

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

75-29,216

FAITH, Ray Edwin, 1944-
MINIMAX BAYES SET AND POINT ESTIMATORS OF A
MULTIVARIATE NORMAL MEAN.

The University of Michigan, Ph.D., 1975
Statistics

Xerox University Microfilms, Ann Arbor, Michigan 48106

MINIMAX BAYES SET AND POINT ESTIMATORS
OF A MULTIVARIATE NORMAL MEAN

by
Ray Edwin Faith

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
1975

Doctoral Committee:

Associate Professor Richard A. Olshen, Co-Chairman
Professor Michael B. Woodroffe, Co-Chairman
Professor Bruce M. Hill
Professor Donald C. Pelz

DEDICATION

This thesis is dedicated to my parents Nora Emigh Faith and
Chalmers C. Faith, Sr.

.

ACKNOWLEDGMENTS

I am indebted to Richard A. Olshen, Michael B. Woodroffe, and Vinayak M. Joshi, each of whom made direct contributions to this paper, and without whose inspiration and guidance it would not have come into being. This research was supported in part by a grant from the Horace H. Rackham School of Graduate Studies at The University of Michigan.

TABLE OF CONTENTS

	<u>Page</u>
DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	v
LIST OF ILLUSTRATIONS	vi
INTRODUCTION	1
CHAPTER ONE: POINT ESTIMATORS	6
General Framework	
Minimax Bayes Estimators	
CHAPTER TWO: CONFIDENCE SETS	16
Background	
Bayes confidence sets	
Convexity of the Bayes sets	
Geometry of C^B	
Volume and Coverage Probabilities	
Comparison of C^B and C^O	
Numerical results	
APPENDIX	66
REFERENCES	69

LIST OF TABLES

	<u>Page</u>
TABLE 1 Volume comparison when $p = 3$	61
TABLE 2 Coverage probability when $p = 3$	62
TABLE 3 Volume comparison when $p = 5$	63
TABLE 4 Coverage probability when $p = 5$	64

LIST OF ILLUSTRATIONS

	<u>Page</u>
Figure 1 Bayes and usual confidence sets when $x = 0$	33
Figure 2 Bayes set when $x \neq 0$	34
Figure 3 Intersection of Bayes set with spheres about x	36
Figure 4 The point where a sphere about x intersects the surface of the Bayes region	42

INTRODUCTION

Let $X = (X_1, \dots, X_p)'$, $p \geq 3$, be a random p -vector drawn from a multivariate normal population having unknown mean vector $\theta = E(X)$ and known covariance matrix $E[(X-\theta)(X-\theta)'] = I$, the $p \times p$ identity matrix.

The classical point estimator of θ is $\hat{\theta}^0(X) = X$ itself. This is the maximum likelihood estimator of θ (Rao, p. 448). Since X is a complete sufficient statistic for θ and $E(X) = \theta$, $\hat{\theta}^0$ is the minimum variance unbiased estimator of θ (Rao, p. 261). For squared Euclidean error loss $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|^2 = \sum_{i=1}^p (\theta_i - \hat{\theta}_i)^2$, $\hat{\theta}^0$ is the best invariant estimator of θ , with respect to the group of translations in $\mathbb{R}^p$, as in the univariate case (see Ferguson, p. 168). It is minimax with respect to this loss (applying for example Theorem 2 of Ferguson, p. 90, using conjugate normal priors).

It is the formal Bayes estimator with respect to the improper prior consisting of Lebesgue measure on $\mathbb{R}^p$ (The posterior density is the likelihood function.). Each component of $\hat{\theta}^0$ is an admissible estimator of the corresponding component of θ , and each pair of components of $\hat{\theta}^0$ is an admissible estimator of the corresponding pair of components of θ , with respect to squared Euclidean error loss for these components (Hodges and Lehmann (1951), Stein (1956)). However, in spite of all of these good properties, $\hat{\theta}^0$ is not an admissible estimator of θ .

The risk function of $\hat{\theta}^0$ is

$$R^0(\theta) = E_{\theta}[L(\theta, \hat{\theta}^0)] = E_{\theta}[|X - \theta|^2] = p$$

for all θ . Stein (1956) introduced the family of estimators

$$\hat{\theta}^1 = \left(1 - \frac{b}{a + |X|^2}\right)X$$

which have risk $R^1(\theta) < p$ for all θ , if the constants a and b satisfy $a \geq 0$ and $0 < b < 2(p-2)$.

In James and Stein (1961) the authors explored a particular case of the above,

$$\hat{\theta}^2 = \left(1 - \frac{p-2}{|X|^2}\right)X \quad .$$

This estimator is better than $\hat{\theta}^0$, but may itself be improved by making the estimate zero whenever the multiplicative scalar is negative. The estimator

$$\hat{\theta}^3 = \left(1 - \frac{p-2}{|X|^2}\right)^+ X \quad (z^+ = \max(0, z))$$

has risk strictly less than that of $\hat{\theta}^2$ (Stein (1966)). For Stein's most recent contributions in this area see Stein (1974).

Although $\hat{\theta}^1$, $\hat{\theta}^2$, $\hat{\theta}^3$ are not Bayes estimators for any prior distribution on θ (Strawderman and Cohen (1971)), they behave somewhat like Bayes estimators in that they compromise between the data point X and an arbitrary but fixed point (the origin) which could be regarded as the "prior mean". They all have risk less than that of $\hat{\theta}^0$ but are not themselves admissible estimators, because all admissible rules for this problem are Bayes with respect to a (possibly improper) prior distribution (Brown (1971)).

Strawderman (1971) demonstrated a family of priors for which the Bayes rule

$$\hat{\theta}^4 = E(\theta | X)$$

has risk less than p , using improper priors for θ when $p = 3, 4$ and proper priors when $p \geq 5$. For $p \geq 5$ these estimators are clearly both minimax and admissible. In chapter one of this paper his results are extended by showing general conditions on the prior distribution under which the Bayes estimator $\hat{\theta}^4$ has risk less than p . In particular it is shown that a family of multivariate Student's t priors for θ meet these conditions.

The principal concern of this paper is not with point estimation, but rather with the formation of confidence sets for θ . The usual confidence procedure C^0 specifies for each X the spherical ball

$$C_X^0 = \{\theta: |X-\theta|^2 \leq r_0^2\}$$

where r_0^2 is chosen so that the coverage probability $1 - \alpha = P_\theta[\theta \in C_X^0]$ is of a predetermined size.

The usual confidence procedure C^0 enjoys many properties similar to those of the usual point estimator $\hat{\theta}^0$. For each X , C_X^0 is the set of values of θ for which the likelihood function exceeds $K_0 = (2\pi)^{-p/2} e^{-r_0^2/2}$ and so is a set of 'high likelihood'. As a consequence of this it is also a Bayes credible region with respect to the formal Bayes prior of Lebesgue measure on $\mathbb{R}^p$. It is unbiased in the sense that for $\theta^1 \neq \theta$, $P_\theta[\theta^1 \in C_X^0] < P_\theta[\theta \in C_X^0]$. For any different procedure C with coverage probability greater than or equal to $1 - \alpha$, and for any θ , there are vectors $\theta^1 \neq \theta$ for which $P_{\theta^1}[\theta \in C_X] > P_{\theta^1}[\theta \in C_X^0]$ where $|\theta^1 - \theta| > 0$ may be chosen at will (Stein (1962)). In this sense, at least, any procedure competitive with C_X^0 will increase the chance of including incorrect values. The usual procedure is best translation invariant in that among all procedures with coverage probability $1 - \alpha$ or greater which are invariant under the group of

translations in $\mathbb{R}^p$ ($C_{X+Y} = C_X + Y$), C_X^0 has minimum volume (Stein (1962)). The usual confidence procedure is minimax in the sense that among all procedures having coverage probability at least $1 - \alpha$, C_X^0 minimizes $\sup_{\theta} E_{\theta}[\text{vol}(C_X)]$ (Stein (1962)).

Stein (1962) attempted to accommodate his results concerning the usual point estimator of θ to the problem of improving on C_X^0 when $p \geq 3$. Brown (1966) showed that the coverage probability could be increased by using spherical sets of radius r_0 , and Joshi (1967) made this point explicit by showing that the spherical sets C_X^1 of radius r_0 centered at the Stein estimator $\hat{\theta}^1$ have higher coverage probability than C_X^0 for all values of θ , whereas their volume is, of course, equal to that of C_X^0 .

In chapter 2 a theory of confidence procedures is developed, using definitions of preference and loss criteria introduced by Joshi (1969). The multivariate Student's t priors employed in chapter 1 are used to form Bayes confidence sets (sets where the posterior density is high). Conditions are specified under which the Bayes confidence sets are convex. Expressions are given for the volume and pointwise coverage probability of these procedures. For the case of $p = 3$ dimensions strong evidence (in the form of theorems) is given that the Bayes sets have lower volume and higher coverage probability than the usual confidence sets. Finally, tabulated values of the volume and coverage probability of the Bayes sets is given for various priors in $p = 3$ and $p = 5$ dimensions.

The Bayes set estimators are interesting in their own right, as minimax admissible decision rules. With regard to the more general problem wherein the covariance structure of the data is unknown, if results similar to the present ones can be obtained, then the Bayes sets

may serve as intermediate steps in the determination of simultaneous confidence intervals for all linear combinations of the parameters in the same way that the usual confidence sets are used in Scheffé's S-method (Scheffé (1953), Scheffé (1959)). As Olshen (1973) has pointed out, if these intervals are only made conditionally on having rejected the null hypothesis that $\theta = 0$ then the coverage probability of the intervals is strictly less than $1 - \alpha$. For intervals based on Bayes confidence sets, this phenomenon might be avoided.

CHAPTER 1

POINT ESTIMATION

1. General Framework. Throughout this paper it is assumed that the observable data consists of $X = (X_1, \dots, X_p)'$, which is a multivariate normal random p -vector having probability density function

$$f(x|\theta) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}|x-\theta|^2\right), \quad x \in \mathbb{R}^p,$$

where $\theta \in \mathbb{R}^p$ is unknown and $|x-\theta|^2$ denotes $\sum_{i=1}^p (x_i - \theta_i)^2$. In more concise notation the distribution of X is $N_p(\theta, I)$. The components of X are independent normal random variables for which $E(X_i) = \theta_i$ and $\text{Var}(X_i) = 1$, $i = 1, \dots, p$.

2. Point Estimation. Consider the problem of estimating θ with respect to squared Euclidean error loss $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|^2$. The risk function of an estimator $\hat{\theta}$ is $R(\theta) = E_{\theta}[|\theta - \hat{\theta}|^2]$.

The usual estimator of θ is $\hat{\theta}^0(X) = X$. This estimator, which has constant risk $R^0(\theta) \equiv p$, is minimax. The standard argument to this effect is that for every $\varepsilon > 0$ there is a conjugate Bayes prior for which the total Bayes risk is greater than $p - \varepsilon$. Although $\hat{\theta}^0$ is minimax for all p , when $p \geq 3$ there are many estimators that have risk strictly lower than that of $\hat{\theta}^0$.

Lemma 1.1. Let $\tilde{\theta} = [1 - r(|X|) \cdot \frac{p-2}{|X|^2}]X$ where $p \geq 3$, and $r(\cdot)$ is any non-decreasing function such that $0 \leq r(\cdot) \leq 2$, $r(\cdot) \neq 0$, and $r(\cdot) \neq 2$. Then the estimator $\tilde{\theta}$ has risk $R(\theta) = E_{\theta}[|\theta - \tilde{\theta}|^2] < p$ for all θ .

Proof. See Baranchik (1970) and Strawderman (1971).

It may be noted that the original Stein estimator

$$\hat{\theta}^1(x) = \left(1 - \frac{b}{a + |x|^2}\right)x$$

satisfies the conditions of this lemma if $a \geq 0$ and $b \leq 2(p-2)$, and that the James-Stein estimator

$$\hat{\theta}^2(x) = \left(1 - \frac{p-2}{|x|^2}\right)x$$

and its positive part version

$$\hat{\theta}^3(x) = \left(1 - \frac{p-2}{|x|^2}\right)^+ x$$

both satisfy the conditions of the lemma.

Strawderman (1971) demonstrated the existence when $p \geq 5$ of proper Bayes (and therefore admissible) estimators of θ which satisfy the conditions of this lemma and which are therefore both admissible and minimax. The following is a generalization of his argument which broadens the class of priors. We assume in what follows that $p \geq 3$. In the following we treat θ as a random vector, denoted by $\tilde{\theta}$.

Let $\tilde{\theta}$ have a multivariate normal conditional prior density

$$\pi(\theta|t) = (2\pi)^{-p/2} t^{p/2} \exp\left(-\frac{t}{2}|\theta|^2\right), \quad \theta \in \mathbb{R}^p, t \geq 0.$$

Each component of $\tilde{\theta}$ has mean zero and variance $\frac{1}{t}$, given that the prior precision $\underline{t}$ takes on the value t . Let the precision $\underline{t}$ itself have marginal prior density $g(t)$, $t \geq 0$, which is permitted to have an infinite integral. If $g(t)$ is not proper then the marginal prior density of $\tilde{\theta}$, $\pi(\theta) = \int_0^\infty \pi(\theta|t)g(t)dt$, will not be proper either, by Fubini's theorem. On the other hand, if $g(t)$ is such that the marginal density $f(x)$ of X is finite for all x , then the posterior density of $\tilde{\theta}$, $\pi(\theta|x) = \frac{f(x|\theta) \cdot \pi(\theta)}{f(x)}$, is a proper density, since the

denominator of this is equal to the integral of the numerator with respect to θ . In examples 1 and 2 priors will be used for which this is the case. It is assumed in what follows that either $g(t)$ is proper or, if it is improper, that $g(t)$ is sufficiently well behaved that $\pi(\theta|x)$ is proper.

Lemma 1.2. Under the above specification of the prior density of θ

$$\pi(\theta) = \int_0^\infty \pi(\theta|t)g(t)dt ,$$

the (possibly formal) Bayes estimator of θ is of the form $\tilde{\theta}$ in Lemma 1.1 with

$$r(|X|) = \frac{|X|^2}{p-2} E\left[\frac{t}{1+t} | X\right] .$$

Proof.
$$\begin{aligned} \hat{\theta}(X) &= E[\tilde{\theta}|X] = E[E[\tilde{\theta}|t, X]|X] \\ &= E\left[\frac{1}{1+t} X | X\right] = E\left[\frac{1}{1+t} | X\right] \cdot X \\ &= (1 - E\left[\frac{t}{1+t} | X\right]) X . \end{aligned}$$

This proof is found in Strawderman (1971).

Theorem 1.3. Let the marginal prior density $g(t)$ of t satisfy $\lim_{t \rightarrow \infty} t^2 g(t) < \infty$. If $g(t)$ is differentiable in t and if

$$h(t) = \frac{g'(t)}{g(t)} t(t+1) - 2t$$

is well defined for $t > 0$, then the (possibly formal) Bayes estimator $\hat{\theta} = E(\tilde{\theta}|X)$ has pointwise risk $R(\theta)$ less than p if $h(t)$ satisfies

i) $h(t) \nearrow$ in t ($h(t)$ is non-increasing in t)

and ii) $h(0+) \leq p/2 - 3$ ($\lim_{\substack{t \rightarrow 0 \\ t > 0}} h(t) \leq p/2 - 3$).

Proof. Let $\tilde{\lambda} = \frac{t}{1+t}$. This is a strictly monotone increasing function of t . The random parameter $\tilde{\lambda}$ has marginal density

$$d(\lambda) = g\left(\frac{\lambda}{1-\lambda}\right) \cdot \frac{1}{(1-\lambda)^2}, \quad 0 \leq \lambda \leq 1 \quad (d(1) \stackrel{\text{df}}{=} d(1-) < \infty).$$

Differentiate this to obtain the function

$$\begin{aligned} e(\lambda) &= \frac{d'(\lambda)}{d(\lambda)} \cdot \lambda = \frac{g'\left(\frac{\lambda}{1-\lambda}\right)}{g\left(\frac{\lambda}{1-\lambda}\right)} \left(\frac{\lambda}{1-\lambda}\right) \left(\frac{1}{1-\lambda}\right) - 2 \frac{\lambda}{1-\lambda} \\ &= h(t(\lambda)) \end{aligned}$$

Observe that conditions i) and ii) are equivalent to

i') $e(\lambda) \nearrow$ in λ

and ii') $e(0+) \leq p/2 - 3$.

From Lemma 1.2

$$\begin{aligned} r(|x|) &= \frac{|x|^2}{p-2} E\left[\frac{t}{1+t} \mid X=x\right] = \frac{|x|^2}{p-2} E[\tilde{\lambda} \mid X=x] \\ &= \frac{|x|^2}{p-2} \frac{\int_0^1 \lambda d(\lambda) f(x|\lambda) d\lambda}{\int_0^1 d(\lambda) f(x|\lambda) d\lambda} \end{aligned}$$

Since $X - \tilde{\theta} \mid \lambda \sim N_p(0, I)$ and $\tilde{\theta} \mid \lambda \sim N_p(0, \frac{1-\lambda}{\lambda} I)$ are independent given $\tilde{\lambda} = \lambda$, it follows that $X \mid \lambda = (X - \tilde{\theta}) + \tilde{\theta} \mid \lambda \sim N_p(0, \frac{1}{\lambda} I)$. Therefore

$$(1.2.1) \quad f(x|\lambda) = (2\pi)^{-p/2} \lambda^{p/2} \exp\left(-\frac{1}{2}\lambda |x|^2\right)$$

and

$$r(|x|) = \frac{|x|^2}{p-2} \frac{\int_0^1 \lambda^{p/2+1} d(\lambda) \exp\left(-\frac{1}{2}\lambda |x|^2\right) d\lambda}{\int_0^1 \lambda^{p/2} d(\lambda) \exp\left(-\frac{1}{2}\lambda |x|^2\right) d\lambda}.$$

Integrate the numerator by parts to obtain

$$\begin{aligned}
 (1.2.2) \quad r(|x|) &= \frac{2}{p-2} \left[\left(\frac{p}{2} + 1 \right) + \frac{\int_0^1 \lambda^{p/2+1} d'(\lambda) \exp(-\frac{1}{2}\lambda |x|^2) d\lambda}{\int_0^1 \lambda^{p/2} d(\lambda) \exp(-\frac{1}{2}\lambda |x|^2) d\lambda} \right. \\
 &\quad \left. - \frac{d(1) \exp(-\frac{1}{2}|x|^2)}{\int_0^1 \lambda^{p/2} d(\lambda) \exp(-\frac{1}{2}\lambda |x|^2) d\lambda} \right] \\
 &= \frac{2}{p-2} \left[\left(\frac{p}{2} + 1 \right) + E[e(\lambda) | X=x] \right. \\
 &\quad \left. - \frac{d(1)}{\int_0^1 \lambda^{p/2} d(\lambda) \exp(\frac{1}{2}(1-\lambda) |x|^2) d\lambda} \right]
 \end{aligned}$$

The claim is that $r(|x|)$ meets the conditions of Lemma 1.1. To begin with, it is non-decreasing. The first term of (1.2.2) is constant and the third term is either zero (if $d(1) = 0$) or is an increasing function of $|x|$. The second term is non-decreasing in $|x|$ because $-e(\lambda)$ is non-decreasing in λ by i') and $f(\lambda|x)$ has monotone likelihood ratio with respect to the parameter $-|x|^2$ (see Lehmann (1959), p. 74). Therefore $r(|x|)$ is non-decreasing in $|x|$.

Secondly $r(|x|) \geq 0$ and $r(|x|) \neq 0$ since $r(|x|) = \frac{|x|^2}{p-2} E[\lambda | X=x]$ and $P[\lambda > 0] = 1$.

Thirdly, since $e(\lambda)$ is non-increasing in λ by assumption i'), it follows from (1.2.2) that

$$r(|x|) \leq \frac{2}{p-2} \left[\left(\frac{p}{2} + 1 \right) + \lim_{\substack{\lambda \rightarrow 0 \\ \lambda > 0}} e(\lambda) \right]$$

which is less than or equal to 2 if ii') holds.

Finally, to show that the conditions of Lemma 1.1 are met, it

remains to show that $r(|x|) \neq 2$. If $r(|x|) \equiv 2$ then by (1.2.2) it must be true that both $d(1) = 0$ and that $E[e(\lambda)|X] \equiv p/2 - 3$. This implies that

$$\lim_{\lambda \rightarrow 1} d(\lambda) = 0$$

and

$$e(\lambda) = \frac{d'(\lambda)}{d(\lambda)}\lambda \equiv p/2 - 3.$$

These conditions are mutually exclusive since the second one implies that

$$d(\lambda) = C \cdot \lambda^{p/2 - 3} \quad \text{where } C > 0.$$

The proof of the above theorem includes the following corollary.

Corollary 1.4. If the marginal prior density $d(\lambda)$ of $\lambda = \frac{t}{1+t}$ is differentiable in t , if $d(1) < \infty$, and if

$$e(\lambda) = \frac{d'(\lambda)}{d(\lambda)}\lambda, \quad 0 \leq \lambda \leq 1$$

satisfies i') $e(\lambda) \nearrow$ in λ

and ii') $e(0+) \leq p/2 - 3$

then $\hat{\theta} = E(\theta|X)$ has risk less than p .

Note: M. Woodroffe provided me with the crucial step in the development of the above results, namely the use of the monotone likelihood ratio argument.

Example 1. Strawderman (1971) used the class of priors for λ

$$d(\lambda) = a\lambda^{a-1}, \quad 0 \leq \lambda \leq 1, \quad 0 < a < \begin{cases} 1 & \text{if } p \geq 6 \\ \frac{1}{2} & \text{if } p = 5. \end{cases}$$

This class of priors is substantially broadened by Corollary 1.4. If $\lambda \sim \text{Beta}(a,b)$, that is, if λ has marginal density

$$d(\lambda) = c \cdot \lambda^{a-1} (1-\lambda)^{b-1},$$

then

$$d'(\lambda) = c \cdot d(\lambda) \cdot \left[\frac{a-1}{\lambda} - \frac{b-1}{1-\lambda} \right]$$

so that

$$e(\lambda) = (a-1) - (b-1) \cdot \frac{\lambda}{1-\lambda}.$$

By Corollary 1.4, the Bayes estimator has risk less than p if

$$i') \quad b \geq 1$$

and $ii') \quad a \leq p/2 - 2.$

Discussion. For the Strawderman priors these conditions are met. Strawderman (1972) showed that when $p < 5$ there exist no proper Bayes minimax estimators of θ . However, as this example shows, there are improper priors for which the Bayes rule is minimax. If $a > -p/2$ and $b > 0$ then the marginal density of X , which by (1.2.1) equals

$$f(x) = \int_0^1 \lambda^{\frac{p}{2}+a-1} (1-\lambda)^{b-1} (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}\lambda|x|^2} d\lambda$$

is finite for all x , so that the generalized Bayes rule $E(\theta|X)$ is well defined and is minimax for parameter values $-p/2 < a \leq p/2 - 2$, $b \geq 1$, even when $p = 3$ or 4 .

The main topic of this paper is confidence procedures rather than point estimators. A Bayes confidence set is a set of values of θ for which the posterior density $\pi(\theta|x)$ is larger than some value $k(x)$. Intuitively, Bayes priors which yield minimax point estimators of θ

are likely to be priors for which the confidence sets do well, in the sense of being small and of frequently including the true value of θ . These ideas will be made more precise in Chapter 2.

In this regard, note that if λ is given a $\text{Beta}(a, b)$ prior density, then the marginal prior density of θ , $\pi(\theta)$, is bounded if and only if $b > p/2 - 1$. Otherwise, for example, if $b = 1$ as in the Strawderman priors, $\pi(\theta)$ has an infinite discontinuity at $\theta = 0$. If such is the case then the posterior density $\pi(\theta|x)$ also has an infinite discontinuity at the origin, so that every Bayes confidence set will include a set of values close to the origin. If $|x|$ is large, the set will be disconnected, with a small part at the origin and another part near x , so that it will not be very widely acceptable as a confidence set.

In this sense the more general class of priors specified above, for which b can take on any non-negative value, are an improvement in that they permit the removal of the infinite discontinuity in $\pi(\theta|x)$ at $\theta = 0$, by choosing $b > p/2 - 1$. This kind of improvement will be exploited in chapter 2, using the priors in the following example.

Example 2. Suppose that the prior precision t has marginally a $\text{Gamma}(a, b)$ density, or, in terms of Theorem 1.3 let

$$g(t) = c \cdot t^{a-1} e^{-bt} ,$$

$$g'(t) = c \cdot g(t) \cdot \left(\frac{a-1}{t} - b \right) ,$$

and

$$h(t) = (a - 1 - bt)(t+1) - 2t$$

$$= (a-1) + (a-3-b)t - bt^2 .$$

By Theorem 1.3 the Bayes point estimator will have risk less than p if

the prior parameters satisfy

$$i) \quad a \leq b + 3$$

and ii) $a \leq p/2 - 2$.

As in the preceding example, the marginal density of X , $f(x)$ is finite for all x if $b > 0$ and $a > -p/2$, so that the posterior density is well defined in this case.

Discussion. This class of priors is more useful than those of Example 1 for the formation of confidence sets because the marginal prior density of θ can be expressed in closed form. Specifically

$$\pi(\theta) = c \cdot (2b + |\theta|^2)^{-(a+p/2)}, \quad \theta \in \mathbb{R}^p,$$

so that θ has a multivariate Student's t prior distribution. If $b > 0$ and $a > -p/2$ this prior clearly has no infinite discontinuity. In Chapter 2 it is shown that if $b > (a + \frac{p}{2})/8$, then, in fact, the posterior density is unimodal and the Bayes sets are convex.

For neither example, is there a closed form expression for the Bayes estimator $\hat{\theta} = E(\theta|X)$. Of course it could be computed numerically, but this problem is not pursued further here.

As an aside it may be of interest to show that Theorem 1.3 is consistent with Strawderman's (1972) result that there is no proper prior if $p = 3$ or 4 for which the Bayes rule is minimax.

Lemma 1.5. Conditions i) and ii) of Theorem 1.3 cannot be met simultaneously by a proper prior if $p = 3$ or 4 .

Proof. Conditions i) and ii) are equivalent to conditions i') and ii') of Corollary 1.4. Assume these are both true. If $p = 3$ or 4 then by ii') it must be true that $e(0+) \leq -1$, and so by i') that $e(\lambda) \leq -1$

for all λ . That is to say

$$\frac{d'(\lambda)}{d(\lambda)} \leq -\frac{1}{\lambda} \quad \text{for all } \lambda > 0 .$$

Integrate this to obtain

$$\log d(\lambda_2) - \log d(\lambda_1) \leq -(\log \lambda_2 - \log \lambda_1)$$

for all $\lambda_2 > \lambda_1 > 0$. Therefore

$$\frac{d(\lambda_2)}{d(\lambda_1)} \leq \frac{\lambda_1}{\lambda_2}$$

and

$$d(\lambda_1) \geq d(\lambda_2) \cdot \frac{\lambda_2}{\lambda_1} = \frac{c}{\lambda_1} ,$$

which is not integrable between 0 and λ_2 .

CHAPTER 2

CONFIDENCE SETS

1. Background. Let $\Theta = \mathbb{R}^p$ denote the parameter space of possible values of θ and $X = \mathbb{R}^p$ denote the data space of possible values of X . As in Chapter 1, for given θ , X is assumed to be distributed as a $N_p(\theta, I)$ random vector. A confidence procedure C is a Borel subset of the Cartesian product space $X \times \Theta = \mathbb{R}^{2p}$ with the usual topology. For each $x \in X$ let C_x , called the confidence set for θ at x , be the Borel subset of Θ defined by

$$C_x = \{\theta: (x, \theta) \in C\} .$$

In a similar fashion the θ section of C is the Borel subset of X

$$C_\theta = \{x: (x, \theta) \in C\} .$$

For a confidence procedure C , the coverage probability is the Borel measurable function

$$P_\theta[C_\theta] = P_\theta[\theta \in C_x] .$$

By $P_\theta[C_\theta]$ is meant the measure of set $C_\theta \subseteq X$ induced by the distribution of X when θ is the true state of nature. This is equal to the usual expression for the coverage probability on the right because of the tautology $x \in C_\theta$ if and only if $\theta \in C_x$. The measurability of $P_\theta[C_\theta]$ follows from Tonelli's theorem.

The confidence procedure C is said to have level $1 - \alpha$ if the coverage probability satisfies

$$P_\theta[C_\theta] \geq 1 - \alpha$$

for all $\theta \in \Theta$. It is said to have size $1-\alpha$ if equality holds in the above for all θ .

The usual size $1-\alpha$ confidence procedure is

$$(2.1.1) \quad C^0 = \{(x, \theta): |x - \theta|^2 \leq r_0^2(\alpha)\}$$

where $r_0^2(\alpha)$ is the $100 \cdot (1-\alpha)^{\text{th}}$ percentile of the χ_p^2 (chi-square with p degrees of freedom) distribution. Some criteria with respect to which C^0 is a good procedure are mentioned in the introduction.

In this paper we use criteria of preference among confidence procedures for which one procedure is better than another if its confidence sets are more likely to include the true value of θ while at the same time they have smaller volume. The volume of C_x (denoted $\text{vol}(C_x)$) is its Lebesgue measure in $\Theta = \mathbb{R}^p$. This is a Borel measurable function of x by Tonelli's theorem.

A partial ordering of confidence procedures is given, following Joshi (1969), by

Definition 2.1. For any two confidence procedures C^1 and C^2 , C^1 is strongly preferable to C^2 if

$$i) \quad P_\theta[C_\theta^1] \geq P_\theta[C_\theta^2] \quad \text{for all } \theta \in \Theta$$

and

$$ii) \quad \text{vol}[C_x^1] \leq \text{vol}[C_x^2] \quad \text{for almost all } x \in X,$$

with strict inequality either in i) for some $\theta \in \Theta$ or in ii) on a subset of positive measure in X .

A weaker partial ordering is obtained from this by replacing volume by expected volume.

Definition 2.2. For any two confidence procedures C^1 and C^2 , C^1 is weakly preferable to C^2 if

$$i) \quad P_{\theta}[C_{\theta}^1] \geq P_{\theta}[C_{\theta}^2] \quad \text{for all } \theta \in \Theta$$

and

$$ii) \quad E_{\theta}[\text{vol}(C_X^1)] \leq E_{\theta}[\text{vol}(C_X^2)] \quad \text{for all } \theta \in \Theta,$$

with strict inequality in either i) or ii) for at least one value of $\theta \in \Theta$. □

Joshi (1969) pointed out that for every confidence procedure C for which the coverage probability is less than 1. for some value of θ (which is an extremely weak condition) there is a procedure which is strongly preferable to it. A strongly preferable procedure is obtained by simply adding a point θ_0 to each confidence set C_X , where $P_{\theta_0}[C_{\theta_0}] < 1$. The new procedure so defined is trivially strongly preferable to C .

If a procedure C were to be defined as admissible only if there were no procedure C^* preferable to it, then the foregoing remarks show that there would exist no nontrivial admissible procedures.

On the other hand, adding sets of measure zero to confidence procedures is of little practical interest, because attention is usually restricted to procedures C for which the confidence sets C_X are homeomorphic to the unit ball in $\mathbb{R}^p$ or are convex sets. We circumvent these issues by adopting the following definitions.

Definition 2.3. Two procedures C and C^* are equivalent if their symmetric difference

$$C \Delta C^* = (C - C^*) \cup (C^* - C)$$

is a (Borel) set of Lebesgue measure zero.

Definition 2.4. A procedure C is strongly admissible if there is no procedure C^* , C^* not equivalent to C , such that C^* is weakly preferable to C .

Definition 2.5. A procedure C is weakly admissible if there is no procedure C^* , C^* not equivalent to C , such that C^* is strongly preferable to C .

Note. A procedure which is strongly admissible is also weakly admissible.

With these definitions it is possible to show the existence of admissible procedures. The only ones considered below will have confidence sets which are homeomorphic to balls.

The major results that have been obtained up to the present concerning the admissibility of the usual confidence procedures C^0 are the following.

1. If $p = 1$ or 2 then C^0 is strongly admissible among $1 - \alpha$ level confidence procedures (Joshi (1969)).
2. If $p \geq 3$ then C^0 is not even weakly admissible among $1 - \alpha$ level confidence procedures. In particular, the procedure

$$C' = \{(x, \theta) : |\hat{\theta} - \theta|^2 \leq r_0^2(\alpha)\}$$

is strongly preferable to C^0 , where

$$\hat{\theta} = \hat{\theta}(x) = \left(1 - \frac{b}{a + |x|^2}\right)x$$

(with b sufficiently small and a sufficiently large) is

the original Stein point-estimator of θ (Joshi (1967)).

It should be noted that the second result was also obtained by Brown (1966) using a different approach.

The parallelism between these results and those for point estimation is apparent. My effort in the ensuing sections is to construct confidence sets based on Bayes priors which are strongly preferable to C^0 , thus paralleling the Strawderman work and the results in Chapter 1 of this thesis regarding minimax Bayes point estimates. Of course attention is restricted to the case $p \geq 3$ in what follows, given Joshi's result for $p = 1$ and 2.

2. Bayes Confidence Sets. Let the parameter θ have conditional prior density

$$(2.2.1) \quad \pi(\theta|t) = \left(\frac{t}{2\pi}\right)^{p/2} \exp\left(-\frac{t}{2}|\theta|^2\right), \quad t \geq 0, \quad \theta \in \mathbb{R}^p$$

where the prior precision t is either a fixed constant (if $t \equiv 0$ set $\pi(\theta|t) \equiv 1$) or else t has a prior density (not necessarily proper) $g(t)$, $t \geq 0$. The conditional distribution of θ given t is multivariate normal with mean zero.

As in Chapter 1, the prior mean can be any other fixed value as well. If another value is used, translate the coordinate system so that the prior mean is the origin and proceed as follows.

Suppose that the datum X has, conditional on $t = t$ and $\theta = \theta$, the multivariate normal distribution $X|\theta, t \sim N_p(\theta, I)$. Note that given θ , X and t are conditionally independent. As in Chapter 1, θ has marginal prior density

$$\pi(\theta) = \int_0^\infty \pi(\theta|t)g(t)dt$$

unless $\tilde{t}$ is a constant, $\tilde{t} \equiv t_0$, in which case $\pi(\theta) = \pi(\theta|t_0)$. In terms of the prior $\pi(\theta)$, conditional on $X = x$ the posterior density $\pi(\theta|x)$ is of the form

$$(2.2.2) \quad \pi(\theta|x) = \frac{\pi(\theta) (2\pi)^{-p/2} \exp(-\frac{1}{2}|x-\theta|^2)}{\int_{\mathbb{R}^p} \pi(\theta) \cdot (2\pi)^{-p/2} \cdot \exp(-\frac{1}{2}|x-\theta|^2) d\theta} \\ = \frac{\pi(\theta) \cdot (2\pi)^{-p/2} \cdot \exp(-\frac{1}{2}|x-\theta|^2)}{E[\pi(\xi)]}$$

where $\xi \sim N_p(x, I)$.

For each Borel set $A \subseteq \Theta$, given $X = x$ the probability that $\tilde{\theta}$ is in A is the posterior coverage probability

$$(2.2.3) \quad P[\tilde{\theta} \in A|x] = \int_A \pi(\theta|x) d\theta \quad .$$

It follows from the Neyman-Pearson lemma that for a fixed posterior coverage probability $1-\alpha$, the set A of minimum volume is the subjectivist region

$$(2.2.4) \quad C_x^S = \{\theta: \pi(\theta|x) \geq k_S(x)\}$$

where $k_S(x)$ is chosen to satisfy $P[\tilde{\theta} \in C_x^S|x] = 1-\alpha$. If $\pi(\theta)$ is regarded as subjective probability density for $\tilde{\theta}$, then C_x^S so defined is the set of smallest volume that, given the datum $X = x$, has subjective probability $1-\alpha$ of including $\tilde{\theta}$. The subjectivist confidence sets C_x^S are traditionally called credible regions. We will not follow tradition in this paper because later on we will be looking at other Bayes confidence procedures which do not have fixed posterior coverage probability $1-\alpha$. The confidence procedure

$C^S = \{(x, \theta) : \theta \in C_x^S\}$ will not generally be a $1 - \alpha$ level procedure since the expected coverage probability is

$$\begin{aligned}
 \int_{\mathbb{R}^p} P_\theta[C_\theta^S] \pi(\theta) d\theta &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} 1_{C^S}(x, \theta) \cdot f(x|\theta) \cdot \pi(\theta) dx d\theta \\
 &= \int_{\mathbb{R}^p} f(x) \cdot \int_{C_x^S} \frac{f(x|\theta) \pi(\theta)}{f(x)} d\theta dx \\
 &= \int_{\mathbb{R}^p} f(x) P[\theta \in C_x^S | x] dx \\
 &= \int_{\mathbb{R}^p} f(x) (1 - \alpha) dx \\
 &= 1 - \alpha .
 \end{aligned}$$

In the above expression $1_{C^S}(x, \theta)$ denotes the characteristic function of C^S .

Unless $P_\theta[C_\theta^S] \equiv 1 - \alpha$ (as, for example, if $\pi(\theta) \equiv 1$, in which case $C^S = C^0$) the coverage probability $P_\theta[C_\theta^S]$ will be less than or greater than $1 - \alpha$ depending on the value of θ . Since our interest is to find Bayes procedures which are preferable to C^0 , the subjectivist procedures are not pursued further here.

A second way to obtain Bayes confidence procedures is with respect to a loss function, the decision theoretic approach. Let the loss when θ is the state of nature and C_x the decision made be

$$(2.2.5) \quad L(\theta, C_x) = k_o \cdot \text{vol}(C_x) - I_{C_x}(\theta) .$$

For this loss the risk function is

$$R^C(\theta) = k_o \cdot E_\theta[\text{vol}(C_x)] - P_\theta[C_\theta] .$$

Lemma 2.6. The usual confidence procedure C^0 is minimax with respect to loss function (2.2.5) when

$$(2.2.6) \quad k_0 = (2\pi)^{-p/2} \cdot \exp(-r_0^2/2) \quad .$$

We take (2.2.6) as the definition of k_0 for the remainder of this paper.

Preliminary to the proof of this lemma we determine the general form of the Bayes rule for loss function (2.2.5). If θ has marginal prior density $\pi(\theta)$ and posterior density $\pi(\theta|x)$ then

$$E[L(\theta, C_x) | X=x] = k_0 \cdot \text{vol}(C_x) - P[\theta \in C_x | X=x]$$

is minimized for fixed $\text{vol}(C_x)$ by a set of the form

$$C_x = \{\theta: \pi(\theta|x) \geq k(x)\}$$

by the Neyman-Pearson Lemma.

The appropriate choice of $k(x)$ is $k(x) = k_0$ by the fact that

$$\frac{d}{d \text{vol}(C_x)} \int_{\pi(\theta|x) \geq k(x)} \pi(\theta|x) d\theta = k(x) \quad .$$

Therefore the Bayes decision theoretic rule is the procedure that specifies for each x the confidence set

$$(2.2.7) \quad C_x^D = \{\theta: \pi(\theta|x) \geq k_0\} \quad .$$

Proof of Lemma 2.6. If the prior precision t is a fixed constant, then

$$\pi(\theta|x) = \left(\frac{1+t}{2\pi}\right)^{p/2} \exp\left(-\frac{1}{2}(1+t)\left|\theta - \frac{x}{1+t}\right|^2\right) \quad .$$

By (2.2.6) and (2.2.7)

C_x^D is the spherical set ,

$$C_x^D = \{\theta: |\theta - \frac{x}{1+t}|^2 \leq \frac{r_o^2 + p \log(1+t)}{1+t}\} .$$

The conditional Bayes risk, given $X = x$ is

$$\begin{aligned} E[L(\theta, C_x^D) | X=x] &= k_o \text{vol}(C_x^D) - P[\theta \in C_x^D | X=x] \\ &= k_o \text{vol}\left(\text{sphere with radius } \left(\frac{r_o^2 + p \log(1+t)}{1+t}\right)^{1/2}\right) \\ &\quad - P\left[\left|N_p\left(0, \frac{1}{1+t}I\right)\right|^2 \leq \frac{r_o^2 + p \log(1+t)}{1+t}\right] \\ &= k_o \text{vol}(C_x^0) \cdot \frac{(r_o^2 + p \log(1+t))^{p/2}}{r_o^2 (1+t)^{p/2}} \\ &\quad - P[\chi_p^2 \leq [r_o^2 + p \log(1+t)]] . \end{aligned}$$

As t approaches zero this tends pointwise to the constant $k_o \text{vol}(C_x^0) - (1-\alpha)$ which is the risk of the usual procedure R^{C^0} . The total Bayes risk of C^D is

$$E[R^{C^D}(\theta)] = E[E[L(\theta, C_x^D) | X=x]] .$$

By bounded convergence $\lim_{t \rightarrow 0} E[R^{C^D}(\theta)] = R^{C^0}$. Therefore C^0 is minimax, since any procedure which has risk strictly less than $R^{C^0}(\theta) - \epsilon$ for some $\epsilon > 0$ has total Bayes risk less than that of a proper Bayes procedure by the usual argument. $\square$

The thread of the above argument was provided to me by Joshi during personal conversations.

The conjugate priors considered in the above proof do not yield procedures which are preferable to C^0 in the sense of Section 1. This is obvious from the following heuristic argument. Since on the

average X is near θ , for large $|\theta|$, $|X|$ will tend to be large. For large $|X|$, C_X^D , a sphere of fixed radius centered at $\hat{\theta} = \frac{1}{1+t}X$, will be far from X , and therefore on the average far from θ . So for $|\theta|$ sufficiently large C^D can have arbitrarily low coverage probability when the prior precision t is constant. However if t is allowed to be random, this phenomenon can be avoided, as was suggested by the minimaxity of the Bayes point estimates of Examples 1 and 2 of Chapter 1, and will be shown in Sections 5, 6 and 7 below.

Consider, in particular, Example 2 of Chapter 1. Let t be distributed as a $\text{Gamma}(a, b)$ random variable. Then θ has marginal prior density

$$\pi(\theta) = c \cdot (2b + |\theta|^2)^{-(a+p/2)}.$$

By (2.2.2) for this prior the Bayes decision theoretic procedure (2.2.7) is

$$\begin{aligned} C_X^D &= \left\{ \theta : \frac{(2b + |\theta|^2)^{-(a+p/2)} \exp(-\frac{1}{2}|x-\theta|^2) (2\pi)^{-p/2}}{E[(2b + \chi^2)^{-(a+p/2)}]} \geq k_0 \right\} \\ &= \left\{ \theta : \left[\frac{\exp(r_0^2)}{\exp(|x-\theta|^2)} \right]^{1/2} \geq E \left[\frac{(2b + |\theta|^2)^{a+p/2}}{(2b + \chi^2)^{a+p/2}} \right] \right\} \end{aligned}$$

where χ^2 is a noncentral chi-square random variable with p degrees of freedom and mean $p + |x|^2$.

This decision rule is interesting, but difficult to work with since it is defined in terms of a non-trivial expectation. Instead we will work with the explicitly defined procedure

$$(2.2.8) \quad C^B = \{(x, \theta) : \left[\frac{\exp(r_0^2)}{\exp(|x-\theta|^2)} \right]^{1/2} \geq \frac{2b + |\theta|^2}{2b + |x|^2}\}$$

where in (2.2.8) k_0 has in effect been replaced by

$$k_0(x) = k_0 \cdot \frac{[2b + |x|^2]^{-(a+p/2)}}{E[(2b + \chi^2)^{-(a+p/2)}]}.$$

For large values of $|x|$ at least, $k_0(x)$ is close to k_0 , so that C_x^B is similar to C_x^D . The procedure C^B is the Bayes rule for the loss function

$$(2.2.9) \quad L(\theta, C_x) = k_0(x) \text{ vol}(C_x) - I_{C_x}(\theta)$$

for the same reason that C^D is the Bayes rule for the loss function (2.2.5).

One good property that all three Bayes rules, C^S , C^D , and C^B have in common is that they all are weakly admissible, when the prior $\pi(\theta)$ is proper. This is shown by the following theorem, which is true in greater generality than the present context.

Theorem 2.7. If $f(x)$, the marginal density of X , is positive for all x , if $\pi(\theta)$ is proper, and if $k(x)$ is any Borel measurable function of x , then the Bayes confidence procedure

$$C = \{(x, \theta) : \pi(\theta|x) \geq k(x)\}$$

is weakly admissible.

Proof. Let C' be a procedure which is strongly preferable to C . By Definition 2.1

$$(2.2.10) \quad \text{vol}(C'_x) \leq \text{vol}(C_x) \quad \text{for almost all } x$$

and

$$(2.2.11) \quad P_\theta[C'_\theta] \geq P_\theta[C_\theta] \quad \text{for all } \theta.$$

Take the expectation of both sides of the second inequality with respect to $\pi(\theta)$ to obtain

$$E_{\pi}[P_{\theta}[C'_{\theta}]] \geq E_{\pi}[P_{\theta}(C_{\theta})] \quad .$$

By Fubini's theorem this may be expressed in terms of $f(x)$ as

$$(2.2.12) \quad \int_{\mathbb{R}^p} \int_{C'_x} \pi(\theta|x) d\theta f(x) dx \geq \int_{\mathbb{R}^p} \int_{C_x} \pi(\theta|x) d\theta f(x) dx \quad .$$

On the other hand by the definition of C_x , and the Neyman-Pearson lemma it follows from (2.2.10) that for almost all x

$$(2.2.13) \quad \int_{C'_x} \pi(\theta|x) d\theta \leq \int_{C_x} \pi(\theta|x) d\theta \quad .$$

By (2.2.12) and the assumed positiveness of $f(x)$, equality must hold in (2.2.13) for almost all x . Since (2.2.10) also holds for almost all x , both (2.2.10) and (2.2.13) hold for almost all x . For such x , it follows from the definition of C_x that $C'_x \Delta C_x$ is a Borel set of Lebesgue measure zero, since C_x maximizes the conditional probability of coverage among sets with volume less than or equal to that of C_x . Consequently, for almost all x , C'_x differs from C_x by a Borel set of measure zero, and hence C' differs from C by a Borel set of Lebesgue measure zero in $X \times \Theta$. That is to say, C' and C belong to the same equivalence class of procedures. Since admissibility is defined only up to equivalence classes, C is weakly admissible.

Discussion. For the present normal framework the marginal density $f(x)$ of x is strictly positive for any prior distribution on θ .

The Borel measurability of $k(x)$ in the above argument is not part of the proof but rather is included to ensure that the Bayes set C

is a Borel set. In the case of C^D , the function $k(x)$ is a constant, which of course is measurable in x . In the case of C^B , $k(x)$ is a continuous function of x . For the subjective rule C^S , $k(x)$ is defined implicitly, but it is in fact continuous for smooth priors $\pi(\theta)$, such as the multivariate Student's t priors of the present context.

For these priors the posterior density $\pi(\theta|x)$ is continuous in θ . This implies that C_x is the closure of an open set $(\{(x, \theta): \pi(\theta|x) > k(x)\})$ and so, at the worst, is homeomorphic to a countable union of balls, or is empty. In the next section we shall see that for a wide range of the prior parameters a and b the Bayes confidence sets C_x^S , C_x^D , and C_x^B will each be convex sets for all x .

3. Convexity of the Bayes Sets. As mentioned in Example 1 of Chapter 1, there are priors for θ which yield minimax Bayes point estimates of θ but which may lead to disconnected Bayes credible regions. The following theorem gives conditions under which this cannot happen.

Theorem 2.8. If θ given t has conditional normal prior density

$$\theta|t \sim N(0, \frac{1}{t}I)$$

and the possibly improper prior density of t , $g(t)$, $0 \leq t < \infty$, is such that, for all $\theta \in \mathbb{R}^p$, $E(t|\theta)$ and $\text{Var}(t|\theta)$ are finite and satisfy

$$(2.3.1) \quad E(t|\theta) + 1 - |\theta|^2 \text{Var}(t|\theta) > 0,$$

then 1) $\pi(\theta|x)$ is unimodal for all x

and 2) $C_x = \{\theta: \pi(\theta|x) \geq k\}$ is convex for all x and all k .

Proof. Assume that (2.3.1) is true. Let x be arbitrary but fixed and θ_1 and θ_2 be any two distinct points in $\Theta = \mathbb{R}^p$. For each $s \in \mathbb{R}$ let $\theta^*(s)$ be the point on the line joining θ_1 and θ_2 defined by

$$\theta^*(s) = s\theta_1 + (1-s)\theta_2 \quad .$$

We will show that the posterior density evaluated at $\theta^*(s)$, $h(s) = \pi(\theta^*(s)|x)$ has no local minimum in s . This in turn implies the conclusions of the theorem since if either conclusion is false, then θ_1 and θ_2 can be chosen so that there is a local minimum on the line connecting them. If conclusion 1) is false let θ_1 and θ_2 be any distinct local maxima. If conclusion 2) is false let θ_1 and θ_2 be any two points in C_x for which some point between them is not in C_x .

If $h(s)$ has a local minimum at s_0 , then at s_0 both $h'(s_0) = 0$ and $h''(s_0) \geq 0$. The existence and finiteness of these two derivatives is implied by the finiteness of $E(\underline{t}|\theta)$ and $\text{Var}(\underline{t}|\theta)$ in the following expressions.

Denote $\theta^*(s)$ by θ^* and express

$$h(s) = \pi(\theta^*|x) = c(x) \cdot \exp\left(-\frac{1}{2}|x-\theta^*|^2\right) \cdot \int_0^\infty t^{\frac{p}{2}-1} e^{-\frac{1}{2}t|\theta^*|^2} g(t) dt \quad .$$

Differentiate this once to obtain

$$h'(s) = h(s) \cdot \left[\frac{d}{ds} \left(-\frac{1}{2}|x-\theta^*|^2 \right) + E(\underline{t}|\theta^*) \cdot \frac{d}{ds} \left(-\frac{1}{2}|\theta^*|^2 \right) \right] \quad .$$

Differentiate again to obtain

$$\begin{aligned}
(2.3.2) \quad h''(s) &= h'(s) \cdot \left[\frac{d}{ds} \left(-\frac{1}{2} |x - \theta^*|^2 \right) + E(\underline{t} | \theta^*) \cdot \frac{d}{ds} \left(-\frac{1}{2} |\theta^*|^2 \right) \right] \\
&\quad + h(s) \cdot \left[\frac{d^2}{ds^2} \left(-\frac{1}{2} |x - \theta^*|^2 \right) + E(\underline{t} | \theta^*) \cdot \frac{d^2}{ds^2} \left(-\frac{1}{2} |\theta^*|^2 \right) \right. \\
&\quad \left. + \frac{d}{ds} \left(-\frac{1}{2} |\theta^*|^2 \right) \cdot \frac{d}{ds} E(\underline{t} | \theta^*) \right] .
\end{aligned}$$

For any fixed vector $v \in \mathbb{R}^p$

$$\frac{d}{ds} \left(-\frac{1}{2} |v - \theta^*|^2 \right) = (\theta_1 - \theta_2)' (v - \theta^*)$$

and

$$\frac{d^2}{ds^2} \left(-\frac{1}{2} |v - \theta^*|^2 \right) = -|\theta_1 - \theta_2|^2 .$$

Also

$$\begin{aligned}
\frac{d}{ds} E(\underline{t} | \theta^*) &= \frac{d}{d \left(-\frac{1}{2} |\theta^*|^2 \right)} E(\underline{t} | \theta^*) \cdot \frac{d}{ds} \left(-\frac{1}{2} |\theta^*|^2 \right) \\
&= \text{Var}(\underline{t} | \theta^*) \cdot \frac{d}{ds} \left(-\frac{1}{2} |\theta^*|^2 \right) .
\end{aligned}$$

Using these relations, if $h'(s) = 0$ then (2.3.2) simplifies to

$$\begin{aligned}
h''(s) &= h(s) \cdot \left[-|\theta_1 - \theta_2|^2 (1 + E(\underline{t} | \theta^*)) + [\theta^*]' (\theta_1 - \theta_2) \right]^2 \text{Var}(\underline{t} | \theta^*) \\
&\leq h(s) \cdot \left[-|\theta_1 - \theta_2|^2 (1 + E(\underline{t} | \theta^*)) + |\theta^*|^2 |\theta_1 - \theta_2|^2 \text{Var}(\underline{t} | \theta^*) \right] \\
&= -h(s) \cdot |\theta_1 - \theta_2|^2 [E(\underline{t} | \theta^*) + 1 - |\theta^*|^2 \text{Var}(\underline{t} | \theta^*)] \\
&< 0
\end{aligned}$$

under the assumptions of the theorem.

Example 2. Returning to the case where $\underline{t}$ has the prior density

$$g(t) = t^{a-1} \exp(-bt)$$

we have that $\underline{t}$ given $\underline{\theta}$ has density

$$g(\underline{t} | \underline{\theta}) = c \cdot t^{a+p/2-1} \exp \left[- \left(b + \frac{|\underline{\theta}|^2}{2} \right) t \right] ,$$

that is to say

$$t|\theta \sim \text{Gamma}(a + \frac{p}{2}, b + \frac{|\theta|^2}{2})$$

This is a proper density if $a > -p/2$ and $b > 0$. Therefore

$$E(t|\theta) = \frac{a + p/2}{b + |\theta|^2/2}$$

and

$$\text{Var}(t|\theta) = \frac{a + p/2}{(b + |\theta|^2/2)^2}$$

so that the left hand side of relation (2.3.1) equals

$$\begin{aligned} & \frac{a + p/2}{b + |\theta|^2/2} + 1 - |\theta|^2 \frac{a + p/2}{(b + |\theta|^2/2)^2} \\ &= \frac{1}{(b + |\theta|^2/2)^2} [((a + \frac{p}{2})b + b^2) + (b - \frac{a + p/2}{2})|\theta|^2 + \frac{|\theta|^4}{4}] \end{aligned}$$

This remains positive for all θ if and only if the term in brackets is always positive. This obtains its minimum value in $|\theta|^2$ when it equals

$$\begin{aligned} & [((a + \frac{p}{2})b + b^2) - (b - \frac{a + p/2}{2})^2] \\ &= [(a + \frac{p}{2})b + (a + \frac{p}{2})b - \frac{1}{4}(a + \frac{p}{2})^2] \end{aligned}$$

which is positive if $a > -\frac{p}{2}$ and $b > \frac{1}{8}(a + \frac{p}{2})$.

If the prior density is of this form then C_x^S , C_x^D , and C_x^B will all be convex sets, since the theorem does not depend on the choice of k . In the remaining sections we will examine more closely the procedure C^B .

In Chapter 1, Example 2, it was shown that this prior yields a minimax point estimator of θ if $a \leq b+3$ and $a \leq \frac{p}{2}-2$. The conditions for convexity, $a > -\frac{p}{2}$ and $a < 8b - \frac{p}{2}$, do not conflict with

these. For example, in three dimensions if $a = -1$ and $b > \frac{1}{16}$ then both conditions are satisfied and in five dimensions if $a \leq \frac{1}{2}$ and $b > \frac{a}{8} + \frac{5}{16}$ then both conditions are met.

Theorem 2.8 is of interest not only because it is desirable that confidence sets be convex, but also because such regions are of interest as a way to construct simultaneous confidence intervals for various linear combinations of the parameters θ_1 , in the same way that Scheffé intervals are based on the usual confidence sets for θ . If the Bayes sets were not convex then some of these intervals would be disconnected.

We will not attempt to pursue this topic further at this time, however.

4. The Geometry of C^B . The Bayes procedure C^B is defined as

$$(2.4.1) \quad C^B = \{(x, \theta): [\frac{\exp(r_o^2)}{\exp(|x-\theta|^2)}] \frac{1}{2a+p} \geq \frac{2b + |\theta|^2}{2b + |x|^2}\}$$

where r_o^2 is the radius of the usual $1-\alpha$ size confidence sets and a and b are arbitrary parameters, subject to the condition that C_x^B always be convex, which is the case when $a > -\frac{p}{2}$ and $b > (a + \frac{p}{2})/8$ (see Section 3). Later, in Section 5, we will impose a somewhat stronger condition to guarantee that C_θ^B is convex for all θ .

It is clear from the above definition that there is a duality between the geometries of C_x^B and C_θ^B .

We begin by describing the confidence sets C_x^B .

1. For each $x \in \mathbb{R}^p$, x is in C_x^B . Replace θ by x in (2.4.1) and observe that the value $\theta = x$ satisfies the inequality.

2. For $x = 0$, C_x^B is a ball about the origin of radius less than r_o . Substitute $x = 0$ in (2.4.1). The inequality holds if and

only if $|\theta|$ is less than some value between 0 and r_0 . This demonstrates that at $x = 0$, $C_x^B \subset C_x^O$. See Figure 1.

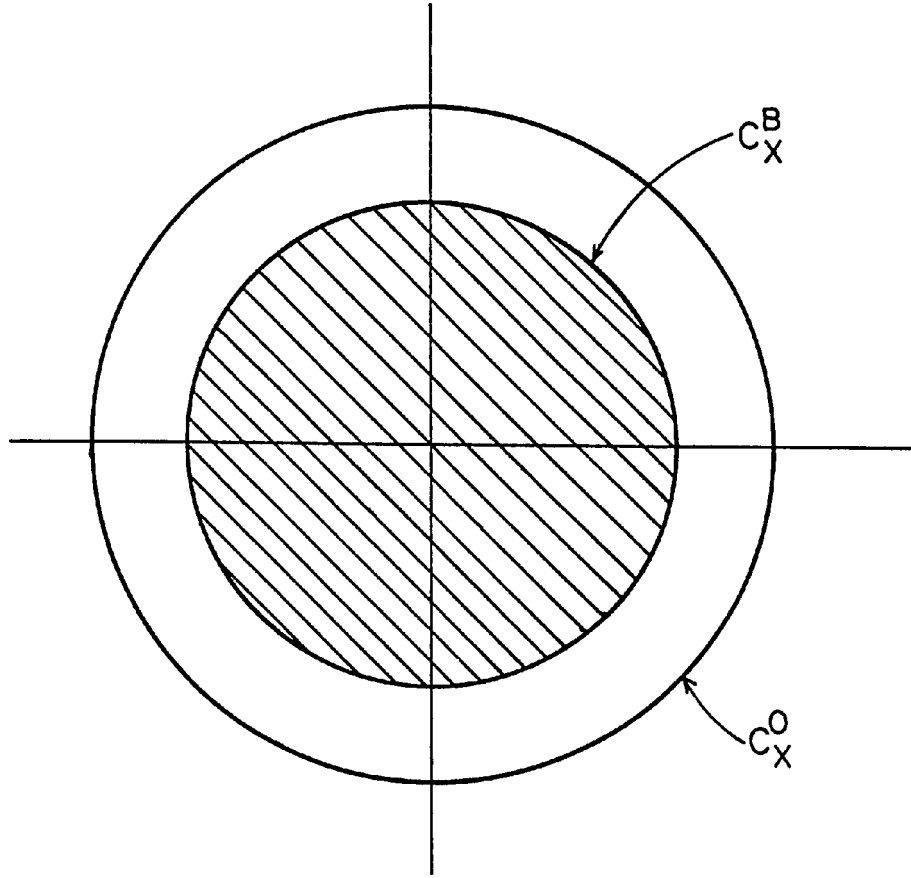


Figure 1. The Bayes confidence set C_x^B and the usual confidence set C_x^O for the case $x = 0$.

3. For fixed $x \neq 0$ let the x -axis be the line through x and the origin. By (2.4.1) the inclusion of θ in C_x^B depends only on the values of $|\theta|$ and $|x-\theta|$. The set of points in $\mathbb{R}^p$ for which these are held fixed is either an empty set (if $|\theta| < (|x| - |x-\theta|)$ or $|\theta| > (|x| + |x-\theta|)$) or else as the intersection of two $p-1$ spheres and so is a $p-2$ sphere centered at some point x^* on the x -axis. For example if $p = 3$ it is a one sphere (a circle) about the x -axis.

Since C_x^B is defined only in terms $|\theta|$ and $|x-\theta|$, it is the solid of rotation formed by rotating a set of points in $\mathbb{R}^2$ about the x -axis, in the same way that a ball is formed by rotating a disc about

an axis through its center. If $\theta \in C_x^B$ then so is its image under such rotations.

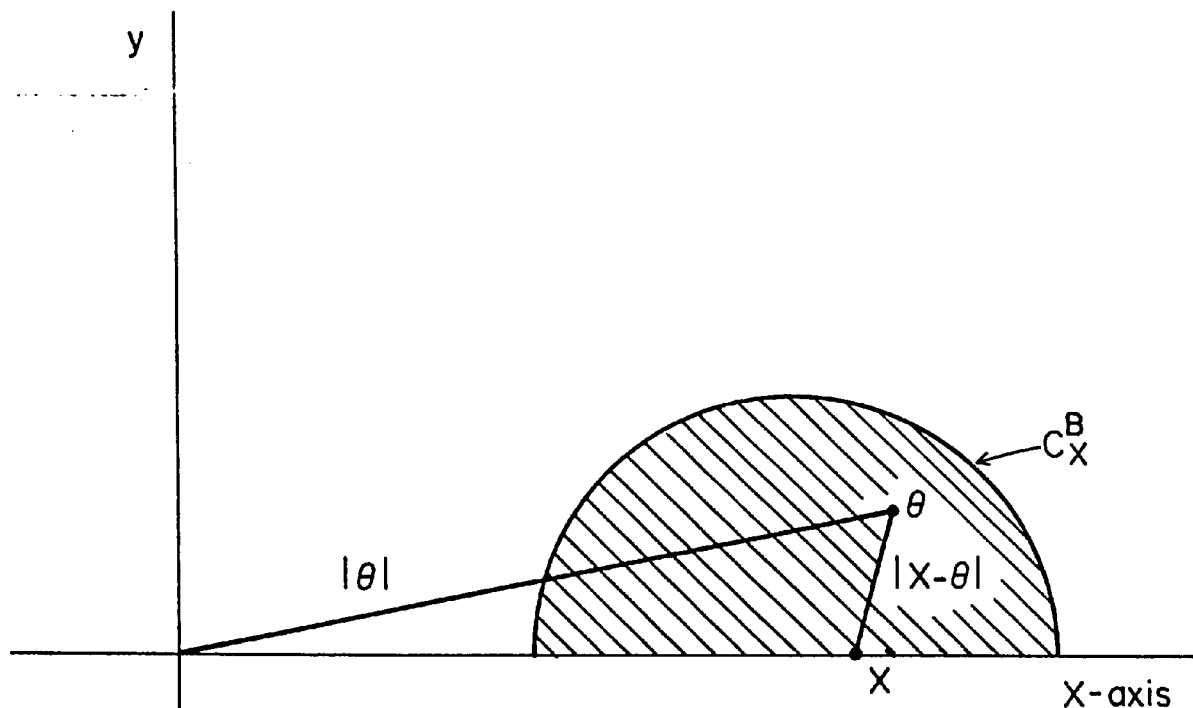


Figure 2. Two dimensional representation of the Bayes confidence set C_x^B when $x \neq 0$. The x-axis is the line connecting x and the origin. The y-axis denotes that component of θ perpendicular to x .

The set C_x^B can be represented as in Figure 2 by making the x-axis the horizontal one, and the vertical axis y represent that component of θ which is perpendicular to the x-axis. Alternatively y could be regarded as an arbitrary but fixed direction perpendicular to the x-axis and C_x^B formed by rotating the set described in Figure 2 about the x-axis.

4. Unfortunately it is not easy to describe C_x^B analytically in terms of the relation between the height of the outline in Figure 2 and the distance along the x-axis. On the other hand C_x^B does lend itself to description in terms of the distances $|\theta|$ and $|x - \theta|$. We look at

the intersection of C_x^B with various spheres centered at x .

It follows from (2.4.1) that the intersection of C_x^B with the sphere

$$S_{x,r} = \{\theta: |x-\theta| = r\}$$

is the set

$$(2.4.2) \quad S_{x,r} \cap \{\theta: |\theta|^2 \leq q(r)\}$$

where

$$(2.4.3) \quad q(r) = (2b + |x|^2) \exp\left(\frac{r_0^2 - r^2}{2a+p}\right) - 2b.$$

This set is either

- a) All of $S_{x,r}$ if $q(r) \geq (|x|+r)^2$ (Figure 3a);
- b) A spherical segment on the side of $S_{x,r}$ nearest the origin if

$$(|x|-r)^2 \leq q(r) \leq (|x|+r)^2 \quad (\text{Figure 3b});$$

or c) The empty set if

$$q(r) < (|x|-r)^2. \quad (\text{Figure 3c})$$

Since $q(r)$ is decreasing in r case a holds for all r less than r_2 , the positive root to the equation

$$q(r) = (|x|+r)^2.$$

The value r_2 is the minimum distance between x and the surface of C_x^B . The minimum distance is attained at the point $\theta_2 = (1 + \frac{r_2}{|x|})x$ on the x -axis.

As r increases beyond r_2 , case b holds. Ultimately, $S_{x,r}$ no longer intersects $\{\theta: |\theta|^2 \leq q(r)\}$. For r large enough in fact,

Figure 3. The intersection of the Bayes confidence set C_x^B with spheres $S_{x,r}$ centered at x of radius r for values of r ranging from r_2 to greater than r_1 .

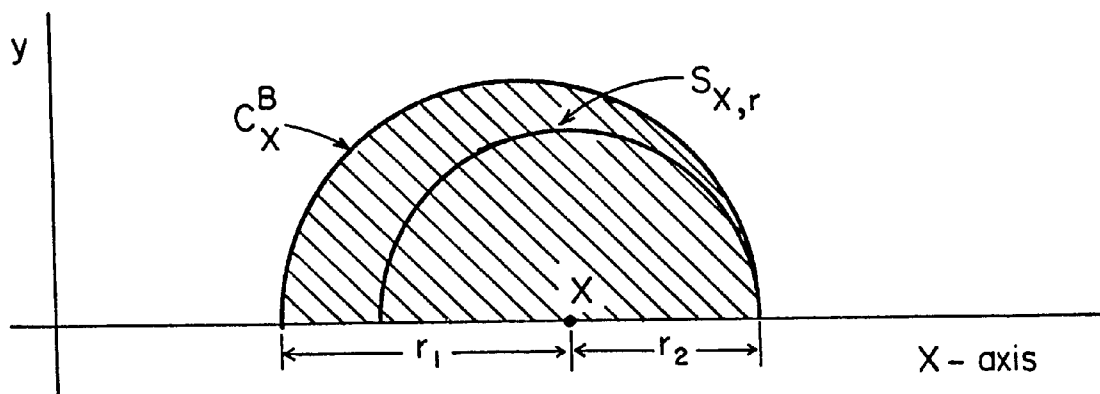


Figure 3a. Case a: $r = r_2$

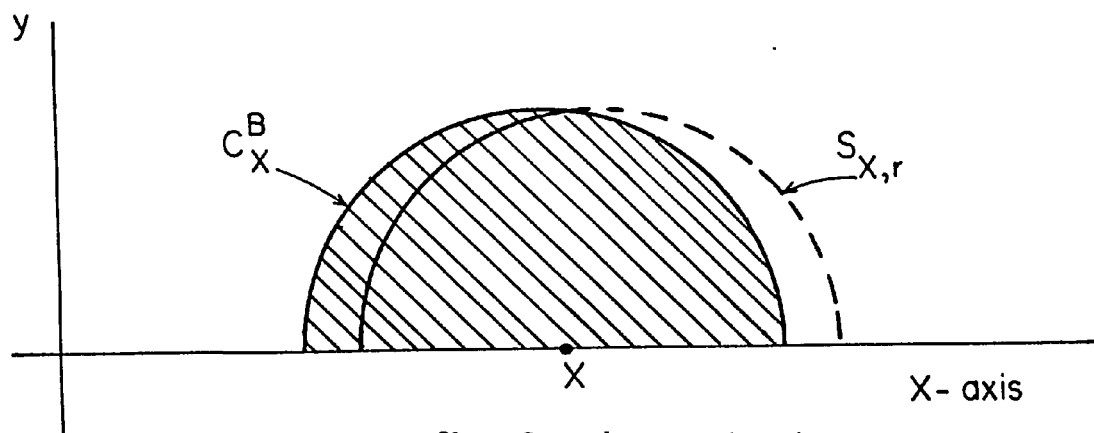


Figure 3b. Case b: $r_2 < r \leq r_1$

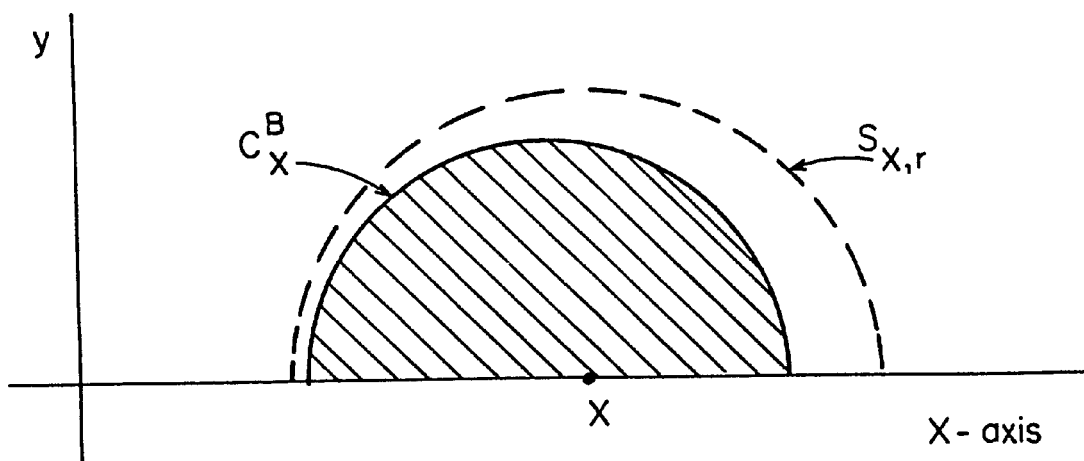


Figure 3c. Case c: $r > r_1$

$q(r)$ is negative and the latter set empty. The intersection can vanish only once, since if it vanished and subsequently reappeared the set C_x^B would not be convex, contrary to assumption. Denote by r_1 the value of r for which the intersection consists of a single point (which necessarily lies on the x -axis). This is the positive root to the equation

$$q(r_1) = (|x| - r_1)^2 = 0.$$

It is the maximum distance between x and the surface of C_x^B , and is attained at the point $\theta_1 = (1 - \frac{r_1}{|x|})x$. For r larger than r_1 case c holds, and the intersection is empty.

5. The above discussion gives bounds on C_x^B . Let $B_{x,r}$ be the ball

$$B_{x,r} = \{\theta : |x - \theta| \leq r\}.$$

Then for all x

$$B_{x,r_2} \subseteq C_x^B \subseteq B_{x,r_1}.$$

6. For $|x| \leq r_0/2$, $C_x^B \subseteq C_x^0$. Evaluate

$$\begin{aligned} q(r_0) &= (|x|^2 + 2b) - 2b = |x|^2 \\ &\leq (|x| - r_0)^2. \end{aligned}$$

This implies that $r_1 \leq r_0$ so that

$$C_x^B \subseteq B_{x,r_1} \subseteq B_{x,r_0} = C_x^0.$$

On the other hand, for $|x| > r_0/2$, $r_1 > r_0$ and C_x^B does not contain C_x^0 .

7. For all x

$$q(r_0) = |x|^2 < (|x| + r_0)^2$$

so that $r_2 < r_0$ for all x .

8. In the limit as $|x|$ tends to ∞ both r_1 and r_2 tend to r_0 . To prove this, observe that by (2.4.3) for any $\epsilon > 0$, $\epsilon < r_0$ and $|x|$ sufficiently large,

$$q(r_0 + \epsilon) < (|x| - (r_0 + \epsilon))^2.$$

Therefore $S_{x, r_0 + \epsilon}$ does not intersect C_x^B for large x . On the other hand for large $|x|$

$$q(r_0 - \epsilon) > (|x| + (r_0 - \epsilon))^2$$

so that $S_{x, r_0 - \epsilon}$ is contained in C_x^B .

Therefore $r_0 - \epsilon < r_2 < r_0 < r_1 < r_0 + \epsilon$ for sufficiently large $|x|$. Both C_x^B and C_x^O are bounded between B_{x, r_2} and B_{x, r_1} so the foregoing shows that the difference between C_x^B and C_x^O

$$C_x^B \Delta C_x^O$$

tends to zero in Lebesgue measure as $|x|$ tends to infinity.

We list the salient points of the preceding discussion.

- i) The exact shape of C_x^B is difficult to specify analytically.
- ii) The endpoints of C_x^B along the x axis are at $(1 - \frac{r_1}{|x|})x$ and $(1 + \frac{r_2}{|x|})x$ where r_1 and r_2 are respectively the absolute values of the negative and positive roots to the equation

$$\exp\left(\frac{r_0^2 - r^2}{2a + p}\right) = \frac{2b + (|x| + r)^2}{2b + |x|^2}.$$

- iii) For all $|x|$

$$B_{x, r_2} \subseteq C_x^B \subseteq B_{x, r_1}.$$

iv) For $|x| \leq r_0/2$, $C_x^B \subseteq C_x^O$.

v) For $r_0/2 < |x| < \infty$

$$B_{x,r_2} \subseteq C_x^O \subseteq B_{x,r_1}.$$

vi) As $|x| \rightarrow \infty$, C_x^B tends to C_x^O .

Now we consider the θ cross sections, C_θ^B , $\theta \in \Theta$. These are of the form

$$(2.4.4) \quad C_\theta^B = \{x: e^{-\frac{|x-\theta|^2}{2a+p}} (2b+|x|^2) \geq e^{-\frac{r_0^2}{2a+p}} (2b+|\theta|^2)\}.$$

Note that for the usual procedure the θ section is simply the ball $C_\theta^O = B_{\theta,r_0}$.

If $b \geq a+p/2 > 0$ then C_θ^B is convex for all θ and r_0 . This fact may be obtained as follows: Fix θ in (2.4.4). Let x_1 and x_2 be any pair of points in C_θ^B . For $0 \leq s \leq 1$ let $x^*(s) = sx_1 + (1-s)x_2$. Define the function

$$h(s) = e^{-\frac{|x^*-\theta|^2}{2a+p}} (2b + |x^*|^2).$$

Then $x^*(s)$ is in C_θ^B for $0 \leq s \leq 1$ if $h(s)$ has no local minimum. Take derivatives as in Section 3 to prove that this is the case regardless of θ when $b \geq a+p/2 > 0$. When $b < a+p/2$ on the other hand and $|\theta|$ is small it is possible for a local minimum to occur in this function so that for some values of r_0 the set C_θ^B may have a hole in it. The coverage probability computations of Section 5 are based on the convexity assumption. We will use the above rather strong condition to ensure that the computations are valid in general. By choosing particular values for r_0 , the condition can be weakened, however.

Without going into details the following results about C_θ^B are obtained in the same way their counterparts were obtained for C_x^B .

By the convexity of C_θ^B , the intersection of C_θ^B with the θ -axis will be an interval. This interval contains θ itself. Its endpoints are $\theta(1 - \frac{r_1}{|\theta|})$ and $\theta(1 + \frac{r_2}{|\theta|})$ where r_1 and r_2 are the absolute values of the negative and positive roots to the equation

$$(2.4.5) \quad e^{-\frac{r^2}{2a+p}}(2b + (|\theta|+r)^2) = e^{-\frac{r_o^2}{2a+p}}(2b + |\theta|^2) \quad .$$

Although the same symbols r_1 and r_2 are being used to denote the endpoints of the θ sections as are used to denote the endpoints of the x sections, it should be observed that they are not the same thing and are related to each other only indirectly. Since we will not deal simultaneously with both C_x^B and C_θ^B there should be no confusion.

In contrast to the case of C_x^B , for fixed θ and $r = |x-\theta|$ the intersection of C_θ^B with the sphere $S_{\theta,r}$ is that portion of the sphere which is most distant from the origin. Therefore C_θ^B contains the sphere S_{θ,r_1} (as well as its interior, by convexity) and is contained in the sphere S_{θ,r_2} . It follows from the definition of r_1 and r_2 that $0 < r_1 \leq r_2$ with equality when $|\theta| = 0$. The set C_θ^B is pushed away from the origin in much the same way that C_x^B is pulled towards it. It may be verified that for $0 < |\theta| < r_o/2$ both r_1 and r_2 are greater than r_o whereas for $|\theta| \geq r_o/2$, $r_1 \leq r_o$ and $r_2 > r_o$. Also as $|\theta|$ tends to infinity both r_1 and r_2 tend to r_o .

If C_θ^B is convex, in particular if $b \geq a+p/2 > 0$ then

- i) The exact shape of C_θ^B is difficult to specify analytically.
- ii) The endpoints of C_θ^B along the θ -axis are $\theta(1 - \frac{r_1}{|\theta|})$ and

$\theta(1 + \frac{r_2}{|\theta|})$ where r_1 and r_2 are defined by (2.4.5).

iii) For all $|\theta|$, $B_{\theta, r_1} \subseteq C_{\theta}^B \subseteq B_{\theta, r_2}$.

iv) For $|\theta| \leq r_0/2$, $C_{\theta}^0 \subseteq C_{\theta}^B$.

v) For $r_0/2 < |\theta| < \infty$

$$B_{\theta, r_1} \subseteq C_{\theta}^0 \subseteq B_{\theta, r_2}.$$

vi) As $|\theta| \rightarrow \infty$ C_{θ}^B tends to C_{θ}^0 .

5. The Volume and Coverage Probabilities.

In this section we derive expressions for the volume $\text{vol}(C_x^B)$ and the coverage probability $P_{\theta}[C_{\theta}^B]$ in the cases of $p = 3$ and $p = 5$ dimensions.

Theorem 2.9. If $p = 3$ and $b > \frac{1}{8}(a + \frac{3}{2}) > 0$ then

$$(2.5.1) \quad \text{vol}(C_x^B) = \frac{2}{3}\pi[r_1^3 + r_2^3] + 2(a + \frac{3}{2})\pi[r_1 + r_2] \\ - \frac{\pi}{|x|}[(a + \frac{3}{2} + b + \frac{|x|^2}{2})(r_1^2 - r_2^2) + \frac{1}{4}(r_1^4 - r_2^4)]$$

if $x \neq 0$ and

$$\text{vol}(C_x^B) = \frac{4}{3}\pi r_2^3$$

if $x = 0$, where r_1 and r_2 are respectively the absolute values of the negative and positive roots to

$$(2.5.2) \quad \frac{r_0^2 - r^2}{e^{2a+p}} = \frac{2b + (|x| + r)^2}{2b + |x|^2}.$$

Proof. When $x = 0$ the result is immediate since, as was shown in Section 4. C_x^B is the ball B_{x, r_2} (and $r_1 = r_2$).

Now suppose x is fixed and $x \neq 0$. As was indicated in the preceding section, $B_{x,r_2} \subset C_x^B \subset B_{x,r_1}$, and for $r_2 \leq r \leq r_1$ the intersection of C_x^B with $S_{x,r}$ is a segment of $S_{x,r}$ on which the distance to the origin is less than a specified value. See Figure 4. The volume of C_x^B is the sum of the volume of B_{x,r_2} and the integral of the surface area of this spherical segment between r_2 and r_1 , integrating with respect to r .

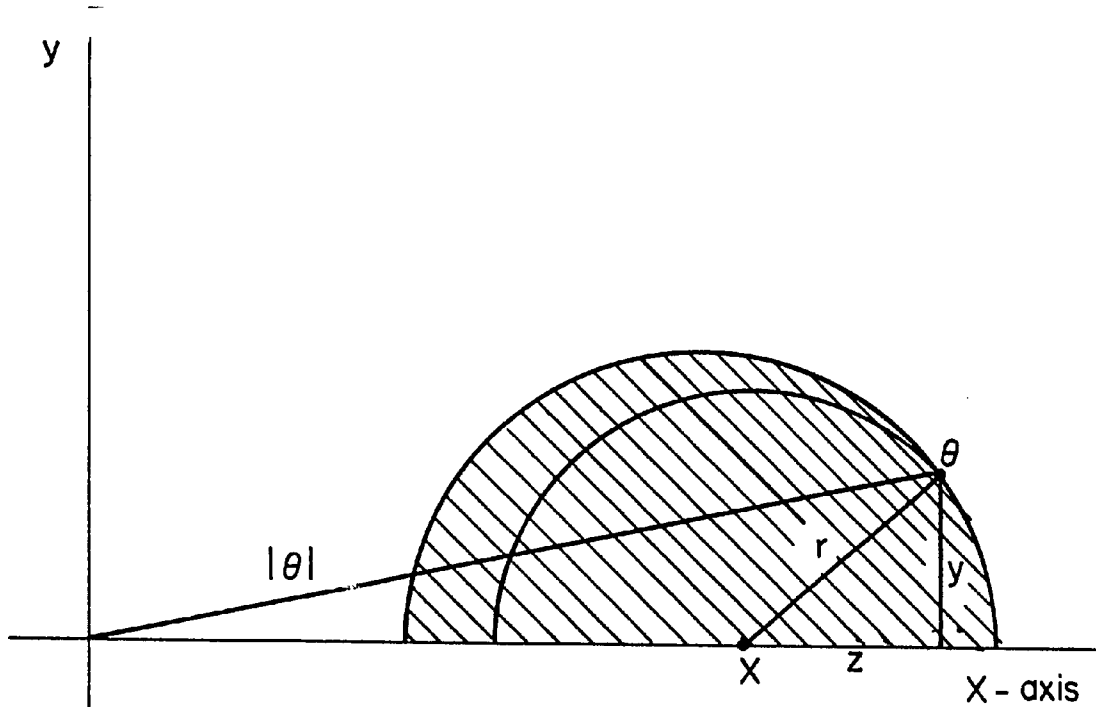


Figure 4. The point θ at which the sphere about x of radius r intersects the surface of the Bayes confidence set C_x^B , when $r_2 \leq r \leq r_1$.

Each $\theta \in \mathbb{R}^p$ may be expressed uniquely as

$$\theta = (1 + \frac{z}{|x|})x + y$$

where $y'x = 0$ and $z \in \mathbb{R}$.

On the sphere $S_{x,r}$

$$(2.5.3) \quad z^2 + |y|^2 = r^2 \quad .$$

By the definition of C^B (2.4.1) $S_{x,r}$ intersects the surface of C_x^B at the set of points for which

$$(2.5.4) \quad 2b + |\theta|^2 = (2b + |x|^2) e^{\frac{r_o^2 - r^2}{2a+p}}.$$

Since $|\theta|^2 = (|x|+z)^2 + |y|^2 = |x|^2 + 2|x|z + r^2$, z is determined uniquely as a function of r by

$$(2.5.5) \quad z(r) = \frac{(2b + |x|^2) [e^{\frac{r_o^2 - r^2}{2a+p}} - 1] - r^2}{2|x|}.$$

In three dimensions the surface area of the segment of $S_{x,r}$ which is bounded on the right by the perpendicular plane z units to the right of x is

$$2\pi r(r+z).$$

The volume of S_{x,r_2} is $\frac{4}{3}\pi r_2^3$, so that

$$\begin{aligned} \text{vol}(C_x^B) &= \frac{4}{3}\pi r_2^3 + \int_{r_2}^{r_1} 2\pi r(r+z) dr \\ &= \frac{2}{3}\pi(r_1^3 + r_2^3) + \int_{r_2}^{r_1} rz dr \end{aligned}$$

which gives the result after integration and substitution using (2.5.2) and $p = 3$.

Corollary 2.10. Under the conditions of Theorem 2.9, if $p = 5$ then

$$(2.5.6) \quad \text{vol}(C_x^B) = \frac{4}{15}\pi^2[r_1^5 + r_2^5] + 2\pi^2 \int_{r_2}^{r_1} [r^3 z - \frac{1}{3}rz^3] dr$$

if $x \neq 0$ and

$$\text{vol}(C_x^B) = \frac{8}{15}\pi^2 r_2^2$$

if $x = 0$, where z is given by (2.5.5) and r_1 and r_2 by (2.5.2).

Proof. The computation is identical to that of the preceding theorem, except that in five dimensions the surface area of the segment of $S_{x,r}$ which is bounded on the right by the perpendicular hyperplane z units to the right of x is

$$2\pi^2 r \left[\frac{2}{3}r^3 + r^2 z - \frac{1}{3}z^3 \right]$$

and the volume of S_{x,r_2} is $\frac{8}{15}\pi^2 r_2^5$. Equation (2.5.6) follows directly from these two formulas.

Discussion. The equation for four dimensions has been left out because the area formula in four dimensions includes an $\arcsine(\frac{z}{r})$ term and therefore would be difficult to integrate. On the other hand in the five dimensions the integral is computable in closed form. By (2.5.5)

$$z = \frac{1}{2|x|} [u - r^2 + ve^{wr^2}]$$

where $u = (2b + |x|^2)$, $v = u \cdot e^{\frac{r_0^2}{2a+p}}$ and $w = \frac{-1}{2a+p}$. It follows that $r \cdot z^3$ and $r^3 z$ in (2.5.6) are both explicitly integrable. Each term of these is either a constant times a power of r or is a constant times $r^n \cdot e^{cr^2}$ where n is an odd integer and $c \neq 0$.

For n odd, repeated integration by parts gives

$$\int r^n e^{cr^2} dr = e^{r^2 c} \left(\frac{r^{n-1}}{2c} - \frac{n-1}{(2c)^2} r^{n-3} + \frac{(n-1)(n-3)}{(2c)^3} r^{n-5} + \dots \right. \\ \left. + (-1)^{\frac{n+1}{2}} \frac{(n-1)(n-3) \cdots (1)}{(2c)^{\frac{n+1}{2}}} \right)$$

so the integral is computable. For details see the Appendix.

Numerically computed values of $\text{vol}(C_x^B)$ in three and five dimensions are given in Section 7.

The coverage probability for the Bayes procedure C^B may be evaluated also.

Theorem 2.11. If $p = 3$, $b > a + \frac{3}{2} > 0$ and $a \neq -\frac{1}{2}$ then

$$(2.5.7) \quad P_\theta[C_\theta^B] = \frac{1}{2} [P[\chi_3^2 \leq r_1^2] + P[\chi_3^2 \leq r_2^2]] \\ + \frac{1}{|\theta|(8\pi)^{1/2}} \cdot [e^{-\frac{r^2}{2}} [2 + r^2 + (|\theta|^2 + 2b) \\ \cdot (1 - \frac{2a+3}{2a+1} e^{\frac{r^2 - r_0^2}{2a+3}})]]^{r_2}_{r_1}$$

if $\theta \neq 0$ and

$$P_\theta[C_\theta^B] = P[\chi_3^2 \leq r_2^2]$$

if $\theta = 0$, where r_1 and r_2 are respectively the absolute values of the negative and positive roots to the equation

$$(2.5.8) \quad e^{\frac{r_0^2 - r^2}{2a+p}} = \frac{2b + |\theta|^2}{2b + (|\theta| + r)^2}.$$

If $a = -\frac{1}{2}$ and $\theta \neq 0$ then the second term of the coverage probability

is

$$(2.5.9) \quad \frac{1}{|\theta| (8\pi)^{1/2}} \cdot [e^{-\frac{r^2}{2}} [2 + r^2 + (|\theta|^2 + 2b)] + (|\theta|^2 + 2b) e^{-\frac{r_0^2}{2}} \cdot \frac{r^2}{2}]^{\frac{r_2}{r_1}}.$$

Proof. When $\theta = 0$ the result is immediate since, as was shown in Section 4. C_θ^B is the ball B_{x, r_1} (at $\theta = 0$, $r_1 = r_2$).

Now suppose θ is fixed and $\theta \neq 0$. By the discussion in Section 4

$$B_{\theta, r_1} \subseteq C_\theta^B \subseteq B_{\theta, r_2}.$$

Conditional on θ and the value r of $\underline{r} = |X - \theta|$ the distribution of X is uniform on the sphere $S_{\theta, r}$. Therefore the conditional coverage probability $P[X \in C_\theta^B | \theta, r]$ is the ratio of the surface area of $S_{\theta, r} \cap C_\theta^B$ to the total surface area of $S_{\theta, r}$. This ratio is 1 for $r \leq r_1$ and 0 for $r > r_2$. For $r_1 < r \leq r_2$ proceed as in Theorem 2.9.

Each $x \in X$ may be expressed uniquely as

$$x = (1 + \frac{z}{|\theta|})\theta + y$$

where $y' \theta = 0$. On the set $S_{\theta, r}$

$$(2.5.10) \quad z^2 + |y|^2 = r^2.$$

This intersects the surface of C_θ^B at the set of points for which

$$(2.5.11) \quad 2b + |x|^2 = e^{\frac{r^2 - r_0^2}{2a + p}} (2b + |\theta|^2).$$

For all x ,

$$|x|^2 = (|\theta| + z)^2 + |y|^2 = |\theta|^2 + 2|\theta|z + r^2$$

so for the points of intersection z is determined by the relation

$$(2.5.12) \quad z(r) = \frac{(2b+|\theta|^2) \left[e^{\frac{r^2-r_0^2}{2a+p}} - 1 \right] - r^2}{2|\theta|}.$$

The surface area of $S_{\theta,r} \cap C_{\theta}^B$ is $2\pi r(r-z)$ since by the discussion in Section 4 this intersection consists of the spherical segment of $S_{\theta,r}$ to the right of the perpendicular plane z units to the right of θ . Therefore

$$P[X \in C_{\theta}^B | \theta, r] = \frac{2\pi r(r-z)}{4\pi r^2} = \frac{1}{2} - \frac{z}{2r}.$$

One component of total coverage probability $P_{\theta}[C_{\theta}^B]$ is the probability given θ that $|X-\theta| < r_1$, which is just $P[\chi_3^2 \leq r_1^2]$, since $|X-\theta|^2 | \theta \sim \chi_3^2$. By change of variable $\underline{s} = \underline{r}^2$, with $\underline{s} | \theta \sim \chi_3^2$ the other component of the coverage probability is the integral between r_1^2 and r_2^2 of the conditional coverage probability with respect to the χ_3^2 density of $\underline{s}$. That is to say

$$(2.5.13) \quad P_{\theta}[C_{\theta}^B] = P[\chi_3^2 \leq r_1^2] + \int_{r_1^2}^{r_2^2} \left(\frac{1}{2} - \frac{z}{2\sqrt{s}} \right) \frac{1}{\Gamma(\frac{3}{2})} \left(\frac{1}{2} \right)^{3/2} s^{1/2} e^{-s/2} ds$$

$$= \frac{1}{2} [P(\chi_3^2 \leq r_1^2) + P(\chi_3^2 \leq r_2^2)] - \frac{1}{(8\pi)^{1/2}} \int_{r_1^2}^{r_2^2} z e^{-s/2} ds$$

where, in terms of s , by (2.5.12)

$$(2.5.14) \quad z(s) = \frac{(2b+|\theta|^2) \left[e^{\frac{s-r_0^2}{2a+p}} - 1 \right] - s}{2|\theta|}.$$

Using the fact that $p = 3$ this integrates to one of the two expressions

in the statement of the theorem, depending on whether a equals $-\frac{1}{2}$.
□

The above proof may be modified to give the coverage probability when $p = 5$.

Corollary 2.12. Under the conditions of Theorem 2.11, if $p = 5$ then

$$(2.5.15) \quad P_{\theta}[C_{\theta}^B] = \frac{1}{2}[P[\chi_5^2 \leq r_1^2] + P[\chi_5^2 \leq r_2^2]] \\ + \frac{1}{\sqrt{32\pi}} \int_{r_1^2}^{r_2^2} \left(\frac{1}{3}z^3 - sz\right) e^{-s/2} ds$$

if $\theta \neq 0$ and

$$P_{\theta}[C_{\theta}^B] = P[\chi_5^2 \leq r_1^2]$$

if $\theta = 0$, where z is given by (2.5.14) and r_1 and r_2 are defined by (2.5.8).

Proof. This computation is done in the same way as that of Theorem 2.11. The surface area of the segment of $S_{\theta,r}$ to the right of the perpendicular plane z units to the right of x is

$$2\pi^2 r \left[\frac{2}{3}r^3 - r^2 z + \frac{1}{3}z^3 \right].$$

The total surface area of $S_{\theta,r}$ is obtained by substituting $z = -r$ in this expression. It equals $\frac{8}{3}\pi^2 r^4$. For $r_1 < r < r_2$ the conditional coverage probability is the ratio of the two, that is

$$P[X \in C_{\theta}^B | \theta, s = r^2] = \frac{1}{2} - \frac{3}{4} \frac{z}{s^{1/2}} + \frac{1}{4} \frac{z^3}{s^{3/2}}.$$

This quantity is integrated with respect to the χ_5^2 density

$\frac{1}{\Gamma(5/2)} \left(\frac{1}{2}\right)^{5/2} s^{3/2} e^{-s/2}$. As in Theorem 2.11, the first term combines with the probability given θ that $|X-\theta| < r_1$ to give $\frac{1}{2}[P[\chi_5^2 \leq r_1^2] + P[\chi_5^2 \leq r_2^2]]$ and what is left is the integral in (2.5.15).

□

Discussion. The integral specified in this corollary is a somewhat complicated function but each term of it is of the form

$$\int s^n e^{cs} ds ,$$

n an integer which is computable by simple integration by parts.

Numerically computed values of $P_\theta[C_\theta^B]$ in three and five dimensions are given in Section 7. For details see the Appendix.

6. Comparison of C^B and C^O .

For $|x| \leq r_0/2$, by the discussion in Section 4 $C_x^B \subseteq C_x^O$, which implies that $\text{vol}(C_x^B) \leq \text{vol}(C_x^O)$ for small $|x|$. On the other hand it was shown in Section 4 that for all x $B_{x,r_2}^B \subseteq C_x^B \subseteq B_{x,r_1}^B$ and that r_1 and r_2 converge to r_0 as $|x|$ tends to infinity, so that for large $|x|$, $\text{vol}(C_x^B)$ tends to $\text{vol}(C_x^O)$. The following theorem shows a class of priors for which $\text{vol}(C_x^B)$ is actually less than $\text{vol}(C_x^O)$ for sufficiently large x .

Theorem 2.13. If $p = 3$, $a = -1$, $b > \frac{1}{8}(a + \frac{p}{2}) = \frac{1}{8}$ and $r_0^2 > 3$ (which is the case if $1 - \alpha > .609$) then for $|x| \leq r_0/2$ and for $|x|$ sufficiently large $\text{vol}(C_x^B) < \text{vol}(C_x^O)$.

Proof. If $|x| \leq r_0/2$ the result is already established. We prove the asymptotic part by showing that for large $|x|$, $\frac{\partial}{\partial |x|} \text{vol}(C_x^B)$

is positive. This will be sufficient for the result because in the limit $\text{vol}(C_x^B)$ tends to $\text{vol}(C_x^O)$. To obtain $\frac{\partial}{\partial |x|} \text{vol}(C_x^B)$ recall from the proof of Theorem 2.9 that

$$\text{vol}(C_x^B) = \frac{4}{3}\pi r_2^3 + \int_{r_2}^{r_1} 2\pi r^2 + 2\pi r z \, dr \quad .$$

(The convexity assumption $b > \frac{1}{8}(a+p)$ was needed to get this result.)

The derivative of this with respect to $|x|$ is

$$\begin{aligned} \frac{\partial}{\partial |x|} \text{vol}(C_x^B) &= 4\pi r_2^2 \frac{\partial}{\partial |x|} r_2 + (2\pi r_1^2 + 2\pi r_1 z(r_1)) \frac{\partial}{\partial |x|} r_1 \\ &\quad - (2\pi r_2^2 + 2\pi r_2 z(r_2)) \frac{\partial}{\partial |x|} r_2 + \int_{r_1}^{r_2} \frac{\partial}{\partial |x|} (2\pi r^2 + 2\pi r z) dr \\ &= 2\pi \int_{r_1}^{r_2} r \frac{\partial}{\partial |x|} z \, dr \end{aligned}$$

since $z(r_1) = -r_1$ and $z(r_2) = r_2$. From (2.5.5)

$$z = \frac{(2b + |x|^2)(e^{\frac{r_0^2 - r^2}{2|x|^2}} - 1) - r^2}{2|x|}$$

so that

$$\frac{\partial z}{\partial |x|} = \frac{(|x|^2 - 2b)(e^{\frac{r_0^2 - r^2}{2|x|^2}} - 1) + r^2}{2|x|^2} \quad .$$

Integrate $2\pi r \frac{\partial}{\partial |x|} z \, dr$ to obtain

$$\begin{aligned} (2.6.1) \quad \frac{\partial}{\partial |x|} \text{vol}(C_x^B) &= \frac{\pi}{2|x|^2} \left[\frac{1}{2} r^4 - (|x|^2 - 2b)(r^2 + e^{\frac{r_0^2 - r^2}{2|x|^2}}) \right]_{r_2}^{r_1} \\ &= \frac{\pi}{2|x|^2} \left[\frac{1}{2} (r_1^4 - r_2^4) - \frac{|x|^2 - 2b}{|x|^2 + 2b} ((r_1^2 - r_2^2)(|x|^2 + 2b + 1) \right. \\ &\quad \left. - 2|x|(r_1 + r_2)) \right] \quad . \end{aligned}$$

The last expression is obtained using (2.5.2) whereby

$$(2.6.2) \quad e^{r_o^2 - r_1^2} = \frac{2b + (|x| - r_1)^2}{2b + |x|^2} = 1 + \frac{r_1^2 - 2|x|r_1}{2b + |x|^2}$$

and

$$(2.6.3) \quad e^{r_o^2 - r_2^2} = \frac{2b + (|x| + r_2)^2}{2b + |x|^2} = 1 + \frac{r_2^2 + 2|x|r_2}{2b + |x|^2} .$$

To show that the derivative is ultimately positive it is sufficient to show that for sufficiently large $|x|$

$$B = \frac{1}{2}(r_1^4 - r_2^4) > (2b + |x|^2 + 1)(r_1^2 - r_2^2) - 2|x|(r_1 + r_2) = A .$$

For large $|x|$ r_1 and r_2 converge to r_o so one can expand the difference of the logs of (2.6.2) and (2.6.3) to obtain

$$(2.6.4) \quad (r_1^2 - r_2^2) = \sum_{k=1}^{\infty} \frac{(2|x|r_1 - r_1^2)^k}{k(2b + |x|^2)^k} - \sum_{k=1}^{\infty} \frac{(-1)^k (2|x|r_2 + r_2^2)^k}{k(2b + |x|^2)^k} .$$

Multiply this by $2b + |x|^2$ to obtain

$$\begin{aligned} A &= (2b + |x|^2 + 1)(r_1^2 - r_2^2) - 2|x|(r_1 + r_2) \\ &= \sum_{k=2}^{\infty} \frac{(2|x|r_1 - r_1^2)^k}{k(2b + |x|^2)^{k-1}} - \sum_{k=2}^{\infty} \frac{(-1)^k (2|x|r_2 + r_2^2)^k}{k(2b + |x|^2)^{k-1}} \\ &= 2(r_1^2 - r_2^2) \frac{|x|^2}{2b + |x|^2} - 2(r_1^3 + r_2^3) \frac{|x|}{2b + |x|^2} + \frac{8}{3}(r_1^3 + r_2^3) \frac{|x|^3}{(2b + |x|^2)^2} + O\left(\frac{1}{|x|^2}\right) \\ &< 2(r_1^2 - r_2^2) + \frac{2}{3}(r_1^3 + r_2^3) \frac{|x|}{2b + |x|^2} + O\left(\frac{1}{|x|^2}\right) \\ &= 4(r_1 + r_2) \frac{|x|}{2b + |x|^2} + \frac{2}{3}(r_1^3 + r_2^3) \frac{|x|}{2b + |x|^2} + O\left(\frac{1}{|x|^2}\right) . \end{aligned}$$

The last expression follows from (2.6.4). Use (2.6.4) a third time to

evaluate

$$\begin{aligned}
 B &= \frac{1}{2}(r_1^4 - r_2^4) \\
 &= \frac{1}{2}(r_1^2 + r_2^2)(r_1^2 - r_2^2) \\
 &= \frac{1}{2}(r_1^2 + r_2^2) \left[2(r_1 + r_2) \frac{|x|}{2b + |x|^2} \right] + o\left(\frac{1}{|x|^2}\right) \\
 &= \left[(r_1^3 + r_2^3) + r_1 r_2 (r_1 + r_2) \right] \frac{|x|}{2b + |x|^2} + o\left(\frac{1}{|x|^2}\right).
 \end{aligned}$$

To show that $B - A > 0$ for large enough $|x|$ it is sufficient to show that the leading term of B is larger than the leading term of A .

Since r_1 and r_2 tend to $r_0 > \sqrt{3}$, for $|x|$ large enough r_1^2, r_2^2 , and $r_1 r_2$ are all greater than $3 + \frac{r_0^2 - 3}{2}$. Apply all three inequalities to the difference $B - A$ to obtain

$$\begin{aligned}
 B - A &> \left[\frac{1}{3}(r_1^3 + r_2^3) - 4(r_1 + r_2) + r_1 r_2 (r_1 + r_2) \right] \frac{|x|}{2b + |x|^2} + o\left(\frac{1}{|x|^2}\right) \\
 &> \left[\frac{1}{3}\left(3 + \frac{r_0^2 - 3}{2}\right)(r_1 + r_2) - 4(r_1 + r_2) + \left(3 + \frac{r_0^2 - 3}{2}\right)(r_1 + r_2) \right] \frac{|x|}{2b + |x|^2} \\
 &\quad + o\left(\frac{1}{|x|^2}\right) \\
 &= \frac{2}{3}(r_0^2 - 3)(r_1 + r_2) \frac{|x|}{2b + |x|^2} + o\left(\frac{1}{|x|^2}\right) > 0
 \end{aligned}$$

for large $|x|$. □

Discussion. It is to be anticipated that the above theorem can be strengthened. There is substantial numerical evidence to support the assertion that if $\text{vol}(C_x^B) < \text{vol}(C_x^O)$ for large $|x|$ then the inequality holds for all $|x|$. In those numerical cases where the volume of C_x^B has dropped below that of C_x^O , it has stayed below this out to $|x| = 100$. The theorem may also be strengthened in that the numerical evidence

suggests that for 90% level confidence sets if $a < -.538$ and for 95% level confidence set if $a < -.69$ then $\text{vol}(C_x^B) < \text{vol}(C_x^O)$ at least out to $|x| = 100$. The numerical results indicate on the other hand that the inequality is not obtainable uniformly for proper prior distributions ($a > 0$) when $p = 3$. On the other hand proper priors seem to work when $p = 5$. Both of these results, neither of which has been proven, are in accord with the Strawderman results concerning minimax Bayes point estimates (see Chapter 1).

Now consider the behavior of the coverage probability. Assume that $b > a + p/2$ so that C_θ^B is always convex. For $|\theta| \leq r_0/2$, $C_\theta^B \supseteq C_\theta^O$, which implies that $P_\theta[C_\theta^B] \geq P_\theta[C_\theta^O]$ for small $|\theta|$. On the other hand, it was shown in Section 4 that $B_{\theta, r_1} \subseteq C_\theta^B \subseteq B_{\theta, r_2}$ and that as $|\theta|$ tends to infinity both r_1 and r_2 tend to r_0 , so that for large $|\theta|$, $P_\theta(C_\theta^B)$ tends to $P_\theta(C_\theta^O)$. The following theorem demonstrates that in three dimensions, for a certain class of improper priors $P_\theta(C_\theta^B)$ exceeds $P_\theta(C_\theta^O)$ over most of the range of $|\theta|$.

Theorem 2.14. If $p = 3$, $a = -1$, and $b \geq \frac{1}{2}$ then for $|\theta| < r_0/2$ or $|\theta| > (2b + r_0^2/2)^{1/2}$

$$P_\theta[C_\theta^B] > P_\theta[C_\theta^O] \quad .$$

Proof. For $|\theta| \leq r_0/2$ this result has been demonstrated already. For the other case, $|\theta| > (2b + r_0^2/2)^{1/2}$, we will show that

$\frac{\partial}{\partial |\theta|} P_\theta[C_\theta^B] < 0$ in a fashion similar to the proof of Theorem 2.13. It is sufficient to show that this derivative is negative since it has already been noted that $\lim_{|\theta| \rightarrow \infty} P_\theta(C_\theta^B) = P_\theta(C_\theta^O) = 1 - \alpha$.

Recall from the proof of Theorem 2.4 that

$$P_{\theta}[C_{\theta}^B] = P[\chi_3^2 \leq r_1^2] + \int_{r_1^2}^{r_2^2} \left(\frac{1}{2} - \frac{z}{2\sqrt{s}} \right) \cdot g(s) ds$$

where $g(s)$ is the χ_3^2 density, $g(s) = c \cdot s^{1/2} e^{-s/2}$. The partial derivative of this with respect to $|\theta|$ is, by the fact that $z(r_1^2) = -r_1$ and $z(r_2^2) = r_2$,

$$\begin{aligned} \frac{\partial}{\partial |\theta|} P_{\theta}[C_{\theta}^B] &= g(r_1^2) \frac{\partial r_1^2}{\partial |\theta|} + 0 \cdot \frac{\partial r_2^2}{\partial |\theta|^2} - g(r_1^2) \frac{\partial}{\partial |\theta|} r_1^2 \\ &\quad + \int_{r_1^2}^{r_2^2} \frac{\partial}{\partial |\theta|} \left(\frac{1}{2} - \frac{z}{2\sqrt{s}} \right) g(s) ds \\ &= -\frac{1}{2} \int_{r_1^2}^{r_2^2} s^{-1/2} g(s) \frac{\partial}{\partial |\theta|} z ds. \end{aligned}$$

From (2.5.14)

$$z = \frac{(2b + |\theta|^2)(e^{s-r_0^2} - 1) - s}{2|\theta|}$$

so

$$\frac{\partial z}{\partial |\theta|} = \frac{(|\theta|^2 - 2b)(e^{s-r_0^2} - 1) + s}{2|\theta|^2}.$$

Substitute this expression for the derivative into the above integral to obtain that $\frac{\partial}{\partial |\theta|} P_{\theta}[C_{\theta}^B] < 0$ if and only if

$$0 > [e^{-s/2}(-(2+s) + (|\theta|^2 - 2b)(1 + e^{s-r_0^2}))] \int_{r_2^2}^{r_1^2}.$$

To show that this inequality holds we first obtain an upper bound

on

$$-(2+s)e^{-s/2} \left[\frac{r_1^2}{r_2^2} \right] .$$

Let $v = e^{-s/2}$ and $h(v) = -2v \log v = se^{-s/2}$. Then

$$h'(v) = -2[\log v + 1] = s - 2$$

and

$$h''(v) = -\frac{2}{v} = -2e^{-s/2} < 0 \quad \text{for all } s.$$

Let $s_1 = r_1^2$ and $s_2 = r_2^2$. Then

$$\begin{aligned} s_1 e^{-s_1/2} - s_2 e^{-s_2/2} &= h(v_1) - h(v_2) \\ &> (v_1 - v_2) h'(v_1) \\ &= (e^{-s_1/2} - e^{-s_2/2}) (s_1 - 2) \end{aligned}$$

so that

$$-(2+s)e^{-s/2} \left[\frac{s_1}{s_2} \right] < -s_1 e^{-s_1/2} \left[\frac{s_1}{s_2} \right] .$$

To show that $\frac{\partial}{\partial |\theta|} P_\theta[C_\theta^B] < 0$ it is sufficient therefore to show that

$$\begin{aligned} 0 &> [e^{-s/2} (-s_1 + (|\theta|^2 - 2b)(1 + e^{s-r_o^2})] \frac{s_1}{s_2} \\ &= e^{-r_o^2/2} \left[e^{\frac{r_o^2 - s_1}{2}} - e^{\frac{r_o^2 - s_2}{2}} \right] \\ &\quad \cdot [(|\theta|^2 - 2b - s_1) - (|\theta|^2 - 2b) \cdot e^{\frac{s_1 - r_o^2}{2}} \cdot e^{\frac{s_2 - r_o^2}{2}}] . \end{aligned}$$

Since $s_1 < r_0^2 < s_2$ this is true if and only if the third factor is negative, that is if and only if

$$(|\theta|^2 - 2b - s_1) < (|\theta|^2 - 2b) \cdot e^{\frac{s_1 - r_0^2}{2}} \cdot e^{\frac{s_2 - r_0^2}{2}}.$$

By (2.5.8) the right side of this inequality equals

$$\begin{aligned} & (|\theta|^2 - 2b) \left[\frac{(2b + (|\theta| - r_1)^2)(2b + (|\theta| + r_2)^2)}{(2b + |\theta|^2)^2} \right]^{1/2} \\ & > (|\theta|^2 - 2b) \left[\frac{(2b + (|\theta| - r_1)^2)(2b + (|\theta| + r_1)^2)}{(2b + |\theta|^2)^2} \right]^{1/2}. \end{aligned}$$

It is sufficient, therefore, that

$$(2.6.4) \quad (|\theta|^2 - 2b - s_1)^2 (|\theta|^2 + 2b)^2 < (|\theta|^2 - 2b)^2 (2b + (|\theta| - r_1)^2) \cdot (2b + (|\theta| + r_1)^2).$$

Let

$$\begin{aligned} A &= (|\theta|^2 - 2b - s_1)^2 \\ &= |\theta|^4 + 4b^2 + s_1^2 - 4|\theta|^2 b - 2|\theta|^2 s_1 + 4bs_1 \end{aligned}$$

and

$$\begin{aligned} B &= (2b + (|\theta| - r_1)^2)(2b + (|\theta| + r_1)^2) \\ &= A + 8|\theta|^2 b. \end{aligned}$$

Inequality (2.6.4) is true if and only if $(B - 8|\theta|^2 b)(|\theta|^2 + 2b)^2 < B(|\theta|^2 - 2b)^2$, that is if and only if $B < (|\theta|^2 + 2b)^2$.

Evaluate

$$\begin{aligned} B &= A + 8|\theta|^2 b \\ &= |\theta|^4 + 4b^2 + s_1^2 + 4|\theta|^2 b - 2|\theta|^2 s_1 + 4bs_1 \end{aligned}$$

so that this inequality in turn is true if and only if

$$|\theta|^2 > 2b + \frac{s_1}{2} \quad .$$

However $s_1 < r_0^2$ so it is sufficient that

$$|\theta|^2 \geq 2b + r_0^2/2 \quad . \quad \square$$

Discussion. Theorem 2.14 is stronger than Theorem 2.13 in that the range of values not covered by the theorem is quite short and can be examined numerically to show with a high degree of certainty that it is true for all $|\theta|$. On the other hand, in order for the theorem to apply it is necessary that C_θ^B contains no holes. At the present time this is assured only if $b > \frac{1}{2}$ which is probably stronger than necessary. Again, Theorem 2.14 is stronger than Theorem 2.13 in that no restrictions are made on r_0 , or equivalently, on $1 - \alpha = P_\theta(C_\theta^O)$.

Both theorems are probably too weak in making the assumption that $a = -1$. This assumption was made primarily in order to simplify the problem. On the other hand there is numerical evidence that the Bayes rules are strongly preferable to the usual 90% level confidence rules if $a < -.538$ and to the 95% level rules when $a < -.69$. In three dimensions, however, proper priors don't seem to work. In five dimensions the picture seems brighter: the numerical results of the next section suggest that proper priors can yield rules which are strongly preferable to C^O . If this turns out to be the case then such Bayes procedures would be weakly admissible in the sense of Definition 2.5 by Theorem 2.7 and minimax with respect to the loss function (2.2.5) by the minimaxity of C^O (Lemma 2.6).

7. Numerical Results

Theorems 2.9 and 2.11 give explicit expressions for computing $\text{vol}(C_x^B)$ for various values of $|x|$ and $P_\theta[C_\theta^B]$ for various values of $|\theta|$ when the dimension p is equal to three. Corollaries 2.10 and 2.12 the same for the case that p is equal to five. In all four cases the formulas are explicit up to the determination of r_1 and r_2 , the endpoints of the respective integrals.

In each case r_1 and r_2 are respectively the absolute values of the negative and positive roots to the equation

$$(2.7.1) \quad e^{mr^2} (2b + (n+r)^2) = e^{mr_0^2} (2b + n^2) \quad .$$

For the computation of $\text{vol}(C_x^B)$, $m = \frac{1}{2a+p}$ and $n = |x|$. For the evaluation of $P_\theta[C_\theta^B]$, $m = -\frac{1}{2a+p}$ and $n = |\theta|$.

In computing the tables of this section the roots to (2.7.1) have been approximated numerically in the following way.

1. An initial guess of the root is made. This guess is $-r_0$ in the case of r_1 and $+r_0$ in the case of r_2 .
2. The guess is made better by repeatedly stepping a constant distance (.1) in the direction in which the fit to equation (2.7.1) is improved until such stepping yields no further improvement.
3. Starting from this point Newton's method is applied repeatedly until the difference between the right and left hand sides of (2.7.1) is less than 10^{-12} .

All computations were performed using double precision arithmetic on the CDC 6400 at the University of California, Berkeley. Numerical values are stored in this mode with an accuracy of greater than 10^{-15} . The FORTRAN double precision library programs were used to evaluate the exponential, logarithm, and square root functions.

In computing the coverage probability it was necessary to evaluate $P[\chi_n^2 \leq s]$ where $n = 3, 5$. This was done using the recursive formula

$$P[\chi_n^2 \leq s] = P[\chi_{n-2}^2 \leq s] - \frac{\left(\frac{s}{2}\right)^{\frac{n}{2}-1} e^{-\frac{s}{2}}}{\Gamma(\frac{n}{2})}.$$

To compute $P[\chi_1^2 \leq s] = 2 \cdot P[N_1(0,1) \leq \sqrt{s}] - 1$ an approximation to $P[N_1(0,1) \leq r]$, $r \geq 0$ was used:

$$P[N_1(0,1) \leq r] = 1 - \frac{1}{\sqrt{2\pi}} e^{-r^2/2} (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \varepsilon(r)$$

where

$$t = \frac{1}{1 + .2316419r}$$

$$b_1 = .319381530$$

$$b_2 = - .356563782$$

$$b_3 = 1.781477937$$

$$b_4 = -1.821255978$$

$$\text{and } b_5 = 1.330274429$$

(Abramowitz and Stegun (1970), p. 932).

According to the reference the error is bounded by

$$|\varepsilon(r)| < 7.5 \times 10^{-8}, \quad 0 \leq r < \infty.$$

There are four tables. In all of them the value of $1 - \alpha$ is equal to .90, so that the Bayes procedures are competing with the usual confidence procedure C^0 which has size .90. The first two tables are for $p = 3$ dimensions, the second two are for $p = 5$ dimensions. In Tables 1 and 3 are given the value of

$$\left(\frac{\text{vol}(C_x^B)}{\text{vol}(C_x^O)} \right)^{1/p},$$

which is the ratio of the radius of a sphere having volume $\text{vol}(C_x^B)$ to that of the sphere C_x^O . (The latter radius is r_o .) This is the ratio of the effective radii of these sets. This, rather than the volume ratio, is used to avoid making the comparison of the procedures dependent upon p . Tables 2 and 4 give the pointwise coverage probability of C^B , that is $P_\theta[C_\theta^B]$.

Table 1

The ratio of the effective radius $(\text{vol}(C_x^B))^{1/3}$ of the Bayes confidence sets in three dimensions to the radius r_0 of the usual confidence set C_x^0 for selected improper priors. In each case $r_0 = 2.500278$ which corresponds to a coverage probability of .90 for C^0 . Table 2 gives the coverage probabilities of the Bayes procedures C^B . It is shown in Theorem 2.13 that for the first four columns the radius ratio tabled here is ultimately less than 1. The last two columns are not covered by the theorem.

prior para- meters	a = -1 b = .5	a = -1 b = 5	a = -1 b = 5	a = -1 b = 10	a = -.54 b = .96	a = -.5 b = 1
x						
0.	.8521482	.8942203	.9627491	.9788238	.8066456	.8039662
1.	.9003526	.9202251	.9666566	.9802049	.8552567	.8523944
2.	.9543239	.9582121	.9753863	.9836408	.9255477	.9234973
3.	.9822848	.9813905	.9839790	.9876816	.9697203	.9687793
5.	.9957041	.9952220	.9938000	.9938693	.9955397	.9956624
10.	.9990878	.9990553	.9988455	.9986670	.9995628	.9996618
50.	.9999653	.9999652	.9999648	.9999643	.9999897	.9999943
100.	.9999913	.9999913	.9999913	.9999913	.9999975	.9999987

Table 2

The pointwise coverage probability $P_\theta(C_\theta)$ of the Bayes procedures C^B in three dimensions. In each column of this table the procedure C^B is the same as that of the corresponding column in Table 1. The coverage probabilities are to be compared with $P_\theta(C_\theta^0) = .90$ for the usual procedure. It is shown in Theorem 2.14 that for the first four columns $P_\theta(C_\theta^B) > .90$ for $|\theta| < 1.25$ and for $|\theta| > 5$. For the fifth column it is not known what happens to $P_\theta(C_\theta^B)$. In the sixth column it falls below .90.

prior para- meters θ	a = -1 b = .5	a = -1 b = 1	a = -1 b = 5	a = -1 b = 10	a = -.54 b = .96	a = -.5 b = 1
0	.9633164	.9506789	.9203385	.9116721	.9787986	.9796694
1	.9481071	.9402322	.9181397	.9108756	.9662891	.9673735
2	.9212676	.9206133	.9131908	.9089060	.9316728	.9323282
3	.9078836	.9087292	.9084077	.9066273	.9109043	.9110347
5	.9021800	.9024120	.9032074	.9046741	.9018885	.9017797
10	.9004650	.9004827	.9005980	.9006932	.9001331	.9000719
50	.9000176	.9000176	.9000179	.9000181	.9000003	.8999973
100	.9000044	.9000044	.9000044	.9000044	.9000001	.8999993

Table 3

The ratio of the effective radius $(\text{vol}(C_x^B))^{1/5}$ of the Bayes confidence sets in five dimensions to the radius r_0 of the usual confidence set C_x^0 for selected proper priors. In each case $r_0 = 3.039136$ which corresponds to a coverage probability of .90 for C_0^0 . Table 4 gives the coverage probabilities of the Bayes procedures C^B .

prior para- meters	a = .25 b = 2.75	a = .25 b = 3.	a = .25 b = 5.	a = .25 b = 10.	a = .5 b = 3.
x					
0	.7681113	.7791478	.8380255	.9003570	.7639364
1	.7978035	.8061356	.8528679	.9062796	.7930201
2	.8583350	.8623018	.8871231	.9213566	.8535499
3	.9135493	.9147126	.9234485	.9398898	.9101524
5	.9745654	.9517631	.9714539	.9706348	.9745447
10	.9980701	.9979554	.9971267	.9955988	.9991753
50	.9999952	.9999919	.9999902	.9999859	1.0000558
100	.9999987	.9999987	.9999985	.9999982	1.0000147

Table 4

The pointwise coverage probability $P_{\theta}(C_{\theta})$ of the Bayes procedure C^B in five dimensions. In each column of this table the procedure C^B is the same as that of the corresponding column in Table 3. The coverage probabilities are to be compared with $P_{\theta}(C_{\theta}^0) = .90$ for the usual procedure.

prior para- meters θ	a = .25 b = 2.75	a = .25 b = 3.	a = .25 b = 5.	a = .25 b = 10.	a = .5 b = 3.
0	.9954613	.9944660	.9848086	.9621316	.9962043
1	.9921067	.9908511	.9802696	.9588063	.9931989
2	.9745416	.9734752	.9652679	.9497543	.9761365
3	.9459596	.9460206	.9446958	.9380427	.9469489
5	.9159628	.9162689	.9179707	.9190330	.9157273
10	.9027112	.9027641	.9031502	.9038789	.9022553
50	.9000729	.9000730	.9000739	.9000763	.9000426
100	.9000176	.9000176	.9000177	.9000178	.9000099

The preceding tables are not complete of course. The original computer output shows that for every value of $|x|$ or $|\theta|$ which is an integer between 0 and 100 the functions $\text{vol}(C_x^B)$ and $P_\theta[C_\theta^B]$ are monotonely approaching their limiting values, except in the cases where the Bayes procedures are not preferable to C^O , such as the right hand columns above. In these cases the functions pass through their limiting value at most once for every set of computed values I have generated so far.

The numerical results are strong evidence in themselves that for a good choice of the prior, the Bayes procedures do better than C^O in three or five dimensions. It may be noted that among the Bayes procedures it is possible, by altering b for example, to increase one's chance of doing well when $|x|$ or $|\theta|$ is near zero, at the expense of doing worse when $|x|$ or $|\theta|$ is large.

Overall, the results show that if an individual is able to pick a prior estimate of θ , and is willing to place a somewhat high degree of prior confidence in it, then he can expect to do substantially better than he would do if he used the usual procedure if his guess is nearly correct, while being assured of doing slightly better than the usual procedure even if his guess is quite wrong. As a final comment about the tables it should be pointed out that in three or four dimension, if the parameter b is allowed to go as low as $\frac{1}{8}(a + \frac{p}{2})$ rather than $a + \frac{p}{2}$ as in these tables then the results are even more dramatic than those shown here for small values of $|x|$ and $|\theta|$. The only problem with this is that the present results do not guarantee that for these parameter values C_θ^B is convex, so that the computations could conceivably be incorrect.

APPENDIX

1. The expression (2.5.6) in five dimensions for $\text{vol}(C_x^B)$ involves the integral

$$(A.1) \quad \int_{r_2}^{r_1} r^3 z - \frac{1}{3} r z^3 \, dr \quad .$$

As in (2.5.5)

$$(A.2) \quad z = \frac{1}{2|x|} [A(g(r^2, B) - 1) - r^2]$$

where $A = (2b + |x|^2)$, $B = \frac{-1}{2a+5}$, and $g(r^2, B) = e^{B(r^2 - r_0^2)}$.

The first term of the integrand is simply

$$(A.3) \quad r^3 z = \frac{1}{2|x|} [Ar^3 g(r^2, B) - Ar^3 - r^5] \quad .$$

The second term is

$$\begin{aligned} (A.4) \quad -\frac{1}{3} r z^3 &= \frac{-1}{24|x|^3} [rA^3(g-1)^3 - 3r^3A^2(g-1)^2 + 3r^5A(g-1) - r^7] \\ &= \frac{-1}{24|x|^3} \{A^3[rg(r^2, 3B) - 3rg(r^2, 2B) + 3rg(r^2, B) - r] \\ &\quad - 3A^2[r^3g(r, 2B) - 2r^3g(r, B) + r^3] \\ &\quad + 3A[r^5g(r, B) - r^5] - r^7\} \quad . \end{aligned}$$

On the computer the integral

$$\int_{r_2}^{r_1} r^n g(r^2, c) \, dr \quad , \quad n \text{ an odd integer, } c \neq 0$$

is evaluated using the recursive formulas

$$\int_{r_2}^{r_1} r g(r^2, c) dr = \frac{g(r^2, c)}{2c} \Big|_{r_2}^{r_1} \quad \text{and for } n \geq 3$$

(A.5)

$$\int_{r_2}^{r_1} r^n g(r^2, c) dr = \frac{r^{n-1} g(r^2, c)}{2c} \Big|_{r_2}^{r_1} - \frac{n-1}{2c} \int_{r_2}^{r_1} r^{n-2} g(r^2, c) dr \quad .$$

2. The expression in five dimensions for the coverage probability (2.5.15) involves the integral

$$(A.6) \quad e^{-\frac{r_0^2}{2}} \int_{r_1}^{r_2} \frac{1}{3} z^3 g(s, -\frac{1}{2}) - s z g(s, -\frac{1}{2}) ds \quad .$$

In this case from (2.5.14),

$$(A.7) \quad z = \frac{1}{2|\theta|} [A \cdot (g(s, B) - 1) - s] \quad ,$$

$A = (2b + |\theta|^2)$, $B = \frac{1}{2a+5}$, and $g(s, B) = e^{B(s-r_0^2)}$. (The principal change here is that B is positive.) Proceeding as for the volume case, the first term of the integrand of (A.6) is

$$\begin{aligned} (A.8) \quad \frac{1}{3} z^3 g(s, -\frac{1}{2}) &= \frac{1}{24|\theta|^3} [A^3 (g(s, B) - 1)^3 - 3A^2 (g(s, B) - 1)^2 s \\ &\quad + 3A (g(s, B) - 1) s^2 - s^3] \cdot g(s, -\frac{1}{2}) \\ &= \frac{1}{24|\theta|^3} \{ A^3 [g(s, 3B - \frac{1}{2}) - 3g(s, 2B - \frac{1}{2}) + 3g(s, B - \frac{1}{2}) \\ &\quad - g(s, -\frac{1}{2})] \\ &\quad - 3A^2 [s g(s, 2B - \frac{1}{2}) - 2s g(s, B - \frac{1}{2}) + s g(s, -\frac{1}{2})] \\ &\quad + 3A [s^2 g(s, B - \frac{1}{2}) - s^2 g(s, -\frac{1}{2})] - s^3 g(s, -\frac{1}{2}) \} \quad . \end{aligned}$$

The second term of the integrand is, of course

$$(A.9) \quad -szg(s, -\frac{1}{2}) = \frac{-1}{2|\theta|} [Asg(s, B-\frac{1}{2}) - Asg(s, -\frac{1}{2}) - s^2 g(s, -\frac{1}{2})] \quad .$$

On the computer, the integral

$$\int_{r_1}^{r_2} s^n g(s, c) ds, \quad n \text{ a non-negative integer, } c \neq 0$$

is evaluated using the recursive formulas

$$\int_{r_1}^{r_2} g(s, c) ds = \left. \frac{g(s, c)}{c} \right|_{r_1}^{r_2}$$

(A.10)

$$\text{and for } n \geq 1, \quad \int_{r_1}^{r_2} s^n g(s, c) ds = \left. \frac{s^n g(s, c)}{c} \right|_{r_1}^{r_2} - \frac{n}{c} \int_{r_1}^{r_2} s^{n-1} g(s, c) ds \quad .$$

REFERENCES

- [1] Abramowitz, M. and Stegun, I.A. (1970). Handbook of Mathematical Functions. Applied Mathematics Series 55. National Bureau of Standards, Washington, D.C.
- [2] Baranchik, A.J. (1970). A family of estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* 41, 642-645.
- [3] Brown, L.D. (1966). Admissibility of invariant estimators. *Ann. Math. Statist.* 37, 1087-1136.
- [4] Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* 42, 927-945.
- [5] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *J. Amer. Statist. Assoc.* 68, 117-130.
- [6] Ferguson, T.S. (1967). Mathematical Statistics, a Decision Theoretic Approach. Academic Press, New York.
- [7] Hodges, J.L. Jr. and Lehmann, E.L. (1951). Some applications of the Cramér-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 1, 13-22.
- [8] James, W. and Stein, C. (1961). Estimation with quadratic loss function. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1, 361-379.
- [9] Joshi, V.M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* 38, 1868-1875.
- [10] Joshi, V.M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* 40, 1042-1067.
- [11] Lehmann, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.
- [12] Olshen, R.A. and Savage, L. (1970). A generalized unimodality. *J. Appl. Prob.* 7, 21-34.
- [13] Olshen, R.A. (1973). The conditional level of the F-test. *J. Amer. Statist. Assoc.* 68, 692-698.
- [14] Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* 40, 87-104.
- [15] Scheffé, H. (1959). The Analysis of Variance. Wiley, New York.

- [16] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 197-206.
- [17] Stein, C. (1961). Confidence sets for the mean of a multivariate normal distribution. J. Roy. Statist. Soc. Ser. B 24, 265-296.
- [18] Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. Festschrift fur J. Neyman. Research Papers in Statistics. Wiley, New York. 351-366.
- [19] Stein, C. (1974). Estimation of the parameters of a multivariate normal distribution, part I. Technical Report No. 63, Department of Statistics, Stanford University.
- [20] Strawderman, W.E. and Cohen, A. (1971). Admissibility of estimators of the mean vector of a multivariate normal distribution with quadratic loss. Ann. Math. Statist. 42, 270-296.
- [21] Strawderman, W.E. (1971). Proper Bayes minimax estimators for the mean of a multivariate normal population. Ann. Math. Statist. 42, 385-388.