dp}}m_3\alpha + \frac{8}{dp}m_8^{\frac{1}{2}}\alpha^{\frac{1}{2}}. \end{aligned} \tag{42}$$

Here we have denoted the n-th moment of $(||\mathbf{X}||^2 - p)$ by $\mu_n(p)$, and the n-th absolute moment of X_1 by m_n. Hence

$$\begin{aligned}\frac{A^2}{p}E_{\boldsymbol{\theta}}\phi(\mathbf{X}) &\geq E_0\{(1-\rho^2-\frac{4}{p})||\mathbf{X}||^2(\psi(\mathbf{X})-\alpha) - \frac{(||\mathbf{X}||^2-p)^3}{4\rho^2 S^2}\psi(\mathbf{X}) \\ &\quad - \frac{4X_1(X_1+\theta_1)(||\mathbf{X}||^2-p)}{S}\psi(\mathbf{X})\} - J_3 \\ &= J_0 - J_1 - J_2 - J_3, \qquad (43)\end{aligned}$$

where the $J's$ denote, respectively, the terms of the right hand of the above inequality.

Next we calculate the values of J_0, J_1, J_2.

$$J_0 = E_0(1-\rho^2-\frac{4}{p})||\mathbf{X}||^2(\psi(\mathbf{X})-\alpha) = (1-\rho^2-\frac{4}{p})p(\tau_p-\alpha), \qquad (44)$$

where

$$\tau_p = \frac{1}{p}E_0||\mathbf{X}||^2\psi(\mathbf{X}). \qquad (45)$$

It is easy to check that $\tau_p > \alpha$, if $\alpha < \frac{1}{2}$. Then, using the *Hölder* inequality, it follows that

$$J_1 \leq \frac{1}{4\rho^2 d^2 p^2}\mu_{12}(p)^{\frac{1}{4}}\alpha^{\frac{3}{4}}, \qquad (46)$$

$$J_2 \leq \frac{2}{\sqrt{dp}}m_4^{\frac{1}{4}}\mu_{12}(p)^{\frac{1}{12}}\alpha^{\frac{2}{3}}. \qquad (47)$$

Combining (41)-(47), we obtain, for sufficiently large d,

$$E_{\boldsymbol{\theta}}\phi(\mathbf{X}) > \frac{p}{(pd+p+||\boldsymbol{\theta}||^2)^2}\{(1-\rho^2-\frac{4}{p})p(\tau_p-\alpha) - O(\frac{1}{\sqrt{d}})\} > 0. \qquad (48)$$

It is clear that d only depends on p and α.

Next we consider the second term in (40). We shall use the inequality that for any constants $a_1, a_2, ..., a_n$,

$$(a_1 + a_2 + ... + a_2)^2 \leq 2a_1^2 + 4a_2^2 + ... + 2^n a_n^2. \qquad (49)$$

Now

$$\begin{aligned}E_{\boldsymbol{\theta}}\frac{b\alpha}{2(d+||\mathbf{X}||^2/p)^2} &= \frac{bp^2\alpha}{2A^2}E_0(\frac{A}{S})^2 \\ &\leq \frac{bp^2\alpha}{2A^2}E_0(1+\frac{p-X_1^2-2X_1\theta_1}{dp+(X_1+\theta_1)^2})^2 \\ &\leq \frac{bp^2\alpha}{2A^2}E_0(1+\frac{p+X_1^2-2X_1(X_1+\theta_1)}{dp+(X_1+\theta_1)^2})^2\end{aligned}$$

$$\begin{aligned} &\le \frac{bp^2\alpha}{2A^2}E_0(2+\frac{4X_1^2}{dp}+8\frac{1}{d^2}+16(\frac{X_1^2}{dp})^2) \\ &\le \frac{bp^2\alpha}{A^2}(1+\frac{2}{dp}+\frac{4}{d^2}+\frac{24}{d^2p^2}) \\ &= \frac{bp^2\alpha}{A^2}(1+O(\frac{1}{d})). \end{aligned} \tag{50}$$

Comparing (48) to (50), we obtain that, for sufficiently large d and for sufficiently small b,

$$E_{\boldsymbol{\theta}}g_o(\mathbf{X}) \ge \frac{p}{pd+p+||\boldsymbol{\theta}||^2}\{(1-\rho^2-\frac{4}{p})p(\tau_p-\alpha)-O(\frac{1}{\sqrt{d}})-bp\alpha O(1)\} > 0. \tag{51}$$

Thus, the lemma is proved. □

Lemma 3.2 *For $0 < a < a_o$, there exists a positive number d_1 such that for $d \ge d_1$ and any $b > 0$,*

$$\frac{\partial}{\partial a}E_{\boldsymbol{\theta}}g_a(\mathbf{X}) > 0, \quad \textit{for all } \boldsymbol{\theta}. \tag{52}$$

Proof. The method of proof is very similar to the proof of Theorem 2.1 of Hwang and Casella (1984). Let

$$h_{\boldsymbol{\theta}}(a) = E_{\boldsymbol{\theta}}\frac{I_{c_a(\mathbf{X})}(\boldsymbol{\theta})}{d+||\mathbf{X}||^2/p}. \tag{53}$$

We only need to show that

$$\frac{\partial}{\partial a}h_{\boldsymbol{\theta}}(a) > 0. \tag{54}$$

Let $r = ||\mathbf{x}||$ and β be the angle between $\mathbf{x}$ and $\boldsymbol{\theta}$. The inequality

$$||\boldsymbol{\theta}-\delta_a(\mathbf{x})||^2 \le c^2 \tag{55}$$

is equivalent to

$$r^2u(r)^2 - 2ru(r)||\boldsymbol{\theta}||\cos\beta + ||\boldsymbol{\theta}||^2 \le c^2 \tag{56}$$

where

$$u(r) = (1-\frac{a}{r^2})^+. \tag{57}$$

A little algebra shows that the set of $\mathbf{x}$ values satisfying (50) equals the region

$$\{\ \mathbf{x}\ : r_- \le r \le r_+ \quad \textit{and} \quad 0 \le \beta \le \beta_0\ \ \}. \tag{58}$$

For $||\boldsymbol{\theta}|| \le c$, we have $\beta_0 = \pi$, $r_- = 0$ and $r_+ = r_+(a, \boldsymbol{\theta}, \beta)$ satisfies

$$r_+u(r_+) = ||\boldsymbol{\theta}||\cos\beta + (c^2 - ||\boldsymbol{\theta}||^2\sin^2\beta)^{\frac{1}{2}}. \tag{59}$$

For $||\boldsymbol{\theta}|| > c$, we have $\beta_0 = \arcsin(c/||\boldsymbol{\theta}||)$ and $r_\pm = r_\pm(a, \boldsymbol{\theta}, \beta)$ are solutions to

$$r_\pm u(r_\pm) = ||\boldsymbol{\theta}|| \cos\beta \pm (c^2 - ||\boldsymbol{\theta}||^2 \sin^2\beta)^{\frac{1}{2}} \stackrel{\text{def}}{=} r_\pm^o, \tag{60}$$

which can be seen to be

$$r_\pm(a, \boldsymbol{\theta}, \beta) = \frac{1}{2}[r_\pm^o + \sqrt{r_\pm^{o\,2} + 4a}]. \tag{61}$$

Now, writing $h_{\boldsymbol{\theta}}(a)$ in terms of r and β, we obtain

$$h_{\boldsymbol{\theta}}(a) = K \int_0^{\beta_0} \int_{r_-}^{r_+} r^{p-1} \sin^{p-2}\beta \frac{f^*(r, \beta)}{d + r^2/p} dr d\beta, \tag{62}$$

where $K = 1$ if $p = 2$ and, if $p \geq 3$,

$$K = 2 \prod_{i=0}^{p-3} \int_0^{\pi} \sin^i t dt, \tag{63}$$

and

$$f^*(r, \beta) = (2\pi)^{-p/2} \exp(-(r^2 - 2r||\boldsymbol{\theta}|| \cos\beta + ||\boldsymbol{\theta}||^2)/2). \tag{64}$$

From the bounded convergence theorem, we can interchange the order of differentiation and integration. Hence for $||\boldsymbol{\theta}|| > c$,

$$\begin{aligned} \frac{\partial}{\partial a} h_{\boldsymbol{\theta}}(a) &= K \int_0^{\beta_0} \frac{\partial}{\partial a} \int_{r_-}^{r_+} \frac{r^{p-1} f^*(r, \beta)}{d + r^2/p} \sin^{p-2}\beta dr d\beta \\ &= K \int_0^{\beta_0} \{ \frac{r_+^{p-1} f^*(r_+, \beta)}{d + r_+^2/p} \frac{\partial}{\partial a} r_+ - \frac{r_-^{p-1} f^*(r_-, \beta)}{d + r_-^2/p} \frac{\partial}{\partial a} r_- \} \sin^{p-2}\beta d\beta \\ &= K \int_0^{\beta_0} \{ \frac{r_+^p f^*(r_+, \beta)}{(d + r_+^2/p)(a + r_+^2)} - \frac{r_-^p f^*(r_-, \beta)}{(d + r_-^2/p)(a + r_-^2)} \} \sin^{p-2}\beta d\beta, \end{aligned} \tag{65}$$

where

$$\frac{\partial}{\partial a} r_\pm = \left[r_\pm (1 + \frac{a}{r_\pm^2}) \right]^{-1}. \tag{66}$$

For the case $||\boldsymbol{\theta}|| \leq c$, we have similarly

$$\frac{\partial}{\partial a} h_{\boldsymbol{\theta}}(a) = K \int_0^{\beta_0} \frac{r_+^p f^*(r_+, \beta)}{(d + r_+^2/p)(a + r_+^2)} \sin^{p-2}\beta d\beta, \tag{67}$$

and it is clear that $\frac{\partial}{\partial a} h_{\boldsymbol{\theta}}(a) > 0$. For the case $||\boldsymbol{\theta}|| > c$, to show that $\frac{\partial}{\partial a} h_{\boldsymbol{\theta}}(a) > 0$, it is sufficient to prove

$$R \stackrel{\text{def}}{=} \frac{r_+^p f^*(r_+, \beta)(a + r_-^2)(dp + r_-^2)}{r_-^p f^*(r_-, \beta)(a + r_+^2)(dp + r_+^2)} > 1, \quad a.e. \tag{68}$$

From (64) we have

$$R = (\frac{r_+}{r_-})^p \exp(-\frac{1}{2}(r_+ - r_-)(r_+ + r_- - 2||\boldsymbol{\theta}|| \cos\beta)) \frac{(a + r_-^2)(dp + r_-^2)}{(a + r_+^2)(dp + r_+^2)}. \tag{69}$$

The equation (60) implies

$$2||\boldsymbol{\theta}|| \cos\beta = r_+ u(r_+) + r_- u(r_-) = r_+ + r_- \frac{a}{r_+} - \frac{a}{r_-}. \tag{70}$$

By substituting (69) into (70), and performing a little algebra, it can be shown that

$$R = S_p(t) \frac{(r_-^2 + a)(r_-^2 + dp)}{(r_+^2 + a)(r_+^2 + dp)}, \tag{71}$$

where, for $m \geq 1$,

$$S_m(t) = t^m \exp(-a(t - t^{-1})/2) \quad \textit{and} \quad t = \frac{r_+}{r_-} \geq 1. \tag{72}$$

It can be shown that for a given constant $0 < \gamma < 1$, there exists $d_1 > 0$ such that for any $0 \leq \beta \leq \beta_o$ and for any $||\boldsymbol{\theta}|| \geq c$,

$$\frac{r_-^2 + dp}{r_+^2 + dp} \geq \gamma. \tag{73}$$

Using (61) we have

$$\frac{r_-^2 + a}{r_+^2 + a} = \frac{r_-}{r_+} \sqrt{\frac{{r_-^0}^2 + 4a}{{r_+^0}^2 + 4a}}. \tag{74}$$

It can be shown that the unique minimum of $[{r_-^0}^2 + 4a]/[{r_+^0}^2 + 4a]$ occurs at $\beta = 0$ and $||\boldsymbol{\theta}|| = (c^2 + 4a)^{1/2}$, which implies that

$$\sqrt{\frac{{r_-^0}^2 + 4a}{{r_+^0}^2 + 4a}} \geq \frac{2\sqrt{a}}{c + \sqrt{c^2 + 4a}}. \tag{75}$$

By (73) and (74), we only need to prove, for $1 \leq t \leq t^*$,

$$R \geq \gamma S_{p-1}(t) \sqrt{\frac{{r_-^0}^2 + 4a}{{r_+^0}^2 + 4a}} > 1, \quad a.e. \tag{76}$$

and, furthermore by (75)

$$R \geq \gamma S_{p-1}(t) \frac{2\sqrt{a}}{c + \sqrt{c^2 + 4a}} > 1, \quad a.e. \tag{77}$$

where

$$t^* = \max_{||\boldsymbol{\theta}|| > c, 0 \leq \beta \leq \beta_0} \frac{r_+(a, \boldsymbol{\theta}, \beta)}{r_-(a, \boldsymbol{\theta}, \beta)}. \tag{78}$$

The value of t^* can be calculated as follows. For fixed $||\boldsymbol{\theta}||$, r_+ is decreasing in β and r_- is increasing in β. Consequently t is decreasing in β, which implies

$$\sup_{0\le\beta\le\beta_0} t = \frac{r_+}{r_-}\Big|_{\beta=0} = \frac{||\boldsymbol{\theta}|| + c + \sqrt{(||\boldsymbol{\theta}|| + c)^2 + 4a}}{||\boldsymbol{\theta}|| - c + \sqrt{(||\boldsymbol{\theta}|| - c)^2 + 4a}} = \exp\Big(\int_{||\boldsymbol{\theta}||-c}^{||\boldsymbol{\theta}||+c} \frac{dy}{\sqrt{y^2+4a}}\Big) \tag{79}$$

is decreasing in $||\boldsymbol{\theta}||$. Then

$$t^* = \frac{c + \sqrt{c^2 + a^2}}{\sqrt{a}}. \tag{80}$$

Straightforward calculation shows that

$$t^* - \frac{1}{t^*} = \frac{2c}{\sqrt{a}}. \tag{81}$$

Clearly, for any m, $S_m(1) = 1$, $S'_m(1) = m - a$, and, for $t \ge 1$, $S_m(t)$ decreases if $a \ge m$; $S_m(t)$ increases to a unique maximum and then decreases to zero if $a < m$. When $a < p - 2$ let t_{**} and t_* be, respectively, the smallest and the largest solutions to $S_{p-2}(t) = 1/\gamma$, for $t > 1$. Suppose $a < p - 4$. Let t_0 be the unique solution to $S_{p-4}(t) = 1, t > 1$. It is easy to check that for sufficiently large $\gamma < 1$, $t_{**} \le t_0$. Therefore $S_{p-4}(t_{**}) \ge 1$ and, for $1 < t \le t_{**}$,

$$R \ge S_{p-4}(t)\frac{(1 + a/r_-^2)(1 + dp/r_-^2)}{(1 + a/r_+^2)(1 + dp/r_+^2)} > 1, \quad a.e.. \tag{82}$$

When $t_{**} \le t \le t_*$, $S_{p-2}(t) \ge 1/\gamma$ and then

$$R \ge S_{p-2}(t)\frac{(1 + a/r_-^2)(r_-^2 + dp)}{(1 + a/r_+^2)(r_+^2 + dp)} \ge S_{p-2}(t)\frac{1 + a/r_-^2}{1 + a/r_+^2}\gamma > 1, \quad a.e.. \tag{83}$$

To prove $R > 1$, it is sufficient to show, for $t_* \le t \le t^*$, that

$$S_{p-1}(t) \ge \frac{c + \sqrt{c^2 + 4a}}{\gamma 2\sqrt{a}}. \tag{84}$$

In fact, we only need to check that the above inequality holds at t_* and t^*. Since $S_{p-2}(t_*) = 1/\gamma$ and $S_{p-1}(t_*) = t_* S_{p-2}(t_*) = t_*/\gamma$, (84) at t_* is equivalent to

$$t_* \ge \frac{c + \sqrt{c^2 + 4a}}{2\sqrt{a}}. \tag{85}$$

Therefore, it is equivalent to

$$S_{p-2}(c + \sqrt{c^2 + 4a}/2\sqrt{a}) \ge 1/\gamma. \tag{86}$$

By direct substitution, it is equivalent to

$$\Big[\frac{c + \sqrt{c^2 + 4a}}{2\sqrt{a}}\Big]^{p-2} e^{-c\sqrt{a}/2} \ge 1/\gamma. \tag{87}$$

By (80) and (81), at t^*, (84) is equivalent to

$$\Big[\frac{2\sqrt{a}}{c+\sqrt{c^2+4a}}\Big]\Big[\frac{c+\sqrt{c^2+a}}{\sqrt{a}}\Big]^{p-1} e^{-c\sqrt{a}} \geq \gamma. \tag{88}$$

Since the left hand sides of (87) and (88) are decreasing in a, it can be shown that, for $a < a_o$, we can choose γ, and then choose d_1, such that (87) and (88) hold. It can be concluded that $R > 1 \quad a.e.$.

Thus the lemma is proved. □

References

[1] BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **4** 716-761.

[2] BERGER, J. (1985a). 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.

[3] BERGER, J. (1985b). In defense of the likelihood principle: axiomatics and coherency. In *Bayesian Statistics II*, J. M. Bernardo et. al. (Eds.). North-Holland, Amsterdam.

[4] BERGER, J. (1985c). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

[5] BERGER, J. (1985d). A review of J. Kiefer's work on conditional frequentist statistics. In *The Collected Works of Jack Kiefer,* L. Brown, I. Olkin, J. Sacks, and H. Wynn (Eds.). Springer-Verlag, New York.

[6] BERGER, J. and WOLPERT, R. L. (1984). *The Likelihood Principle.* Institute of Mathematical Statistics, Hayward, California.

[7] BROWN, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Ann. Statist.* **6**, 59-71.

[8] GLESER, L. J. and HWANG, J. T. (1987). The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variable and related models. *Ann. Statist.,* To Appear.

[9] HWANG, J. and CASELLA, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10**, 868-881.

[10] HWANG, J. and CASELLA, G. (1984). Improved set estimators for a multivariate normal mean. *Statistics and Decisions,* Supplement Issue No. **1**(1984), 3-16.

[11] JOHNSTONE, I. (1987). On the inadmissibility of Stein's unbiased estimate of loss. In *Statistical Decision Theory and Related Topics IV*, S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

[12] KIEFER, J. (1975). Conditional confidence approach in multi-decision problems. In *Multivariate Analysis IV,* P. R. Krishnaiah (Ed.). Academic Press, New York.

[13] KIEFER, J. (1976). Admissibility of conditional confidence procedures. *Ann. Math. Statist.* **4**, 836-865.

[14] KIEFER, J. (1977). Conditional confidence statements and confidence estimators (with Discussion). *J. Amer. Statist. Assoc.* **72**, 789-827.

[15] LU, K. L. and BERGER, J. (1987). Estimation of normal means: estimated loss frequentist approach. Technical Report #87-16, Department of Statistics, Purdue University.

[16] ROBINSON, G. K. (1979a). Conditional properties of statistical procedures. *Ann. Statist.* **7**, 742-755.

[17] ROBINSON, G. K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist.* **7**, 756-771.

[18] RUKHIN, A. L. (1987). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics IV*, S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

[19] SANDVED, E. (1968). Ancillary statistics and estimation of the loss in estimation problems. *Ann. Math. Statist.* **39**, 1756-1758.

[20] STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135-1151.